UCGE Reports
Number 20224

Department of Geomatics Engineering

Quality Assessment of Ikonos and Quickbird Fused Images for Urban Mapping

(URL: http://www.geomatics.ucalgary.ca/links/GradTheses.html)

by

Valarmathy Meenakshisundaram

June 2005

UNIVERSITY OF CALGARY

Quality Assessment of Ikonos and Quickbird Fused Images for Urban Mapping

by

Valarmathy Meenakshisundaram

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF GEOMATICS ENGINEERING

CALGARY, ALBERTA

JUNE, 2005

Abstract

New series of very high spatial resolution (VHR) satellites Ikonos and Quickbird have enabled mapping and updating of GIS databases of urban areas that is presently carried out using field surveys and aerial images. Satellites provide higher spatial resolution in panchromatic (PAN) mode compared to that in multispectral (MS) mode. High spatial and high spectral resolution are desirable for urban mapping as high spatial resolution provides better geometric quality while high spectral resolution provides better object identification. Image fusion techniques aim at increasing the spatial resolution of MS images using information from PAN image. However, fusion methods alter the spectral content of the original images. This is not desirable in applications requiring spectral information such as visual interpretation or classification procedures that depend on the spectral information of MS images. In this study, fused images obtained for Ikonos PAN and MS and Quickbird PAN and MS images by the standard methods namely IHS (Intensity-Hue-Saturation) and PCA (Principal Component Analysis), and simple wavelet methods namely, IHS with wavelet (IHS+W), PCA with wavelet (PCA+W), Wavelet Addition (WA) and Wavelet Substitution (WS) and complex ARSIS ("Amelioration de la Resolution Spatiale par Injection of Structures") methods are compared and analysed visually and statistically for urban mapping. Since PAN is less correlated with the Blue band, it results in high spectral error in the fused Blue band of IHS, IHS+W, WA and WS methods. The ARSIS models aim at synthesizing the images at high resolution close to reality. However, it is found that the ARSIS models produce similar results to the WA and WS methods in some bands and introduce more error in the NIR band compared to other methods. The ARSIS M2 method provides similar results as the PCA method. Based on the subjective (visual) assessment, of all the methods ARSIS M2 and PCA method provides good spatial quality while best preserving the colour of objects. Thus, these fused MS images are better for visual interpretation and mapping.

VHR images have inadequate spectral resolution for complete discrimination of urban classes: roads and buildings. The high within-class spectral variance in VHR images

results in misclassifications. The problem is increased in the fused images where there are more spatial details compared to the original MS images. Also, the spectral variance in each class is further increased by fusion methods resulting in more misclassifications. However, because of the high redundancy in the MS bands, the classified fused images of different methods do not show much difference. Considering other pre and post-processing steps involved in automated urban feature extraction, classification is only a part of the whole process. Future Worldview satellite from Digital Globe will provide higher spatial resolution for PAN (0.5 m) and 8-MS bands (2 m). With such a very high resolution, the need for the fusion of PAN and MS images has to be further investigated especially for automatic feature extraction procedures.

Acknowledgements

I would like to express my sincere thanks to Dr. Isabelle Couloigner for supporting, guiding and supervising my graduate studies. I thank Dr. Thierry Ranchin for providing ARSIS synthesized images for the study and also for being my external thesis reader. I thank Ms. Claire Thomas of the Ecole des Mine de Paris for her time and interest in discussing her ideas and providing clarifications on quality assessment.

I would like to acknowledge all the sponsors and collaborators for the project: GEIODE, NSERC, Center for Topographic Information (Ottawa), the City of Fredericton (NB, Canada), and Ecole des Mines de Paris (Remote Sensing Group).

I personally thank our group members Mr. Santhosh Phalke, Dr. Qiaoping Zhang, and Ms. Wen-Ya Chiu for providing valuable suggestions, discussions during the entire course of my studies and also for sharing the computing resources.

Table of Contents

List of Tables

List of Figures and Illustrations

List of Abbreviations

Symbol	Definition
CC	Correlation coefficient
DIV	Difference in Variance
SDD	Standard deviation of difference image
HIS	Intensity-Hue-Saturation
IHS+W	IHS with wavelet
IK D1	Ikonos data set 1
IK D2	Ikonos data set 2
MS	Multispectral
M2	ARSIS MSM3M2 model
PAN	Panchromatic
PCA	Principal Component Analysis
PCA+W	PCA with wavelet
QB D1	Quickbird data set 1
QB D2	Quickbird data set 2
VHR	Very High Spatial Resolution

Chapter 1 Introduction

Mapping of urban features (e.g. roads and buildings) from satellite images has gained enormous research interest with the launch of the Ikonos and Quickbird satellites that provide very high spatial resolution (VHR) panchromatic images (PAN) of 1 m and 0.7 m respectively and multispectral images (MS) in four bands of 4 m and 2.8 m respectively. Sensor limitations in acquiring images with high spatial as well as high spectral resolution have led to the research in image fusion techniques to obtain images with high spatial as well as high spectral resolution. Because of the complexity of an urban environment and the high level of spatial details in VHR images, different fusion techniques to combine complementary data sets such as PAN, MS, Lidar and hyperspectral data is currently of interest in the field of urban feature extraction.

Many image fusion methods have been developed in the last two decades for integrating images of different characteristics (e.g. SAR and Optical) and of different spatial resolutions (e.g. SPOT PAN and Landsat TM) to exploit the complementary data sets to obtain better information, interpretation and mapping. The need for increasing the interpretability of “low” spatial resolution images such as Landsat TM (30 m) initiated the research in PAN and MS fusion. The earlier fusion methods were developed based on simple pixel by pixel addition, subtraction, band arithmetic, ratio (Price, 1987, Munechika, 1993), IHS (Intensity-Hue-Saturation) (Welch *et al*., 1987; Carper, 1990), PCA (Principal Component Analysis), and high pass filters (Chavez *et al*., 1991). In the past few years, the wavelet tools have been extensively used and several new methods have been proposed. Currently, the research in image fusion focuses on applying the existing methods on images from different sensors and also on evaluating the quality of the fused images. In this chapter, the general motivation for the fusion of VHR PAN and MS images, the need for this research and the research objectives are presented.

1.1 General Motivation

Very high spatial resolution enables an accurate description of shapes, features and structures while high spectral resolution enables better identification and classification of the features (Couloigner *et al*., 1998). Even though Ikonos and Quickbird images provide MS images with a very high spatial resolution of 4 m and 2.8 m respectively, this spatial resolution is insufficient for an operational level mapping. For urban mapping, map scales of 1: 5000 to 1:10 000 and 1:1000 to 1:2000 are desired for tactical and operational levels respectively (Weber *et al*., 2003). To obtain maps in these scales, the spatial resolution requirements of remote sensing images are of the order of 50 cm to 5 m for the tactical level and of the order of 20 to 50 cm for the operational level (Weber *et al*., 2003). Ideally, improving the spatial resolution of the MS images should enable more detailed and more accurate urban maps. As we can see from Figure 1.1 a, the roads marked 3, 4 and 5 are difficult to map using 4 m MS images. Figure 1.1 b shows the fused images in true colour composite. The increased spatial resolution of the fused images enables a better interpretation and an easier mapping. The fused MS images have an additional advantage of colour (object identification) over a PAN image at VHR, thereby reducing time in photo interpretation as well as errors in feature identification and mapping. Similarly, in automatic urban mapping, fused MS images have high spatial resolution and have the spectral information of objects that allows automatic classification of objects.

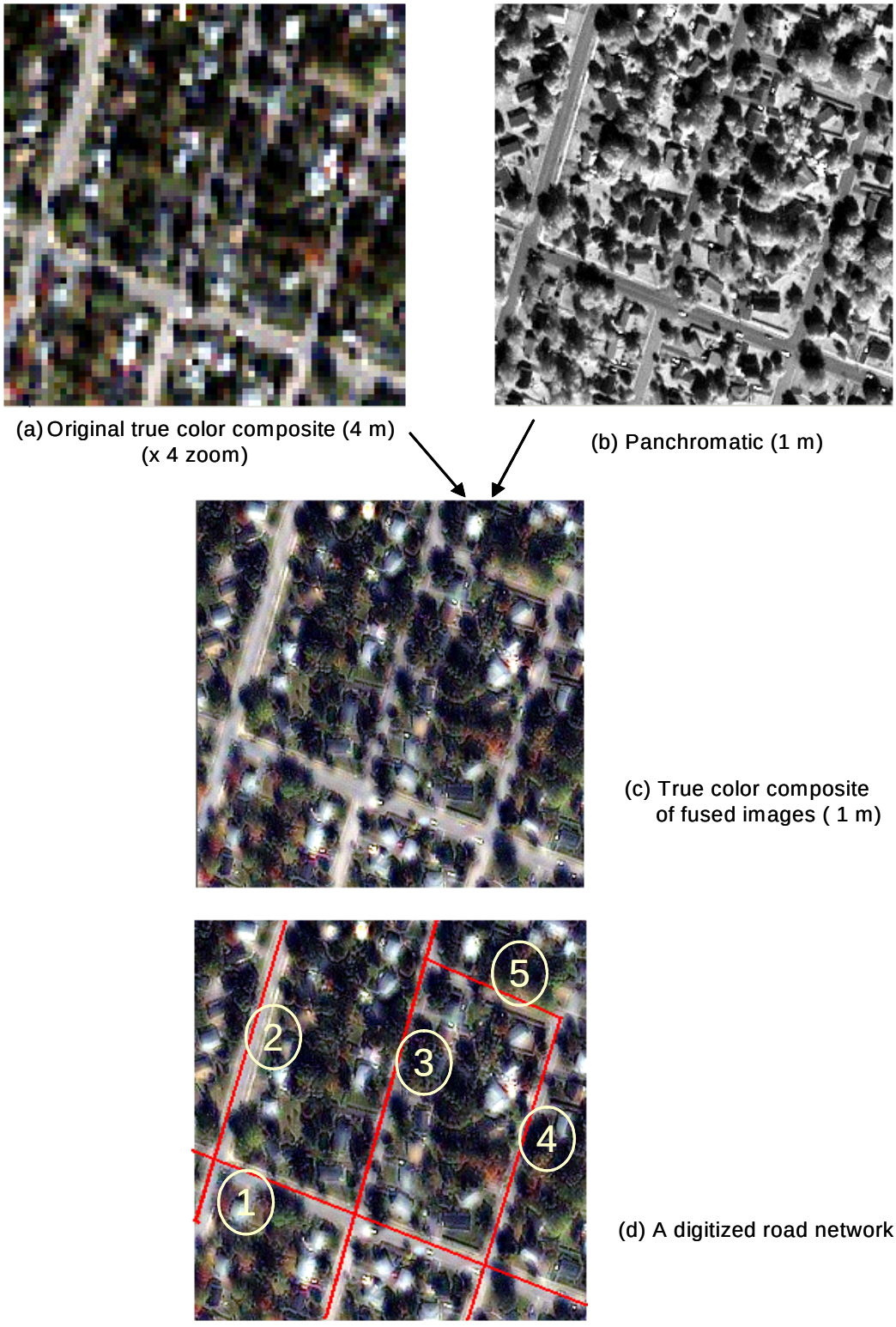


Figure 1.1 Original and fused Ikonos images of a sub-urban area

1.2 Need for Research

Most of the existing methods were developed for the fusion of “low” spatial resolution images such as SPOT and Landsat TM. They may or may not be suitable for the fusion of VHR images (Zhang, 2002). Hence, the existing methods have to be evaluated for VHR images.

Results obtained with one fusion method may vary depending on the scene complexity and the application. For example, a method “A” that is “superior” to a method “B” for a certain data set may not be superior for another data set even if the data sets are from the same sensors. Therefore, a number of experiments on different data sets are required before conclusions can be drawn on the most suitable method of fusion.

There is also a lack of measures for assessing the objective and subjective quality of the fusion methods. The quantitative measures are based on assessing the objective quality: spectral preservation while increasing the spatial resolution. They do not reflect the subjective quality of the images: visual quality for photo interpretation and preservation of spectral variance required in classification. In other words, the objective and subjective quality measures have low correlation (Cornet & Binard, 2004). Therefore, a subjective quality as well as an objective quality assessment is required.

1.3 Research Objectives

The primary objective of this research is to evaluate the visual and spectral quality of different fused images obtained from Ikonos PAN and MS images and from Quickbird PAN and MS images in urban areas. The secondary objective is to analyze the usefulness of the fused images in automatic urban feature extraction.

This research is mainly concerned with the evaluation of some pixel-based fusion methods for VHR images. The research outcomes and conclusions will contribute towards understanding the suitability of the existing fusion methods for Ikonos and Quickbird images. However, the results and conclusions are based on specific data sets and firm and global conclusions cannot be drawn. Several other experiments on different data sets are necessary to make final conclusions on fusion methods.

1.4 Thesis Outline

In Chapter 2, data fusion, common terminologies, architectures and fusion levels are presented. A review of standard and wavelet based methods with a brief introduction of the wavelet transform is provided.

In Chapter 3, some of the relevant and useful quantitative measures for the quality assessment of fused images are discussed.

In Chapter 4, the data sets and the fusion methods used are presented. The fused images obtained by the different methods are presented. The fused images have been evaluated based on the visual and statistical criteria discussed in Chapter 3.

In Chapter 5, the classification of VHR images in urban environment and of the fused images obtained by the different methods is discussed. The relevance of the statistical quality parameters for classification is discussed.

In Chapter 6, conclusions are drawn and future scope of the research is discussed.

Chapter 2 Fusion Process

2.1 Introduction

In this chapter, a general overview of data fusion, fusion architectures and fusion levels are presented first. Then, some pixel-based standard and wavelet-based fusion methods are discussed. In the final part, the sensor characteristics of Ikonos and Quickbird sensors are discussed.

2.2 Data Fusion

Wald (1999) proposed a general definition for data fusion in the context of earth data.

> "*Data Fusion is a formal framework in which are expressed means and tools for the alliance of data originating from different sources. It aims at obtaining information of greater quality; the exact definition of 'greater quality' will depend upon the application"*

This definition equally emphasizes the tools for combining the data and the quality of the result. Merging, combination, data assimilation, and integration are other terms that are used to refer to data fusion. Image fusion is a sub domain of data fusion referring to the fusion of two or more images. Pohl and Van Genderen (1998) defined image fusion as:

> "*[..] the combination of two or more different images to form a new image by using a certain algorithm*"

It can refer to any fusion process involving images from sensors of same satellites or different satellites having different spatial, spectral and temporal characteristics (e.g. SPOT PAN with Landsat TM, SPOT PAN and SPOT XS, ENVISAT ASAR with SPOT Vegetation).

2.3 Fusion Architecture

Fusion architecture describes the general scheme for combining and processing the inputs in the fusion process. The selection of a suitable architecture depends on the nature of the problem, the characteristics of the data, the availability of computing power and other factors. They are generally categorized into centralized, decentralized and hybrid architectures.

The centralized architecture (Figure 2.1) takes all the available input simultaneously in order to derive the information. The D1, D2 and D3 are data from different sources (e.g. Images, DEM, GIS, other ancillary data like maps or ground truth) entering the fusion process. The input data set may comprise multi-temporal images, images of different spatial and spectral resolution or any other auxiliary data sets (e.g. D1 and D2 can be two images obtained at different dates and D3 may be GIS data for change detection). Since all the sources are taken in one fusion process, it offers a minimal loss of information. One drawback of this architecture is that if one dataset is of poor quality, it will affect the quality of the final result. Another disadvantage is the requirement of high processing power and computer memory.

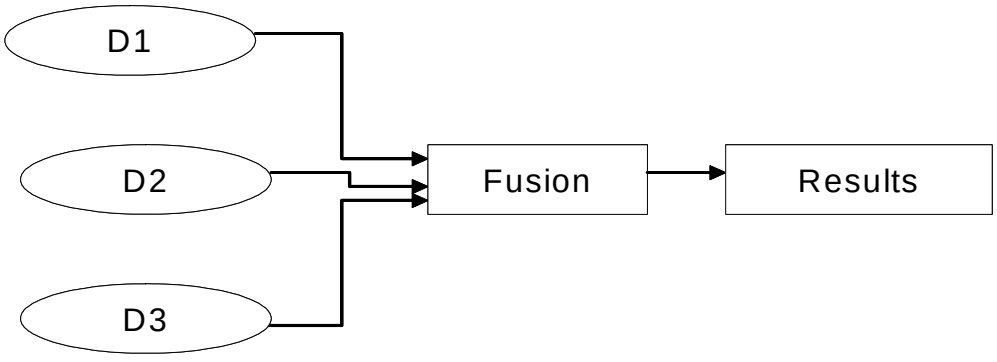


Figure 2.1 Centralized architecture

In the decentralized architecture (Figure 2.2), the inputs are processed in different fusion processes. In Figure 2.2, the sources D1 and D2, D3 and D4 are processed in different fusion processes and the results are combined using another fusion process. The decentralized architecture offers a greater flexibility in processing.

Hybrid architecture is one that combines centralized and decentralized architectures. An illustration is shown in Figure 2.3. These architectures might need different processing stages and levels.

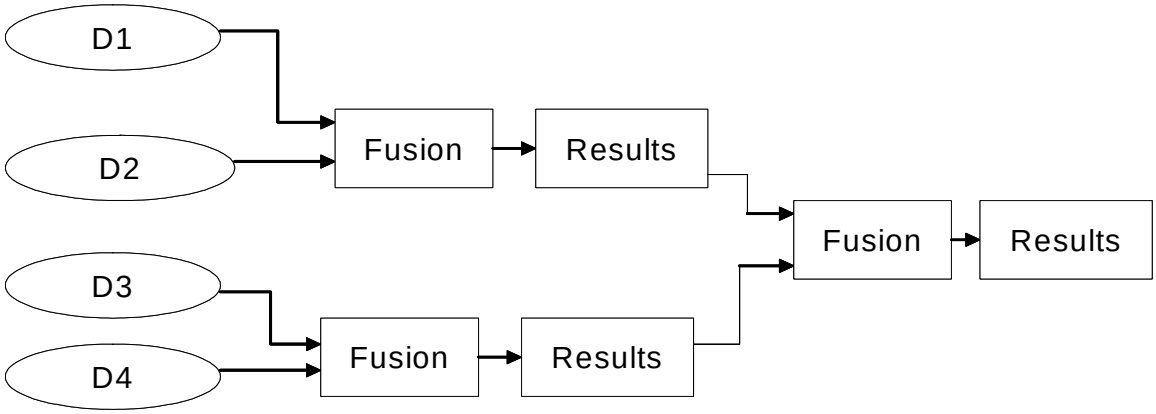


Figure 2.2 Decentralized architecture

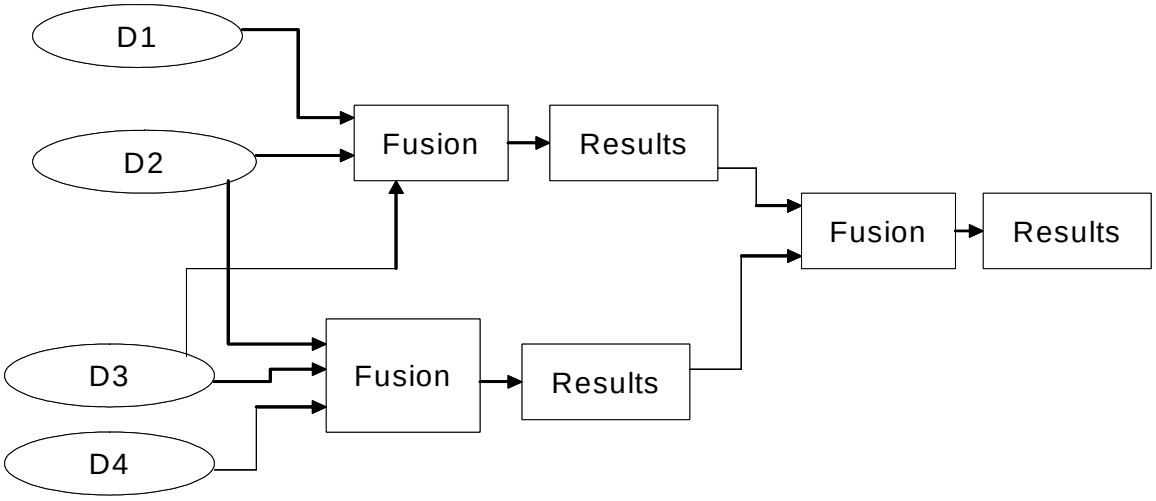


Figure 2.3 Hybrid architecture

2.4 Fusion Levels

Fusion level describes the level at which the fusion takes place. Fusion can be either at the pixel, feature or decision level. The following description and illustrations of fusion levels are given in the context of feature extraction.

2.4.1 Pixel level fusion

Pixel level fusion requires the least amount of pre-processing. It uses the DN or radiance values of each pixel from different sources in order to derive the useful information. Geometric registration and time difference in the acquisition of the inputs should be taken into consideration. Classification of multispectral or hyperspectral images along with other sources like PAN and DEM (Digital Elevation Model) for land use mapping is a good example to explain pixel level fusion. An illustration is provided in Figure 2.4. Here, data fusion refers to the use of a pixel vector composed of MS images, a PAN image and a DEM to derive the information. The pixel vector obtained from the different sources is used to obtain the result but there is no actual manipulation of the pixel values. The pixel based fusion of PAN and MS is also a pixel level fusion where new values are created or modeled from the DN values of PAN and MS images.

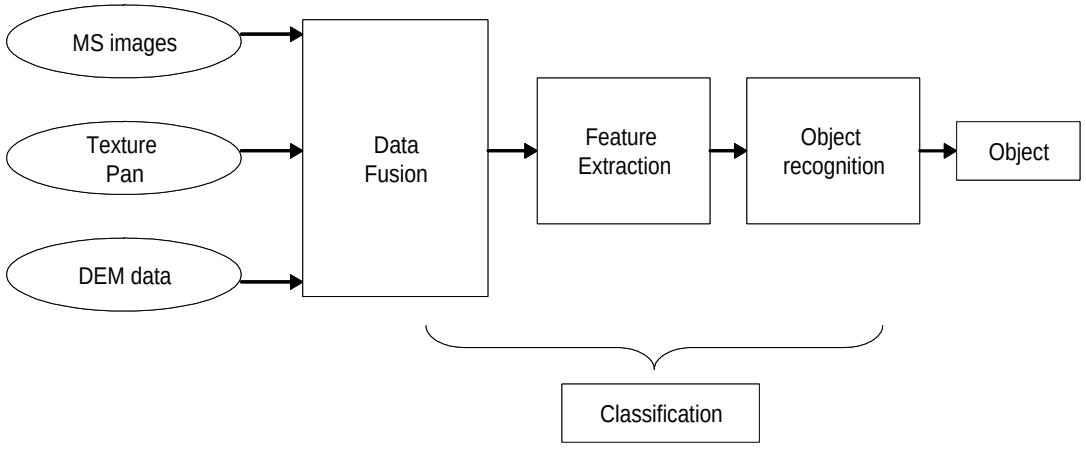


Figure 2.4 Pixel level fusion

2.4.2 Feature level fusion

Feature level fusion involves the extraction of feature primitives like edges or regions by segmentation procedures from different images. These extracted features are then combined using rule-based (fuzzy approaches) or knowledge-based approaches using Artificial Neural Networks (ANN), object-oriented or statistical approaches. This involves a higher level of processing. This fusion level is increasingly used in urban feature extraction. Figure 2.5 represents a feature level fusion. The regions and edges extracted from the different sources like MS images, PAN and Lidar data are combined to obtain a more meaningful representation of the objects of interest (e.g. 2D or 3D building models).

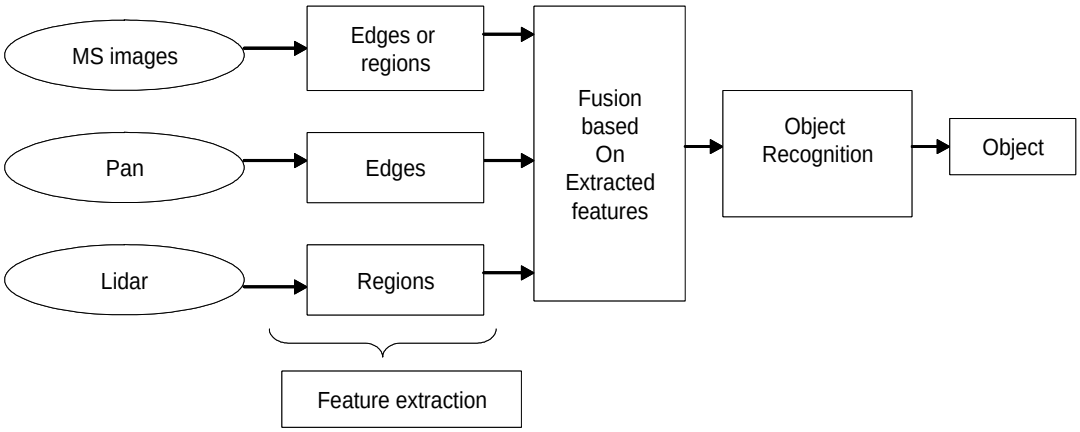


Figure 2.5 Feature level fusion

2.4.3 Decision/Object level fusion

In decision level fusion, the results are derived from the combined knowledge from different sources. Figure 2.5 is an example of an object level fusion. In this example, objects are derived independently from each source and decision rules are framed to use the information obtained from different sources to make a final decision about the objects of interest.

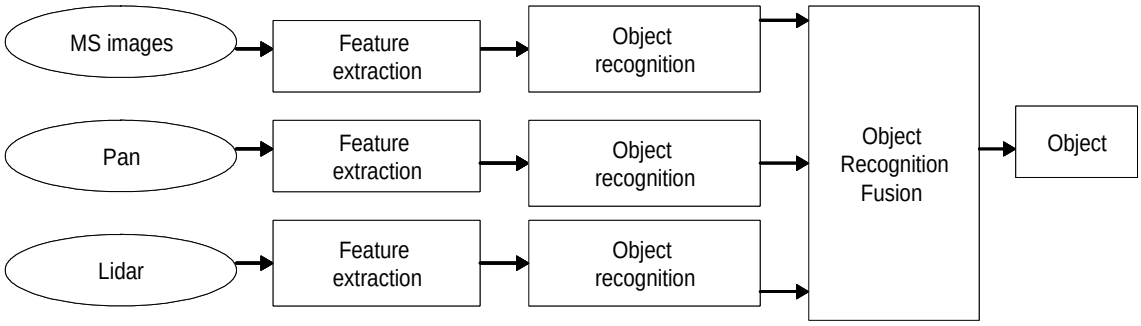


Figure 2.6 Decision level fusion

2.5 Pixel based Fusion

Pixel-based fusion methods can be categorised under the pixel level fusion. Pixels corresponding to the same spatial objects from two different sources are manipulated to obtain the resultant image. Before fusing two sources at a pixel level, it is necessary to perform a geometric registration and a radiometric adjustment of the images to one another. When images are obtained from sensors of different satellites as in the case of fusion of SPOT and Landsat, the registration accuracy is very important. But registration is not much of a problem with simultaneously acquired images as in the case of Ikonos/Quickbird PAN and MS images. The PAN images have a different spatial resolution from that of MS images. Therefore, resampling of MS images to the spatial resolution of PAN is an essential step in some fusion methods to bring the MS images to the size of PAN especially if an undecimated wavelet transform is used. Bicubic interpolation is considered a good choice for resampling (Wald, 2002). The resultant fused images may not be satisfactory when the two images entering the fusion have differences in their dynamic ranges. Hence, for some fusion methods such as IHS and PCA, pre-processing steps should include mean matching, variance matching or histogram matching to adjust the images entering the fusion process.

Pixel based image fusion methods can be grouped into three categories namely:

1. Projection and Substitution methods
2. Spectral Contribution methods

3. Frequency Filtering / Modeling methods

A few methods are described here. The frequency modeling methods are promising for VHR urban areas as most of the objects are well represented in VHR MS images except for some spatial details. Many recent works in fusion demonstrate that wavelet-based methods provide better results. Therefore, only standard and wavelet-based methods are used in the study. The examples and equations provided in the next sections correspond to the fusion of PAN and MS images although they can be applied to fusion of images from other sensors with or without slight modifications.

