

**SUPER RESOLUTION METHODS FOR
REMOTE SENSING IMAGES**



M.Sc. THESIS

Çağlayan TUNA

**Department of Communication Systems
Satellite Communication and Remote Sensing**

JANUARY 2017

**SUPER RESOLUTION METHODS FOR
REMOTE SENSING IMAGES**

M.Sc. THESIS

**Çağlayan TUNA
(705141016)**

**Department of Communication Systems
Satellite Communication and Remote Sensing**

Thesis Advisor: Prof. Dr. Elif SERTEL

JANUARY 2017

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ BİLİŞİM ENSTİTÜSÜ

**UZAKTAN ALGILAMA GÖRÜNTÜLERİ İÇİN
SÜPER ÇÖZÜNÜRLÜK METOTLARI**

YÜKSEK LİSANS TEZİ

**Çağlayan TUNA
(705141016)**

İletişim Sistemleri Anabilim Dalı

Uydu Haberleşmesi ve Uzaktan Algılama

Tez Danışmanı: Prof. Dr. Elif SERTEL

OCAK 2017

Çağlayan TUNA, a M.Sc. student of ITU Informatics Institute Engineering and Technology 705141016 successfully defended the thesis entitled “SUPER RESOLUTION METHODS FOR REMOTE SENSING IMAGES”, which he/she prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Thesis Advisor : **Prof. Dr. Elif SERTEL**
Istanbul Technical University

Jury Members : **Prof. Dr. Elif SERTEL**
Istanbul Technical University

Prof. Dr. Bülent BAYRAM
Yıldız Technical University

Assoc. Prof. Dr. Gözde ÜNAL
Istanbul Technical University

Date of Submission : **30 January 2017**

Date of Defense : **04 January 2017**





To My Parents,



FOREWORD

I would like to express my sincere thanks to my thesis supervisor Prof. Elif Sertel for her great supervision, guidance and encouragement during my graduate studies. I consider myself lucky to have a chance to work with her throughout this project.

I would like to express my gratitude to Assoc. Prof. Gzde nal for helping me throughout my project, leading my way with her knowledge, kindly encouraging me to learn from my mistakes.

Above all, I would like to express my deepest thanks to my family members for their endless love, support and trust at all stages of my life.

January 2017

ađlayan TUNA

TABLE OF CONTENTS

	<u>Page</u>
FOREWORD	ix
TABLE OF CONTENTS	xi
ABBREVIATIONS	xiii
SYMBOLS	xv
LIST OF TABLES	xvii
LIST OF FIGURES	xix
SUMMARY	xxi
ÖZET	xxiii
1. INTRODUCTION	1
1.1 Purpose of Thesis	2
1.2 Literature Review	2
1.3 Thesis Outline.....	4
2. SUPER RESOLUTION	5
2.1 Observation Model and Operators.....	5
2.2 Single Frame Super Resolution	8
2.3 Multi Frame Super Resolution	8
2.3.1 Image Registration.....	9
2.3.2 Sub-Pixel Shifts Estimation.....	10
2.3.3 Data and Study Area.....	12
3. TIKHONOV AND TV BASED REGULARIZATION METHODS	15
3.1 Super Resolution As An Inverse Problem	15
3.2 Tikhonov Regularization	16
3.3 Total Variation Regularization.....	17
3.4 Iterative Solution	17
4. SUPER RESOLUTION BY DEEP LEARNING	19
4.1 Deep Learning	19
4.1.1 Convolutional Neural Network.....	19
4.2 Super Resolution By Convolutional Neural Network	20
4.3 Super Resolution By Very Deep Networks	22
4.4 Color Transformation By IHS	24
5. RESULTS	27
5.1 Quality Metrics	27
5.2 Multi Frame Super Resolution Experiments	29
5.2.1 Experiments on Synthetically Generated Images.....	30
5.2.2 Experiments on Real Data	31
5.3 Single Frame Super Resolution Experiments.....	32
5.3.1 Experiments on Synthetically Generated Images.....	32

5.3.2 Experiments on Real Data 34

6. CONCLUSIONS AND RECOMMENDATIONS 39

6.1 Summary and Contributions 39

6.2 Future Work 40

REFERENCES 41

CURRICULUM VITAE 45



ABBREVIATIONS

CNN	: Convolutional Neural Network
HR	: High Resolution
IHS	: Intensity Hue Saturation
KML	: Keyhole Markup Language
LR	: Low Resolution
PSNR	: Peak Signal Noise Ratio
RS	: Remote Sensing
SAM	: Spectral Angle Mapper
SR	: Superresolution
SSIM	: Structural Similarit Index
TV	: Total Variation
VDSR	: Very Deep Super Resolution



SYMBOLS

B	: Bias
D	: Downsampling Operator
H	: Blur Operator
F	: Warping Operator
W	: Convolutional Filters





LIST OF TABLES

	<u>Page</u>
Table 5.1 : Comparison of results with respect to two image similarity measures PSNR and SSIM.....	31
Table 5.2 : Results for Pleiades Multispectral.....	33
Table 5.3 : Results for Spot Multispectral.....	33
Table 5.4 : Results for Pleiades Panchromatic	33
Table 5.5 : Results for SPOT Panchromatic.....	34





LIST OF FIGURES

	<u>Page</u>
Figure 1.1 : Spatial resolution effect.....	1
Figure 2.1 : SPOT MS original image (left) and super-resolution (SR) result (right).	5
Figure 2.2 : Low and high resolution images, adapted from [24].....	9
Figure 2.3 : Image registration example, adapted from [27].....	10
Figure 2.4 : Multi frame image super-resolution process, adapted from [16]	11
Figure 2.5 : Satellite position while tri-stereo acquisition	12
Figure 2.6 : Pleiades and SPOT satellites, adapted from [30].....	13
Figure 2.7 : Difference of stereo and tri-stereo acquisition, adapted from [30].....	13
Figure 4.1 : Convolutional neural network, adapted from [36].....	20
Figure 4.2 : SRCNN, adapted from [14].....	21
Figure 4.3 : VDSR, adapted from [41].....	23
Figure 4.4 : Flowchart for SRCNN and VDSR implementation	24
Figure 5.1 : First synthetic low resolution images (LR) are created from original image on the left. The HR image reconstructed from the LR images (middle) is shown on the right.....	30
Figure 5.2 : SPOT multispectral and its VDSR result.....	31
Figure 5.3 : SPOT original PAN image (left) and SR result (right).	31
Figure 5.4 : Pléiades original MS image (left) and SR result (right).	32
Figure 5.5 : Pléiades original PAN image (left) and SR result (right).	32
Figure 5.6 : SPOT multispectral and its VDSR result.....	35
Figure 5.7 : Pleiades multispectral and VDSR result.....	35
Figure 5.8 : SPOT panchromatic and its VDSR result.....	35
Figure 5.9 : Pleiades multispectral and its VDSR result.....	36
Figure 5.10 : Pleiades panchromatic and Its VDSR result.....	36
Figure 5.11 : SPOT multispectral and its VDSR result.....	36
Figure 5.12 : SPOT panchromatic and its VDSR result.....	37
Figure 5.13 : Original Pleiades panchromatic image (left) and VDSR result of multispectral Pleiades image (right) by a scale factor of 4.	37
Figure 5.14 : Original Pleiades panchromatic image (left) and VDSR result of multispectral Pleiades image (right) by a scale factor of 3.	37



SUPER RESOLUTION METHODS FOR REMOTE SENSING IMAGES

SUMMARY

This thesis aims at enhancing the spatial resolution of single or multi frame remote sensing images. This process is referred to as super resolution (SR) image reconstruction in the literature.

Application of multi frame super resolution to real data presents challenges due to the very high dimensions of the datasets, such as those obtained in RS satellite imaging and due to the impacts of atmospheric, topographic, land cover characteristics and sun-sensor-geometry during the data acquisition. Besides, multi-frame super resolution for remote sensing images is not common because of lack of data. In this thesis, a novel application of tri-stereo RS satellite images based on regularization method to the super resolution problem is proposed. Since the tri-stereo RS images of the same observation area are acquired from three different viewing angles along the flight path of the satellite, these RS images are properly suited to a SR application. For multi frame super resolution, regularization methods which are Total Variation and Tikhonov are used to preserve edges.

The second contribution of the thesis is deep learning based single frame super resolution. Two pre-trained convolutional neural network models, SRCNN and VDSR, are applied to remote sensing images after IHS transform. Since this implementation is single frame super resolution, disadvantages of topography or moving objects doesn't effect the results such multi frame super resolution does; therefore, method can be applied to all kind of areas such as mountainous regions. Methods are implemented to Intensity channel only, then colour information is restored with an inverse IHS transformation.

As data, SPOT and Pleiades satellites images which are acquired by ITU Center for Satellite Communications and Remote Sensing (CSCRS) are used. Pleiades 1A, 1B RS satellite images having 0.5 m product spatial resolution for panchromatic mode and 2.0 m product spatial resolution for multispectral mode that include 4 spectral bands (Red, Green, Blue and IR); and SPOT 6 , SPOT 7 RS satellite images, which have 1.5 m and 6.0 m product spatial resolution for panchromatic and multispectral modes respectively. Structural Similarity Index (SSIM), Peak Signal Noise Ratio (PSNR), Spectral Angle Mapper (SAM) and ERGAS values are calculated for a quantitative evaluation of the methods, which showed satisfactory results with synthetic low resolution images. Finally, the method which gave better results, is tested within a real scenario, i.e. with original low resolution images and the obtained high resolution images demonstrated visible qualitative enhancements.