2.5.1 Projection and Substitution methods

The methods under this category involve the transformation of the input (MS) images into new components. The IHS (Intensity-Hue-Saturation) and PCA (Principal Component Analysis) transformations fall under this category. These two methods have become standard methods in image fusion.

2.5.1.1 The IHS fusion method

The IHS method is based on the human colour perception parameters. It separates the spatial (I) and spectral (H, S) components of a RGB image. Intensity refers to the total brightness of the colour. Hue refers to the dominant wavelength. Saturation refers to the purity of the colour relative to gray. In fusion, the IHS transformation is used to convert three bands of an MS image from the RGB colour space to the IHS colour space. The I component is related to the spatial frequencies and is highly correlated with the PAN image. However, PAN has higher spatial frequencies than the MS images. These high frequencies represent the finer details present in the PAN image. Therefore, replacing the I component with the PAN image and transforming back to the RGB colour space will introduce high frequencies from PAN into the MS image. The PAN is usually contrast stretched or histogram matched to the I component it replaces. There are different algorithms for the computation of the IHS components. These algorithms differ in the

computation of the I component; however they tend to produce the same values for H and S (Nuñez *et al*., 1999).

A simple model for the IHS transformation is given in Pohl and Van Genderen (1998). This model is implemented in many commercial software (Wald, 2002). This is the model used to test the suitability of the IHS method for VHR images.

The conversion equations are:

$$\begin{pmatrix} I \\ v1 \\ v2 \end{pmatrix} = \begin{pmatrix} 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \\ 1/\sqrt{6} & 1/\sqrt{6} & -2/\sqrt{6} \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix} \tag{2.1}$$

and $H = \tan^{-1}(v2/v1)$ with H not defined if R + G = 2B,

$S = \sqrt{v1^2 + v2^2}$.

The IHS to RGB transform equations are:

$$v1 = S\cos(H)$$

$$v2 = S\sin(H)$$

$$\begin{pmatrix} R \\ G \\ B \end{pmatrix} = \begin{pmatrix} 1/\sqrt{3} & 1/\sqrt{6} & 1/\sqrt{2} \\ 1/\sqrt{3} & 1/\sqrt{6} & -1/\sqrt{2} \\ 1/\sqrt{3} & -2/\sqrt{6} & 0 \end{pmatrix} \begin{pmatrix} I \\ v1 \\ v2 \end{pmatrix} \tag{2.2}$$

2.5.1.2 The PCA (Principal Component Analysis) fusion method

The PCA method is based on statistical parameters. It transforms a multivariate data set of inter-correlated variables into new uncorrelated linear combinations of the original values. This method is also based on the assumption that the first PC (Principal Component) is highly correlated with PAN. The PCA method is very similar to IHS except that it is the first PC (PC1) that is replaced by PAN. As with the IHS method, PAN is stretched or histogram matched to PC1.

The equation to compute the PC components from 3 bands of the MS image is given by:

$$\begin{bmatrix} PC1 \\ PC2 \\ PC3 \end{bmatrix} = \begin{bmatrix} \phi_{11} & \phi_{12} & \phi_{13} \\ \phi_{21} & \phi_{22} & \phi_{23} \\ \phi_{31} & \phi_{32} & \phi_{33} \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{2.3}$$

where each row in the transformation matrix ϕ represents the eigen vectors of the covariance matrix ρ. The transformation matrix satisfies the relationship $\phi\rho\phi^T = \Lambda$ where $\Lambda = diag(\Lambda_1, \Lambda_2, \Lambda_3)$ are the eigen values corresponding to ϕ in the descending order. R, G, and B are the input images.

To convert back to the RGB space, the transformation is given by:

$$\begin{bmatrix} R_{fused} \\ G_{fused} \\ B_{fused} \end{bmatrix} = \begin{bmatrix} \phi_{11} & \phi_{21} & \phi_{31} \\ \phi_{12} & \phi_{22} & \phi_{32} \\ \phi_{13} & \phi_{23} & \phi_{33} \end{bmatrix} \begin{bmatrix} PAN \\ PC2 \\ PC3 \end{bmatrix} \tag{2.4}$$

The PCA method can be used for more than three bands.

A schematic representation of the IHS and PCA methods is given in Figure 2.7. When the correlation between the I or PC1 component with PAN is not high, the results of these methods are not generally good. The IHS method can handle only three input images while the PCA method can be applied to any number of images. Since all the finer details of PAN will be introduced, the resulting fused image will appear spatially enhanced but less real. Modifications of the IHS and PCA methods involve the injection of the high frequency components of PAN corresponding to the missing high frequencies in the MS image. These modifications are discussed later.

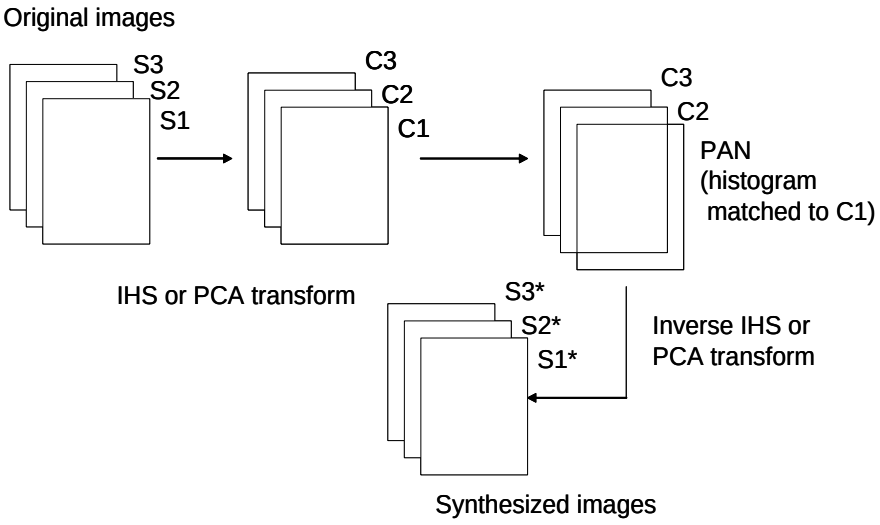


Figure 2.7 Representation of the IHS and PCA methods

2.5.2 Spectral Contribution methods

In these methods, the relationship between the PAN and MS bands are used. This assumes that the spectral range of the PAN image covers the spectral range of the sum of the MS bands' spectral range. Figure 2.8 shows the spectral response curve for the SPOT sensors.

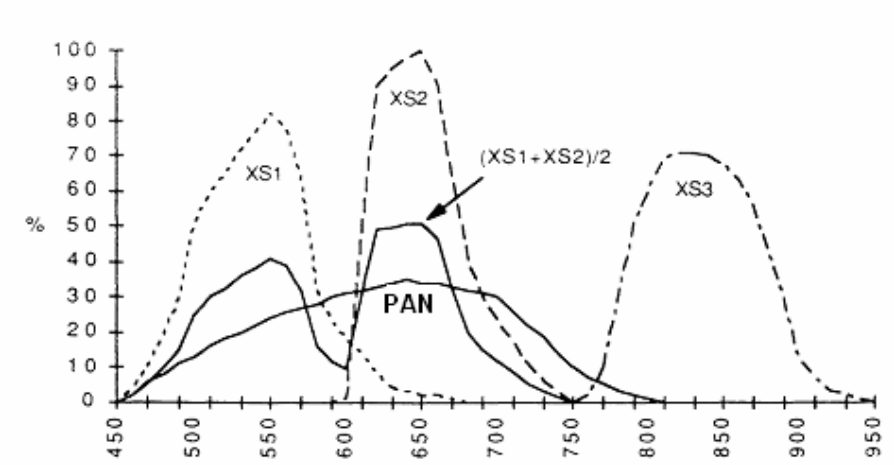


Figure 2.8 Relative spectral responsivity of SPOT sensors

2.5.2.1 SPOT P+XS method

The SPOT P+XS method is specifically designed for SPOT images. The SPOT MS bands are referred by XS1, XS2 and XS3. For earlier SPOT 1, 2 and 3, the panchromatic spectral range covers the XS1 and XS2 bands. The XS3 band is in the NIR part of the electromagnetic spectrum and therefore it is not possible to use the PAN for improving the spatial resolution of XS3. This method is based on the assumption that the half-sum of the radiances in XS1 and XS2 is equal to the radiance in the PAN. The equations for XS1 and XS2 are:

$$XS_{1h}^{*} = 2 * PAN_{h} * \frac{XS_{1l}}{XS_{1l} + XS_{2l}} \tag{2.5}$$

$$XS_{2h}^{*} = 2 * PAN_{h} * \frac{XS_{2l}}{XS_{1l} + XS_{2l}} \tag{2.6}$$

where XS_{1h}^{*} is the fused XS1 band,

XS_{2h}^{*} is the fused XS2 band,

PAN_{h} is the PAN at resolution *h*,

XS_{1l}, XS_{2l} are XS1 and XS2 bands respectively at the spatial resolution *l*.

2.5.2.2 Relative spectral contribution methods

These methods also model the relationship between PAN and MS bands. These models assume that there is high correlation between the PAN and each of the MS bands. Eqn 2.7 represents the Brovey transform. In the following equations, PAN is the PAN image, MS is the original image, MS^{interp} is the interpolated MS image, N is the number of bands, k is the band under consideration, h denotes the high resolution, and l denotes the low resolution. General equations for computing are:

$$MS_{kh}^{*} = \frac{MS^{\text{interp}}{}_{kh} * PAN_{h}}{\sum_{j=1}^{N} MS^{\text{interp}}{}_{jh}} \tag{2.7}$$

$$MS^{"}_{kh} = \frac{MS^{*}_{kh} * m(MS_{kl})}{m(MS^{*}_{kh})} \tag{2.8}$$

where $MS^{"}_{kh}$ is the mean adjusted fused image,

$m(MS_{kl})$ and $m(MS^{*}_{kh})$ are the mean values of the images MS_{kl} and MS^{*}_{kh}.

The "Colour Normalization" method (eqn. 2.9) is a modification of the Brovey transform.

$$MS^{*}_{kh} = \frac{3(MS^{\text{interp}}_{kh} + 1)(PAN_h + 1)}{3 + \sum_{j=1}^{N} MS^{\text{interp}}_{jh}} - 1 \tag{2.9}$$

The Pradines', Price's, Local correlation modeling, Local Mean and Variance matching, and Synthetic Variable Ratio methods are based on similar modeling of the relationship between the PAN and MS images with slight variations.

a) Pradines' method: The model is given by

$$MS^{*}_{kh} = PAN_h * \frac{MS_{kl}}{PAN_l} \tag{2.10}$$

A modification is to apply the model (eqn. 2.11) to the interpolated image.

$$MS^{*}_{kh} = PAN_h MS^{\text{interp}}_{kh} / < PAN_h >_{l/h} \tag{2.11}$$

where $< PAN_h >_{l/h}$ is the average over the window size defined by *h/l*.

b) The Local Correlation Modeling and the Price method: A linear relationship is searched between the moving windows centered on the current pixel at the spatial resolution *l* for both images. The equations are:

$$MS^{*}_{kh} = MS^{\text{interp}}_{kh} + a(PAN_h - (PAN_l)^{\text{interp}}_{h}) \tag{2.12}$$

where the coefficient *a* is computed by linear regression $MS_{kl} = a * PAN_l + b$ at the spatial resolution *l*.

c) The LMVM method (Local Mean and Variance Matching): The mean and variance are adjusted locally over a moving window. The equations are:

$$MS^*_{kh} = (PAN_h - < PAN_h >_s) * \text{stdev}(MS^{\text{int}erp}_{kh}) / \text{stdev}(PAN_h) + < MS^{\text{int}erp}_{kh} >_s \quad (2.13)$$

where $< PAN_h >_s$ and $\text{stdev}(PAN_h)$ are the mean and standard deviation over the window of size s,

$< MS^{\text{int}erp}_{kh} >_s$ and $\text{stdev}(MS^{\text{int}erp}_{kh})$ are the mean and standard deviation over the window of size s.

d) SVR (Synthetic Variable Ratio) Method

The SVR method was proposed by Munechika *et al*. (1993) based on the Pradines (1986) method and Price (1987) methods. The merged MS image is calculated using the equation:

$$MS^*_{kh} = PAN_h * \frac{MS_{kl}}{PAN_{lSyn}} \quad (2.14)$$

The fused image (PAN_{lSyn}) is calculated by:

$$PAN_{lSyn} = \sum_{k=1}^{4} \varphi_i MS_{kl} \quad (2.15)$$

The parameters φ_i are calculated by regression with the values simulated using an atmospheric model that accepts target reflectance and relative spectral responses. Zhang

(1999) simplified the SVR method. The equation for synthesizing the MS image at a higher resolution is given by:

$$MS^{*}_{kh} = PAN_{h} * \frac{MS^{\text{int}erp}_{kh}}{PAN_{hSyn}} \quad (2.16)$$

The PAN_{hSyn} is calculated as in eqn. 2.15 but using multiple regression at resolution *h*.

2.5.3 Frequency Filtering/Modeling methods

The frequency filtering/modeling methods use high pass filters, Fourier transform or wavelet transform to model the frequency components between the PAN and MS images. They are based on the assumption that the difference between PAN and MS is only the lack of high frequencies in MS image that are present in PAN. High frequencies correspond to the spatial details (edges, small details) in the images. As mentioned earlier, PAN has a better spatial resolution and hence it has more high frequency information compared to the MS image. Thus, these methods aim at modeling the frequency components and introducing them into the MS image.

Chavez *et al*. (1991) introduced the HPF (High Pass Filtering) method for PAN and MS fusion. In HPF, the high frequencies in PAN are extracted using high pass filters. The extracted high frequencies are then introduced into one band of the MS image by simple addition. The high frequency is introduced equally without taking into account the relationship between the MS and PAN images.

High pass filtering forms the basis for many of the wavelet-based image fusion methods. Several wavelet transforms such as Haar, Daubechies and à-trous wavelets have been used for image fusion. In the next section, a brief introduction to the wavelet transform is given.

2.5.3.1 Wavelet Transform

Any signal or image can be decomposed into several components representing different frequencies for a better analysis, description and for further processing. Wavelet transform (WT) provides a good localisation in both frequency and space. Wavelet transforms allow a decomposition of the signal (also called analysis) as well as a perfect reconstruction of the signal (also called synthesis). This property is highly useful in image fusion. The analysis and synthesis are explained using multiresolution analysis within a filter bank structure.

The wavelet transform of a continuous 1D function *f (t)* can be expressed as

$$WT(f)(a,b) = |a|^{-1/2} \int_{-\infty}^{+\infty} f(t)\psi\left(\frac{t-b}{a}\right)dt \qquad (2.17)$$

where *a* and *b* are the scaling and translation parameters, respectively. Each base function $\psi\left(\frac{t-b}{a}\right)$ is a scaled and translated version of a function $\psi(t)$ called mother wavelet. With the different scaled versions of the mother wavelet, it is possible to analyze the signal at different scales. This is referred to as the multiresolution analysis. Figure 2.9 shows a good representation of a multiresolution (or multiscale) analysis.

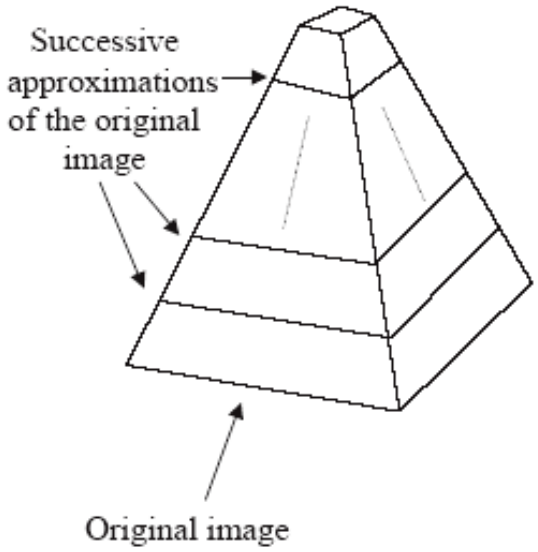


Figure 2.9 Multiresolution Analysis

Multiresolution analysis can be performed either using a Generalised Laplacian Pyramids (GLP) or using wavelet transform (with or without decimation). In Figure 2.9, the original image forms the base of the pyramid and the successive approximations are placed at the top of the original image giving rise to pyramidal structure. As we go up the pyramid, the approximation images have coarser and coarser spatial resolution. The difference between two successive approximations constitutes the detail images or the wavelet coefficients. The original images can be reconstructed from the final approximation and all the detail images if the process of multiresolution analysis is inverted. This is the synthesis property of wavelets. Of the many discrete wavelet transforms, the most common implementations in image fusion are the Mallat's algorithm and the à-trous algorithm. Figure 2.10 shows an implementation of the Mallat algorithm using a filter bank structure. The filter bank structure consists of a high pass filter G and a low pass filter H. In the first level (*j+1*) of the analysis, the original image $f_j(x,y)$ is decomposed into an approximate image $f_{j+1}(x,y)$, horizontal $CH_{j+1}(x,y)$, vertical $CV_{j+1}(x,y)$ and diagonal $CD_{j+1}(x,y)$ details by successively applying H and G filters. $\downarrow 2$ denotes sub-sampling the image by a factor of 2 that gives rise to the pyramidal structure. In the second level (*j+2*) of the analysis, $f_{j+1}(x,y)$ is decomposed

into $f_{j+2}(x,y)$, $CH_{j+2}(x,y)$, $CV_{j+2}(x,y)$ and $CD_{j+2}(x,y)$. At each level, the size of the image is reduced by half resulting in a pyramidal structure. For synthesizing the original image from the final approximate and all the detail images, the complementary filters $\overline{H}$ and $\overline{G}$ of *H* and *G* respectively, are used. The filters are applied as shown in Figure 2.10.

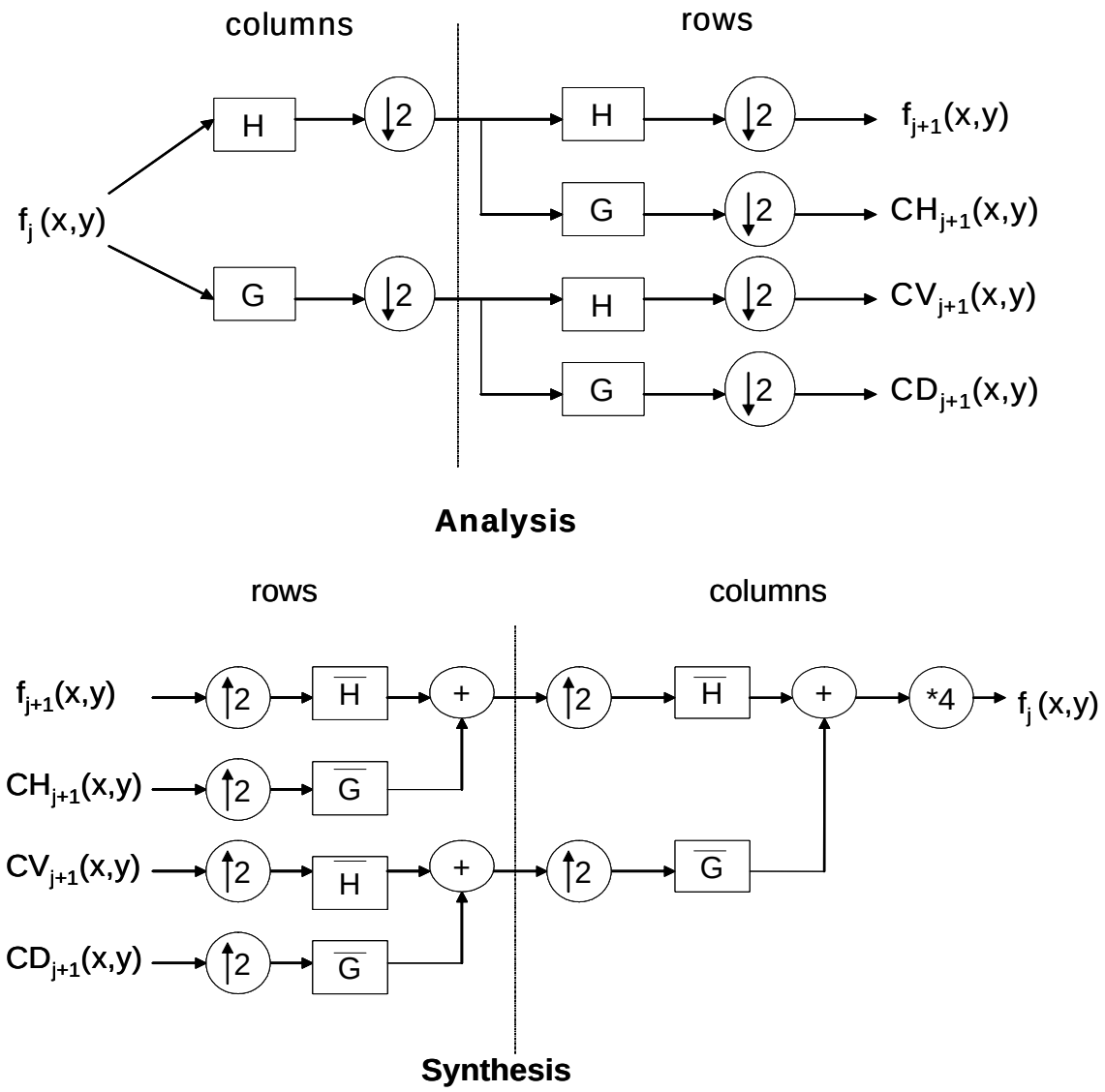


Figure 2.10 Filter bank structure for implementing the Mallat algorithm

The filters H, G, $\overline{H}$ and $\overline{G}$ can be designed using the Daubechies wavelet coefficients for example. The four Daubechies wavelet coefficients are given in Table 2.1. The filters H, G, $\overline{H}$ and $\overline{G}$ are shown in Table 2.2.

Table 2.1 Daubechies filter coefficients

H(0)	***H(1)***	***H(2)***	***H(3)***
0.482962913145	0.836516303738	0.224143868042	-0.129409522551

The coefficients are divided by $\sqrt{2}$ for normalization. The filter H is a low pass filter and G is a high pass filter.

Table 2.2 Filter masks for analysis and synthesis filters

Filter	**Filter Coefficients**			
H	H(3)	H(2)	H(1)	H(0)
G	-H(0)	H(1)	-H(2)	H(3)
$\overline{\text{H}}$	H(0)	H(1)	H(2)	H(3)
$\overline{\text{G}}$	H(3)	-H(2)	H(1)	-H(0)

An example of WT analysis and synthesis using the Daubechies wavelet coefficients is presented in Figure 2.11. The original image has been decomposed to approximate and detail images at the first level.

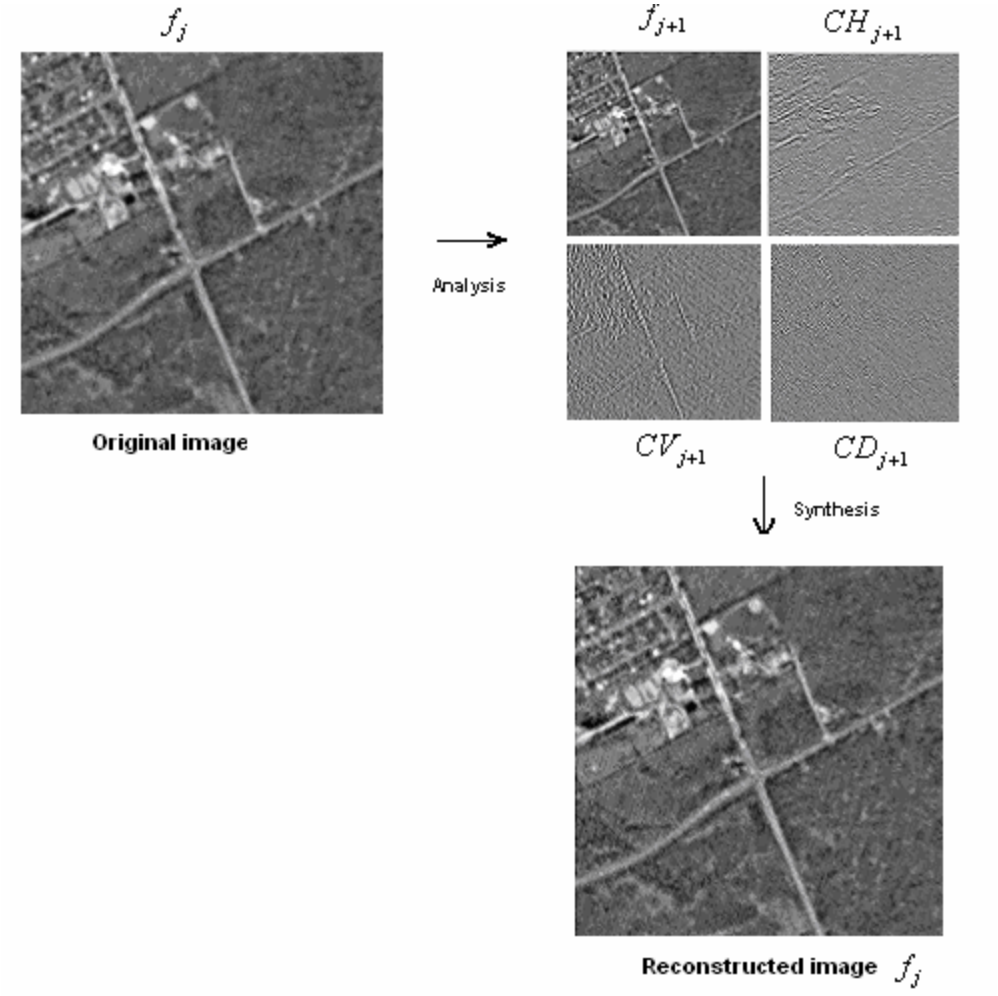


Figure 2.11 WT analysis and synthesis using a Daubechies wavelet

The "à trous" wavelet transform is a non-orthogonal, shift-invariant, symmetric, dyadic, undecimated, discrete redundant wavelet transform (Dutilleux, 1987). The sampled data at each level are the scalar products of the function $f(x)$ with the scaling function $\phi(x)$ which corresponds to a lowpass filter. The scaling function can be a triangle scaling function or a bicubic spline scaling function. The wavelet coefficients are usually calculated as the difference between two consecutive approximations.

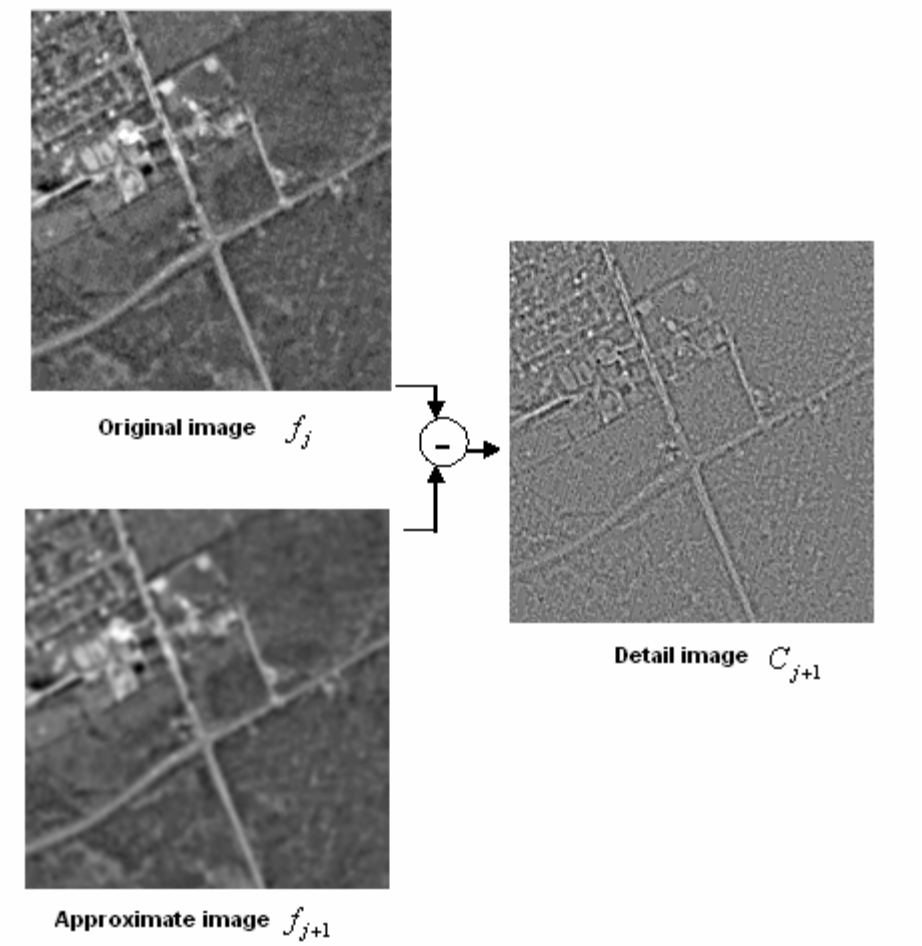


Figure 2.12 Approximate and detail images using the à-trous algorithm

The mask for the B_3-Spline Scaling function is (1/16, 1/4, 3/8, 1/4, 1/16) in one dimension (1D). This mask can be extended to two dimensions (2D) by assuming separability and applying the 1D mask along row and then along column of image or by applying the 2D mask given by:

$$\begin{pmatrix} 1/256 & 1/64 & 3/128 & 1/64 & 1/256 \\ 1/64 & 1/16 & 3/32 & 1/16 & 1/64 \\ 3/128 & 3/32 & 9/64 & 3/32 & 3/128 \\ 1/64 & 1/16 & 3/32 & 1/16 & 1/64 \\ 1/256 & 1/64 & 3/128 & 1/64 & 1/256 \end{pmatrix} \tag{2.18}$$

This algorithm produces a band stack of approximate images that have the same size as the original image (*i.e.* no decimation). At each level of decomposition, one approximate and one detail image is produced. In this case, the detail image is isotropic. An example of the decomposition using the à-trous algorithm is presented in Figure 2.12. The original image represents the image at level *j* and the B_3-spline function is applied to obtain the approximate image at level *j+1*. The detail image at *j+1* is the pixel-based difference between the image at level *j* and the approximate image at *j+1*.

2.5.3.2 IHS+W and PCA+W

Modifications of the IHS method were proposed by Nuñez *et al*. (1999). Instead of replacing the I by PAN, the high frequency components from PAN are modeled using the à-trous with a B_3-Spline scaling function and injected into the I component resulting in a new intensity I^*. Then the I^*, *H* and *S* components are retransformed into the RGB colour space. This way, the dominance of PAN in the fused MS images is substantially reduced resulting in a better spectral preservation. This combined use of wavelet and IHS is referred in the thesis as the IHS+W method. On the same principle, González-Audícana *et al*. (2004) proposed a Mallat's undecimated algorithm for improving the IHS and PCA methods. The undecimated Mallat algorithm uses the same filter bank structure shown in Figure 2.10 but without the sub-sampling by 2. The PCA+W is used in the thesis to refer to the PCA method in combination with wavelet. In the PCA+W method, the high frequency components of PAN are introduced into the PC1 component to obtain a new *PC1** and then an inverse PCA is carried out with the new *PC1**. Both the à-trous and Mallat undecimated algorithm with the Daubechies wavelet coefficients were used in IHS+W for one data set. But only the results of the à-trous algorithm have been used for comparison in the chapter 4. Some results of the Mallat undecimated algorithm are presented in the Appendix 1.

Figure 2.13 illustrates the wavelet analysis and synthesis with decimation in PAN and MS image fusion. In IHS+W and PCA+W methods, the I or the PC1 component are decomposed in a similar way and a new *I** or *PC1** are synthesized.

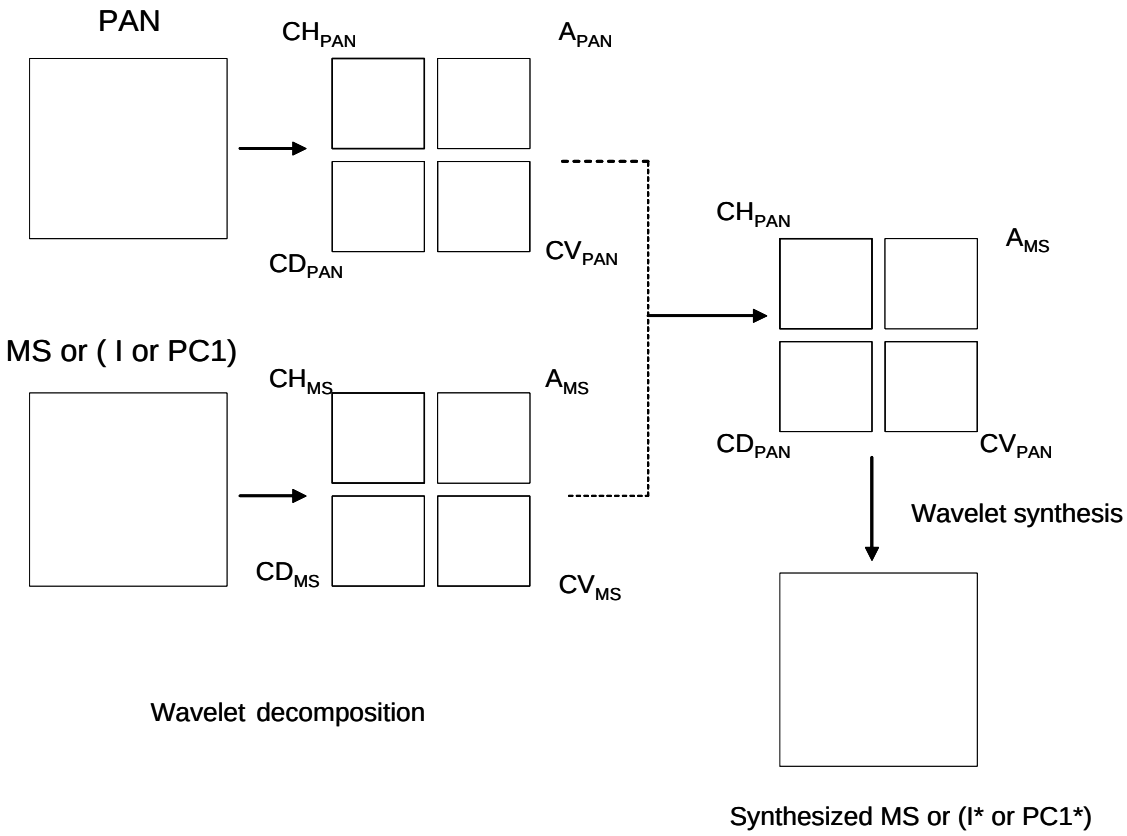


Figure 2.13 Diagram to illustrate wavelet transform in image fusion

2.5.3.3 ARSIS Concept

The ARSIS concept is based on wavelet transform and multiresolution analysis. It was also developed on the assumption that the difference between the PAN and MS images is the lack of the high frequency components in the MS image. It tries to model these missing frequencies using a multiresolution analysis to synthesize the image at a higher resolution (Ranchin *et al*., 2000).

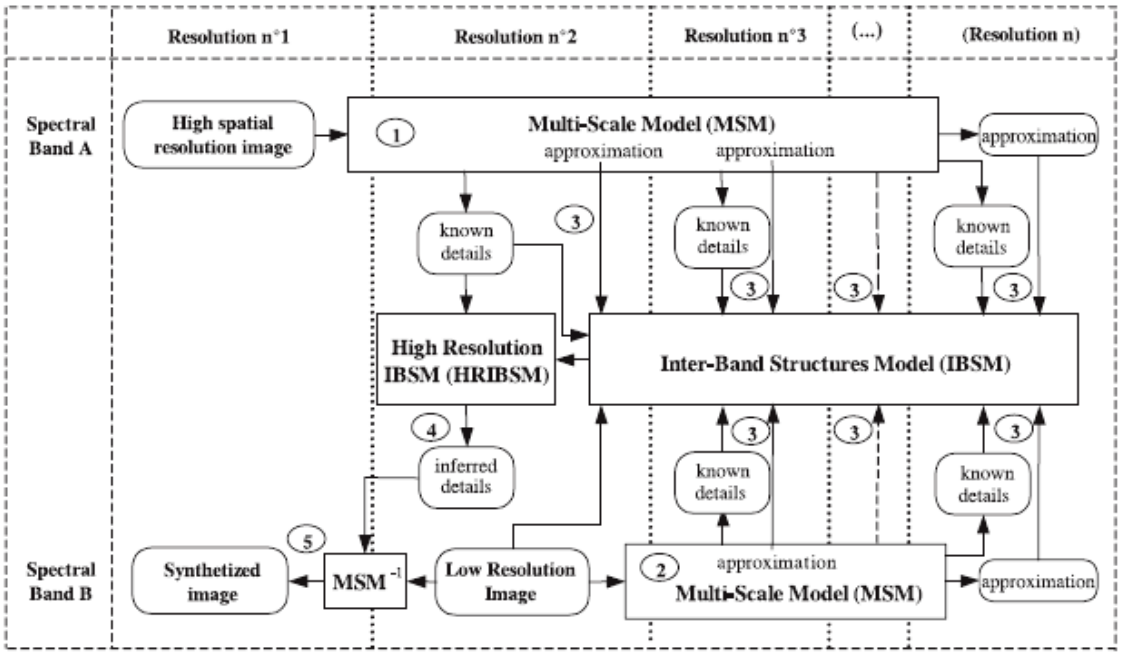


Figure 2.14 General scheme for ARSIS concept

(Source: Ranchin *et al*., 2000)

Figure 2.14 illustrates a general scheme for the ARSIS concept. There are three models in the ARSIS scheme: MSM (MultiScale Model), IBSM (Inter-Band Structure Model) and HRIBSM (High Resolution Inter-Band Structure Model). The MSM model describes the spatial structures in an image at different spatial resolutions. This can be either implemented using the Mallat algorithm combined with a Daubechies wavelet or the à-trous algorithm. The IBSM model describes the relationship between the spatial structures with change in spectral bands. The HRIBSM model is the IBSM model which models the transformation with change in spatial resolution. The HRIBSM model is identical as the IBSM model as the relationship between the spatial resolutions of different modalities is not known exactly.

The general scheme (Figure 2.14) consists of the following steps:

1. Multiresolution analysis of the PAN and MS image using MSM to obtain approximate and detail images at different spatial resolutions.

2. Modeling the relationship between known detail images of PAN and MS at each spatial resolution using the IBSM model.
3. Inferring the missing high frequency information in the MS image (HRIBSM model) from the known modeled relationship.
4. Inverting the MSM model of MS image taking into account the transformation parameters computed in the IBSM and HRIBSM models to synthesize the MS image at a higher spatial resolution, *i.e.* the one of PAN.

There are three IBSM models proposed by Ranchin et *al.* (2000). The simplest model is the Identity model (Model 1) where the detail images of MS image are assumed to be equal to PAN. The wavelet addition (WA) and wavelet substitution (WS) methods proposed by Nuñez *et al*. (1999) can be categorized as a Model 1 technique. PAN and MS images are decomposed into approximate and detail images at different levels say n = 1 to 4 represented in the following equations:

$$PAN = \sum_{i=1}^{n} w_{PAN} + A_{PAN} \tag{2.19}$$

$$MS_k = \sum_{i=1}^{n} w_{MS_k} + A_{MS_k} \tag{2.20}$$

where k refers to the multispectral band under consideration,
w refers to the detail images and
A refers to the approximate image.

In the wavelet ***addition method*** (WA), the detail images of the PAN image are directly added to the MS image (eqn. 2.21). The detail images of one or two levels are usually added. In the wavelet ***substitution method*** (WS) (Figure 2.13), both the PAN and MS image are decomposed into approximate and detail images. Depending on the level of decomposition, the detail images of MS images are substituted by the corresponding detail images of PAN (eqn 2.22).

$$MS_{kh}^{*} = MS_{kh}^{\text{int}erp} + \sum_{i=1}^{n} w_{PAN} \tag{2.21}$$

$$MS_{kh}^{*} = A_{MS_k} + \sum_{i=1}^{n} w_{PAN} \tag{2.22}$$

Model 2 (M2) is based on mean and variance adjustments between the details of PAN and MS. In this model, the missing detail is given by:

$$w_{MS(h-l)} = a * w_{PAN(h-l)} + b \tag{2.23}$$

where $w_{MS(h-l)}$ is the detail of MS between the scales *h* and *l*,

$w_{PAN(h-l)}$ is the detail of PAN between the scales *h* and *l*,

$$a = \sqrt{\frac{\text{variance}(w_{MS(ph-pl)})}{\text{variance}(w_{PAN(ph-pl)})}}$$

$$b = \text{mean}(w_{MS(ph-pl)}) - a * \text{mean}(w_{PAN(ph-pl)})$$

p is the ratio of two successive scales in the multiscale model.

Model 3 (M3) is different from Model 2 in the computation of the coefficients *a* and *b*. The coefficients are calculated based on least square adjustments and axes of inertia. The RWM model is another IBSM model named after the authors Ranchin, Wald and Mangolini. More details on these models can be found in (Ranchin *et al*., 2000; Ranchin *et al*., 2003).

2.5.4 General Remarks

The IHS and PCA methods only aim at increasing the spatial resolution of the images. These methods generally provide good results when the PAN spectral range covers the spectral range of MS images. In other words, they perform well when the PAN is highly correlated with the MS images. The IHS+W, WA and WS methods depend on the correlation between the high frequency components of PAN and MS images. But they do not try to model the frequency relationship between PAN and each MS band. Instead, the same amount of high frequency component is introduced into all the bands of the MS image irrespective of the wavelength bandwidth of each band of the MS image. The

ARSIS M2 and M3 models try to model the relationship between the high frequency components of PAN and MS images. These models have been shown to produce better results for many different data sets from different sensors. The ARSIS concept has a good theoretical framework to synthesize high resolution MS images that would be obtained if a MS sensor at high resolution existed. But, the synthesis at high resolution depends on the sensor characteristics of both the PAN and MS sensors.

Figure 2.15 and Figure 2.16 show the spectral response curves for the Ikonos and Quickbird sensors respectively. It can be seen that the spectral response of the PAN sensor is not uniform in the entire wavelength. As we can see in Figure 2.15, the spectral response of Ikonos PAN is very low in the Blue band and maximum in the Green, Red and NIR bands. The PAN spectral response curve extends beyond 0.90 μm. Similarly, the Quickbird PAN sensor has low spectral response in the Blue band, and maximum in the Green-Red bands (Figure 2.16). Even though the spectral ranges of the PAN sensors are provided as 0.45-0.90 μm, the spectral sensitivity is not uniform over the MS bands. Thus, the fusion methods may encounter problems in the Blue and NIR bands.

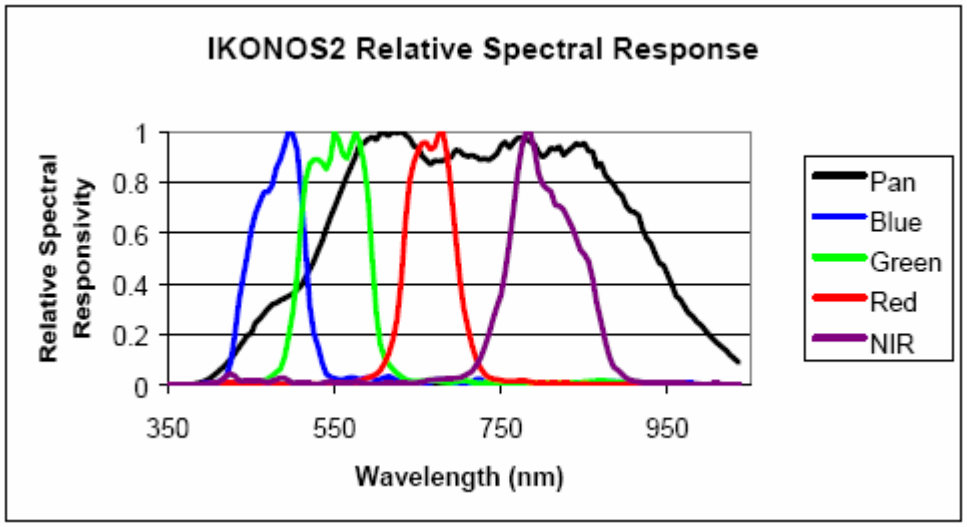


Figure 2.15 Ikonos spectral response

(Source: Space Imaging)

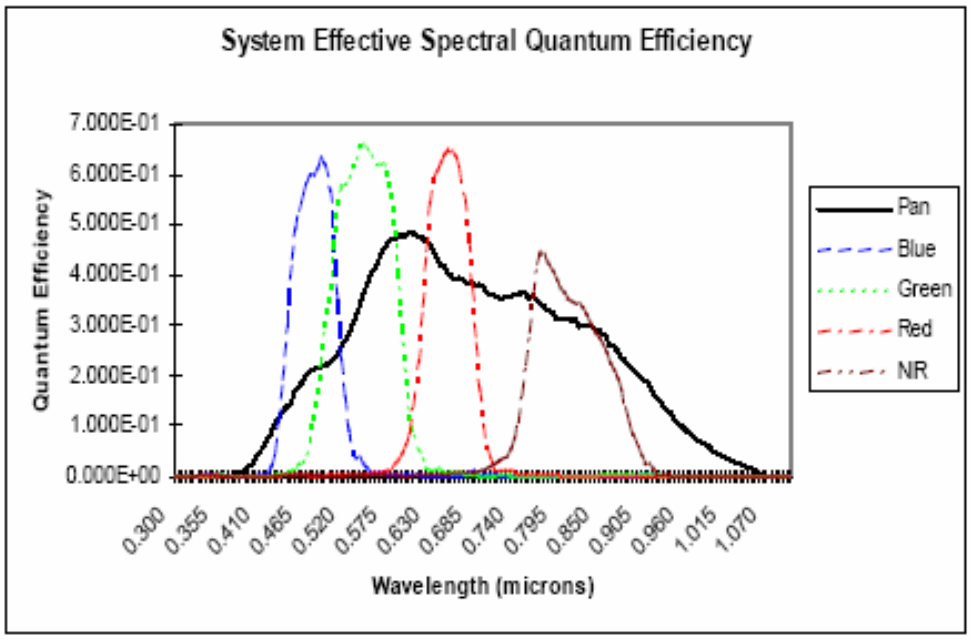


Figure 2.16 Quickbird spectral response

(Source: NASA Library)

2.6 Summary

In this chapter, concepts of data fusion and some pixel based fusion methods were discussed. Any pixel-based fusion method modifies the spectral values of the original MS images. Several applications like photo-interpretation and classification depend on the spectra of objects and high error in synthesis may result in inaccurate mapping. Therefore quality assessment is essential in image fusion. Some of the existing quality measures are discussed in the next chapter.

Chapter 3 Quality Assessment

3.1 Introduction

Quality refers to both the spatial and spectral quality of images. Image fusion methods aim at increasing the spatial resolution of the MS images while preserving their original spectral content. Spectral content is very important for applications such as photo interpretation and classification that depend on the spectra of objects. The lack of standard methods and tools for assessment has led to poor knowledge about the fusion methods and their suitability for different data sets and landscapes. Several efforts have been taken to frame a standard protocol for evaluating quality. This chapter discusses the properties of the fused images, the limitations in quality assessment and some quantitative criteria for the quality assessment of fused images.

3.2 Properties of Fused Images

As formulated by Wald (1997), the properties of fused images are:

1. Any fused image once downsampled to its original spatial resolution should be as identical as possible to the original image.
2. Any fused image should be as identical as possible to the image that a corresponding sensor would observe with the same high spatial resolution.
3. The MS set of fused images should be as identical as possible to the MS set of images that a corresponding sensor would observe with the same high spatial resolution.

These three properties have been reduced to two properties: consistency property and synthesis property (Thomas & Wald, 2004). The consistency property is same as the first property and the synthesis property combines the second and third properties defined by Wald (1997). The synthesis property emphasizes the synthesis at an actual higher spatial

and spectral resolution. These properties cannot be tested directly due to the lack of reference images at the higher spatial resolution.

3.3 Reference Images

Reference MS images at a higher spatial resolution are not available for assessing the quality of the fused images. The only available reference images are the original MS images at the "low" spatial resolution. Wald *et al.* (1997) proposed a protocol for quality assessment and several quantitative measures for testing the three properties. The consistency property is verified by downsampling the fused image at the higher spatial resolution *h* to their original spatial resolution *l* using suitable filters such as bi-cubic spline. The synthesis properties of fused images need a reference image. Since they are not available, the original PAN at resolution *h* and MS at resolution *l* are downsampled to their lower resolutions *l* and *v* respectively. Then, PAN at resolution *l* and MS at resolution *v* are fused to obtain fused MS at resolution *l* that can be then compared with the original MS image. The quality assessed at resolution *l* is assumed to be close to the quality at resolution *h*. This reduces the problem of reference images. However, we cannot predict the quality at higher resolution from the quality of lower resolution (Wald *et al*., 2002). The quality at higher resolution can be better or worse depending on the high frequencies introduced and it is difficult to predict the variability of the quality with respect to the spatial resolutions.

Although neither the consistency property nor the synthesis property can provide the true quality of the fused images, they can be used to infer the quality of the fused images to some extent. The consistency property used for testing the quality of fused images in preserving the original spectral content seems reasonable compared to the actual synthesis. Nevertheless, the conclusions might be similar for both approaches.

3.4 Quality Assessment

3.4.1 Visual Quality

The objective of fusion is to increase the spatial resolution of MS images. Therefore, visual analysis is a necessity to check if the objective of fusion has been met. The general visual quality parameters are: image quality (geometric shape, size of objects), spatial details and local contrast. Other visual quality parameters for testing the properties are:

1. Spectral preservation of features in each multispectral band: Based on the appearance (high or low spectral values) of objects in the original MS images, the appearance of the same objects in the fused images are analysed in each band.
2. Multispectral synthesis in fused images: Fusion should not distort the original spectral characteristic of objects. The multispectral characteristics of objects at higher spatial resolution should be similar to that in the original images. Analysing different colour composites of the fused images and comparing them with that of original images can help verifying this property.
3. Synthesis of images close to actual images at high resolution as defined by the synthesis property of fused images: This property cannot be actually verified but can be analysed from our knowledge of spectra of objects in the lower spatial resolutions.

3.4.2 Statistical Quality

Some measures often used to evaluate the quality of fused images are presented in this section. MS_{kl} refers to the multispectral band *k* at the spatial resolution *l*, $(MS^*_{kh})_l$ is the downsampled fused image at resolution *l* and MS^*_{kl} refers to the fused image created at resolution *l*. For simplicity, MS^*_{kl} is used in the following equations. This is applicable for the synthesis property. For the consistency property, it is replaced by $(MS^*_{kh})_l$.

Bias, standard deviation of the difference image (SDD), difference in variance (DIV) and correlation coefficient (CC) are criteria that are used to verify both the consistency and synthesis (second property) properties of fused images. Some criteria for multispectral quality includes computing the correlation coefficient between the fused images with PAN; the correlation coefficient between the fused images in different bands; and the frequency of pixels in each dominant spectra in the image and classification.

Let $\overline{MS_{kl}}$ and $\overline{MS_{kl}^{*}}$ be the mean of MS_{kl} and MS_{kl}^{*} respectively and $\overline{\overline{MS_{kl}}}$ and $\overline{\overline{MS_{kl}^{*}}}$ the variance of MS_{kl} and MS_{kl}^{*} respectively.

1. **Bias** is the difference between the means of the original image and of the fused image. The value is given relative to the mean value of the original image. The ideal value is zero.

$$Bias = \frac{\overline{MS_{kl}} - \overline{MS_{kl}^{*}}}{\overline{MS_{kl}}} = 1 - \frac{\overline{MS_{kl}^{*}}}{\overline{MS_{kl}}} \tag{3.1}$$

2. **Difference in Variance** (DIV) is the difference between the variances of the original image and of the fused image. It indicates the amount of information added or lost during fusion. A positive value indicates a loss of information and a negative value some added information. It is given relative to the variance of the original image. The ideal value is zero.

$$\text{DIV} = \frac{\overline{\overline{MS_{kl}}} - \overline{\overline{MS_{kl}^{*}}}}{\overline{\overline{MS_{kl}}}} = 1 - \frac{\overline{\overline{MS_{kl}^{*}}}}{\overline{\overline{MS_{kl}}}} \tag{3.2}$$

3. **Correlation Coefficient** (CC) measures the correlation between the original and the fused images. The higher the correlation between the fused and the original images, the better the estimation of the spectral values. The ideal value of correlation coefficient is 1.

$$\mathrm{CC} = \frac{\sum (MS_{kl} - \overline{MS_{kl}})(MS_{kl}^{*} - \overline{MS_{kl}^{*}})}{\sum (MS_{kl} - \overline{MS_{kl}}) \sum (MS_{kl}^{*} - \overline{MS_{kl}^{*}})} \tag{3.3}$$

4. **Standard deviation** is the standard deviation of the difference image (SDD) relative to the mean of the original image. It indicates the closeness of the fused image to the original image at a pixel level. The ideal value is zero.

$$\mathrm{SDD} = \frac{\text{Standard deviation}(MS_{kl} - MS_{kl}^{*})}{\overline{MS_{kl}}} \tag{3.4}$$

5. **Correlation between different bands**: The correlations between the fused image and PAN (MS_{kl}^{*} and PAN) and between the fused bands are computed. This quantity indicates the correlation between the different bands. A correlation coefficient close to that in the original images indicates a preservation of the multispectral integrity between the two bands under consideration.

6. **Number and frequency of spectra:** The number of different spectra in the original MS image and in the fused MS image is calculated using the pixel vector composed of all the MS bands. However, even if the number of spectra is identical, this measure does not ensure that the spectra are identical in both images as they should. The other criterion is the number of occurrences of each spectra in the original and fused images. This can accurately describe the multispectral synthesis of the fusion method, but does not provide any spatial information about the occurrences of pixels. Thus, the results obtained may be misleading.

7. **Classification:** Fused images are often classified to obtain land-use maps. Classification is similar to the frequency of each spectrum but only the most dominant spectra are considered in classification. Classification results of fused images can be analysed to understand the effect of spectral error(s) in the fused images relative to different fusion methods. Standard deviation and mean differences between different objects in original and fused images has been used

to characterize spectral distortion in different objects (Terrattaz, 1997; Weber, 2003). Weber (2003) analysed the mean for different objects in original and fused images to characterize the spectral distortion in those objects. An unsupervised classification is used to classify the original image into clusters. The mean, standard deviation and root mean square error (RMSE) are calculated between the original and fused image for the pixels in each cluster. Clusters (Pixels) of only the dominant spectra are considered.

The RMSE is given by:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(MS_{kl}(i) - MS_{kl}^{*}(i)\right)^2}{n}} \tag{3.5}$$

where n is the number of pixels,

MS_k is the original image for the band k and

MS_k^* is the fused image in band k.

8. **ERGAS** (from the French acronym "Erreur Relative Globale Adimensionnelle de Synthèse") is a simplified quantity proposed by Wald (2002) that summarizes the error(s) in all the bands. The lower the error(s) in the bands, the better the quality of the fused images.

   It is given by:

$$ERGAS = 100\frac{h}{l}\sqrt{\frac{1}{N}\sum_{k=1}^{N}\left[RMSE(MS_k)^2 \Big/ \overline{MS_{kl}}^2\right]} \tag{3.6}$$

   where RMSE is calculated using eqn. 3.5 using every pixel of the image,

   N is the number of bands,

   h/l is the ratio of spatial resolutions of original PAN and MS images.

3.5 Related Work

Much literature is available for the fusion of SPOT and Landsat TM images. The ARSIS models have been tested on SPOT P and XS images, SPOT XS and KVR-1000 image, Landsat TM bands, and SPOT P and Landsat TM bands (Ranchin *et al*., 1997; Terretaz, 1997) and they generally provide better statistics and preserve multispectral content.

Only very little research has been done using VHR images. The ARSIS M2 and M3, PCA with wavelet (PCA+W), IHS (IHS+W) with wavelet have been shown to provide better results for Quickbird images compared to standard IHS and PCA fusion methods (Reyes *et al*., 2004). Weber *et al*. (2003) presented better results with the UWT-M2 method that uses the ARSIS concept compared to a method based on correlation. The UWT-M2 method uses undecimated wavelet transform (the à-trous algorithm) for MSM and ARSIS M2 for IBSM. Among the different combinations of MSM and IBSM models in the ARSIS concept, GLP-AABP (Gaussian Laplacian Pyramid (MSM model) - AABP (IBSM model) named after the authors Aiazzi, Alparone, Baronti and Pippi)) provided better results for the fusion of Ikonos PAN and MS (Ranchin *et al*., 2003). Other methods based on least squares (Zhang, 2002) and IHS combined with Fourier filtering (Ehlers, 2005) are shown to better preserve multispectral content in fusion of VHR images.