UZAKTAN ALGILAMA GÖRÜNTÜLERİ İÇİN SÜPER ÇÖZÜNÜRLÜK METOTLARI

ÖZET

Uzaktan algılama görüntülerinde çözünürlük, mekansal, spektral, radyometrik ve zamansal çözünürlük olmak üzere dört gruba ayrılır. Bu tez, tek veya çok çerçeveli düşük çözünürlüklü uzaktan algılama görüntülerinin mekansal çözünürlüğünü arttırmayı hedeflemektedir. Mekansal çözünürlük, algılayıcı tarafından kaydedilen görüntünün en küçük detayı olarak tanımlanabilir. Yani, uydu görüntülerinde, yeryüzündeki ayırt edilebilen en küçük objenin boyutu mekansal çözünürlüktür. Sensörlerin boyutunu arttırarak veya lensleri iyileştirerek kamera teknolojisini geliştirmek görüntülerin daha iyi algılanmasını sağlar ve görüntü çözünürlüğünü arttırmada birincil sıradaki seçenektir. Ancak uydu görüntüleri uygulaması için, fırlatılmış bir uydunun optiğini daha iyi bir teknolojiyle değiştirmek imkansıza yakındır. Bu nedenle yazılım tabanlı bir geliştirme gereklidir. Bir bölgeyi daha fazla pikselle göstermeye çalışma düşüncesi teoride doğrudur ancak yeterli değildir. Yani, sadece görüntünün boyutunu arttırmak tek başına bir çözüm değildir. İşte bu çözünürlük arttırma problemi için süper çözünürlük metotları mevcuttur. Bu çalışma literatürde, süper çözünürlükle görüntü iyileştirme olarak kullanılmaktadır.

Bu tez kapsamında ilk önce, uydu görüntülerine çok çerçeveli süper çözünürlük uygulanmıştır. Çok çerçeveli süper çözünürlük uygulaması, özellikle uzaktan algılama verilerinin yüksek boyutları nedeniyle zordur. Ayrıca, hem birden fazla uydu görüntüsü sağlamanın maliyeti hem de bir bölgenin yakın tarihli birden fazla kez çekilmiş uydu görüntülerinin azlığının neden olduğu veri kısıtlaması nedeniyle uzaktan algılama görüntüleri için çok çerçeveli süper çözünürlük çok yaygın değildir. Bu yöntemin uygulanabilirliğinin olması için, aynı alana ait farklı açılardan çekilmiş görüntülerin temin edilmesi gerekir. Yani, görüntüler aynı yeri farklı bakış açısıyla çekerek, birbirinden farklı bilgilere sahip olmalıdırlar. Bu farklı bilgiler, alt piksel kayması denilen görüntüler arası kayıklığın hesaplanmasıyla elde edilir. Bu tezde, özgün bir veri uygulaması olarak tri-stereo görüntülerle süper çözünürlük sunulmuştur. Tri-stereo uzaktan algılama görüntüleri, bir uydunun yörüngesi boyunca aynı bölgeden, üç farklı açıdan alındığı için, çok çerçeveli süper çözünürlük uygulaması için uygundur. Uygulama için, öncelikle bu üç görüntüden bir tanesi referans görüntü olarak seçilir. Referans görüntü, algılandığı açığa göre belirlenir. NADIR'e en yakın olan görüntü yapılan her uygulamada referans görüntüdür. Bunun sebebi, Yer Ornekleme Mesafesi NADIR'e yakın olan görüntüde en iyi olduğundandır. Daha sonra bu referans görüntünün diğer görüntülerle olan alt piksel kayması hesaplanır. Bu tezdeki çok çerçeveli süper çözünürlük uygulamaları sırasında, ölçü faktörü olarak "2" seçilmiştir. Yani, görüntünün çözünürlüğü "2" katına çıkarılmaya çalışılmıştır. Süper çözünürlük teorisine göre, düşük çözünürlük görüntüler, yüksek çözünürlüklü görüntünün alt örneklenmiş, bulanıklaşmış ve kaydırılmış halidir. Bu durumda düşük çözünürlüklü görüntüler, yüksek çözünürlüklü görüntü cinsinden yazılabilir.

Alt örnekleme değeri ölçü faktörüyle eşit ve "2"dir. Bulanıklaşma algılayan optiğin lensinden kaynaklandığı için bulanıklaşma değeri bütün görüntüler için eşittir çünkü görüntüler aynı optikle algılanmaktadır. Kayma parametreleri tam sayı olmamalı, yani görüntüler birbirlerinden elde edilememelidirler. Bu bilgiler ışığında, süper çözünürlük teorisine göre operatörler üretilmiş ve ters işlem gerçekleştirilerek yüksek çözünürlüklü görüntüye ulaşılmaya çalışılmıştır. Bu operatörler seyrek matrislerdir ama boyutları çok yüksek olduğundan bu uygulamanın hesaplama süresi yüksektir. Çünkü her piksel için ayrı ayrı işlem yapılmaktadır. Diğer yandan uydu görüntülerinin boyutları da normal kamerayla çekilen görüntülerle kıyaslanamayacak kadar büyüktür. Ayrıca, bu tezde kullanılan multispektral uydu görüntüleri 4 bantlıdır ve matematiksel süreç bu 4 banda ayrı ayrı uygulanmaktadır. Pankromatik görüntü uygulamalarında aynı durum geçerli değildir, çünkü pankromatik görüntüler yalnızca tek bantlıdır. Bu nedenle pankromatik görüntülerde yapılan uygulamalar daha hızlıdır. Çok çerçeveli süper çözünürlük uygulamasında kenarları korumak amacıyla düzenleme metotları olan Toplam Varyans ve Tikhonov kullanılmıştır. Uygulama, metotun doğruluğunu ölçmek amacıyla önce sentetik verilerle denenmiştir. Bunun için önce bir uydu görüntüsü alt örneklenip, bulanıklaştırılıp ve farklı parametrelerle kaydırılarak düşük çözünürlüklü görüntüler yaratılmıştır. Daha sonra mevcut yöntem bu düşük çözünürlüklü görüntüye uygulanmıştır. Bu sayede, yeni oluşturulan görüntü ve ilk baştaki orijinal görüntü birbirleriyle karşılaştırılabilir duruma gelmiştir. Bu yöntem görüntü füzyonu uygulamalarında da sıkça kullanılmaktadır. Görüntüler çeşitli metrikler yardımıyla karşılaştırılır ve metrik değerleri ikna edici bir seviyeye ulaştıktan sonra metodun başarılı olduğu sonucuna varılır. Bu tür uygulamalarda genellikle biküçük interpolasyonla genişletilmiş görüntüyle karşılaştırma yapılır. Bu tezde de bu görüntüyle karşılaştırılmıştır. En son, metot, gerçek verilere uygulanmıştır. Bu da görüntülerin çözünürlüğünü görsel olarak iyileştirmeye yaramaktadır. Yani tezin amacına ulaştığı sonuçlar bu sayede elde edilmektedir.

Tezin ikinci katkısı derin öğrenme ile tek çerçeveli süper çözünürlük metodunun uydu görüntülerine uygulanmasıdır. Derin öğrenme, makine öğrenme algoritmalarının en gelişmiş hali olarak kabul edilir ve son yıllarda bilim ve teknolojide en çok uygulaması yapılan konulardan bir tanesidir. Derin öğrenmenin diğer algoritmalarından en büyük farkları, öğrenilen parametrelerin sayısının ve eğitime giren veri sayısının çok fazla olmasıdır. Bu yöntem, görüntü sınıflandırma, video analizi, konuşma tanıma ve doğal dil öğrenme süreci dahil üzere çeşitli uygulamalarda çığır açmıştır. Son olarak, bir derin öğrenme metodu olan konvolüsyonel sinir ağları, iki farklı süper çözünürlük uygulamasında kullanılmış ve başarılı olmuştur. Bu uygulamaların farklılığı ağ modellerinin derinliği, filtre sayıları ve filtre boyutlarının farklılığından kaynaklanmaktadır. Derin öğrenme uzaktan algılamada da bir çok uygulamada kullanılmaya başlanmıştır. Ancak, uydu görüntüleri için konvolüsyonel sinir ağları ile süper çözünürlük ilk kez bu tezde kullanılmıştır. Konvolüsyonel sinir ağları, yapay sinir ağları ile belli boyuttaki filtrelerin eğitilmesiyle oluşturulur. Derin öğrenmede ağların mimarisi çok fazla değişkenlik göstermektedir. Buradaki filtre sayıları ve filtre boyutları uygulamadan uygulamaya değişmektedir. Bu süper çözünürlük uygulamalarında kullanılan iki ağ modeli birbirlerinden farklıdır. Eğitim aşamasından sonra oluşturulmuş bu modeller bu tezde kullanılmıştır. Kullanılan modeller tek kanallı uydu görüntüsüne uygulanmaya uygun olduğu için, uzaktan algılama verilerine IHS dönüşümü yapılmış, daha sonra önceden eğitilmiş konvolüsyonel sinir ağları modelleri uzaktan algılama görüntülerine uygulanmıştır. Metotlar sadece Yoğunluk kanalına

uygulanmış, daha sonra ters IHS dönüşümüyle renkli görüntü tekrardan elde edilmiştir. Bu sayede, hesaplama yükü azaltılmıştır. Kullanılan modellerin ikisi de görsel olarak iyi sonuç verse de bir tanesi diğerine göre metrik değerlerine bakıldığında çok daha iyi sonuç vermiştir. Metrik olarak SSIM, PSNR, SAM ve ERGAS kullanılmıştır. Çok çerçeveli süper çözünürlük de olduğu gibi, bu uygulamada da sentetik veri üretilip, ulaşılan sonuç ilk görüntüyle karşılaştırılmıştır. Daha sonra görsel iyileştirme sonuçları elde edilmiştir. Çok çerçeveli çözünürlük uygulamasında çözünürlük artırma kat sayısı 2 iken, bu uygulamada 2, 3 ve 4 kat uygulamaları yapılmıştır. İyi sonuç veren modelin, ağ mimarisi diğerine göre daha derindir. Bir tanesi yalnızca 3 katman oluşuyorken, daha derin model 20 katmandan oluşmaktadır. 3 katmanlı olan yalnızca SSIM değerlerinde iyi sonuçlar vermiştir.