3.6 Conclusion

Based on several experiments, statistical measures - bias, standard deviation of difference image, difference in variance and correlation coefficient - are found to be suitable for evaluating the quality of fused images. The numbers of spectra and of occurrences of the spectra are not used in this research because of the high radiometric resolution (dynamic range) of the VHR images. The spatial and multispectral quality of fused images is often assessed by visual analysis. Results of visual quality assessment may differ depending on the human perception of quality and it will also differ depending on the application (Cornet and Binard, 2004).

Chapter 4 Results and Analysis

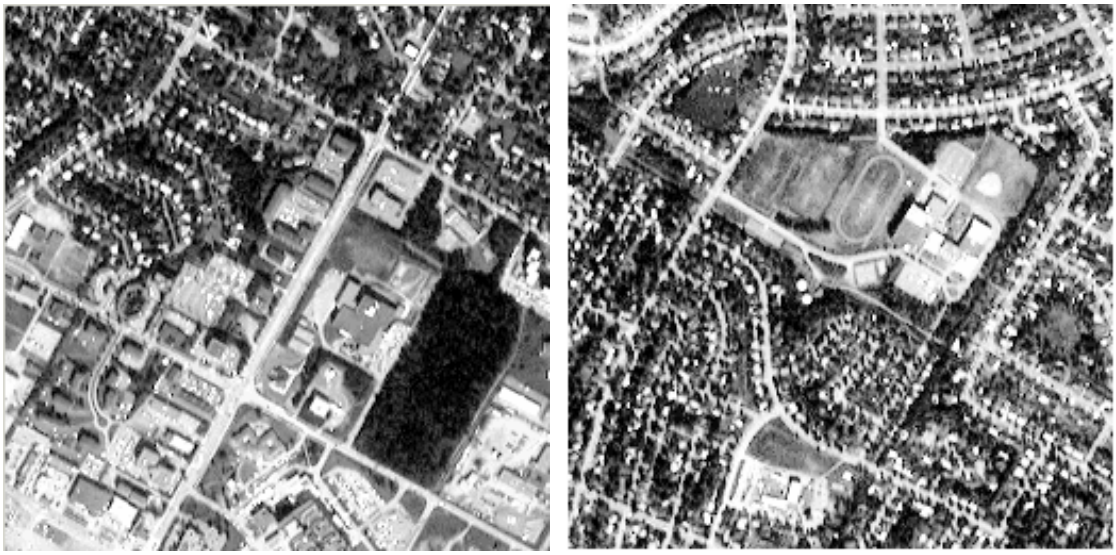
In this chapter, the fused images obtained by different methods are evaluated based on the quality measures described in chapter 3. First, the data sets and the methods used in the study are presented. Then, the quality of the fused images obtained for each data set is analyzed visually and statistically. Finally, some general discussions and conclusions are drawn based on the analyses. Classifications of the fused images are discussed in the next chapter.

4.1 Data

Our study consists of ortho-rectified subset images of the City of Fredericton, New Brunswick, Canada acquired by the Ikonos (IK) satellite of Space Imaging Inc. in October 2001 and by the Quickbird (QB) satellite of Digital Globe in August 2002.

The IK PAN and MS images are provided at a spatial resolution of 1 and 4 m respectively. The QB PAN image is provided at a spatial resolution of 0.7 m and MS images at 2.8 m. The IK and QB PAN images are acquired in the spectral range of 0.45 to 0.90 μm and the MS images are acquired in four spectral ranges of (Blue) 0.45-0.52μm, (Green) 0.52-0.60μm, (Red) 0.63-0.0.69μm, and (NIR) 0.76-0.90μm.

Two Ikonos data sets and two Quickbird data sets were selected for the experiments. The data sets under consideration consist of urban and suburban areas. The test sites of the data sets are shown in Figure 4.1 and Figure 4.2. The MS images are of size 256 x 256 pixels and PAN images are of size 1024 x 1024 pixels.



(a) Ikonos dataset 1 (IK D1) (b) Ikonos dataset 2 (IK D2)

Figure 4.1 Image showing study areas in Ikonos data sets

(a) Quickbird data set 1 (QB D1) (b) Quickbird data set 2 (QB D2)

Figure 4.2 Image showing study areas in Quickbird data sets

The study area of Ikonos dataset 1 (IK D1) shown in Figure 4.1a consists of a complex residential area with houses, streets and dense vegetation in its upper right portion and an industrial area with large buildings, parking lots, and vehicles in its lower portion.

The test area of Ikonos dataset 2 (IK D2) shown in Figure 4.1b consists of a sub-urban area with small houses and streets.

The test areas for Quickbird dataset 1 (QB D1) (Figure 4.2 a) and data set 2 (QB D2) (Figure 4.2 b) consists of complex residential and industrial areas. The study area of QB D2 is identical as that of IK D1.

4.2 Methods

Bicubic resampling technique is used for resampling the MS images to the size of PAN and also for downsampling the MS images in quality assessment. The following methods were selected for evaluation:

1. IHS
2. PCA
3. IHS+W
4. PCA+W
5. Wavelet Addition (WA)
6. Wavelet Substitution (WS)
7. ARSIS M2 model

The IHS method was carried out for different combinations of bands. The ARSIS models were only available for IK D1 and QB D1. The ARSIS fused images of different IBSM models using different filter sizes in the MSM model were provided by Dr. Ranchin (Ecole des Mine de Paris, France). The statistics computed for all these images are presented in Appendix 1. The M2 model (called the M2 method) using a filter size of 3x3 for the MSM model was selected for comparison with other methods as the fused images of other models do not have good geometric quality.

4.3 Results of IK D1

The subsets of the fused images obtained by the different methods are shown in Figure 4.3 for the Blue band. Figure 4a shows the original image. Figure 4.3b, c, d, e, f, g and h show the fused images by the IHS, PCA, IHS+W, PCA+W, WA, WS and M2 methods respectively. The increase in spatial details (edges and small objects) is clearly visible in the fused images.

4.3.1 Visual Quality

The IHS image in Blue band shows high spatial details (Figure 4.3b). The fused image appears brighter than the original one and the vegetated areas appear coarser. The PCA fused image (Figure 4.3c) is slightly brighter compared to the original image and it has less spatial details than the IHS one. Visually, the fused images of IHS+W, WA and M2 (Figure 4.3d, e and f) appear to have less sharp spatial details compared to IHS. Similar to the Blue band, the fused images in Green and Red bands also have small differences in the sharpness of the spatial details introduced. The differences can be seen in the WA and M2 fused subset in Green band and of WS and M2 in Red band presented in Figure 4.4 and Figure 4.5 respectively. Edges, small objects and spatial structures in buildings appear sharper in the fused NIR images by PCA and M2 compared to IHS and IHS+W (Figure 4.6).

All the methods have introduced spatial details but the degree of sharpness varies in the fused images. When the spatial details (edges of objects) are less sharp, it results in bright areas surrounding the objects and blurry edges. Visually, all the methods seem to have preserved the relative spectra of the features in the individual bands.

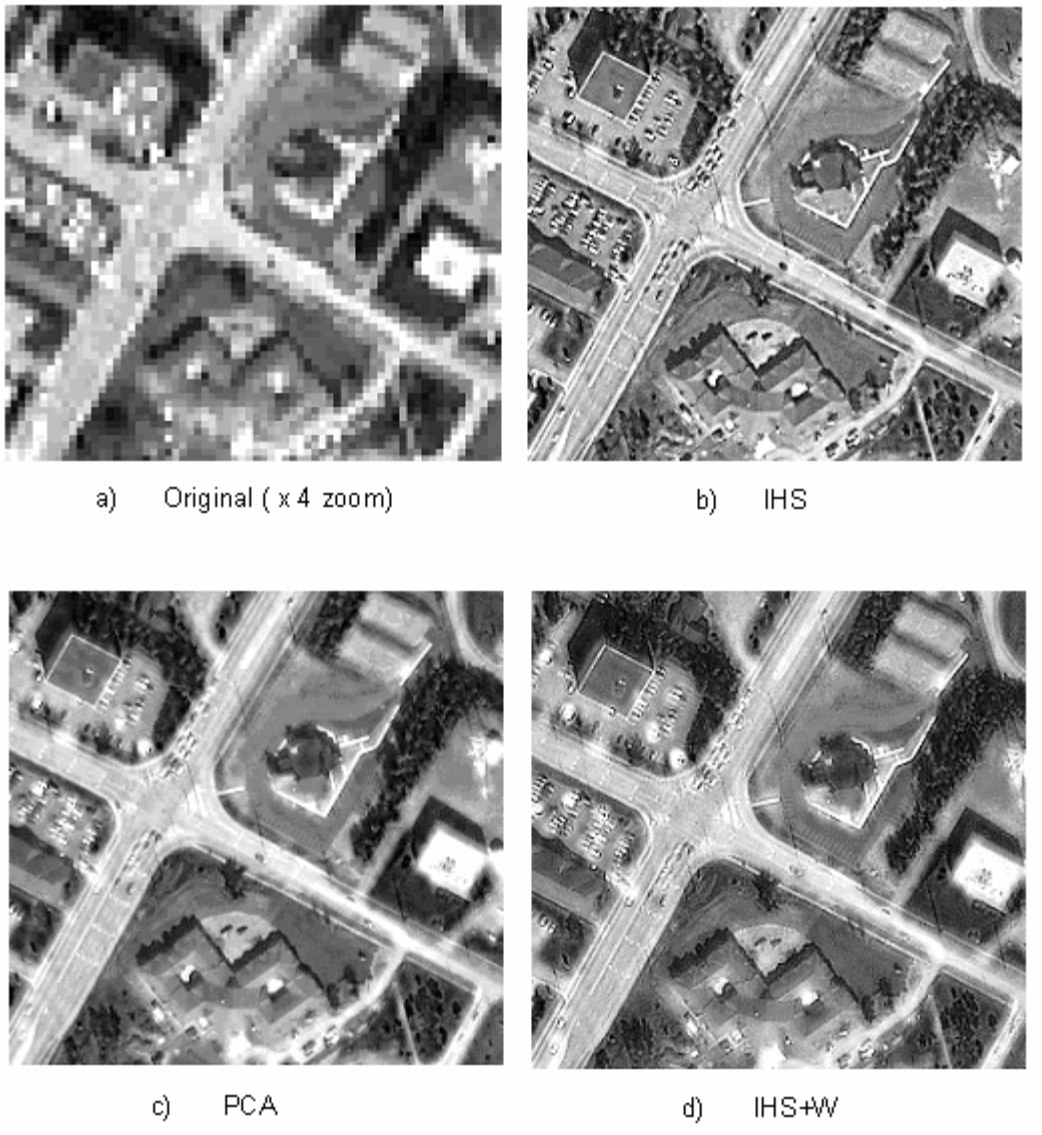


Figure 4.3 Fused images in Blue band (IK D1)

e) PCA+W

f) WA

g) WS

h) M2

Figure 4.3 Fused images in Blue band (IK D1)

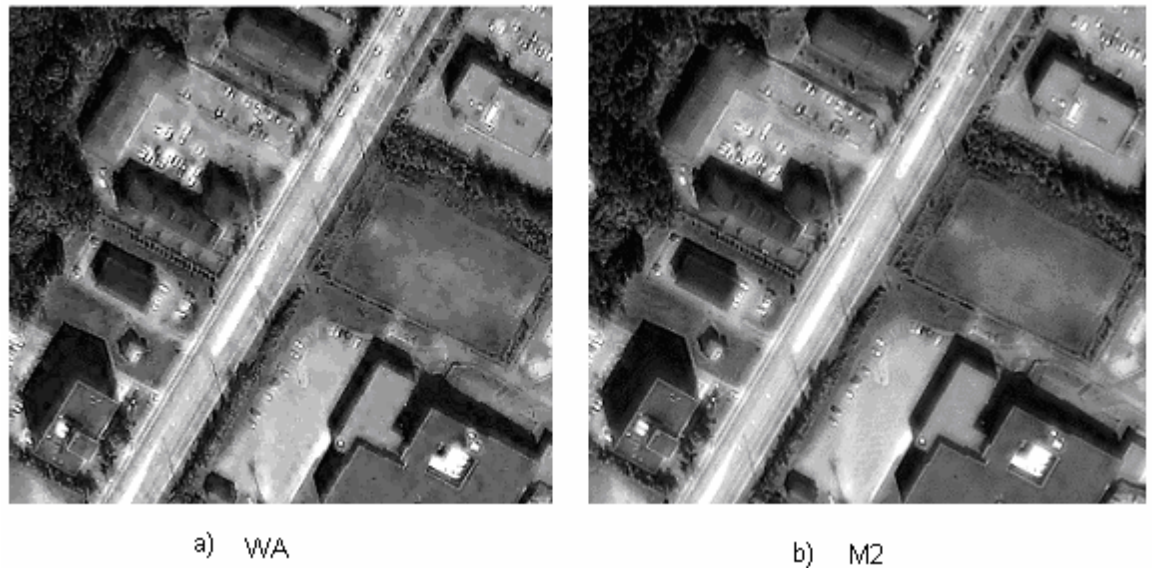


Figure 4.4 Fused images in Green band (IK D1)

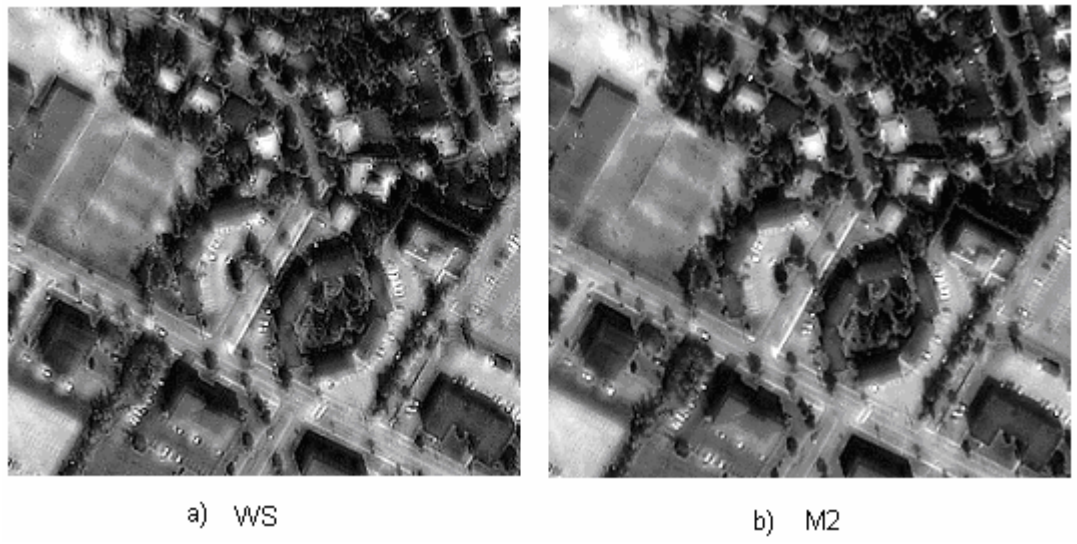


Figure 4.5 Fused images in Red band (IK D1)

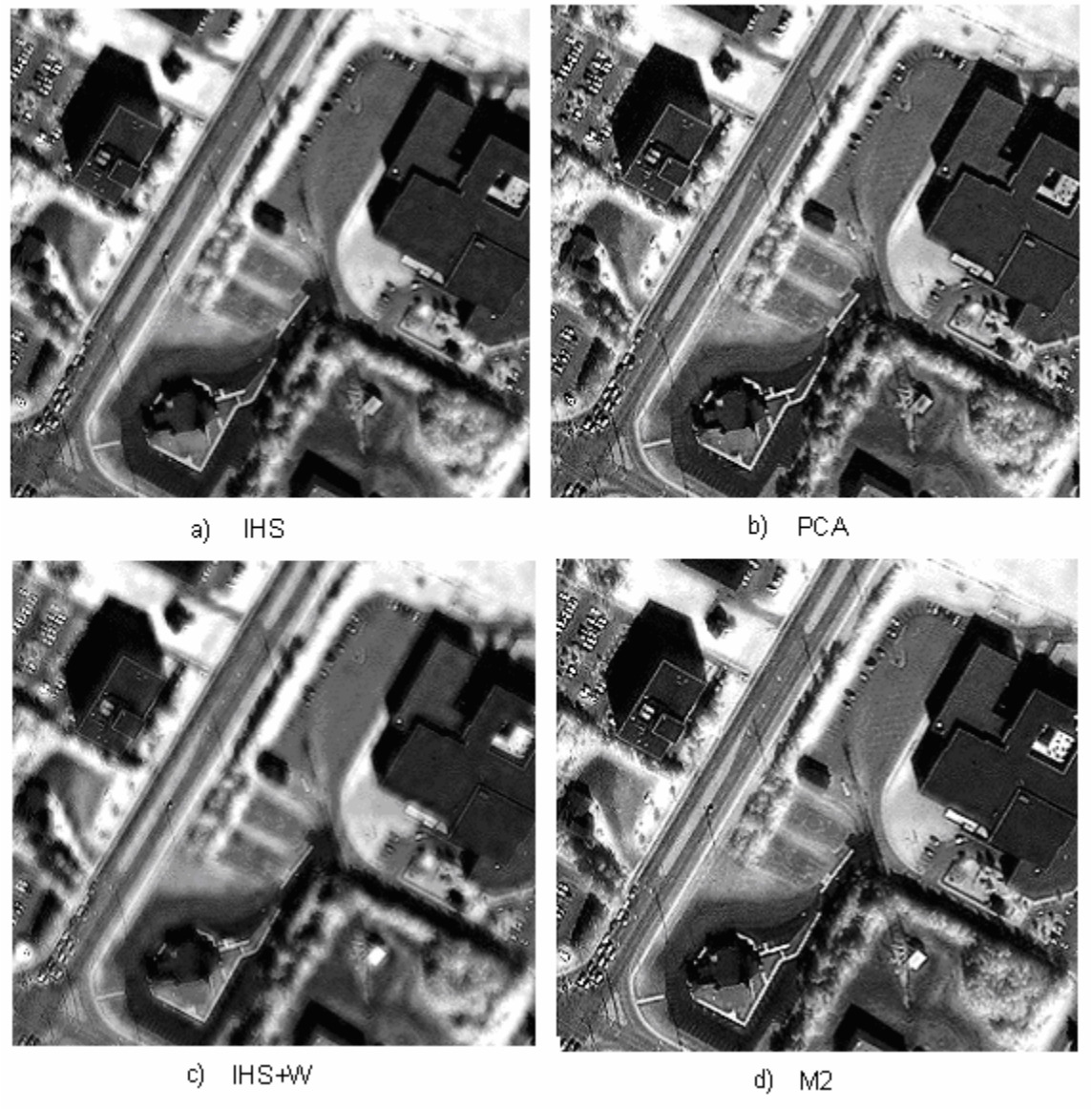


Figure 4.6 Fused images in NIR band (IK D1)

4.3.1.1 Multispectral quality

Visual analysis is also essential for assessing the multispectral synthesis. Several colour composites of the fused images obtained by the different methods were analysed. First a mosaic of the fused MS bands obtained by different methods is created (Figure 4.7). This assures that the same LUTs are applied to create comparable colour composites.

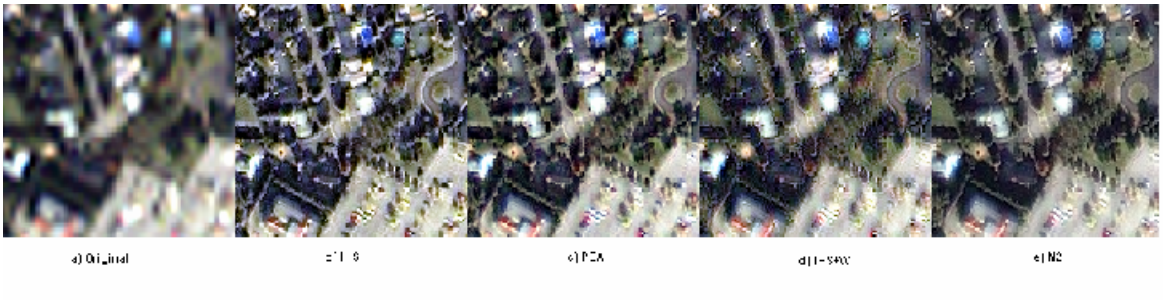


Figure 4.7 True color composite of different methods

The colour composites of the fused images in the Green, Red and NIR bands do not show significant colour differences whatever the method applied. The false colour composite of IHS (NIR band in Red, Red band in Green, Green band in Blue) is as good as the M2 method (Figure 4.8). A subset of the colour composites (Red band in Red, Green band in Green, Blue band in Blue) is shown in Figure 4.9. Figure 4.9a shows the original colour composite. Figure 4.9b shows the colour composite of the IHS fused bands. The colour distortion in the shadow pixels and in vegetation can be clearly seen. The spectral distortion in vegetation seems to be less for IHS+W (Figure 4.9d) compared to IHS. Similar to IHS+W, distortion was also observed in WA and WS colour composites. It seems that there is high error in the synthesis of the Blue band by the IHS, IHS+W, WA and WS methods. The PCA and M2 colour composites (Figure 4.9c and e) appear to have preserved the original spectral content. The colour composite of PCA+W is similar to PCA and it seems to preserve the multispectral content of the original images. In conclusion we can say that of all the methods the M2, PCA and PCA+W methods seem to better preserve the original multispectral content.

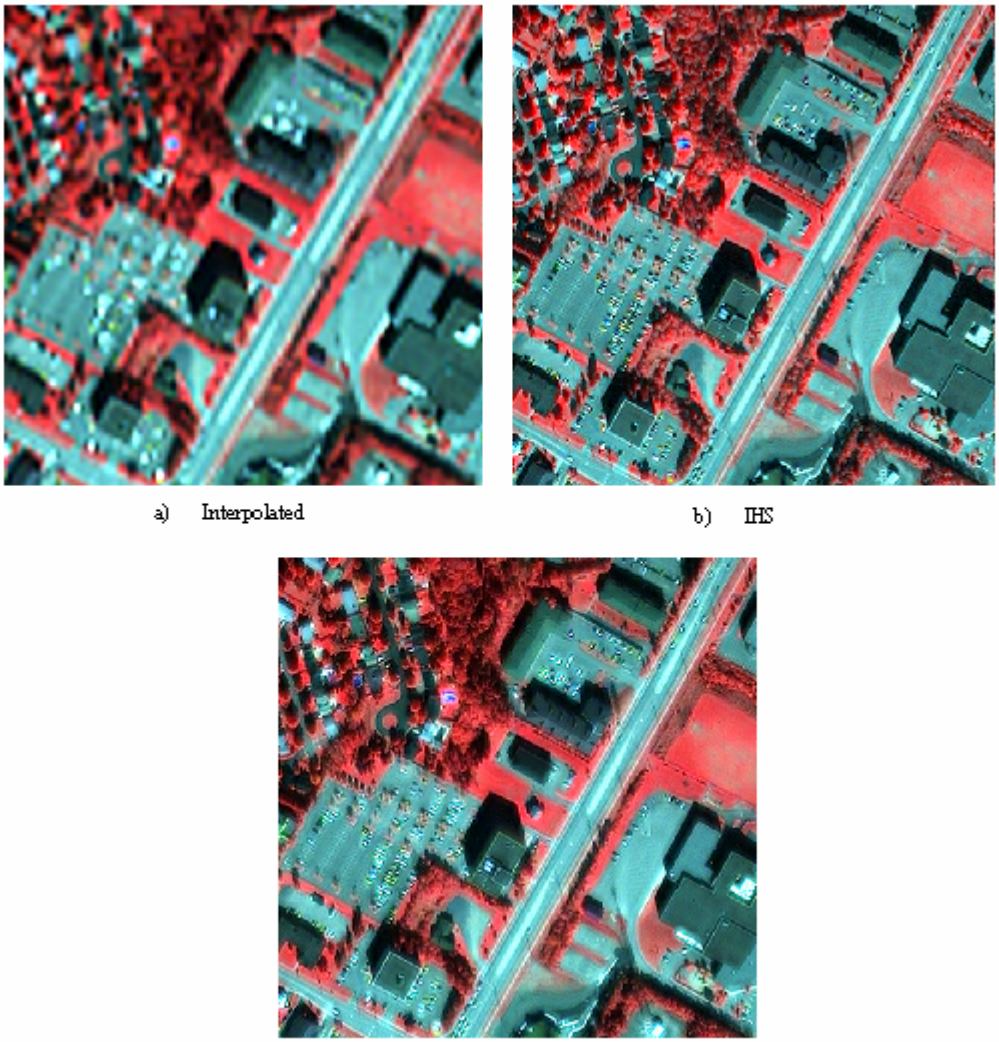


Figure 4.8 False colour composites (IK D1)

(Colour Composite: NIR band -Red, Red band – Green, Green band – Blue)

Figure 4.9 True colour composites (IK D1)
(Colour composite: Red band- Red, Green band - Green, Blue band - Blue)

4.3.2 Statistical Quality

Since the IHS method can take only three input images, different combinations were tested. The band number 4, 3, 2 and 1 refers to the NIR, Red, Green and Blue bands respectively. All the combinations with the NIR band provide similar statistical results (See Appendix 2). Hence, the fused images in the 431 combination and the fused images in the 421 combination are only considered here for the IHS method.

For the PCA method, all four bands were used. The first principal component (PC1) accounts for 57 percent, PC2 for 42 percent, PC3 for 0.47 percent and PC4 for 0.10 percent of the total variance, which indicates high information content in both the PC1 and PC2 components.

Statistics obtained for the different methods are given in Table 4.1. The bias and standard deviation of the difference image are given in percentage relative to the mean of the original image; and variance is given in percentage relative to the variance of the original image. The bias quantifies the first property of the fused images and is less than 0.5 for all the methods.

4.3.2.1 Standard deviation of the difference image (SDD) (Table 4.1)

Standard deviation of the difference image indicates the difference between the original and the fused image at the pixel level. A lower value generally indicates the closeness of the fused image to the original image and thereby implies a better synthesis by a certain method. On the general observation of the statistics for this dataset, the PCA+W fused images have low SDD of 5, 8 and 11 percent for the Blue, Green and Red bands respectively. The IHS+W fused image has the lowest SDD in the NIR band. Thus, PCA+W produces less error in Blue, Green and Red bands and IHS+W less error in NIR band.

4.3.2.2 ERGAS

ERGAS is a global measure that summarizes the error in all the bands fused by a certain method. A plot of the ERGAS values computed for each method is presented in Figure 4.10. A lower value generally indicates a better quality of the fused product. PCA+W produces a low ERGAS of 2.3 followed by IHS+W with 2.5. Both the WA and WS methods have a same ERGAS value of 3.7. The M2 method has an ERGAS of 4.3 and the PCA method has a slightly higher value of 4.5. The IHS method produces the highest ERGAS (equal to 5). Thus, based on the ERGAS values, the PCA+W and IHS+W methods seem to provide better quality results.

Table 4.1 Statistics for IK D1

Bias, SDD (standard deviation of the difference image), DIV (difference in variance) in %, CC (correlation coefficient)

	Band	**IHS**	**PCA**	**IHS+W**	**PCA+**	**WA**	**WS**	**M2**
Bias	*Blue*	-0.32	-0.15	-0.02	-0.01	0.01	0.01	0.17
	Green	-0.3	-0.25	-0.01	-0.01	0.01	0.01	0.16
	Red	-0.4	-0.34	-0.02	-0.02	0.01	0.01	0.21
	NIR	-0.23	-0.35	-0.01	-0.02	0.01	0.01	0.1
SDD	*Blue*	19.59	9.39	9.65	4.74	14.97	14.93	9.83
	Green	18.92	15.62	9.42	7.89	14.59	14.58	14.06
	Red	24.66	21.37	12.32	10.79	18.83	18.81	18.86
	NIR	14.3	22.33	7.05	11.27	10.9	11.28	23.41
DIV	*Blue*	2.97	5.81	-9.5	-1.32	-	-30.85	4.59
	Green	15	6.49	-3.08	-1.88	-	-12.18	3.94
	Red	15.49	5.84	-2.88	-2.03	-	-10.66	3.78
	NIR	1.82	6.57	-0.05	-1.94	-9.25	1.94	2.28
CC	*Blue*	0.7	0.93	0.93	0.98	0.87	0.86	0.92
	Green	0.85	0.9	0.97	0.98	0.93	0.92	0.92
	Red	0.87	0.9	0.97	0.98	0.94	0.93	0.93
	NIR	0.95	0.87	0.99	0.97	0.97	0.97	0.86

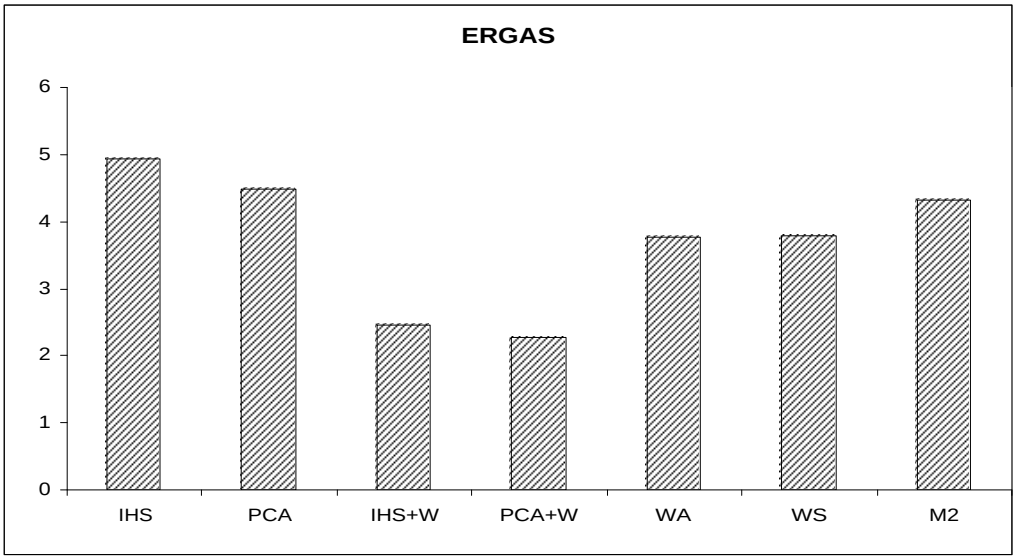


Figure 4.10 Plot of the ERGAS values

4.3.2.3 Difference in Variance (DIV) (Table 4.1)

DIV indicates the information loss or gain in each band. A positive value indicates an information loss and a negative value an information gain. Except for the IHS, WA, and WS methods, all the other methods have a low variance. It is very high for the WA and WS fused images in the Blue band. This indicates more PAN information introduced to the fused images in the Blue band. Considering all four bands, the PCA+W method has the lowest variance (in the range [-1; -2]), followed by M2 in the range [2; 5] and PCA in the range [6; 7].

4.3.2.4 Correlation Coefficient (CC) (Table 4.1)

The CC is over 0.90 for most of the fused images except for the IHS fused Blue band. The M2 and PCA fused NIR bands have a lower correlation of around 0.86 and 0.87 against a value of over 0.97 for the IHS+W, PCA+W, WA and WS fused images. A higher SDD and a lower correlation in the PCA and M2 fused NIR bands indicate that more PAN information has been introduced in the NIR band. Similarly, the IHS, WA and WS fused Blue bands have high SDD compared to the other methods indicating that these methods introduce more PAN information in the Blue band.

4.3.2.5 Correlation between PAN and fused MS images and between fused MS images in different bands

The correlation obtained for the fused images are provided in Table 4.2 and 4.3. All the methods increase the correlation of the MS images with PAN. The M2 fused NIR band has the highest correlation of 0.85 with PAN against a 0.64 with the original image. This explains the low correlation coefficient in NIR discussed in the previous section. The IHS and PCA images in the Blue, Green and Red bands have decreased their correlation with the NIR band (Table 4.3). The WA, WS and M2 fused images in the Blue, Green and Red bands have increased their correlation with NIR. The IHS+W fused images have a correlation close to that with the original images. From these statistics, it appears that the IHS+W and PCA+W methods preserve the spectral integrity between the different bands better than the other methods.

Table 4.2 Correlation Coefficient between PAN and the MS bands (IK D1)

Band	Original	IHS	PCA	IHS+W	PCA+W	WA	WS	M2
Blue	*0.49*	0.68	0.59	0.62	0.56	0.70	0.70	0.65
Green	*0.58*	0.76	0.70	0.68	0.66	0.73	0.74	0.73
Red	*0.58*	0.76	0.70	0.69	0.67	0.74	0.74	0.74
NIR	*0.64*	0.72	0.78	0.70	0.74	0.74	0.75	0.85

Table 4.3 Correlation between the MS bands (IK D1)

Band	Original	IHS	PCA	IHS+W	PCA+W	WA	WS	M2
1&2	*0.97*	0.94	0.97	0.97	0.98	0.97	0.97	0.98
1&3	*0.95*	0.93	0.95	0.95	0.95	0.95	0.95	0.97
1&4	*0.07*	0.04	0.01	0.09	0.09	0.21	0.21	0.21
2&3	*0.98*	0.99	0.98	0.98	0.98	0.99	0.99	0.99
2&4	*0.20*	0.15	0.15	0.21	0.22	0.30	0.29	0.32
3&4	*0.21*	0.17	0.16	0.22	0.22	0.30	0.29	0.32

4.3.2.6 Discussion on SDD and ERGAS

The SDD and ERGAS values are based on the spectral error introduced in each pixel. They do not give any information about the spatial distribution of the error. Analyses of several profiles of lines revealed that the differences in the spectral values synthesized by different methods are very small for certain objects. A profile of line is provided in Figure 4.11 for the Blue and NIR bands. The DN values in the “building” object are similar for all the methods.

To understand the local distribution of errors in different urban objects such as roads, buildings and vegetation, the RMSE for the pixels in vegetation and built-up classes were calculated and analyzed (Figure 4.12). The differences in PCA and M2 are found only for the “tree” and “grass” clusters. In general, the IHS+W and PCA+W methods have the lowest RMSE values. From the analysis of the profiles and the RMSE values, it seems that the very small differences in SDD between two methods are due to the differences in the spatial enhancement (edges and small objects) introduced in the fused images. Thus, the SDD and ERGAS values should be carefully interpreted with respect to the composition and high frequency components in the data set. The differences in the spatial information between the original PAN and MS images have a higher influence on the ERGAS values. This is applicable for VHR images in urban areas where the spatial resolution of MS images is high enough already to represent most of the urban objects. The only difference between the PAN and MS images exists in the spatial details corresponding to edges and very small objects. In these cases, the ERGAS values may therefore be misleading. It is known that the error in synthesis increases as the spatial resolution increases (Wald, 2002), as it is not possible to synthesize both the spatial and spectral characteristics of the spatial details by fusion methods. Also, because the sensor characteristics of the VHR PAN and MS sensors (Figure 2.15 and 2.16) are different, the synthesis of spectra in the Blue and NIR bands is questionable for VHR images. Thus, the SDD values can only be used to describe similarities and differences. Absolute values do not have any significance for this data set.

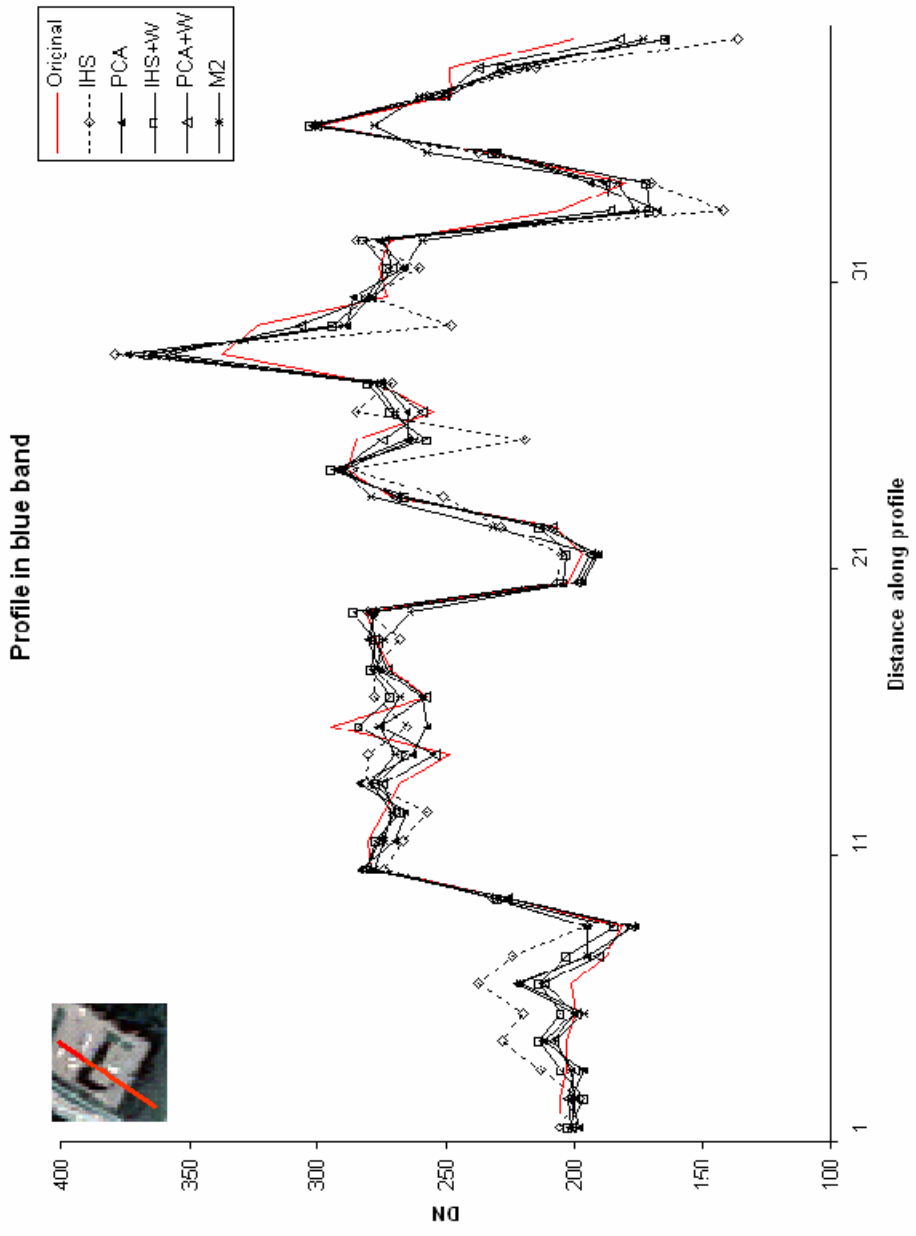


Figure 4.11 Spectral Profiles (IK D1)

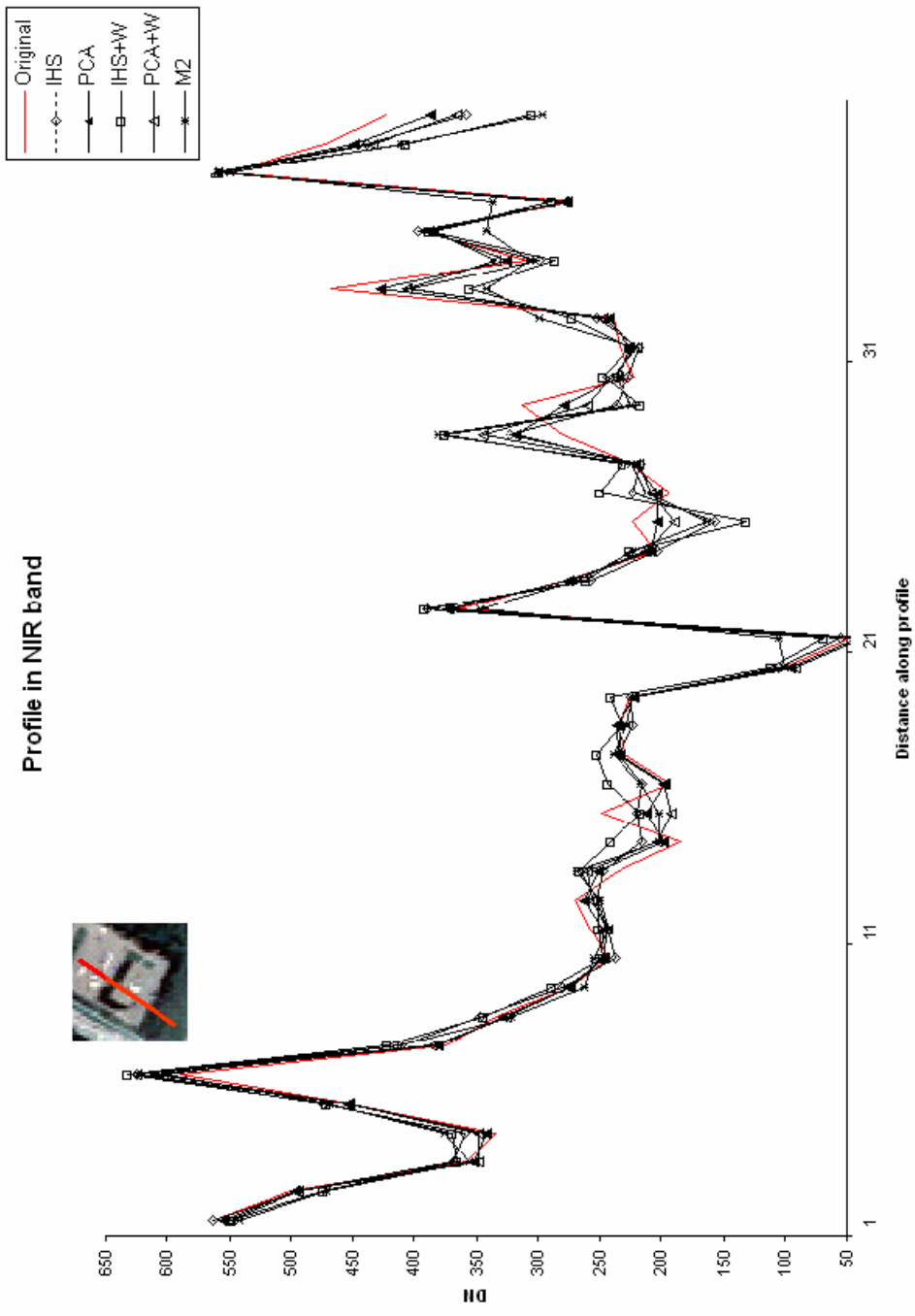


Figure 4.11 Spectral Profiles (IK D1)

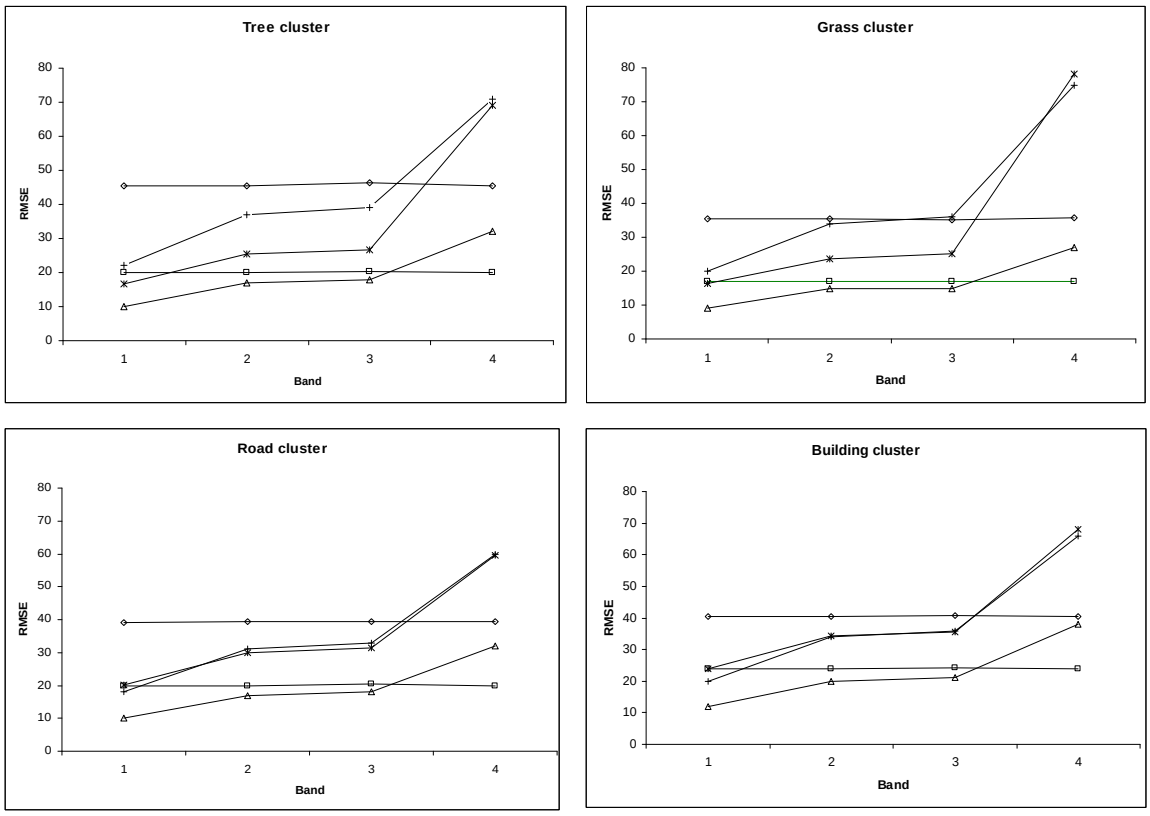


Figure 4.12 Plot of the RMSE (IK D1)

The PCA and M2 methods follow a similar trend in their SDD values (Figure 4.13) indicating a similar spatial enhancement in all the bands. It can also be seen (Table 4.1) that the WA, WS, and M2 methods produce same SDD in the Green and Red bands. This indicates a similar spatial enhancement in the Green and Red bands by these methods. The differences in the ERGAS values for the WA, WS and M2 methods are only due to the Blue and NIR bands. Since the spatial details are too much enhanced by the M2 method in the NIR band, the SDD is quite high compared to the WA and WS methods. Based on the ERGAS parameter, it can only be said that the IHS+W and PCA+W (lower than 3), WA and WS (around 3.7) and M2 and PCA (around 4.5) provide similar spatial enhancement. The IHS method (higher than 5) introduces too much spatial details in the Blue band resulting in a high ERGAS. Thus, a subjective analysis is required for interpreting the statistics.

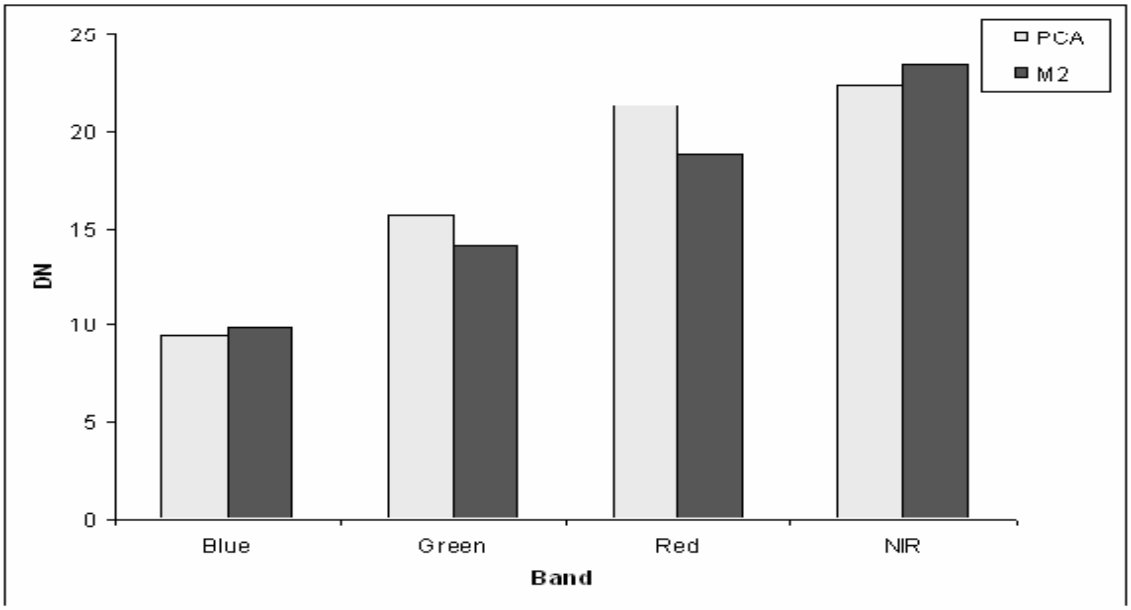


Figure 4.13 Plot of SDD (IK D1) (Values from Table 4.1)

4.3.3 Summary

The IHS fused images present sharp spatial details in all the bands. However the objects in the IHS fused images seem to have a significant loss of spectral information. This is obvious in the colour composites of the IHS fused images. Thus, IHS seems to produce a poor multispectral quality. The PCA fused images show high levels of spatial details but they are less sharp than the IHS fused images in all the bands except in the NIR band. The multispectral content seems to be better preserved in the PCA fused images. The IHS+W, WA and WS images are less sharp than the IHS and PCA images. They have slight distortions in their spectral content. PCA+W preserves the original spectral content, but the fused images are less sharp than the PCA images. The M2 fused images present sharp spatial details as well as preserve the multispectral quality of the original images. Thus visually, the PCA and M2 fused images seem to be better compared to the other methods' results.

From the statistical analysis, the PCA+W and IHS+W methods provide better results for all the measures. It seems that the quality indicated by the ERGAS values is inversely related to the spatial details introduced. A high ERGAS indicates high spatial detail while a low ERGAS indicates a poor spatial detail. This confirms the visual analysis of the fused images that we made. Thus, a "good" ERGAS value depends on the application. While considering the variance in all the bands fused by a certain method, PCA+W has the lowest variance. IHS+W shows low variance for the Green, Red and NIR bands but a relatively high variance in the Blue band. This relatively high variance in one band seems to indicate an error in the spectral synthesis. Most of the statistics only show that a different amount of PAN information is introduced according to the method applied. The IHS, IHS+W, WA and WS methods introduce more PAN information in the Blue band while the M2 and PCA methods introduce more PAN information in the NIR band. However, the correlation between the fused images indicates that the IHS+W and PCA+W methods preserve the spectral integrity of the original images better than the other bands.

Thus statistical quantities can be used to describe similarities and differences in the methods applied but they cannot be interpreted without a visual analysis. Based on our visual analysis, the M2 and PCA methods provide sharp images as well as preserve the spectral content of the original images. The ERGAS values of these methods lie between the IHS, and the IHS+W, PCA+W, WA and WS methods.

4.4 Results for IK D2

The study area is a sub-urban area mostly composed of vegetation and residential houses. The PCA method is not suitable for this data set as the correlation between the original NIR and Blue bands is very low (-0.0003). The fused images obtained for some of the methods applied are shown in Figure 4.14 for the Green band.

4.4.1 Visual Quality

Figure 4.14a, b and c present the original image, the IHS, IHS+W and WS fused images respectively. Similar to the IHS fused Blue band of IK D1, the IHS image is enhanced too much in the Blue band. Too much spatial detail introduced in the vegetated areas by IHS (Figure 4.14) give a poor visual appearance. Visually, not much difference is seen between the IHS+W, WA, and WS fused images in the individual fused bands. An analysis of the multispectral images shows a poor quality for the IHS method and can be clearly seen in the true colour composite presented in Figure 4.15b. Thus, it is obvious from Figure 4.15 that the IHS method introduces high values from PAN for the shadow and vegetation pixels in the Blue band. The IHS+W fused images have also very slight distortions in the shadow and vegetation pixels that is not so obvious in the true colour composite.

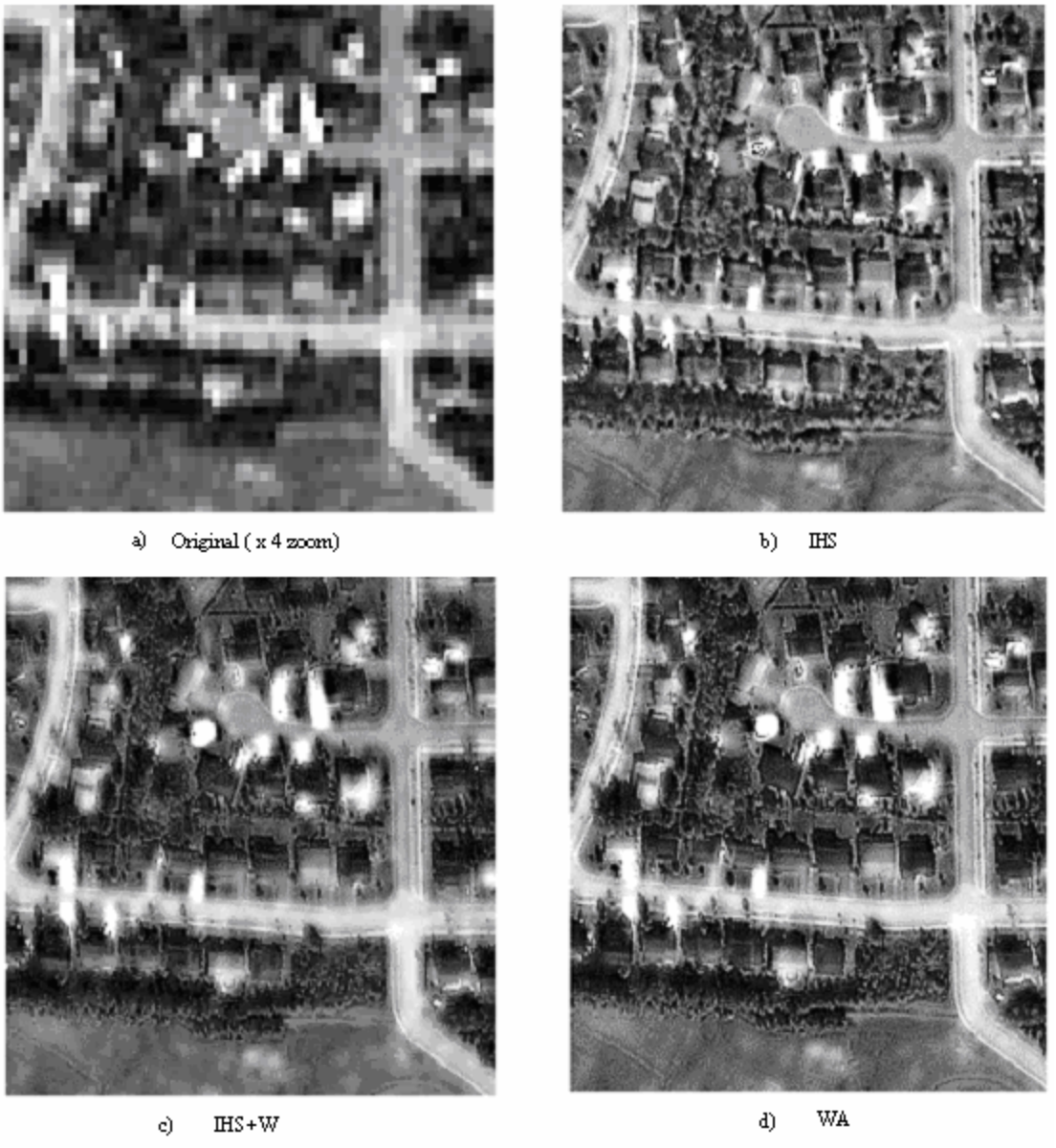


Figure 4.14 Fused images in Green band (IK D2)

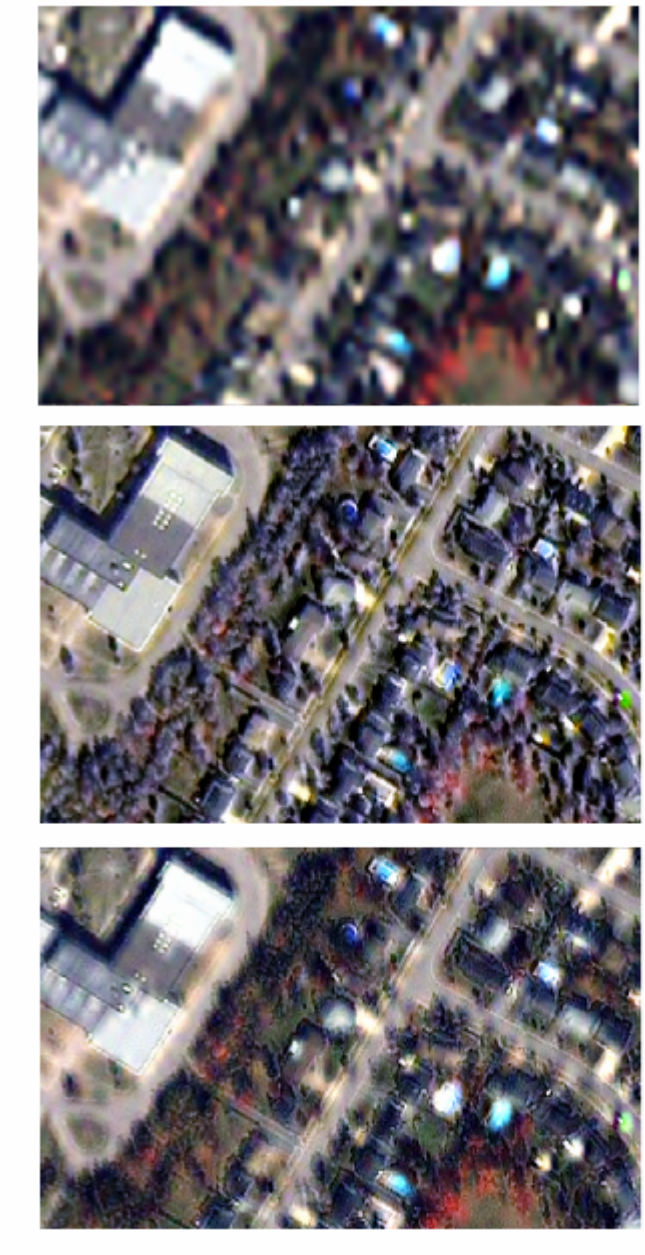


Figure 4.15 True colour composites (IK D2)

(Colour composite: Red band- Red, Green band - Green, Blue band - Blue)

4.4.2 Statistical Quality

Table 4.4 presents the statistics computed for IK D2. The bias is less than 0.5 percent for all the methods applied. The SDD is very high for IHS in the Blue, Green and Red bands compared to the other methods. This indicates that too much information has been introduced from PAN resulting in spatially enhanced fused images with spectral errors. All the SDD values follow a similar trend than for IK D1. Similar to the analysis of IK D1, IHS, IHS+W, WA, and WS show high variance in the Blue band. The spectral synthesis in the Blue band is expected to be poor because of the large variance. This is confirmed by the visual analysis of the colour composites (Figure 4.15). The spectral distortion is less obvious in IHS+W than in the WA and WS methods. Therefore, considering both SDD and variance, the IHS+W method seems to be better comparatively. The PCA statistics are not valid as the fused images do not show any improvement in spatial details due to the uncorrelated MS data set (Table 4.6). The correlation of the IHS+W fused images with PAN is closer to the correlation of the original PAN and MS images (Table 4.5). Table 4.6 presents the correlation between the original MS bands and between the fused MS bands. It can be seen that the IHS method decreases the correlation of the fused images in the Blue, Green and Red bands with the NIR band. The WA and WS methods have increased the correlation of the fused images between the Blue, Green and Red bands with the NIR band. The IHS+W method shows a correlation close to the original, showing that the spectral integrity is better preserved in the bands than for the IHS, WA and WS methods. Thus, based on visual and statistical analysis, the IHS+W method is comparatively better than the other methods.