Veri olarak, İTÜ Uydu Haberleşmesi ve Uzaktan Algılama Merkezinin sağlamış olduğu SPOT ve Pleiades uydularının görüntüleri kullanılmıştır. SPOT multispektral görüntüsünün mekansal çözünürlüğü 6 metre, pankromatik görüntüsünün 1.5 metredir. Pleiades'inki ise multispektral için 2 metre, pankromatik için 0.5 metredir. Pleiades için yer örnekleme aralığı yaklaşık 0.70'tir. Örneğin bir Pleiades tri stereo üçlüsü için başlangıç açıları sırasıyla 2.802, 6.846 ve 7.188'dir. Bu uygulamalarda bazı kısıtlamalar da vardır. Uydu, tri-stereo görüntüleri yaklaşık 30 saniye farkla almaktadır. Özellikle kalabalık ve trafiği çok olan şehirlerde, 30 saniyede bile, görüntüler arası hareket göze çarpmaktadır. O nedenle elde edilen yüksek çözünürlüklü görüntülerde kaymalar meydana gelebilmektedir. Örneğin bir araba iki farklı yerde gözükabilir. Ayrıca dağlık alanlarda, açı farkından dolayı görüntülerde büyük farklılıklar oluşmaktadır. Örneğin bir görüntü dağın bir yamacına bakarken diğeri başka bir yamacına bakıyor olabilir. Bu da görüntülerin birbirinden tamamen farklı görüntülermiş gibi algılanmasına sebep olabilir ve çok çerçeveli çözünürlük için kısıtlayıcı bir özelliktir. Aynı kısıtlamalar tek çerçeveli süper çözünürlük için geçerli olmadığı için tek çerçeveli süper çözünürlük uygulaması daha kolaydır. Ancak görsel sonuçlar, çok çerçeveliye göre daha varsayıma dayalıdır. Bu tezde, derin öğrenmeyle süper çözünürlüğün uydu görüntülerinde kullanılabileceği gösterilmiştir. Gelecek çalışması için, uydu görüntülerinin kullanıldığı bir eğitim aşamasıyla beraber oluşturulmuş konvolüsyonel sinir ağı modeli, bu tezdeki sonuçlardan daha iyi sonuç verebileceği öngörülmektedir.



1. INTRODUCTION

Digital images consist of small components which are called as pixels and a pixel size determine spatial resolution. Spatial resolution is a crucial measure of quality in satellite images providing information about the ability of separating small details on the viewed surface. Large pixel sizes compose lower resolution image because one pixel indicates smallest detail in a digital image. One of the aim of the spatial resolution enhancement methods is to increase details. Increasing the number of pixels per unit area could be thought as viable solution, but unfortunately it has limits [1]. In figure 1.1, effects of spatial resolution difference can be seen for same observation area.



Figure 1.1 : Spatial resolution effect

Improving the sensor and optic manufacturing technique is another method of getting high resolution image. As the more advanced sensors or optics are used, more detailed images can be obtained. Actually this is an expensive method because, beating the physical limitations of the sensors is requiring manufacturing costs. Due to the limitations of various sensors, acquisition of higher resolution images requires overcosting especially from satellites because, after satellites are launched, updating imaging devices is very difficult. Therefore, software solutions to lower this high cost as much as possible are desired. In today's world, higher resolution images are required in most applications such as medical imaging, military surveillance, satellite imaging etc. One of the key factors for success in image analysis is the amount of available informative data, which is determined by the resolution concept in imaging.

As the resolution increases, the quality of the analysis gets higher. Moreover, the demand for further quality would never end as the applications gets more sophisticated and talented. Although many image enhancement techniques have been developed, methods customized to remote sensing image data are required due to different properties such as spatial, spectral and radiometric resolutions. In this study, spatial resolution enhancement is implemented. High spatial resolution for multispectral or panchromatic images is required for accurate information extraction such as building detection, target identification and land cover/use mapping.

Image fusion is an alternative solution to increase spatial resolution. Multiple images with different spatial resolution of the same scene are collected and then fused to attain a high spatial resolution image [2]. However, this image fusion-based approach requires multiple images of the same scene, or the high-resolution PAN image being available. In practice, these images may not exist. Even if they are available, it is a nontrivial task to fuse them to precisely increase the spatial resolution due to errors in geometric registration and radiometric normalization [3]. On the other hand, it may be necessary to enhance resolution of panchromatic images in some applications however, image fusion does not enhance spatial resolution of panchromatic images.

1.1 Purpose of Thesis

Iterative gradient descent solution to the SR inverse imaging problem using two types of regularizers based on the Laplacian and the total variation terms and secondly pre-trained convolutional neural network models are applied to remote sensing images in this thesis. One of the goal of this thesis is to solve problem of remote sensing image super-resolution, which is the processed by multi frame low-resolution images to create an image with a higher spatial resolution. Another goal of this thesis is implementing pre-trained convolutional neural networks for remote sensing image super resolution.

1.2 Literature Review

SR image enhancement approaches have received a lot of interest in the past two decades, and a variety of SR algorithms have been proposed in the literature. Medical image processing such as [4], text image super resolution such as [5], video

enhancement such as [6], facial super resolution such as [7] and finally satellite imaging such as [8] are the some of the application areas of SR algorithm.

Super resolution methods are divided into 3 sub-categories: namely frequency domain approaches, reconstruction based and example based methods [1]. Frequency domain approaches, such as [9], are computationally efficient, but limited in their abilities to handle more complicated image degradation models and include various image priors as proper regularization [10]. Later works on super-resolution have been almost exclusively in the spatial domain with reconstruction based methods [11] and they preserve edges but do not reconstruct details effectively. These approaches have computational cost due to pixel-wise gradient solutions. Moreover, remote sensing images sizes are very huge comparing to camera photographs and this situation cause to get results very slowly. Latest trends in super resolution intensify on example (learning) based algorithms which contain a training phase between HR images and their LR versions [12]. These methods assume that the high-frequency details lost at the decreased resolution can be predicted by learning an SR model from an external LR and HR image pair dataset. Learning based image processing, including learning-based SR, always explores image pair priori which describes the dependency between training image pairs by image pair analysis. Sparse coding [13] and CNN [14] based methods are recent example based SR techniques. According to the image priors, the example-based methods achieve the state-of-the-art performance.

SR algorithms can be classified based on the number of the low resolution images included. [15] explains details of the single frame SR algorithm and [10] is one of the most important article for multi frame image super resolution.

The idea of applying SR techniques to remote sensing imaging has been developed for decades [16]. Though data satisfying the demand for SR are not easy to obtain, there have been a few successful applicable examples for real data. In [17], proposed a MAP algorithm and tested it with moderate-resolution imaging spectroradiometer (MODIS) remote sensing images. Satellites can acquire multi-temporal or multi-view images for the same area thus provide the possibility for SR [[18] [19] and [20].

1.3 Thesis Outline

The thesis contains six chapters. A brief introduction to resolution enhancement and literature review of super resolution are given in Chapter 1.

In Chapter 2, single and multi frame superresolution difference is explained according to observation model separately. Then, image registration, sub pixel shifts and data are introduced in the section multi-frame superresolution. As data tri-stereo remote sensing images and their property are presented.

Regularization based superresolution methods for multi frame superresolution are presented in Chapter 3. As a regularizer, we used Tikhonov and total variation and their mathematical details are analyzed. Later in Chapter 4, single frame superresolution by deep learning is described in details.

Chapter 5 includes implementations results for synthetic and real data for single frame and multi frame separately.

Chapter 6 is the conclusion part. First, thesis is summarized, and the contributions are highlighted. Moreover, a discussion is presented to show future directions of the problem.

2. SUPER RESOLUTION

Super resolution (SR) refers to generation of a High Resolution (HR) image from a decimated, blurred, low-resolution (LR) image set, which can be either a single frame or multi-frame that contains a collection of several images acquired from slightly different views of the same observation area. Purpose of SR is to increase the high-frequency components and remove the degradations caused by the imaging process of the low-resolution camera. It is assumed that low resolution images can be generated from the high resolution image, in other words low resolution images are the downsampled, blurred and shifted form of the high resolution image as it can be seen in figure 2.1.

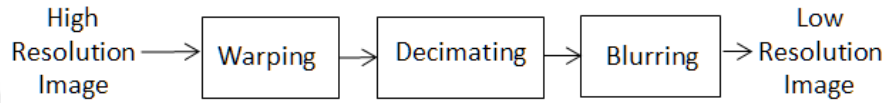


Figure 2.1 : SPOT MS original image (left) and super-resolution (SR) result (right).

Two types of super resolution approaches according to the number of low resolution input images are: single frame and multi frame that contains a collection of several images acquired from slightly different views of the same observation area [10], [21]. Multi-frame superresolution for remote sensing images is not common because of lack of data.

2.1 Observation Model and Operators

The first step to comprehensively analyze the SR image reconstruction problem is to formulate an observation model that relates the original HR image to the observed LR images. In imaging systems, traditional inverse problem formulations have been used with an imaging model given in equation 2.1.

$$Y_k = DHF_k X + n_k, \quad (2.1)$$

Here, Y_k represents low resolution images with k as the index of LR images. F_k is the operator of sub pixel shifts for each low resolution image. In this observation model, the main difference between single frame and multi frame super resolution formulation is apparent by number of available low resolution images; $k=1$ or $k=1, \dots, N$, respectively. N is the number of low resolution images. Down sampling operator is denoted by D and H is the blurring operator. D and H operators are the same for every low resolution image, because scale factor and lens which causes blurring are common features for each low resolution image. As a blur operator following filter is used;

$$H = \frac{1}{16} \begin{bmatrix} 1 & 2 & 1 \\ 1 & 4 & 1 \\ 1 & 2 & 1 \end{bmatrix} \quad (2.2)$$

These D , H and F operators can be combined and a system matrix could be created. X is the unknown and expected high resolution image. Size of these operators are very important in the solution process. Suppose HR image size is $L_1 N_1$ times $L_2 N_2$. L_1 and L_2 demonstrate scale factor of resolution enhancement. We chose scale factor as 2 for both dimensions. Therefore, our HR image dimension is $2N_1$ times $2N_2$ while LR images size is N_1 times N_2 . F_k shift matrix and blur operators sizes are $4N_1 N_2 \times 4N_1 N_2$, decimation operator size is $N_1 N_2 \times 4N_1 N_2$. For example, if LR image size is 100×100 , F_k and H operators sizes are 40000×40000 and operator of downsampling size is 10000×40000 while HR image size becomes only 200×200 . These huge matrices must be sparse to overcome high computational costs. For remote sensing images or for multispectral images these computations multiply for the number of available bands. Therefore, operations on panchromatic images which have only one band is always faster than multispectral images.

Creating these operators is very important step in the SR process. These matrices are sparse matrices but their sizes are very huge. For mathematical solution, images matrices is transformed to the vectorized form. Considering size informatin given above, downsampling operator of a 4×4 image is shown by;

$$D = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (2.3)$$

Downsampling or decimation operator in equation 2.3 creates 4x1 vectorized form of 2x2 low resolution image when multiplying with 16x1 vectorized form of 4x4 high resolution image.

If it is assumed that there is 0.4 pixel shift between two images in horizontal right direction, to warping images to another one this operator may be used;

$$F_x = \begin{pmatrix} 0.6 & 0.4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.6 & 0.4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.6 & 0.4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.6 & 0.4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.6 & 0.4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.6 & 0.4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.6 & 0.4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.6 & 0.4 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.6 & 0.4 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.6 & 0.4 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.6 & 0.4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.6 & 0.4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.6 & 0.4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.6 \end{pmatrix} \quad (2.4)$$

0.6 and 0.4 values are selected as weight coefficients of pixels. Actually, when a pixel is shifted 0.4, it is still close to first situation comparing to second pixel, as expected. Therefore; first pixel's coefficient is 1-0.4=0.6, because effect of first pixel is more than second pixel when shift value is less than 0.5.

If an image is shifted by 0.4 pixel from the reference image in vertical upward direction, to warp image downward, this operator in equation 2.5 may be used.

$$F_y = \begin{pmatrix} 0.6 & 0 & 0 & 0 & 0.4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.6 & 0 & 0 & 0 & 0.4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.6 & 0 & 0 & 0 & 0.4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.6 & 0 & 0 & 0 & 0.4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.6 & 0 & 0 & 0 & 0.4 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.6 & 0 & 0 & 0 & 0.4 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.6 & 0 & 0 & 0 & 0.4 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.6 & 0 & 0 & 0 & 0.4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.6 & 0 & 0 & 0 & 0.4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.6 & 0 & 0 & 0 & 0.4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.6 & 0 & 0 & 0 & 0.4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.6 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.6 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.6 & 0 \end{pmatrix} \quad (2.5)$$

Generally, both shifts in horizontal and vertical directions exist, therefore; this operators are combined by multiplying them;

$$F_1 = F_x \cdot F_y \quad (2.6)$$

F_1 in equation 2.6 is warping operator of only one image. Creating process of warping operator is repeated for every low resolution image respectively.

2.2 Single Frame Super Resolution

Single frame super resolution, as the name implies, is the reconstruction high resolution image from only one low resolution image. In the real world, sufficient images with supplementary information are some times difficult to acquire [16]. Interpolation methods such as bilinear, nearest and bicubic, are the most intuitive single frame superresolution implementations but it doesn't produce an ideal effect, since essentially there is no increase of image information [22] and the lost frequency components can not be recovered. Equation 2.1 is same for single frame super resolution except F_k because there is an only one image, translation is out of the topic. The main idea behind the single frame SR is to sharp edges smoothly.

Interpolation is the process of estimating the values of a continuous function from discrete samples. In most of the experiments, first, single frames are interpolated, then regularization or learning based methods are applied to extended image. In fact, nearest neighbour interpolation method just copies existing values. Hence, it is not convenient method for creating upsampled images. Bilinear interpolation determines weighted average of the four closest pixels and assigns output values according to these 4 pixels. On the other hand, bicubic interpolation uses 16 closest pixels in two dimensions which are horizontal and vertical. Bicubic interpolation gives the best results in terms of image quality according to the literature [23]. Therefore, in this thesis, as an interpolation method, bicubic is utilized for every application. Single frame super resolution is implemented by convolutional neural networks which will be explained in Chapter 4. Observation model and operators are only used for multi frame super resolution.

2.3 Multi Frame Super Resolution

Purpose of SR image reconstruction is to recover lost high frequency components, which are details in the image. In multi frame super resolution operations, low

resolution images are fused in a sense. This can only be achieved if the input LR images include different information, in other words, a different look at the same scene. There should be corresponding subpixel shifts between LR images in order to use them in SR image reconstruction process. In the implementation process, one of low resolution images is selected as a reference image and other images are used for extra information to create HR image. In figure 2.2, LR and HR image difference can be observed by super resolution example with "Cameraman" image.

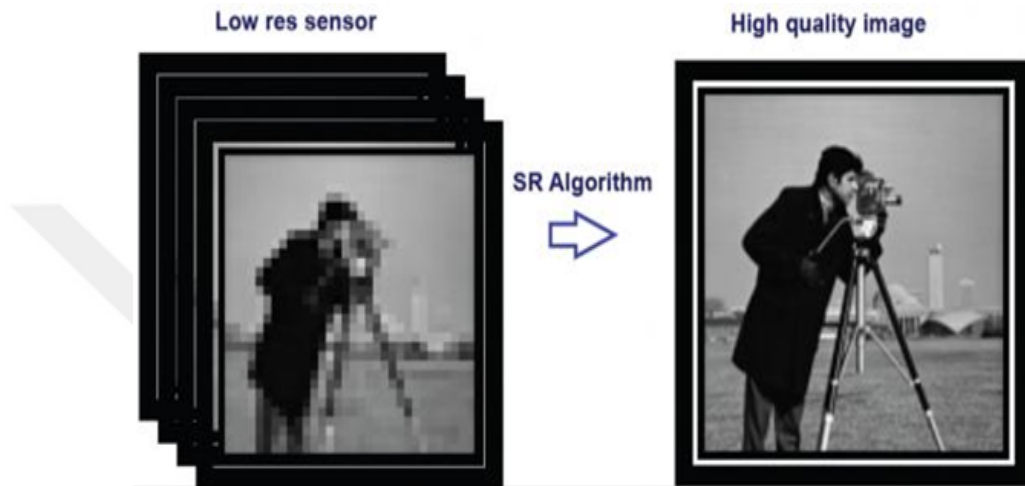


Figure 2.2 : Low and high resolution images, adapted from [24]

The super resolution problem is highly ill-posed due to insufficient number of the observed low-resolution frames [25]. Fortunately, this thesis has novelty in remote sensing are about data usage. Classical approach to reconstruct high resolution image is called regularization. In Chapter 3, detailed information will be given about regularization based super resolution. In this chapter preprocessing applications are presented which are registration and subpixel estimation. Then, data and its suitability to multi frame image super resolution is demonstrated.