Table 4.4 Statistics for IK D2

Bias, DSD (standard deviation of the difference image), DIV (difference in variance) in %, CC (correlation coefficient)

	Band	IHS	PCA	IHS+W	WA	WS
Bias	*Blue*	-0.21	-0.01	0.07	0.07	0.07
	Green	-0.2	-0.05	0.07	0.07	0.07
	Red	-0.27	-0.07	0.09	0.09	0.09
	NIR	-0.12	-0.26	0.04	0.04	0.04
DSD	*Blue*	23.42	1.55	10.19	14.72	14.72
	Green	23.09	6.1	10.04	14.52	14.57
	Red	30.57	8.79	13.3	19.2	19.2
	NIR	13.33	32.29	5.81	8.35	9.07
DIV	*Blue*	-29.43	14.66	-12.73	-45.34	-34.98
	Green	5.4	-10.27	-4.12	-24.44	-14.29
	Red	6.16	-9.66	-3.9	-23.4	-13.47
	NIR	7.99	-3.4	1.18	-6.44	5.63
CC	*Blue*	0.58	1	0.91	0.86	0.84
	Green	0.75	0.98	0.96	0.92	0.91
	Red	0.76	0.98	0.96	0.92	0.92
	NIR	0.95	0.72	0.99	0.98	0.98
ERGAS		5.85	4.26	2.55	3.68	3.71

Table 4.5 Correlation Coefficient between PAN and the MS bands (IK D2)

Band	Original	IHS	PCA	IHS+W	WA	WS
Blue	*0.34*	0.53	0.35	0.49	0.58	0.58
Green	*0.45*	0.65	0.48	0.56	0.63	0.63
Red	*0.46*	0.66	0.49	0.60	0.63	0.64
NIR	*0.67*	0.77	0.96	0.72	0.75	0.76

Table 4.6 Correlation between the MS bands (IK D2)

Band	Original	IHS	PCA	IHS+W	WA	WS
1&2	*0.96*	0.92	0.96	0.96	0.96	0.96
1&3	*0.92*	0.89	0.93	0.92	0.93	0.94
1&4	*-0.003*	-0.08	0.21	0.003	0.14	0.14
2&3	*0.97*	0.99	0.98	0.98	0.98	0.98
2&4	*0.16*	0.06	0.35	0.16	0.25	0.25
3&4	*0.17*	0.08	0.36	0.17	0.26	0.26

4.5 Results of QB D1

The area of QB D1 is a complex environment composed of residential areas and industrial areas. There is a lot of vegetation in the area. A subset of the IHS, PCA, IHS+W and M2 fused Blue bands is presented in Figure 4.16.

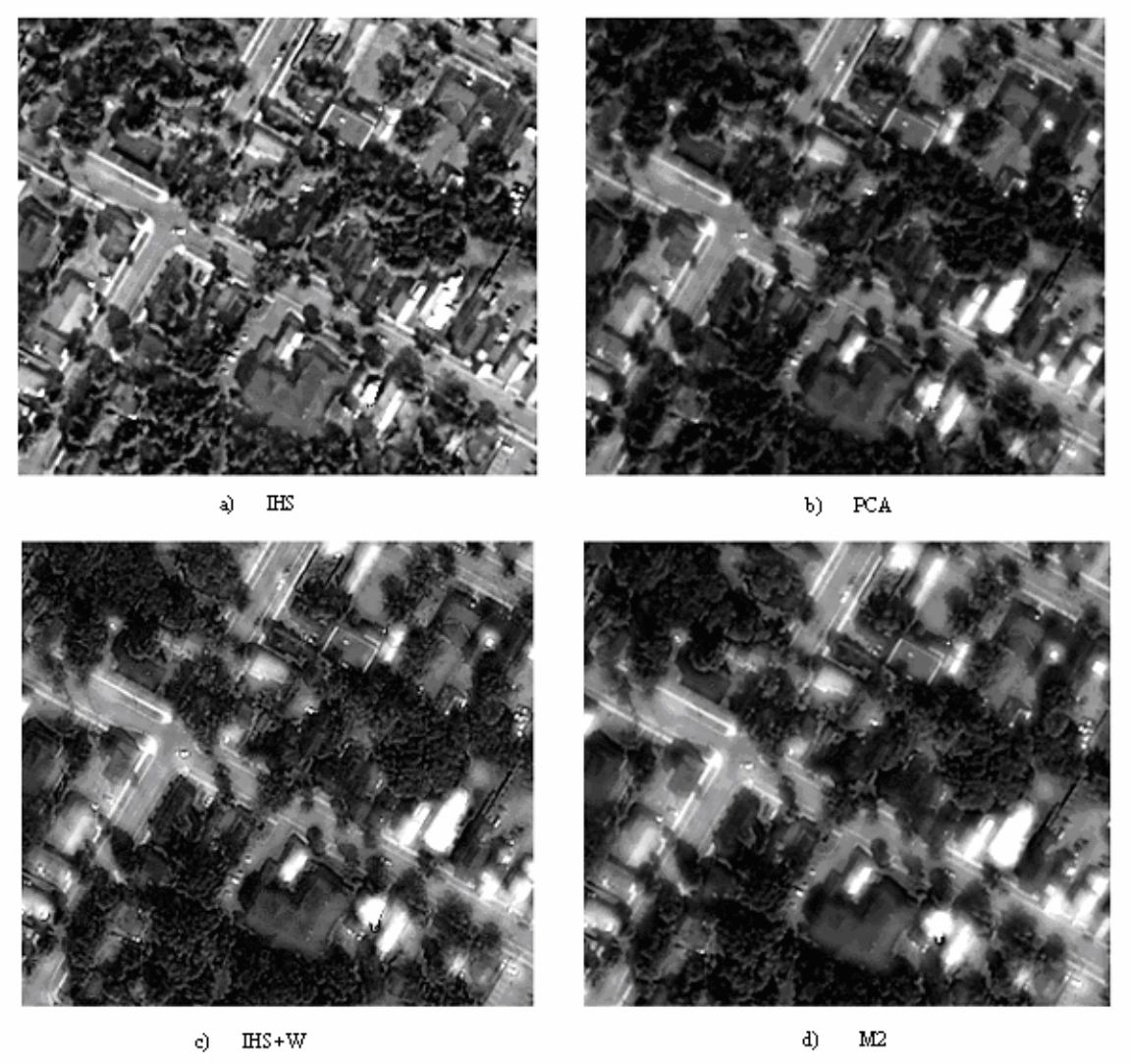


Figure 4.16 Fused image in Blue band (QB D1)

4.5.1 Visual Quality

Figure 4.16a, b, c and d present the IHS, PCA, IHS+W and M2 fused Blue bands respectively. The IHS Blue band is very sharp compared to the other methods (Figure 4.16). For certain buildings, the PCA fused images have sharp edges compared to the M2 images (Figure 4.16b). The relative spectral values of the different features have been preserved by all the methods. The spatial details in IHS+W, PCA+W, and WS are less sharp compared to the PCA and M2 images (Figure 4.17). The analysis of different colour composites shows that the IHS method does not synthesize the spectra accurately (Figure 4.17b). The distortion is more obvious near the shadow pixels of vegetation and near buildings. Figure 4.17d presents the colour composite of the IHS+W. The WA and WS methods are similar to IHS+W and show less spectral distortion compared to IHS. The PCA and M2 methods best preserve the spectral quality of the original images. The PCA+W colour composites are similar to the PCA colour composites. The IHS+W, WA and WS methods provide a better spectral synthesis than IHS but a poorer one compared to the M2, PCA+W and PCA methods.

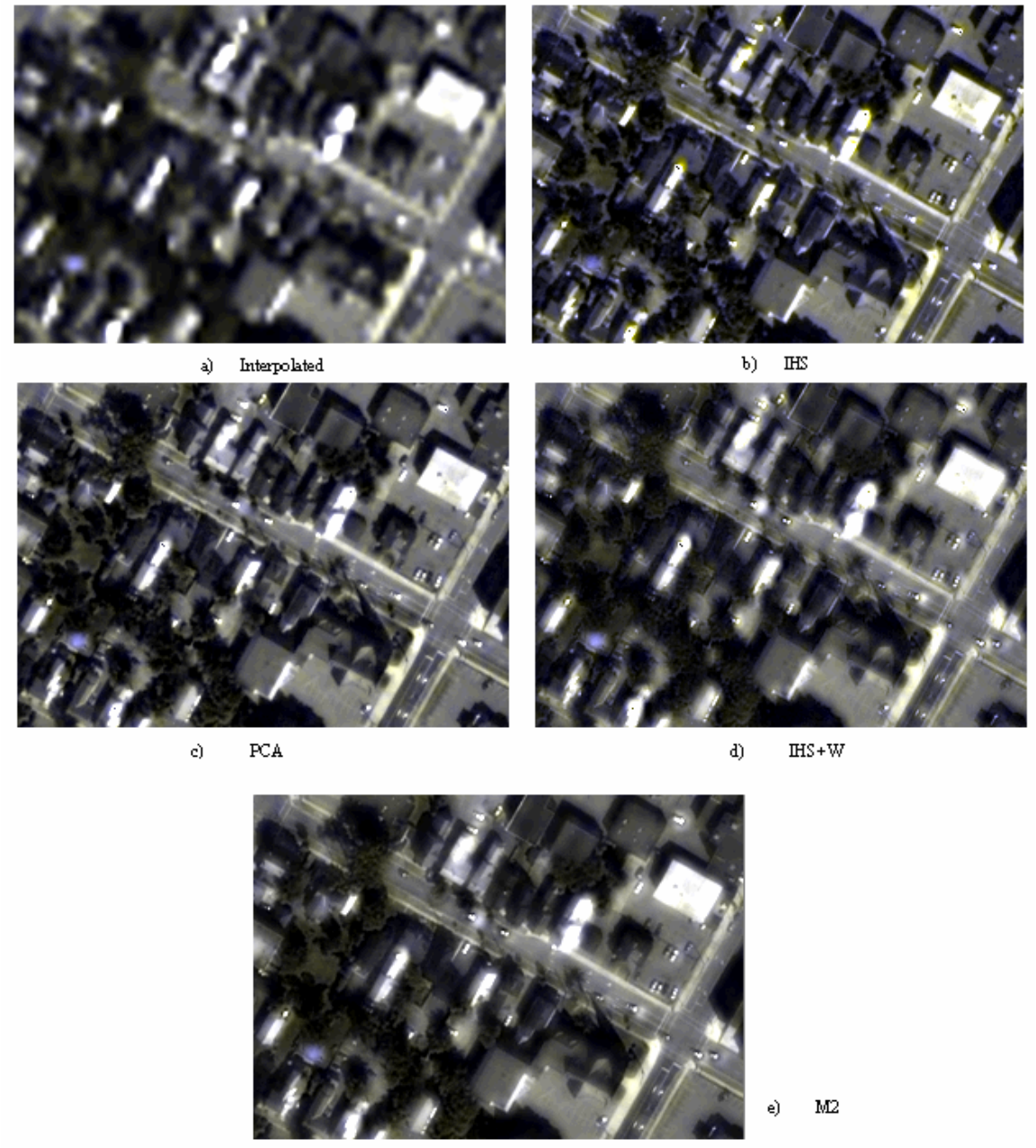


Figure 4.17 True colour composites (QB D1)

(Colour composite: Red band- Red, Green band - Green, Blue band - Blue)

4.5.2 Statistical Quality

The statistics computed for the fused images of the different methods are provided in Table 4.7. The bias is less than 0.5 percent in all the bands and a little bit less for the IHS+W, PCA+W, WA and WS methods. This shows that all methods satisfy the first property.

Table 4.7 Statistics for QB D1

Bias, SDD (standard deviation of the difference image), DIV (difference in variance) in %, CC (correlation coefficient)

	Band	**IHS**	**PCA**	**IHS+W**	**PCA+W**	**WA**	**WS**	**M2**
Bias	*Blue*	-0.27	-0.13	0	0	0	0	0.18
	Green	-0.21	-0.19	0	0	0	0	0.17
	Red	-0.33	-0.26	0	0	0	0.01	0.19
	NIR	-0.16	-0.24	0	0	0	0	0.1
DSD	*Blue*	23.8	12.39	9.73	4.94	15.08	14.77	11.63
	Green	19.25	17.96	7.86	7.15	11.38	11.13	15.33
	Red	28.93	23.99	11.82	9.56	18.31	17.89	20.63
	NIR	14.33	23.04	5.84	9.18	9.06	8.97	17.1
DIV	*Blue*	-4.09	5.36	-14.75	-0.33	-29.75	-20.65	9.94
	Green	11.06	5.03	-7.27	-0.56	-12.51	-3.91	9.66
	Red	7.96	4.57	-8.06	-0.5	-15.55	-7.22	9.34
	NIR	4.62	5.22	-3.83	-0.89	-6.89	0.89	9.81
CC	*Blue*	0.76	0.93	0.96	0.99	0.92	0.92	0.9
	Green	0.91	0.92	0.99	0.99	0.97	0.97	0.94
	Red	0.89	0.92	0.98	0.99	0.96	0.96	0.94
	NIR	0.96	0.9	0.99	0.99	0.99	0.99	0.95

4.5.2.1 SDD and ERGAS

A general observation for all the methods is that the fused Red band has the highest SDD. The IHS fused images have higher SDD in the Blue, Green and Red bands compared to the other methods. This indicates a relatively high error introduced in the fused images by the IHS method in these bands. The PCA and M2 fused NIR bands have a very high SDD compared to the other methods. This indicates a high error in the PCA and M2 fused NIR

bands. As discussed under IK D1, the differences in SDD are highly related to the sharpness of the spatial details (spatial enhancement) in the fused images. Therefore, high or low value only indicates better or poor spatial details compared to another method. To summarize the SDD in all the fused images by a method, the ERGAS values are plotted in Figure 4.18. The IHS+W and PCA+W methods produce the lowest ERGAS values of 1.98 and 2.27 respectively. The WA and WS methods have an ERGAS of 3.4. The IHS method has the highest ERGAS value of 5.6. The PCA and M2 methods have ERGAS values of 5 and 4 respectively. These values verify our visual analysis that the PCA fused images are sharper compared to the M2 method. The similarities in the ERGAS values indicate a similar synthesis of the spatial details. Thus, the ERGAS values in the increasing order of spatial detail are for IHS+W, PCA+W, WA and WS, M2, PCA, and IHS. This confirms the visual analysis of the spatial details.

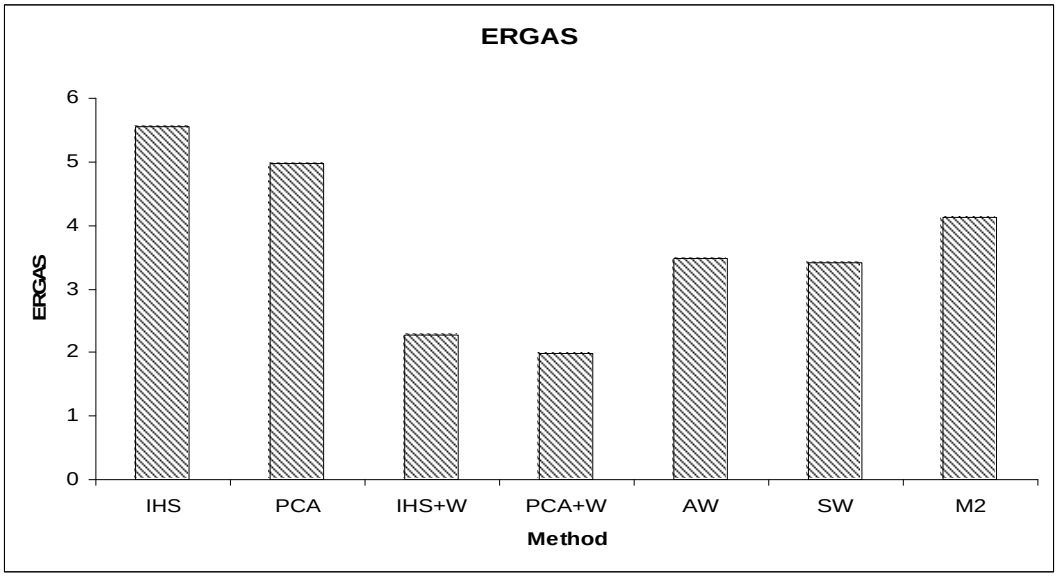


Figure 4.18 Plot of ERGAS values (QB D1)

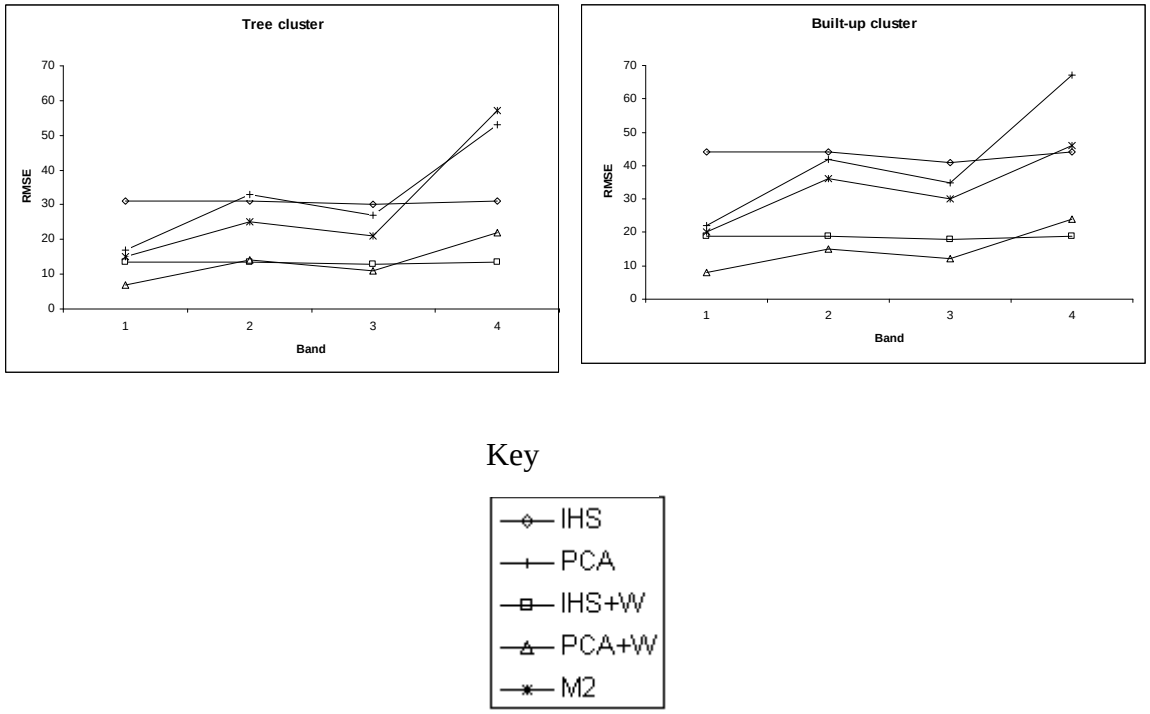


Figure 4.19 Plot of RMSE (QB D1)

The RMSE for the "tree" and "built-up" clusters are analyzed (Figure 4.19). PCA follows the same trend as the M2 model with a slightly higher error in the all the bands for the "built-up" areas. This explains the better spatial detail in some buildings in the PCA fused images. PCA+W has similar RMSE than IHS in the Green and Red bands. PCA+W shows a similar RMSE than IHS+W in the Green and Red bands. PCA+W have the lowest RMSE values in the Blue, Green and Red bands and a slightly higher RMSE than IHS in the NIR band for both clusters. All these values confirm that there are local differences (objects) in the synthesis of the spatial details by different methods.

4.5.2.2 Difference in Variance (DIV)

DIV is highly negative in the WA and WS fused Blue bands (Table 4.7). The WA fused images in the Green and Red bands have highly negative variance. The PCA+W fused images have a low variance between 0 and -1 percent followed by PCA with 4-5 percent and M2 with 9-10 percent. The IHS fused images have low variance but the fused Blue

band has a negative variance while in the other bands it is positive. This relative difference in variances in the fused images resulting from a certain method seems to indicate an error in the spectral synthesis. This also has been observed in the visual analysis of the colour composites of the IHS, IHS+W, WA and WS methods.

4.5.2.3 Correlation Coefficient (CC)

CC indicates the degree of relationship between the original and the fused images. If more information from PAN is introduced during fusion, it results in a low correlation coefficient. The CC is around 0.7 for the Blue band and 0.9 for the other bands with the IHS method. This indicates a high error in the Blue band. It is around 0.9 for the PCA NIR band which indicates that high spatial information has been introduced in the NIR band compared to the other methods. PCA+W has a high correlation of 0.99 in all the bands indicating that the fused images are close to the original images.

4.5.2.4 Correlation between PAN and fused MS images and between fused MS images in different bands

The correlation coefficient between PAN and MS and between the MS bands is given in Table 4.8 and 4.9 respectively. Similar to the observations for IK D1, all the methods have increased the correlation of MS images with PAN. In the correlation between the fused MS bands, IHS and PCA have decreased the correlation between the Blue, Green and Red band with the NIR band. The WA, WS and M2 methods produce highly correlated fused images. The IHS+W and PCA+W methods maintain the original relationship between the MS bands.

Table 4.8 Correlation Coefficient between PAN and the MS bands (QB D1)

Band	Original	IHS	PCA	HIS+W	PCA+W	WA	WS	M2
Blue	*0.62*	0.74	0.71	0.70	0.66	0.75	0.75	0.72
Green	*0.66*	0.80	0.77	0.73	0.72	0.76	0.76	0.77
Red	*0.68*	0.77	0.75	0.71	0.70	0.76	0.76	0.75
NIR	*0.70*	0.76	0.79	0.73	0.75	0.76	0.75	0.78

Table 4.9 Correlation between the MS bands (QB D1)

Band	Original	IHS	PCA	IHS+W	PCA+W	WA	WS	M2
1&2	*0.99*	0.95	0.98	0.98	0.98	0.98	0.98	0.99
1&3	*0.97*	0.95	0.97	0.97	0.97	0.97	0.97	0.98
1&4	*0.21*	0.16	0.17	0.21	0.22	0.30	0.29	0.25
2&3	*0.99*	0.99	0.98	0.99	0.99	0.99	0.99	0.99
2&4	*0.29*	0.26	0.25	0.29	0.30	0.35	0.33	0.32
3&4	*0.26*	0.21	0.22	0.26	0.28	0.33	0.31	0.29

4.5.3 Summary

Similar to the conclusions for IK D1, the PCA+W, PCA and M2 methods produce a better multispectral quality in their results. Statistically, the PCA+W and IHS+W methods provide low ERGAS values. However, considering the variance, the PCA+W, PCA and M2 methods are better compared to the other methods. Overall, the PCA and M2 fused images are better in terms of details as well as multispectral content.

4.6 Results of QB D2

The area in this data set is the same as the IK D1 one. The correlation between the original NIR and Blue bands is 0.08, between the Red and NIR band -0.01, and between the Green and NIR band 0.02 (Table 4.12).

4.6.1 Visual and Statistical Quality

Visually, the IHS+W and WA fused images have poor spatial details and the blur near the edges of the buildings are very obvious (Figure 4.20). The WS fused images are similar to the fused images produced by the IHS+W and WA methods. The IHS method provides better enhanced fused images than the other methods. However, the visual quality is not satisfactory. Figure 4.21 presents the colour composites of the IHS and IHS+W methods. It is clearly observed from the colour of the vegetation that the Blue band is inaccurately synthesized by the IHS method. The IHS+W method has a better multispectral quality but the spatial details are poorer.

Figure 4.20 Fused images in Red band (QB D2)

Figure 4.21 True colour composites (QB D2)

(Blue band- Blue, Green band - Green, Red band –Red)

The statistics computed for this data set are given in Table 4.10. The statistics are poor for the IHS method. The correlation of the IHS fused Blue band is 0.64, which is very low compared to the other bands. This indicates that too much information has been added from PAN. Among the IHS+W, WA and WS methods, it is the IHS+W fused images that are statistically better based on SDD, DIV, CC and ERGAS. The PCA fused images in the Blue, Green and Red bands do not show any improvement in spatial details because of the low inter-band correlation. Therefore, the statistics given are not valid. Table 4.11 and Table 4.12 show the correlations between the PAN with the fused images and between different pairs of MS bands. It can be seen that IHS increases the correlation between PAN and MS bands to a greater extent than the other methods. However, the IHS images are better visually as the spatial details in IHS+W, WA and WS are not sufficient to provide a satisfactory visual quality.