2.3.1 Image Registration

Transforming different satellite images into a common coordinate system is called registration. Generally, before all sorts of remote sensing image processing applications, registration is implemented to obtain geometrically aligned images. Registration is a very important step to the success of the SR image reconstruction. In this study, first, images are registered onto a reference image, which is selected as the closest to Nadir having approximately zero incidence angle to minimize

geometric distortions caused by perspective view, then sub pixel shifts are calculated. Image-to-image registration is performed in ENVI version 5.2 [26]. Base image is chosen as a reference image and other two images are chosen as warp images. 120 tie points are used. As a warping method, polynomial and as resampling method cubic convolution are chosen. Despite registration, sub pixel shifts remain due to differences of viewing angles.

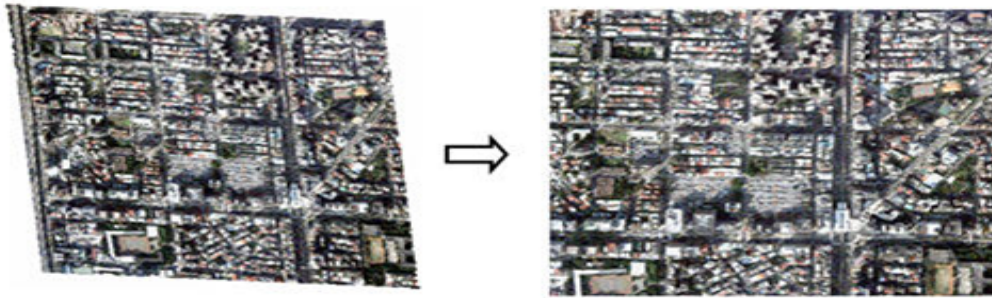


Figure 2.3 : Image registration example, adapted from [27]

In figure 2.4, the grids on the left side represent the LR images of the same scene with sub-pixel alignment, thus the HR image can be acquired by fusing the complementary information with SR methods [16].

2.3.2 Sub-Pixel Shifts Estimation

The key idea is strongly related to the fact that every in multi frame super resolution, low-resolution image contains different information on the same scene and the fusion of this information pieces, makes it possible to extract the subpixel information on the low-resolution image. The subpixel information means that new pixels are present among our existing pixel values that lead us to a higher resolved image. If only the low-resolution images have subpixel shifts, extra information of the scene is at hand. The new information within the low-resolution images can be exploited to get a higher resolution copy of the scene.

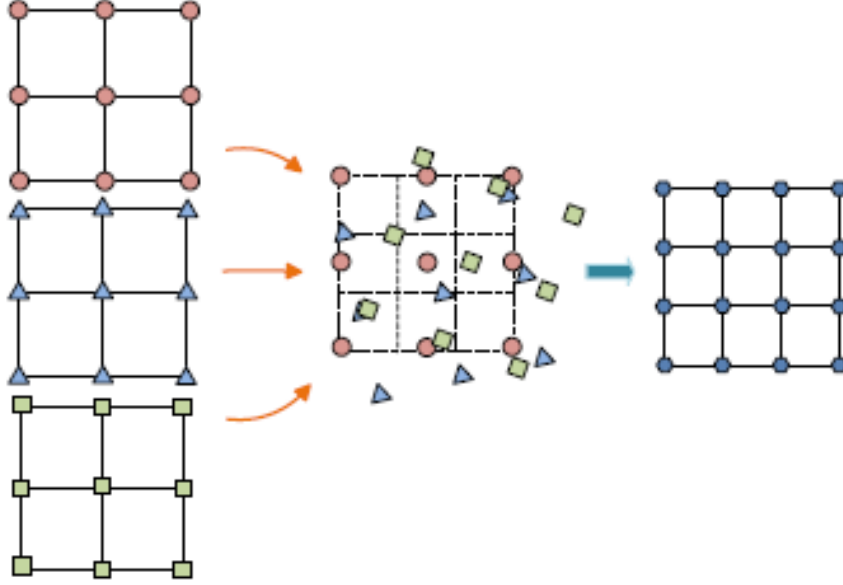


Figure 2.4 : Multi frame image super-resolution process, adapted from [16]

Sub pixel shift of low resolution images has to be different from each other. If the pixels are shifted by the integer values, then each low resolved image contain the same information and they will be the shifted version of the same image. Thus, to obtain a high resolution image different sub pixel shifts and aliasing are needed.

After registration, we estimated motion parameters by Fourier algorithm in [28], which estimates for the horizontal and vertical shifts. In this method, after the Fourier transform is applied to images then, phase differences with unknown slopes for all frequencies are computed between reference and other images.

$$Q(a,b) = \frac{F_1(a,b)F_2^*(a,b)}{|F_1(a,b)F_2^*(a,b)|} = e^{j2\pi(ux_0+vy_0)} \quad (2.7)$$

In equation 2.7, $Q(a,b)$ is the normalized cross-power spectrum by elemntwise multiplication in fourier domain. $F_1(a,b)$ and $F_2(a,b)$ are the fourier transforms of images at a and b points. If inverse fourier transform applied to this equation becomes;

$$Q(x,y) = \delta(x-x_0,y-y_0) \quad (2.8)$$

In equation 2.8, x_0 and y_0 shift values between two images because the peak position in Q is the desired offset.

2.3.3 Data and Study Area

Application of multi-frame super resolution to real data presents challenges due to the very high dimensions of the data sets, such as those obtained in RS satellite imaging [29]. In this study, we propose a novel application of tri-stereo RS satellite images to the super resolution problem. KML file can be seen in figure 2.5, tri-stereo images of the same area are collected from three different viewing geometry of the sensor which are called before, nadir and after. Since the tri-stereo RS images of the same observation area are acquired from three different viewing angles along the flight path of the satellite, these RS images are properly suited to a SR application.

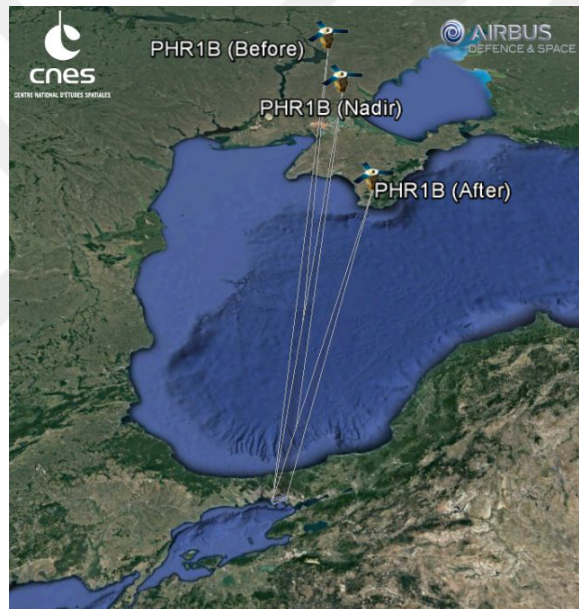


Figure 2.5 : Satellite position while tri-stereo acquisition

The following RS satellite data are provided by ITU Center for Satellite Communications and Remote Sensing (CSCRS), the very first satellite ground receiving station of Turkey: Pléiades 1A and 1B RS satellite images having 0.5 m product spatial resolution for panchromatic mode and 2.0 m product spatial resolution for multispectral mode that include 4 spectral bands (Red, Green, Blue and IR); and SPOT 6 and 7 RS satellite images, which have 1.5 m and 6.0 m product spatial resolution for panchromatic and multispectral modes, respectively. Both satellite images which are used in this thesis are from Istanbul. Spot acquisition time is 2013-12-20 and Pléiades acquisition time is 2014-10-21. In detail, Spot acquisition

time for tri stereo images are 08:31:39, 08:31:56 and 08:32:12 respectively. Istanbul is located north-western Turkey and it is the biggest city with a lot of buildings in Turkey. Therefore, its structure is generally complicated. Figure 2.5 shows that Pleiades and SPOT satellites are twin which means they can deliver very-high-resolution optical data products in record time and offer a daily revisit capability to any point on the globe.

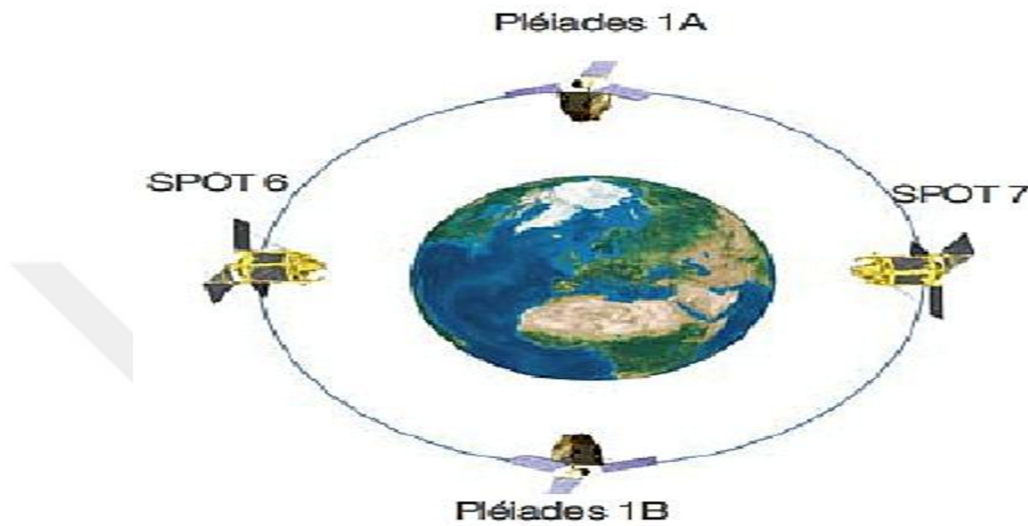


Figure 2.6 : Pleiades and SPOT satellites, adapted from [30]

To perform super resolution, low resolution image set should include different details from each other and as you see in figure 2.7, tristereo imaging system have more details than stereo imaging system for same observation area. We choose reference image for every application according to acquisition position. Being close to NADIR is the primary choice, for instance in this situation number 2 may be reference image, hence shift parameters between one and three and between two and three may be calculated separately.

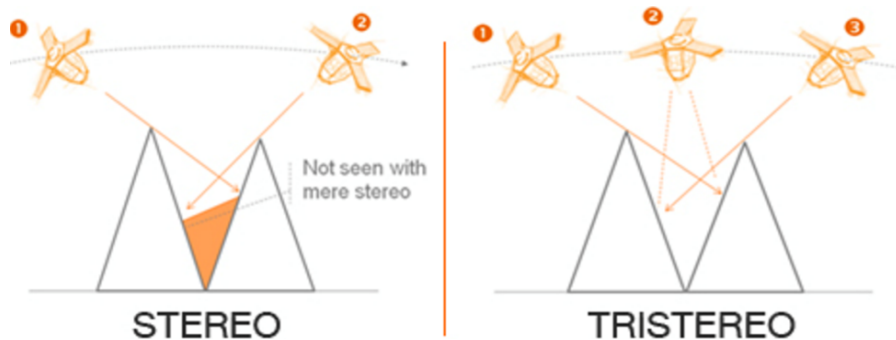


Figure 2.7 : Difference of stereo and tri-stereo acquisition, adapted from [30]

First number 2 may be interpolated to upsampled image, then new information may come from number 1 and number 3 images. Acquisition of tri-stereo lasts one and a half minute approximately and this is an advantage for to use them in super resolution process, because images are not affected by changement.



3. TIKHONOV AND TV BASED REGULARIZATION METHODS

SR is a typically ill-posed inverse problem that is hard to solve without the introduction of some prior image information. Thus, a number of regularization-based SR approaches have been proposed by incorporating the prior knowledge of the unknown high-resolution image in the regularization strategy [31]. In this Chapter, first, how inverse problem is solved will be explained, then detailed information will be given about regularization methods.

3.1 Super Resolution As An Inverse Problem

A matrix representation of imaging model in equation 2.1 reads as equation (3.1).

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \cdot \\ \cdot \\ Y_k \end{bmatrix} = \begin{bmatrix} D & H & F_1 \\ D & H & F_2 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ D & H & F_k \end{bmatrix} X + n_k, \quad (3.1)$$

As it is seen, operators are used as one under the other in image order. It is irrelevant to write HR, X, in matrix form, because there is only one image desired to reach. Decimation and blur operator is the same for every line because decimation scale and blur effect are the same for every image. Decimation scale is the sampling rate for reducing image size. All low resolution images must have same size to solve problem. Blur effect is caused by camera lens but in multi frame super resolution scenes are captured by same cameras. Moreover, using same satellite optic is requisite in our implementation. If different kind of satellite images are used, blur effect and any other features will be different for every image. In this equation while operators are used as in matrix form separately, HR and LR images are used as vectorized. The inverse problem in equation 3.1 can be solved by the minimization approach as given by:

$$X = \underset{X}{\operatorname{argmin}} \left[\sum_k^N \|(DHF_k X - Y_k)\|_p^p + \lambda U(X) \right], \quad (3.2)$$

In equation 3.2, p demonstrates the type of norm and it changes between 1 and 2. If p is 2, formulation becomes the least-squares method. Expansion of argmin is

"argument minimum" which defines smallest function value at a point. The first term is the data fidelity term which measures closeness of the estimated HR image X to the measurements, i.e. the LR images Y_k , the second term denotes the regularizer, which imposes a desired typically smoothness constraint on the unknown X and λ is the regularization parameter balancing these two terms. The minimization problem is solved by taking the gradient of the cost in equation 3.2 whereas data fidelity term varies according to norm selection:

$$\sum_k^N [F_k^T D^T H^T \text{sign}(F_k D H X - Y_k)] = 0, \quad (3.3)$$

$$\sum_k^N [F_k^T D^T H^T (F_k D H X - Y_k)] = 0, \quad (3.4)$$

Equation 3.3 and equation 3.4 represents L_1 and L_2 norm solutions for the data fidelity terms respectively. L_1 norm is also known as least absolute errors. It is basically minimizing the sum of the absolute differences between the target value and the estimated values. L_2 norm is also known as least squares. It is basically minimizing the sum of the square of the differences between the target value and the estimated values. Secondly, a regularization term should be selected. $U(x)$ is the regularization cost function. As a regularization Tikhonov and Total Variation methods are implemented in this thesis.

3.2 Tikhonov Regularization

One of widely used regularization term is Tikhonov cost function which is a highpass operator [17].

$$U(X) = \|\Gamma X\|_2^2, \quad (3.5)$$

Purpose of this operation is spatial smoothness by penalizing the derivatives of the Laplacian of the unknown image. Γ is usually a highpass operator such as derivative, Laplacian, or even identity matrix. As the noisy and edge pixels both contain high-frequency energy, they will be removed in the regularization process. As a Laplacian kernel, following matrix is used:

$$\Gamma_{kernel} = \frac{1}{8} \begin{bmatrix} 1 & 1 & 1 \\ 1 & -8 & 1 \\ 1 & 1 & 1 \end{bmatrix}, \quad (3.6)$$

Although other kernel filters can also be used as a regularizer in the super resolution problem, the kernel in equation 3.6 led to better results. This kernel is typically edge detection matrix. In the solution process, this kernel leads to signalize edges by adding to images which will be explained in the next section.

3.3 Total Variation Regularization

Another widely used regularization term is the total variation (TV) [32]. The TV function is utilized successfully in a number of image recovery methods including denoising, blind deconvolution, inpainting, and super resolution [32]. According to this principle, reducing the total variation of the signal subject to it being a close match to the original signal, removes unwanted detail whilst preserving important details such as edges. The TV criterion penalizes the total amount of change in the image as measured by the norm of the magnitude of the gradient and is defined as

$$U(X) = \|\Delta X\|_1^1, \quad (3.7)$$

where Δ is the gradient operator. TV criterion is powerful in the reconstruction due to preserving property its edge. Based on the spirit of TV criterion, Bilateral TV is introduced in [10].

$$\gamma_{BTV}(X) = \sum_l^P \sum_m^P [a^{(|m|+|l|)} (\|X - S_x^l S_y^m X\|)] \quad (3.8)$$

The bilateral TV regularization function is given in equation 3.8 as where S_x and S_y are the shift operators for X by l and m pixels in horizontal and vertical directions respectively. The scalar weight "a" value is between 0 and 1 and it is applied to give a spatially decaying effect to the summation of the regularization terms.

3.4 Iterative Solution

Combining the components in equation 3.3, equation 3.4, equation 3.6 and equation 3.8, the optimization iterations that updates the unknown HR image can be set up for as:

$$X_{(i+1)} = X_i - \sum_k^N [F_k^T D^T H^T \text{sign}(F_k D H X - Y_K)] + \Gamma(X), \quad (3.9)$$

$$X_{(i+1)} = X_i - \sum_k^N [F_k^T D^T H^T (F_k D H X - Y_K)] + \Gamma(X), \quad (3.10)$$

$$X_{(i+1)} = X_i - \sum_k^N [F_k^T D^T H^T \text{sign}(F_k D H X - Y_K)] + \sum_l^P \sum_m^P [a^{(|m|+|l|)} (||X - S_x^l S_y^m X||)], \quad (3.11)$$

$$X_{(i+1)} = X_i - \sum_k^N [F_k^T D^T H^T \text{sign}(F_k D H X - Y_K)] + \sum_l^P \sum_m^P [a^{(|m|+|l|)} (||X - S_x^l S_y^m X||)], \quad (3.12)$$

In the iterations, which are L_1 +Laplacian in equation 3.9, L_2 + Laplacian in equation 3.10, L_1 +BTV in equation 3.11 and L_2 +BTV in equation 3.12 respectively, as an initial HR image X_i , the interpolated reference LR image which is interpolated by bicubic method is selected.



4. SUPER RESOLUTION BY DEEP LEARNING

4.1 Deep Learning

Recent advances in computing power led to a burst of renewed interest in neural-network based machine learning algorithms. Deep learning refers to multi- and deep layer neural network algorithms which can be used for both feature extraction and classification. Deep learning methods successfully and favorably perform in many problems of the artificial intelligence field and its applications where a large number of training data is available. A significant difference of deep learning and other machine learning algorithms is surplus of trained parameters in deep learning. A recent publication by LeCun [33] set the direction of learning from data research towards deep learning, which involves different classes of deep belief networks (DBNs), deep Boltzmann machines (DBMs), SAEs, stacked denoising AEs (SDAEs) restricted Boltzmann machines (RBMs) and CNNs. Among these approaches, the most widely used is the CNN approach [34].