Table 4.10 Statistics for QB D2

Bias, SDD (standard deviation of the difference image), DIV (difference in variance) in %, CC (correlation coefficient)

	Band	HIS	PCA	IHS+W	WA	WS
Bias	Blue	-0.37	0.03	-0.03	-0.06	-0.06
	Green	-0.28	0.01	-0.03	-0.05	-0.05
	Red	-0.42	0.03	-0.04	-0.07	-0.07
	NIR	-0.19	-0.41	-0.02	-0.03	-0.03
SDD	Blue	26.63	3.43	7.77	11.78	11.9
	Green	20.78	1.07	6.12	8.63	8.89
	Red	30.67	2.58	8.96	13.57	13.86
	NIR	14.04	40.95	4.12	6.21	6.82
DIV	Blue	-36.79	10.39	-3.91	-18.85	-14.7
	Green	4.92	2.72	-0.21	-6.93	-2.52
	Red	-3.79	4.73	-0.64	-8.94	-4.8
	NIR	7.54	0.55	1.07	-3.07	3.19
CC	Blue	0.64	0.99	0.96	0.93	0.92
	Green	0.83	0.99	0.98	0.97	0.97
	Red	0.8	0.99	0.98	0.96	0.96
	NIR	0.95	0.56	0.99	0.99	0.99
ERGAS		5.97	5.15	1.75	2.61	2.68

Table 4.11 Correlation Coefficient between PAN and the MS bands (QB D2)

Band	Original	IHS	PCA	IHS+W	WA	WS
Blue	0.39	0.62	0.36	0.49	0.55	0.55
Green	0.45	0.68	0.45	0.52	0.55	0.56
Red	0.44	0.65	0.43	0.51	0.55	0.56
NIR	0.54	0.69	0.96	0.59	0.61	0.61

Table 4.12 Correlation between the MS bands (QB D2)

Band	Original	IHS	PCA	IHS+W	WA	WS
1&2	0.98	0.93	0.97	0.97	0.97	0.97
1&3	0.95	0.92	0.95	0.95	0.95	.095
1&4	-0.08	-0.11	0.34	-0.09	-0.002	-0.01
2&3	0.98	0.98	0.98	0.98	0.98	0.98
2&4	0.02	-0.04	0.44	0.01	0.06	0.05
3&4	-0.009	-0.07	0.41	-0.02	0.04	0.03

4.7 Conclusion

The results of the visual and statistical analyses are dependent on the application. The statistics do reveal the similarities and differences between the methods. However, it is very difficult to interpret the statistical results without a visual analysis. For urban mapping from VHR images, the objective of the fusion is to have an increased spatial resolution while preserving the spectral information. Therefore, spatial details as well as multispectral quality are important. Some methods that provide poor spatial details give better ERGAS values. Thus quality provided by ERGAS values seem to be inversely related to the spatial details introduced.

Based on the results and analyses for IK D1 and QB D1, it seems that an ERGAS value between 4 and 5 indicates a good spatial quality. For these data sets, a value smaller than that indicates low spatial quality. Relative high or low variances in one band with respect to the others seem to indicate a poor multispectral synthesis by the method. It is difficult to conclude based only on one statistical measure. Collective analyses of all the measures can be used to categorize similar methods but conclusions are difficult to draw when the

differences in statistics are too small between two methods. This holds also for the visual analysis. For example, the PCA and M2 fused images have similar statistics. The visual analysis showed that the PCA fused images have sharper edges in buildings compared to M2 while some edges (pavements) are blurrier in PCA compared to M2. Thus, quality is very subjective in this aspect. If sharper edges are desired, PCA is better and if better image quality is desired, M2 is better.

The IHS method assumes that the PAN information is highly correlated with all the MS bands. The IHS+W, WA and WS methods assume that the high frequencies of PAN are correlated with the high frequencies in the MS bands. This is not true for the Ikonos and Quickbird sensors where the Blue band is less correlated with PAN. Therefore, the IHS, IHS+W, WA and WS fused images contain spectral distortion in the Blue band. Depending on the amount of PAN information introduced, the spectral distortion may or may not be obvious in the visual analysis. This is the case with the IHS+W method that introduces less high frequency information compared to the WA and WS methods and the spectral distortions are not as obvious in the colour composites as with the WA and WS methods.

The PCA and PCA+W methods provide better results for the IK D1 and IK D2 data sets. However, they are not suitable for uncorrelated data sets such as IK D2 and QB D2. For QB D2, the spatial details introduced at the first level are not sufficient for a good visual quality.

To conclude, the PCA and M2 methods provide better spatial and spectral qualities which are required for photo-interpretation and mapping. The IHS+W method is comparatively better than the WA and WS methods. The WA and WS methods are better than IHS.

Chapter 5 Classification

5.1 Introduction

Image fusion techniques alter the spectra of pixels. Any changes in the spectral signatures of the classes will affect the classification. The visual and statistical qualities that were analyzed and discussed in Chapter 4 are more relevant for photo-interpretation and automatic mapping. The visual and statistical discrepancies in different bands and between different methods may or may not be relevant for automatic classification procedures. In this chapter, a general discussion on the classification of VHR images and an analysis of class separabilities in fused images of different methods are presented.

5.2 Classification

Although classification accuracy is expected to be high as the spatial resolution increases, VHR MS images have more small objects such as vehicles, structures on building roofs and signalization on roads that are not of interest. This may lead to poor classification accuracy (Martino *et al.*, 2003). The classification results could be worse with fused MS images because we have increased the spatial details, i.e. higher variance within a class.

Although VHR MS bands have a reasonable spectral resolution, it is insufficient for completely discriminate the urban objects such as roads and buildings based only on their spectral information as certain road materials have similar spectral signatures as certain building materials. Figure 5.1 shows a spectral plot of some road and building pixels that have similar spectral characteristics. An example of the supervised maximum likelihood classification obtained with all the four original MS bands is shown in Figure 5.2.

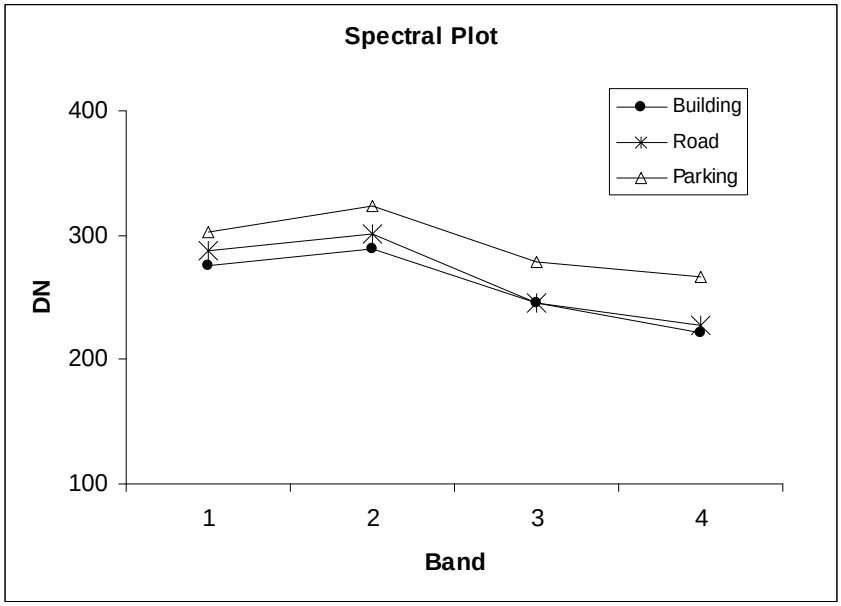


Figure 5.1 Spectral signatures of misclassified road and building pixels

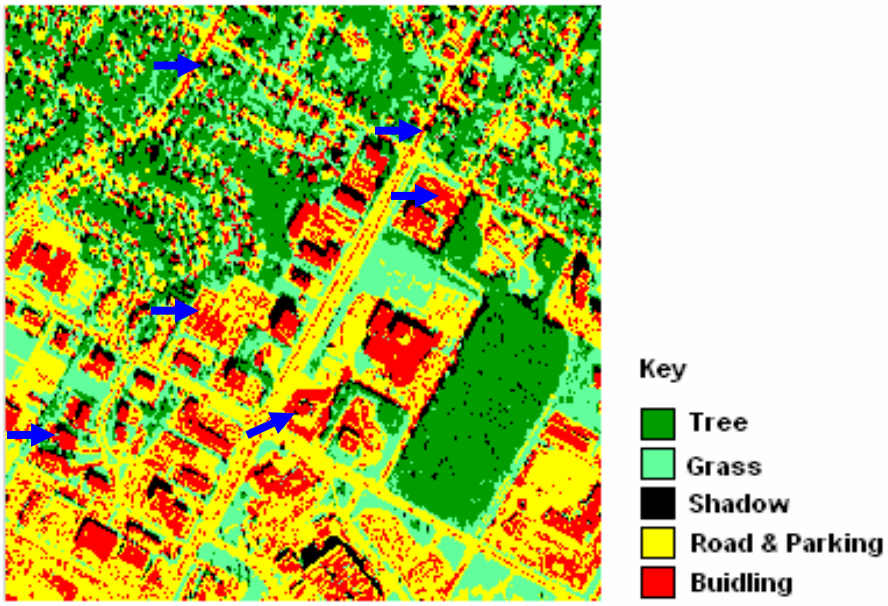


Figure 5.2 Maximum Likelihood Classification of the original MS images

An unsupervised ISODATA classification was used to aid the selection of training sites as homogeneous areas were difficult to find for the road and tree classes because of

shadows, signalization on roads, vehicles and so on. Five major classes - tree, grass, shadow, road and buildings - were used. There are three different classes of buildings in this data set. In a classification based only on the spectral information, parking objects are difficult to differentiate from the road pixels Two sub classes road (1) and road (2) were used for the road class and three sub-classes building (1), building (2) and building (3) were used for the building class. The error matrix for the training pixels is given in Table 5.1. The accuracy of road (2) is very poor (around 63%). The road (2) and building (3) class have similar spectral signatures (Figure 5.1) and it is very difficult to separate them based only on their spectral information. The blue arrows in Figure 5.2 show pixels of road and parking lots that have been classified as building. The tree, grass and shadow classes have a very high accuracy (over 98%). The accuracies of the road (1), building (1) and building (2) classes are greater than 80%. It should be noted that the problem is due to the spectral limitation of the VHR images and not with the training sites.

Table 5.1 Error matrix obtained for training pixels

	Classified data							
	Tree	**Grass**	**Shadow**	**Road(1)**	**Road(2)**	**Building(1)**	**Building(2)**	**Building(3)**
Tree	98.98	1.02	0	0	0	0	0	0
Grass	0	100	0	0	0	0	0	0
Shadow	0	0	100	0	0	0	0	0
Road (1)	0	0	0	82.81	9.9	0	0.52	6.77
Road (2)	0	0	0	9.30	62.79	0	0	27.91
Building (1)	0	0.24	0	0.48	0.72	97.84	0	0.72
Building (2)	0	0	0	4.9	0	0	100	0
Building (3)	0	0	0	5.71	6.67	1.90	0	85.71
User's accuracy (%)	*100*	*97*	*100*	*92*	*48*	*99*	*99*	*76*

Incorporating the spatial information using texture measures shows an improvement in the classification of the multispectral data of urban areas (van der Sande *et al*., 2003; Martino *et al*., 2003; Shackelford *et al*., 2003). Some experiments on IK D1 with different textures (e.g. mean, variance, homogeneity, entropy) have shown improvements in classification. However, some textures while increasing the accuracy for some classes decrease the accuracy for other classes. Some examples and classification results of M2 fused images with different textures are shown in Appendix 2. For dense urban areas, apart from the spectral and textural information, surface area, length, width, morphological shape criteria and shadow information are important to discriminate between road, building, and a non road impervious surface (e.g. parking) (Shackelford *et al*., 2003). And the classification approaches for VHR and dense, complex areas are still in research and this limits the testing of the fused images based on classification for feature extraction. Since the scope of this research is related to fusion techniques, a Maximum Likelihood Classifier (MLC) was used to test if the fused images have enough spectral separability to differentiate the major classes.

The spectral variances in the vegetation and built-up classes are generally high enough in the VHR MS bands that any fusion method will allow for an easy discrimination of the vegetation class from the built-up class. A few building and road pixels were selected (Figure 5.3) to check the percentage of correctly classified pixels. The results are presented in Table 5.2. Nearly 80 percent of the road pixels are correctly classified by whatever fusion method applied. For the building pixels, IHS has the lowest percentage (67 percent). Some building pixels along the edges have been classified as road pixels and shadow pixels. However, the classified image of the IHS fused MS image shows that IHS provides significant spectral variances to discriminate the major classes (Figure 5.4).

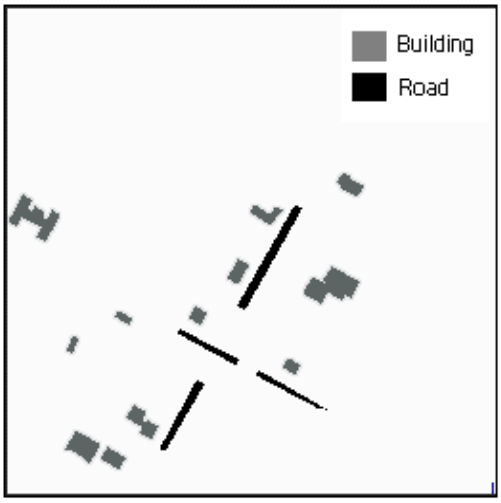


Figure 5.3 Road and building pixels used for Table 5.2

Table 5.2 Percentage of correctly classified pixels

	Road (%)	**Building (%)**
Original	*89.1*	*87.5*
IHS	79.5	67.3
PCA	84.0	81.6
IHS+W	81.3	72.7
PCA+W	84.9	83.1
M2	83.7	86.7

Fused MS images have more spatial details than the original MS image. The spatial structures on some roofs, small objects like vehicles on roads and parking lots, signalization on roads do not have a characteristic spectral signature. They are often classified as either road or building. The classification results for the fused MS image by the different methods are shown in Figure 5.4. A few buildings have been classified partially as road. Differences in classification can only be seen in those pixels corresponding to edges, boundary pixels, pavement pixels and so on. Therefore, a comparison of the classification accuracies does not seem like a valid approach for evaluating the spectral preservation in the fused images.

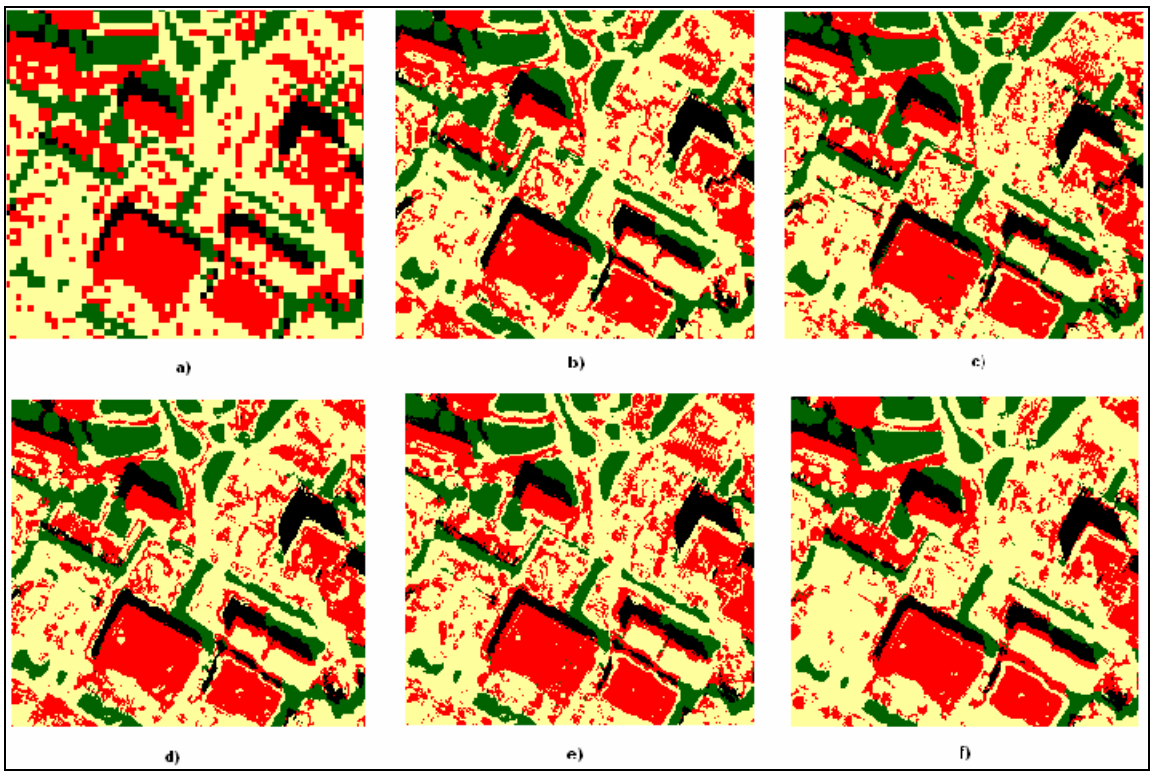


Figure 5.4 Subset of the classified images

a) Original MS image b) IHS c) PCA d) IHS+W e) PCA+W e) M2 fused images

The class separability obtained with the Bhattacharya distance was analyzed for the training as well as the reference pixels (Table 5.3a and b). This distance measure produces a maximum of 2.0 which means a good separability. Class separabilities may increase or decrease in the fused images depending on the error introduced during the fusion. The IHS+W and M2 methods provide a better class separability between all the pairs of classes. For both sets, the class separability between the tree-grass classes is low with the PCA method. The PCA+W and M2 methods also show a low separability compared to IHS+W between the tree-grass classes. Between the road(1)-building(1) classes, all the methods have a separability over 1.8. Considering the urban classes, road(1)-building(1), road(1)-building(2), building(1)- building(3) that have good separabilites in the original MS image, the class separabilities are also high for the IHS+W, PCA+W, PCA and M2 fused MS images. However, it should be noted that there

may or may not be significant differences in the classification even if the class separabilities are low.

Table 5.3 Class Separabilities (Bhattacharya distance)

a) Training pixels

	Original	IHS	PCA	IHS+W	PCA+W	M2
Tree-Grass	*1.91*	1.47	0.98	1.61	1.37	1.51
Road (1) –Building (1)	*1.99*	1.88	1.95	1.96	1.99	1.96
Road (1) –Building (2)	*1.82*	1.30	1.60	1.56	1.62	1.81
Road (1) – Building (3)	*1.49*	0.86	0.98	1.27	1.12	1.13
Road (2) – Building (3)	*0.92*	0.55	0.39	0.89	0.51	0.61
Building(1)–Building (3)	*1.75*	1.64	1.61	1.70	1.89	1.85

b) Reference pixels

	Original	HIS	PCA	IHS+W	PCA+W	M2
Tree-Grass	*1.92*	1.68	1.19	1.71	1.59	1.59
Road (1) –Building (1)	*1.99*	1.85	1.90	1.94	1.97	1.95
Road (1) –Building (2)	*1.89*	1.53	1.80	1.75	1.87	1.91
Road (1) – Building (3)	*1.54*	1.13	0.98	1.19	1.13	0.99
Road (2) – Building (3)	*0.97*	0.60	0.49	0.70	0.63	0.56
Building(1)–Building (3)	*1.76*	1.36	1.66	1.51	1.71	1.86

Classification accuracies cannot be used to conclude whether one method is better than the others considering the complexity of the urban features and the differences in the syntheses created by the different methods. It seems that all the methods have spectral variance to separate the major classes in the image and classified images also prove that there is no significant difference in classification.

5.3 Fusion in Automated Urban Mapping

The benefits of image fusion for urban mapping have been demonstrated by Couloigner *et al*. (1998a, 1998b) for SPOT5 PAN and XS data fusion. Road extraction algorithms and classification approaches that were suitable for the "low" spatial resolution images are not suitable for the feature extraction in urban environment due to high level of details and complexity (Couloigner & Ranchin, 2000). Only few works have been done using VHR fused MS images (Zhang & Wang, 2004; Shackelford and Davis, 2003a, 2003b; Zhang and Couloigner, 2005) in urban areas.

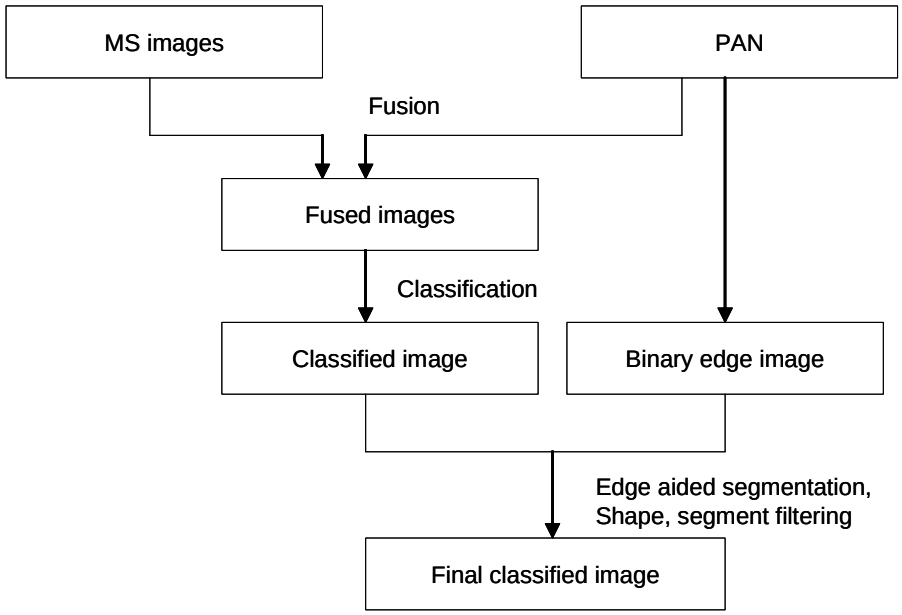


Figure 5.5 Road extraction method

(Zhang & Wang, 2004)

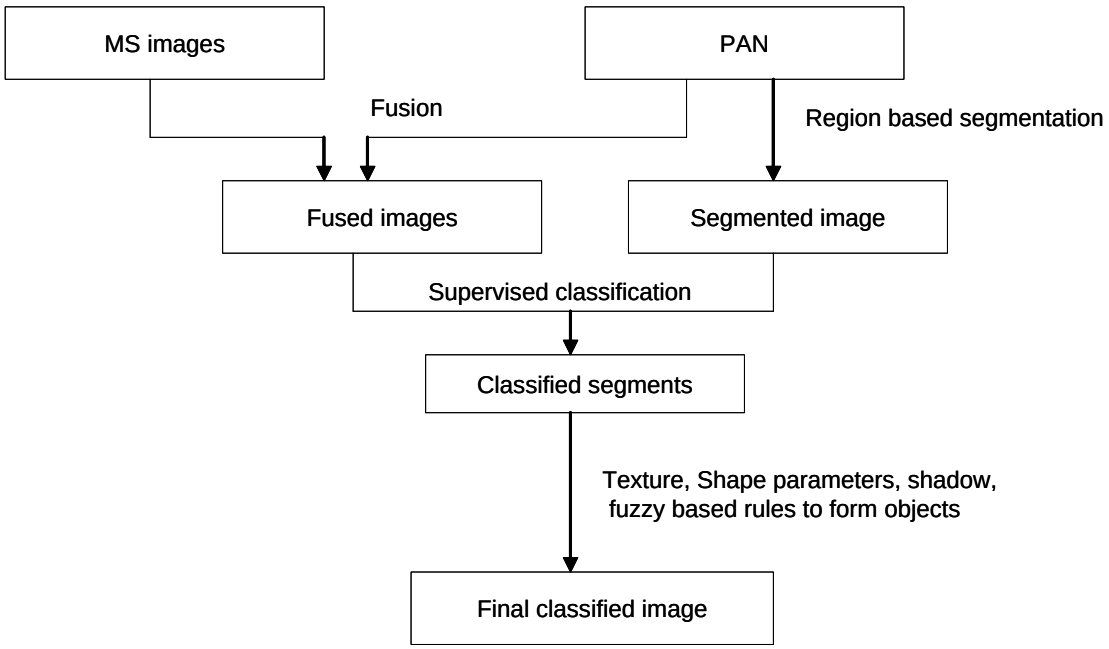


Figure 5.6 Segmentation based classification

(Shackelford & Davis, 2003b)

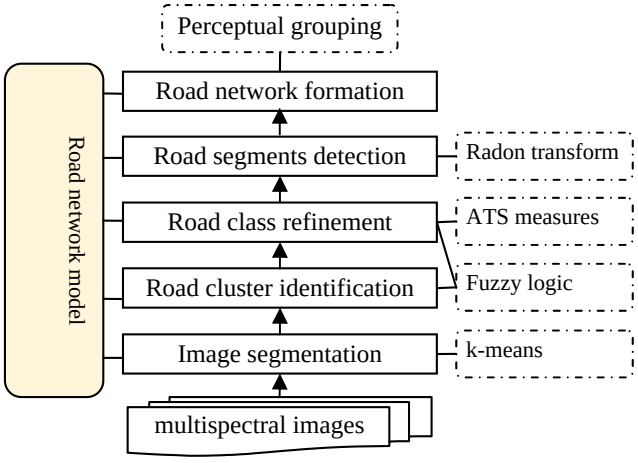


Figure 5.7 Angular texture signature based road network extraction

(Zhang and Couloigner, 2005)

Urban feature extraction algorithms either focus on extracting one feature of interest (e.g. roads) or on improving the classification approach for extracting different features (e.g. roads, buildings etc.). In either case, supervised or unsupervised classification alone is insufficient for a successful extraction due to the spectral similarities of the urban

features and the complexity of the urban environment. An example is provided in Figure 5.8. Figure 5.8a and b present a sub-urban area and its unsupervised ISODATA classified image. The challenge in extracting roads from these areas lies in the post-processing steps like a texture or a shape analysis. The problem increases in dense urban areas due to spectral similarities and more spatial details. Thus, several other processing steps such as edge detection, texture filtering, morphological shape or other contextual information like shadow are essential in the pre-processing or post-processing procedure(s) for a successful urban feature extraction from VHR images.

a) Original image b) Built-up area

Figure 5.8 Built-up areas by unsupervised classification

Automated urban mapping requires both high spatial and high spectral resolution. But the importance of having high spatial details in MS images decreases as original PAN images are available at higher and higher spatial resolution. Most edge detection and region based algorithms to obtain the edges of homogeneous segments are based on single images. The PAN image has high better spatial details (edges) as well as smooth texture that are more advantageous for segmentation. Thus spatial detail (sharp edges, small objects) in the MS images is of less importance for urban feature extraction.

5.3.1 Interpolation Vs. Fusion

The benefit of fusion in automated mapping depends on the objects in the image and the purpose of the fusion. Fusion generally provides better edges compared to interpolation. However, classification results of fused images and of interpolated images may not have differences in certain cases such as the one shown in Figure 5.9. Figure 5.9b and c show the road class obtained from the fused and interpolated images. They are quite similar except that the latter appears more nosiy. Thus the benefit of fusion for automated urban mapping depends on the area under consideration, the algorithm and methodology applied for the feature extraction.