Deep Learning is used for RS data in many different areas such as image preprocessing, pixel-based classification, target recognition, feature extraction and scene understanding [35]. With the big data of RS images, deep learning techniques will be crucial in the future studies. Unlike natural scene images, HR RS images include various types of objects with different sizes, colors, rotations, and locations in a single scene, this situation causes difficulty of learning from RS images. The application of deep learning-based SR to remote sensing images does not exist in the literature. This is the goal of the current thesis.

4.1.1 Convolutional Neural Network

CNN consist of convolution, activation, pooling (subsampling) and fully connected layers and in each layer, filter weights and bias parameters are trained. In figure 4.1, a sample CNN network is demonstrated. In the traditional CNN, activation layer

simply consists of a pointwise nonlinearity function applied to each component in a feature map. This layer is generally chosen as rectified linear unit (RELU) which is $f(x)=\max(0,x)$. The pooling layer reduces the spatial size of the representation to reduce the amount of parameters and computation in the network, this provides to control overfitting.

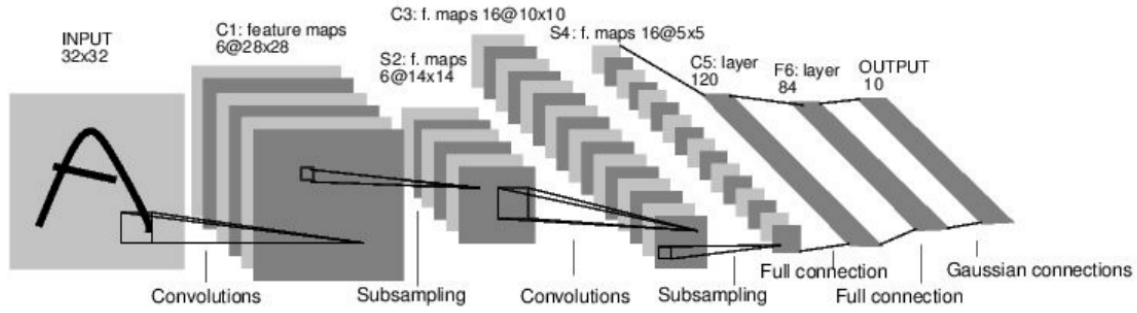


Figure 4.1 : Convolutional neural network, adapted from [36]

Convolutional Neural Networks are very similar to ordinary Neural Networks. CNNs have different kind of layers as it mentioned above and in that layers a lot of parameters, depending on the amount and size of filters, are trained. CNN architecture changes application to application or author to author. Different CNN architectures are utilized in the literature such as AlexNet [37] and GoogleNet [38]. In various computer vision areas such as classification, object detection and face recognition, CNN have recently shown superiority above other methods. In [14] and [39], CNN were applied to super resolution problem where they demonstrated very good results. In following sections, used models in this thesis are explained.

4.2 Super Resolution By Convolutional Neural Network

Firstly, SRCNN model, which is published in [14], is applied in this thesis. In order to enhance resolution of a low resolution image by a CNN, a series of adaptations are required. Goal of this implementation is to learn the mapping from a low resolution satellite image patch to a high resolution image pixel value, the input and output image sizes must be matched in order to produce the pixel intensity values of the HR image. Therefore, the size of low resolution images are scaled up by an interpolation method to the same size as the HR image. Then, filters are trained between the low and high resolution images.

In that model, the authors trained 91 images as a small training set and a large training set consists of 395,909 images from the ILSVRC 2013 ImageNet detection training partition. The size of training sub-images is 33 and the label images, is the image before downsampling, size is 21. As a result, the 91-image dataset can be decomposed into 24,800 sub-images, which are extracted from original images with a stride of 14, whereas the ImageNet provides over 5 million sub-images even using a stride of 33.

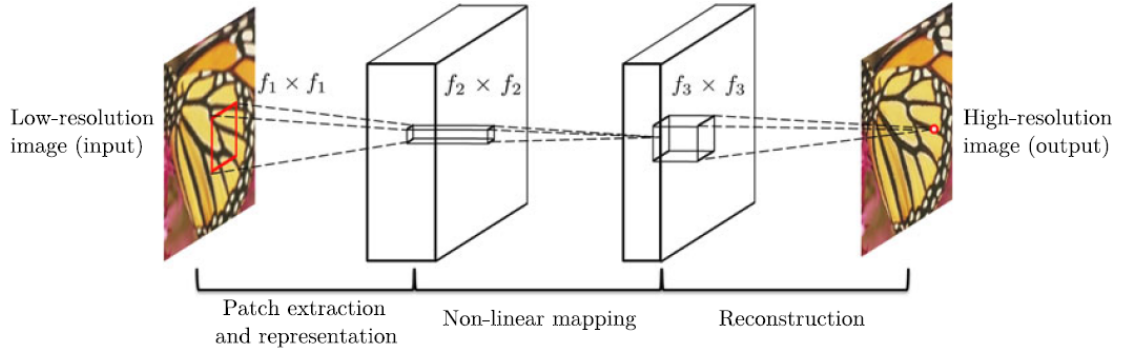


Figure 4.2 : SRCNN, adapted from [14]

Patch extraction and representation, non-linear mapping and reconstruction are operations of super resolution by CNN technique as it can be seen in figure 4.2. Every layer in network complete these processes respectively. After a convolution filter is applied in every layer, as an activation function, Rectified Linear Unit (RELU) is applied [40]. At the top layers of the network are the fully-connected layers, which provide the final learned mapping from an input LR (though upscaled) image patch to a HR image pixel value. Hence, the HR image is constructed at the output of the CNN network. After training, model can be applied to test images, because learned filters parameters are appropriate for the all kind of shapes in the images to sharp edges.

For validation, firstly low resolution images are created in this thesis. After creating the downsampled LR images, they are transformed through a IHS transform. The Intensity (I), Hue (v_1), and Saturation (v_2) matrices are up scaled by 2, 3 and 4 by bicubic interpolation, separately. Because, in this and following model, training phases are implemented for these scales separately. Visual enhancement flowchart can be seen in figure 4.4. To test this model, following equations are used;

$$X_1 = \max(0, W_1 * I + B_1), \quad (4.1)$$

where W_1 are the convolutional filters with size of $1 \times f_1 \times f_1 \times N_1$ and B_1 is N_1 -dimensional trainable basis parameter. X_1 is the first layer output feature map obtained by convolution. This operation equals to extract an N_1 -dimensional feature from each image patch by convolutional filters. The non-linear mapping layer can be expressed as

$$X_2 = \max(0, W_2 * X_1 + B_2), \quad (4.2)$$

here W_2 is the convolutional filter with size $N_1 \times 1 \times 1 \times N_2$ ($N_2 < N_1$) and B_2 is a N_2 -dimensional bias vector. X_2 is the output feature map after non-linear mapping. This operation equals to map each N_1 -dimensional vector into a N_2 -dimensional one. In the reconstruction layer, the high-resolution patches from the non-linear mapping layer are aggregated to generate the final high-resolution image. The third layer is presented as:

$$HR = \max(0, W_3 * X_2 + B_3), \quad (4.3)$$

where HR is the reconstructed image. W_3 and B_3 represent the averaging filters and the bias respectively. W_3 is the convolutional filter with size $N_2 \times f_3 \times f_3 \times 1$ and B_3 is a 1-dimensional bias vector.

4.3 Super Resolution By Very Deep Networks

This implementation, which is called VDSR [39], uses a very deep convolutional network inspired by VGG-net used for ImageNet classification. According to results, increasing network depth indicates critical improvement in accuracy. 20 weight layers, 3×3 for each layer, are used in the network which means it is very deep and information used for reconstruction (receptive field) is much larger.

64 filter of the size $3 \times 3 \times 64$, where a filter operates on 3×3 spatial region across 64 channels (feature maps) are utilized for the training. The first layer operates on the input image. The last layer, used for image reconstruction, consists of a single filter of size $3 \times 3 \times 64$. The network takes an interpolated low-resolution image (to the desired size) as input and predicts image details. Modelling image details is often used in super-resolution methods and CNN-based methods can benefit from this domain-specific knowledge.

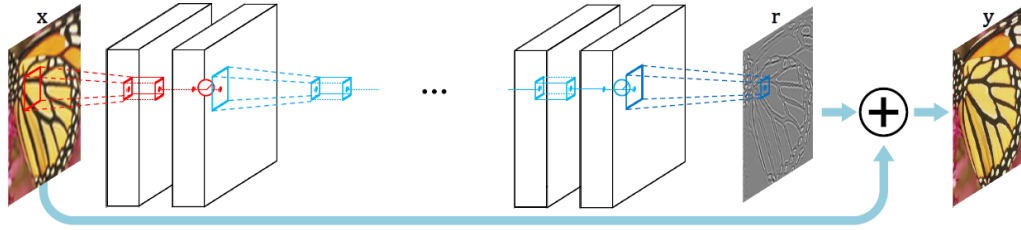


Figure 4.3 : VDSR, adapted from [41]

For training, SRCNN directly models high resolution images but VDSR is approaching in a different way. A high-resolution image can be decomposed into a low frequency information and high frequency information. Low and high resolution images have the same low-frequency information. This demonstrates that SRCNN aims carrying the input to the end layer and reconstructing residuals. Conversely, VDSR models the residual images. According to [39], residual learning converges much faster than non-residual learning.