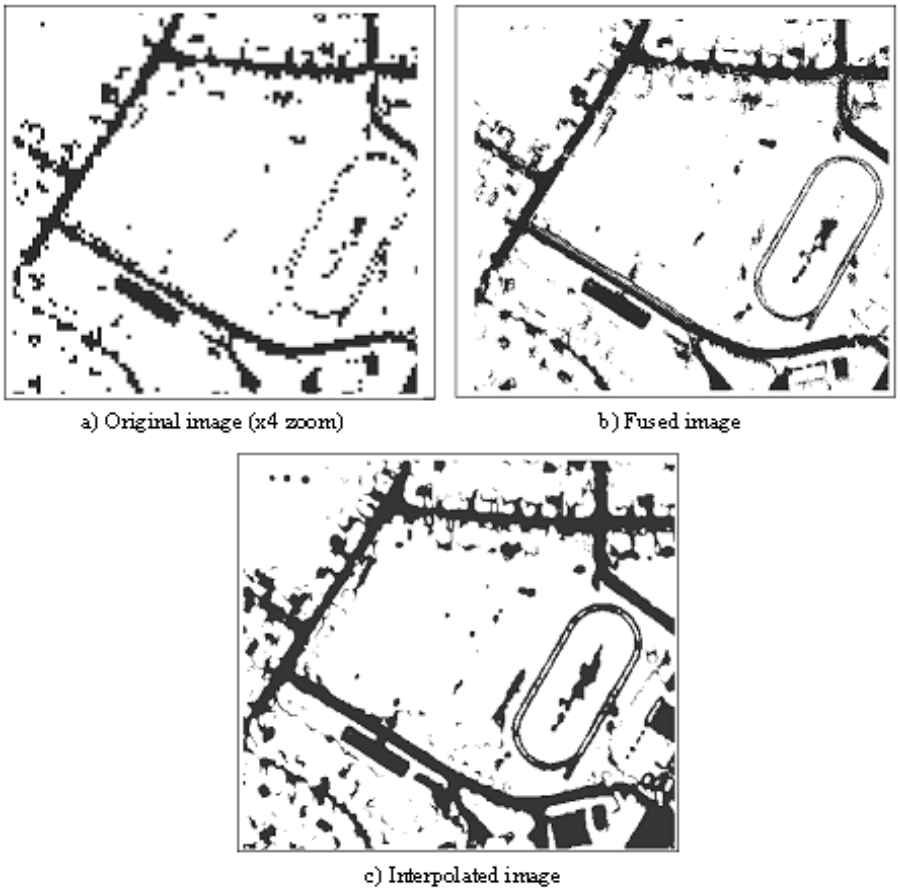


Figure 5.9 Built-up areas from unsupervised classification

5.4 Summary

The visual and statistical differences in the fused images of the different methods do not seem to have much effect on their classification. Considering the class separabilities, it can be only be said that the PCA, PCA+W, IHS+W and M2 methods provide a better separation of the urban classes compared to the IHS method. Spectral and textural information are important for urban feature extraction; however successful feature extraction from VHR images depends not only on the spectral information but mainly depends on the methodology (segmentation and other pre- and post-processing) of feature extraction.

Chapter 6 Conclusions and Future scope

The main objective of this study was to evaluate some pixel based fusion methods on VHR images for urban mapping. The methods were tested on Ikonos and Quickbird data sets covering residential and industrial areas.

The PAN of the VHR images has a low correlation with the Blue band and a high correlation with the NIR band. Thus, even though the spectral range of PAN covers the entire MS range, there are problems in the synthesis of the Blue and NIR bands. Based on the statistics of all the data sets, it is expected that the IHS, IHS+W, WA and WS methods provide better statistics for the Green, Red and NIR bands as they are highly correlated with PAN. The problem is only with the Blue band. The PCA method provides better results in all the bands for certain datasets, but they highly depend on the inter-band correlations. In VHR images, the Blue and the NIR bands have very low correlation that makes PCA not reliable for VHR images. The M2 method provides better statistics for the Blue, Green and Red bands. Since NIR is highly correlated with PAN, the M2 method overestimates the spectral values in the NIR band. The M2 method provides better results even when the inter-band correlations are low as the fusion is carried out separately for each band.

Reyes *et al.* (2004) compared different models and the ERGAS is reported to be low for the PCA-M1, ARSIS M2, PCA-M2, ARSIS-M3, and IHS-M2 methods. Thus it seems that IHS and PCA combined with wavelet provides better statistics comparable to the ARSIS models. The IHS+W and PCA+W methods in this study also provide overall better statistics.

Based on the visual analysis of the fused images for IK D1 and QB D1, the PCA and M2 preserve the colour of the objects in original images while providing high spatial details. For IK D2, the IHS+W method is better than the WA, WS and IHS methods in preserving the colour of objects. For QB D2, the fused images of IHS+W, WA and WS

present a very poor visual quality. The introduced high frequency information seems to be very low to meet the objective of the fusion. Thus, IHS seems better in terms of spatial detail and spectral content in the Red, Green and NIR bands for the QB D2 data set.

Classification results show that irrespective of the differences in the high frequencies introduced, all the fusion methods provide sufficient variances to discriminate between the vegetation and the urban objects such as roads and buildings that are also spectrally separated in the original images. Based on some preliminary study, it is found that it is not possible to obtain good discrimination between different building types and between certain building types and roads and parking lots by only conventional classification procedures.

Assessment protocol

The assessment protocol proposed by Thomas & Wald (2004) addresses two properties: consistency and synthesis properties. Both the properties involve degradation of the images and the effect of interpolation algorithm should be considered in evaluating the fusion methods. In the consistency property, any fusion method that introduces less high frequency will have better statistics and more important is that an interpolated image will have closer values to ideal. In the synthesis property, the original images have to be degraded to their respective lower resolutions and then fused to obtain the MS images at their original spatial resolution. The degradation process should simulate the image at lower resolution but seldom is the case with interpolation algorithms. And, this property also has a drawback that the quality at a higher resolution can not be predicted from the quality at a lower resolution.

The synthesis property may hold for lower resolutions (15 m, 30 m or less), in homogeneous areas and when the spectra of objects in MS bands are more predictable than from PAN. In degrading the images of VHR urban areas consisting of small and large objects, more than 50 percent of the information is lost. It is not possible to synthesize the spectra of lost objects or poorly represented objects by any fusion method.

Fusion methods may appear to synthesize the images at high resolution in certain bands when there is not much difference in the spectra of objects while there is change in spectral range and resolution. Even though the ARSIS models have a good theoretical frame for synthesis, the spectral values for details like edges and small objects cannot be synthesized close to the sensor observed ones as these models do not consider any physical laws governing the sensor characteristics.

The objective assessment criteria such as bias, SDD, DIV and CC for both properties only provide information about the amount of spectral distortion in each band by different methods. Thus, statistics are only useful for comparison and absolute quality can not be derived from the existing quantitative measures. Subjective assessment is always required in addition to an objective assessment. In this case, doing two fusions one at a high resolution and one at a degraded resolution do not seem to be necessary as the statistics from both will provide the same information. Spatial quality and spectral error are related to each other. When a method provides a good spatial quality, it also indicates that there might be high spectral error. Depending on the application, a trade-off between spatial and spectral quality is necessary.

Future Scope

The future Worldview satellite from Digital Globe will provide higher spatial resolution for PAN (0.5 m) and the 8-MS bands (2 m). With such a very high resolution, the need for the fusion of PAN and MS images has to be further investigated especially for automatic feature extraction procedures. No doubt that fusion increases the spatial resolution of the images. But due to the difficulty in developing operational feature extraction algorithms for VHR images, higher and higher spatial resolution further increases the problems in automatic mapping, especially in urban areas. Thus, pixel-level fusion of PAN and MS images is only a part of the extraction procedure, and successful urban feature extraction in fact needs feature- and object-level fusion of multiple sources.

References

Carper, W.J., Lillesand, T.M., and Keiffer, R.W., 1990. The use of Intensity-Hue-Saturation transformation for merging SPOT Panchromatic and multispectral image data. *Photogrammetric Engineering and Remote Sensing*, 56, 459-467.

Chavez, P.S., Stuart, J., Sides, C., Anderson, J.A., 1991. Comparison of three different methods to merge multiresolution and multispectral data: Landsat TM & SPOT Panchromatic. *Photogrammetric Engineering and Remote Sensing*, 57, 295-303.

Cornet, Y and Binard, M., 2004. Which metrics to assess and compare the quality of image fusion products? Proceedings of the 24th symposium of the European Association of Remote Sensing laboratories, Dubrovnik, Croatia, 25-27 May 2004 in New strategies for European Remote Sensing, Oluic (ed) 2005.

Couloigner, I., Ranchin, T. and Wald, L., 1998. Benefit of data fusion to urban mapping. Proceedings of the 2nd International conference on 'Fusion of Earth Data: merging point measurements, raster maps and remotely sensed data', pp. 183-190.

Couloigner, I., Ranchin, T., 2000. Mapping of Urban Areas: A Multiresolution Modeling Approach for Semi-Automatic Extraction of Streets. *Photogrammteric Engineering & Remote Sensing*, 66(7), pp. 867-874.

Dutilleux P. 1987. An implementation of the "algorithme à trous" to compute the wavelet transforms. In Actes du congrès ondelettes et méthodes temps-fréquence et espace des phases, Marseille, 14-18 décembre 1987, Springer Verlag Editors, pp.298-304.

Ehlers, M, 2005. Urban Remote Sensing: New Developments and Trends. Proceedings of URS 2005, March 14-16, 2005.

Herold, M., Gardner, M.E., Roberts, D.A., 2003. Spectral mapping requirements for mapping urban areas. *IEEE Transactions on GeoScience and Remote sensing* , 41(9), 1907-1919.

María González-Audícana, José Luis Saleta, Raquel García Catalán, and Rafael García, 2004. Fusion of Multispectral and Panchromatic Images Using Improved IHS and PCA

Mergers Based on Wavelet Decomposition. *IEEE Transactions on GeoScience and Remote Sensing*, 42(6), 1291-1299.

Martino, M., Causa, F., Serpico, S.B., 2003. Classification of optical high resolution images in urban environment using spectral and textural information. *IEEE International Geoscience and Remote Sensing Symposium*.

Meenakshisundaram, V., Couloigner, I., 2005. Quality Assessment of fusion methods for high resolution images. Proceedings of the URS 2005 conference, March 14-16, 2005.

Meenakshisundaram, V., Couloigner, I., 2004. Image Fusion of Ikonos pan and multispectral for classification of urban environment. Proceedings of the 24th symposium of the European Association of Remote Sensing laboratories, Dubrovnik, Croatia, 25-27 May 2004 in New strategies for European Remote Sensing, Oluic (ed) 2005.

Munechika, C.K., Warnick, J.S., Salvaggio, C., Schott, J.R., 1993. Resolution enhancement of multispectral image data to improve classification accuracy. *Photogrammetric Engineering and Remote Sensing*, 59, 67-72.

Núñez, J., Otazu, X., Fors, O., Prades, A., Palà, V., Arbiol, R., 1999. Multiresolution-based image fusion with Additive Wavelet Decomposition. *IEEE Transactions on Geoscience and Remote Sensing*, 37, 1204-1211.

Pohl. C, Van Genderen. J.L, 1998. Multisensor image fusion in remote sensing: concepts, methods and applications. *International Journal of Remote Sensing*. 19, 823-854.

Price, J.C., 1987, Combining panchromatic and multispectral imagery from dual resolution satellite instruments. *Remote Sensing of Environment*, 21, 119-128.

Ranchin, T., Aiazzi, B., Alperone, L., Baronti, S., Wald, L., 2003. Image fusion-the ARSIS concept and successful implementation schemes. *ISPRS Journal of Photogrammetry & Remote Sensing*, 58, 4-18.

Ranchin, T., Wald, L., 2000. Fusion of high spatial and spectral resolution images: The ARSIS concept and its Implementation. *Photogrammetric Engineering and Remote Sensing*, 66, 47-61.

Ranchin.T, Wald.L, 1997. Sensor fusion to improve the spatial resolution of images: The ARSIS method. Proceedings of the 17th EARSeL symposium "Future Trends in Remote Sensing" Lyngby, Denmark, 17-19 June. pp. 445-452.

Ranchin, T., Wald, L., Mangolini, M., Penicand, C., 1996. On the assessment of merging processes for the improvement of the spatial resolution of multispectral SPOT XS images. Fusion of Earth data, Cannes, France, 6-8, February.

Raptis, V.S., Vaughan, R.A., Hatzopoulos, I.N., Papapanagiotou, V., 1997. The use of data fusion for the classification of dense urban environments: The Mytilene case. Proceedings of 17th Symposium of EARSeL, Future Trends in Remote Sensing, pp. 435-443.

Reyes, R.A., Gutierrez, M.J., Fernandez, S., Thomas, C., Ranchin, T., and Wald, L., 2004a. Evaluation of the quality of Quickbird fused products. Proceedings of the 24th symposium of the European Association of Remote Sensing laboratories, Dubrovnik, Croatia, 25-27 May 2004 in New strategies for European Remote Sensing, Oluic (ed) 2005.

Shackelford, K.A., & Davis, H.S., 2003a. A Hierarchical fuzzy classification approach for high-resolution multispectral data over urban areas. *IEEE Transactions on GeoScience and Remote sensing*, 41, 1920-1932.

Shackelford, K.A., & Davis, H.S., 2003b. A Combined fuzzy pixel-based and object-based approach for classification for high-resolution multispectral data over urban areas. *IEEE Transactions on GeoScience and Remote sensing*, 41, 2254-2363.

Terretaz, P., 1997. Comparison of different methods to merge SPOT P and XS data: Evaluation in an urban area. Proceedings of 17th Symposium of EARSeL, *Future Trends in Remote Sensing*, Lyngby, Denmark, 17 – 20 June, pp. 435–443.

Wald, L., 2002. *Data Fusion: Definitions and Architectures-Fusion of Images of Different Spatial Resolutions*. Les Presses de l'Ecole des Mines, Paris.

Wald, L., 1999. Some terms of reference in data fusion. *IEEE Transactions on Geoscience and Remote Sensing*, 37, 3, 1190-1193.

Wald, L., Ranchin, T., Mangolini, M., 1997. Fusion of satellite images of different spatial resolutions: assessing the quality of resulting images. *Photogrammetric Engineering and Remote Sensing*, 63 (6), pp.691–699.

Weber, C., Ranchin, T., Puissant, A. and Serradj, A., 2003. Extraction of urban features in Strasbourg, France: Comparison of two fusion algorithms for Quickbird MS and Pan Data (manuscript accepted for publication).

Welch, R., Ehlers, M., 1987. Merging Multiresolution SPOT HRV and Landsat TM Data. *Photogrammetric Engineering and Remote Sensing.* 53, 301-303.

Zhang, Y., 2002. Problems in the fusion of commercial high-resolution satellite images as well as Landsat 7 images and initial solutions. *International Archives of Photogrammetry and Remote Sensing (IAPRS)*, 34 (4).

Zhang, Y., 1999. A new merging method and its spectral and spatial effects. *International Journal of Remote Sensing*, 20 (10), 2003-2014.

Zhang Q. & Couloigner I., 2005. Cluster analysis for road network extraction from multispectral imagery. To appear in *Proceedings of the 25th EARSeL Symposium: Global Developments in Environmental Earth Observation from Space*, Porto, Portugal - 6 - 11 June 2005.

Zhang, Y., and Wang, R., 2004. Multi-resolution and multi-spectral image fusion for urban object extraction. XXth ISPRS Congress.

Appendix 1

1. The MALLAT Undecimated Algorithm Vs. the "à-trous" Algorithm.

The results obtained for IK D1 are given here. The statistics of IHS+W and WA are slightly better than the one obtained with the à-trous algorithm. The variance is low in all the bands except in the WA fused NIR band. The spatial quality is poor compared to the M2 and PCA methods. A colour composite is shown in Figure 1.

Table 1 Statistics for IHS+W

Band	Bias	SDD	DIV	CC
Blue	0.24	5.98	-1.93	0.97
Green	0.26	5.85	0.47	0.98
Red	0.31	7.58	0.52	0.99
NIR	0.18	4.50	2.19	0.99

Table 2 Statistics for WA (1 plane)

Band	Bias	SDD	DIV	CC
Blue	0.23	9.07	-1.65	0.94
Green	0.24	9.27	4.57	0.97
Red	0.17	11.84	5	0.97
NIR	0.32	8.77	13	0.98

Figure 1 Fused colour composite by the IHS+W method

2. The ARSIS M2 and RWM models

In the ARSIS scheme, different models for MSM and IBSM exist. The statistics have been computed and are presented in Table 3 for the MSM model with filter size 3x3 and the IBSM models - Model 2 (M2), RWM (named after Ranchin, Wald, and Mangolini) - with different filters denoted as R7 (7x7 mask), R15 (15x15 mask), R21 (21x21 mask). The statistics for all the other models vary by only a few (1-4) percent. The visual quality of the fused images for R7, R15, R21 are not satisfactory in the Blue, Green and Red bands. The images appear to be noisy in certain areas and smoothed in others. A fused image using R7 in the Green band is given Figure 2.

Table 3 Statistics for different MSM and IBSM models for the ARSIS scheme

	Bias	SDD	DIV	CC
MSM3M2				
Blue	0.17	9.83	4.59	0.92
Green	0.16	14.06	3.94	0.92
Red	0.22	18.86	3.78	0.93
NIR	0.1	23.41	2.28	0.86
MSM3R7				
Blue	-0.07	9.89	-4.84	0.93
Green	-0.08	13.8	-2.82	0.93
Red	-0.12	19.18	-4.27	0.93
NIR	0	19.74	2.24	0.9
MSM3R15				
Blue	0.11	8.9	1.56	0.94
Green	0.13	13.52	-0.58	0.93
Red	0.15	18.72	-1.54	0.93
NIR	0.09	21.46	0.43	0.89
MSM3R21				
Blue	0.15	8.53	5.66	0.94
Green	0.16	13.21	2.72	0.93
Red	0.21	18.18	1.89	0.93
NIR	0.08	22.05	2.17	0.88

Figure 2 R7 fused image in Green band

Table 4 Statistics for IHS method for IK D1

	Band	431	421	432	321
Bias	*Blue*	-0.32	-0.31	-	-0.31
	Green	-	-0.3	-0.33	-0.3
	Red	-0.4	-	-0.42	-0.39
	NIR	-0.23	-0.23	-0.24	-
SDD	*Blue*	19.59	19.39	-	32.04
	Green	-	18.92	21.67	31.25
	Red	24.66	-	27.97	40.33
	NIR	14.3	14.16	16.21	-
Variance	*Blue*	2.97	3.37	-	-12.81
	Green	-	15	17.59	12.22
	Red	15.49	-	17.9	13.23
	NIR	1.82	2.4	-2.92	-
Correlation coefficient	*Blue*	0.7	0.7	-	0.64
	Green	-	0.85	0.81	0.6
	Red	0.87	-	0.83	0.26
	NIR	0.95	0.95	0.94	-
ERGAS		4.99	4.41	5.61	8.69

Note: 431, 421, 432 and 321 represent the bands. NIR Band – 4, Red Band -3, Green Band – 2, Blue Band - 1.

a) IHS **b) WA (1 plane)** **c) WA (2 plane)**

Figure 3 Fused Blue bands - QB D2

Appendix 2

Classification via Texture Analysis

Some results obtained with the classification of high resolution 4-m MS image and 1-m fused ARSIS M2 MS images are presented here. Those results are preliminary works to demonstrate the need of new classification methods in VHR images of urban areas to discriminate roads from parking/buildings (see Zhang and Couloigner, 2005). The classification accuracies obtained for the training pixels are given in Table 1.

Table 1

Maximum-Likelihood Classification results for the training pixels

Image	**Overall Accuracy**	**Kappa Coefficient**
4-m MS	89.11	0.866
1-m fused MS	90.65	0.886

The overall accuracy and the kappa coefficient obtained with the 4-m MS image do not vary much from those obtained with the 1-m fused MS image. The classification accuracy of the fused image was evaluated with the reference pixels and the confusion matrix obtained for the classes shown in Table 2.

Table 2

Confusion Matrix for the reference pixels obtained by the maximum-likelihood classification

	Reference pixels				
	Grass	**Tree**	**Road**	**Building**	**Parking**
Grass	11133	1387	69	0	396
Tree	1683	48789	0	1	0
Road	0	11	2951	479	8514
Building	0	178	551	7600	5319
Parking	0	0	1695	1061	10809
Shadow	0	6966	2	123	1
Pavement	0	0	632	24	2387
Vehicle	0	64	429	509	7566
Total	**12816**	**57395**	**6329**	**9797**	**34992**
User accuracy %	*86.86*	*97.21*	*63.42*	*78.83*	*59.33*

The percentage accuracy for *Road, Parking,* and *Building* were calculated with the *Pavement, Vehicle,* and *Shadow* pixels included for the *Road* class, *Shadow* pixels included for the *Building* class, *Pavement* and *Vehicle* pixels included for the *Parking* class. For the *Tree* and *Grass* classes, the accuracy obtained was more than 85%. The largest source of errors was due to the misclassification between the *Road* and *Parking* classes. The classification accuracy obtained for the *Parking* class is 59 % and for the *Road* class 63 %. Nearly 26% of *Road* class is classified under *Parking* class and 24% of *Parking* class is classified under *Road* class. Shadow pixels are more associated with *Building* class and *Tree* class. Even though there was a separate class for shadows, it was impossible to obtain a separation between the shadows and the sides of the building. The pixels around tree shadows and some vehicles are classified as *Building* in many areas. This is also a reason for misclassification between the *Road* and *Parking* class with *Building* class.

Textural classification of ARSIS fused image

Texture measures such as entropy, homogeneity, and contrast can be used to obtain a better separation of the classes. Texture measures were derived from the 1-m fused NIR image and 1-m PAN for window sizes of 25 x 25, 15 x 15 and 5 x 5. The classification accuracy was better for the texture measures derived from 1-m NIR than for the texture measures derived from 1-m PAN. The classification accuracies obtained with different texture measures are presented in Table 3.

The texture measures - homogeneity, angular second moment, and entropy - resulted in high classification accuracies for the *Grass* and *Tree* classes. It can be seen that the classification accuracies are better for the *Grass* and *Tree* classes with a 15 x 15 window. When the variance 5 x 5 texture increased the accuracy of the *Parking* class to 66 %, the accuracy of the *Road* class dropped to 62%. The classification accuracies increased for the *Road, Building,* and *Parking* classes to 80%, 80% and 57 % respectively with the mean 25 x 25 texture, whereas the accuracies of the *Grass* and *Tree* classes decreased considerably.

The classification accuracies are very high for the entropy 15 x 15 and the homogeneity

15 x 15 texture measures for the *Grass* and *Tree* classes. The entropy 15 x 15 measure was found to be best for differentiating the *Grass-Tree* class by visual analysis. Classified images obtained for the entropy 15 x 15 texture is shown in Figure 1 (c). Classification result obtained with only four spectral bands is shown in Figure 1 (b). It is obvious that the entropy texture measure has improved the differentiation between the *Grass-Tree* classes.

Table 3

Classification Accuracies obtained with maximum likelihood classification with different texture measures

	Grass	Tree	Road	Building	Parking
Homogeneity 25 x 25	99.9	98.3	75.5	80.8	42.5
Contrast 25 x 25	94.6	98.55	70.24	80.83	50.1
Mean 25 x 25	90.94	98.7	84.33	81.23	57.5
Variance 25 x 25	91.8	96.5	82.43	80.7	43.5
Angular second moment 25 x 25	99.9	97.42	87.7	81.9	45.8
Entropy 25 x 25	99.9	98.7	85.7	81.73	43.41
Homogeneity 15 x1 5	99.9	99.2	80.7	80.7	49.9
Contrast 15 x 15	96.2	98.2	76.26	81.5	54.8
Mean 15 x 15	97.5	97.6	82.9	81.02	42.9
Variance 15 x 15	89.6	98.6	82.9	79.3	44
Angular second moment 15 x 15	99.9	96.16	84.26	83.01	47.6
Entropy 15 x 15	99.9	99.06	86.16	82.1	46.2
Homogeneity 5 x 5	92.4	98.2	73.1	78.9	59.0
Contrast 5 x 5	89.5	97.1	61.7	68.8	64.5
Mean 5 x 5	91.2	96.6	55.5	77.3	59.37
Variance 5 x 5	91.8	95.6	62.1	74.3	65.9
Angular second moment 5 x 5	96.6	97.05	75.13	76.12	56.1
Entropy 5 x 5	94.1	97.9	74.5	77.01	54.9

With the enormous spatial information present in VHR images, it is possible to overcome the limitations of the spectral range and spectral resolution of these images to some extent. The accuracies obtained by the textural classification vary for different texture measures and window sizes. When a texture measure increases the accuracy of one class, it decreases the accuracy of the other classes for most of the cases. Thus, the selection of a suitable texture measure and window size is very important to obtain good results in textural classification. The classification accuracies obtained for certain classes can be compromised depending on the application at hand. For example, when the purpose of the classification is road extraction, the classification accuracies for the *Grass* and *Tree*

classes could be compromised to some extent and a texture measure that helps to better differentiate the *Road* class from the *Building* and *Parking* classes could be selected. However, specific texture measures like length-width texture (Shackelford, 2003) or angular texture measures are highly useful for road extraction and need to investigate in more detail.

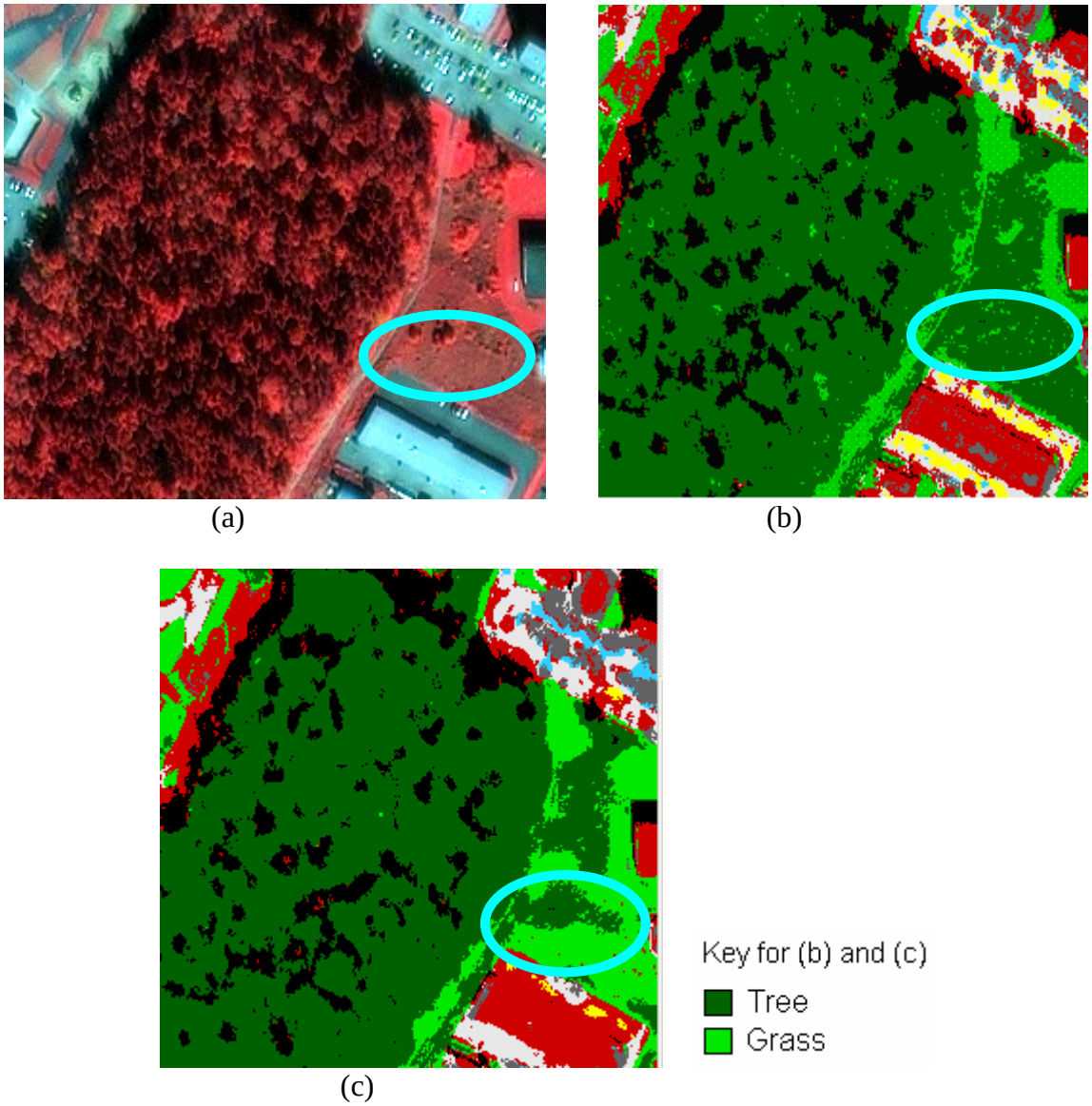


Figure 1. (a) Colour composite (RGB=432) of the 1-m M2 fused image. Maximum Likelihood classification of the 1-m M2 fused image (b) with 4 MS bands and (c) with 4 MS bands and the entropy 15 x 15 texture measure.