Residual-learning CNN and extremely high learning rates are a way to speed up algorithm. As LR image and HR image share the same information to a large extent, explicitly modelling the residual image, which is the difference between HR and LR images, is advantageous. VDSR network structure provides efficient learning when input and output are highly correlated. Moreover, VDSR's initial learning rate is 104 times higher than that of SRCNN. This is enabled by residual-learning and gradient clipping.

In SRCNN, the network model contains only 3 layers but VDSR uses 20 layers which is a very deep network comparing to SRCNN as it can be seen in figure 4.3. A large receptive field means the network can use more context to predict image details. As SR is an ill-posed inverse problem, collecting and analyzing more neighbor pixels give more clues. In most cases, performance increases as depth increases. To quantify the SR quality, the whole pipeline that is presented in figure 4.4 is used for a validation scenario.

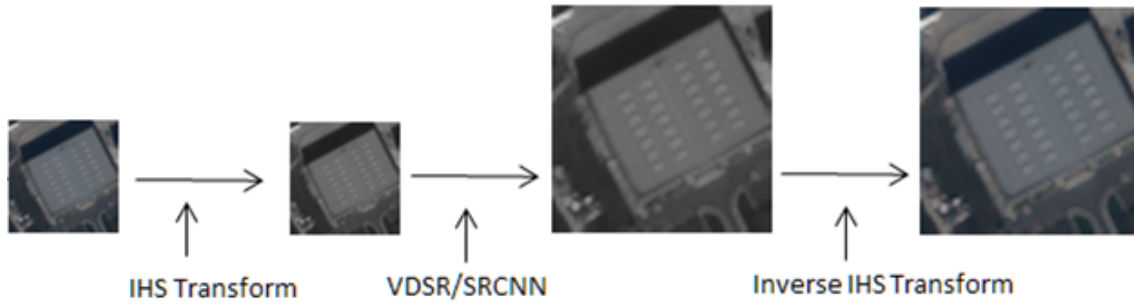


Figure 4.4 : Flowchart for SRCNN and VDSR implementation

4.4 Color Transformation By IHS

IHS transform converts RGB color space to the IHS color space. In the IHS space, spectral information is mostly reflected on the hue and the saturation. Intensity refers to the total brightness or dullness of a color, hue refers to what is perceived as color or the dominant wavelength of light, and saturation refers to the purity of the color [33]. IHS transform is widely used in image fusion in remote sensing applications [42] such as pan-sharpening, which is a kind of image fusion.

Among several forms of forward IHS and inverse IHS transforms, the following transform coefficients are utilized in this paper:

$$\begin{bmatrix} I \\ v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ -1/\sqrt{6} & -1/\sqrt{6} & 2/\sqrt{6} \\ 1/\sqrt{6} & -1/\sqrt{6} & 0 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}, \quad (4.4)$$

In equation 4.4, RGB image converted into I , v_1 and v_2 components, whereas the latter two are components of Hue and Saturation, respectively. I is the Intensity and that matrix retains the spatial details of the image. The proposed method that is described in the next section is applied to the I image only. Later, the inverse IHS transform in equation 4.7 is utilized to obtain the final RGB SR image. By equation 4.5 and 4.6, hue and saturation matrices can be obtained, but not requisite in this experiment. Because, v_1 and v_2 matrices are enough to get RGB image as it can be seen in equation (4). In this equation, interpolated v_1 and v_2 matrices are used according to scale factor.

$$H = \tan^{-1}(v_2/v_1), \quad (4.5)$$

$$S = \sqrt{v_1^2 + v_2^2}, \quad (4.6)$$

$$\begin{bmatrix} R' \\ G' \\ B' \end{bmatrix} = \begin{bmatrix} 1 & -1/\sqrt{6} & 3/\sqrt{6} \\ 1 & -1/\sqrt{6} & -3/\sqrt{6} \\ 1 & 2/\sqrt{6} & 0 \end{bmatrix} \begin{bmatrix} I \\ v_1 \\ v_2 \end{bmatrix}, \quad (4.7)$$

The proposed method is realized for panchromatic and multispectral images separately. For panchromatic images, IHS tranform is not required because those data have only a single band of gray image.





5. RESULTS

Image enhancement or improving the visual quality of a digital image can be subjective therefore; better quality image could vary from person to person. For this reason, it is necessary to establish quantitative measures to compare the effects of image enhancement algorithms on image quality and visual enhancement results should be evaluated by many image interpretation experts.

5.1 Quality Metrics

The structural similarity (SSIM) image quality paradigm is based on the assumption that the human visual system is highly adapted for extracting structural information from the scene, and therefore a measure of structural similarity can provide a good approximation to perceived image quality [43]. In this paper SSIM is chosen as one of image quality metrics. The SSIM index is a decimal value between 0 and 1 and a value of 0 implies zero correlation with the original image whereas a 1 implies the exact same image as the original image. Therefore, SSIM result which is close to 1, is the better result comparing to others. SSIM equation is given in equation 5.1.

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2\mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (5.1)$$

where μ_x , μ_y , σ_x , σ_y and σ_{xy} are the local means, standard deviations, and cross-covariance for images x, y.

The term peak signal-to-noise ratio (PSNR) is an expression for the ratio between the maximum possible value (power) of a signal and the power of distorting noise that affects the quality of its representation. PSNR is measured in decibels (dB). The higher the PSNR, the better the quality of the reconstructed image. The Mean Square Error (MSE) and the Peak Signal to Noise Ratio (PSNR) are the two error metrics used to compare image compression quality. The MSE shows the cumulative squared error between the compressed and the original image, whereas PSNR performs a measure of the peak error.

$$\text{MSE} = \frac{\sum_{j=1}^N \left(\sum_{i=1}^M (X_{i,j} - Y_{i,j})^2 \right)}{MN} \quad (5.2)$$

In the previous equation, M and N are the number of rows and columns in the input images, respectively. Then the block computes the PSNR using the following equation:

$$\text{PSNR} = 10 \log \frac{(\text{Max. Pixel Value})^2}{\text{MSE}} \quad (5.3)$$

ERGAS was proposed by [44] as a multi-modal index to characterize the quality of process and, presents the normalized average error of each band of processed image. Increasing in ERGAS index may be the result of degradation in images. Lower ERGAS value denotes more similar images.

$$\text{ERGAS}(A,B) = r100 \sqrt{\frac{1}{D} \sum_{c=1}^D \frac{\text{RMSE}(A_c, B_c)}{B_c}} \quad (5.4)$$

where A, B are the reconstructed image and reference image respectively, r is the ratio of resolution between images and D is the number of image bands. RMSE is the Root Mean Square Error which is defined as

$$\text{RMSE}(A,B) = \frac{1}{N} \sqrt{\frac{1}{D} \sum_{i=1}^n (a_i - b_i)^2} \quad (5.5)$$

where N is the total number of pixels. Spectral Angle Mapper (SAM) is a tool that permits rapid mapping of spectral similarity of image spectra to reference spectra [44]. The algorithm attempts to obtain the angle formed between the reference spectrum and the processed spectrum treating them as vectors in space by dimensionality equal to the number of bands, consequently SAM is a multi modal image quality metric.

$$\text{SAM}(A,B) = \arccos \left(\frac{a \cdot b}{\|a\|_2 \|b\|_2} \right) \quad (5.6)$$

The final SAM score is the average of all SAM values over the image. The SAM computes for spectral distortion and avoids radiometric distortion. if a, b are parallel but have different lengths their value will be the same. If SAM value is equal to zero, it means there is no spectral distortion.

There are a lot of quality metrics used in remote sensing studies. The reason of choosing above mentined metrics is that they are mostly used in literature.

5.2 Multi Frame Super Resolution Experiments

The registration between the chosen reference LR image and the other LR images are first performed to calculate the sub pixel shifts among the LR images. Sub pixel shifts values are found as 0.282 and 0.8334 in horizontal and 0.6337 and 0.1341 in vertical for SPOT tri stereo images. On the other hand, Pleiades tri stereo images subpixel shift values are 0.059 and 0.55 in horizontal and 0.4334 and 0.88 in vertical. Then, the warping, blurring, and down sampling matrix operators are created as sparse matrices to avoid high memory and computational requirements, which would otherwise make the RS-SR solution impractical. Finally, the overall system matrix, which is constructed based on the obtained operator matrices, is used to obtain the estimate HR image in one step in each iteration of the SR algorithm. We incorporated both the Laplacian and total variation regularizers separately into our algorithm.

For validation, we created synthetically down sampled, warped low resolution images from one of the tri-stereo images selected as the original-HR image, and performed the SR reconstruction described above. The result is compared with the original-HR image using the Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity (SSIM) measure. The latter is based on the assumption that the human visual system is highly adapted for extracting structural information from the scene, and therefore a measure of structural similarity can provide a good approximation to the perceived image quality [43]. The proposed SR methods are also applied to the real RS data. According to expert evaluations, the real data results show clear qualitative resolution improvements in the obtained HR images.

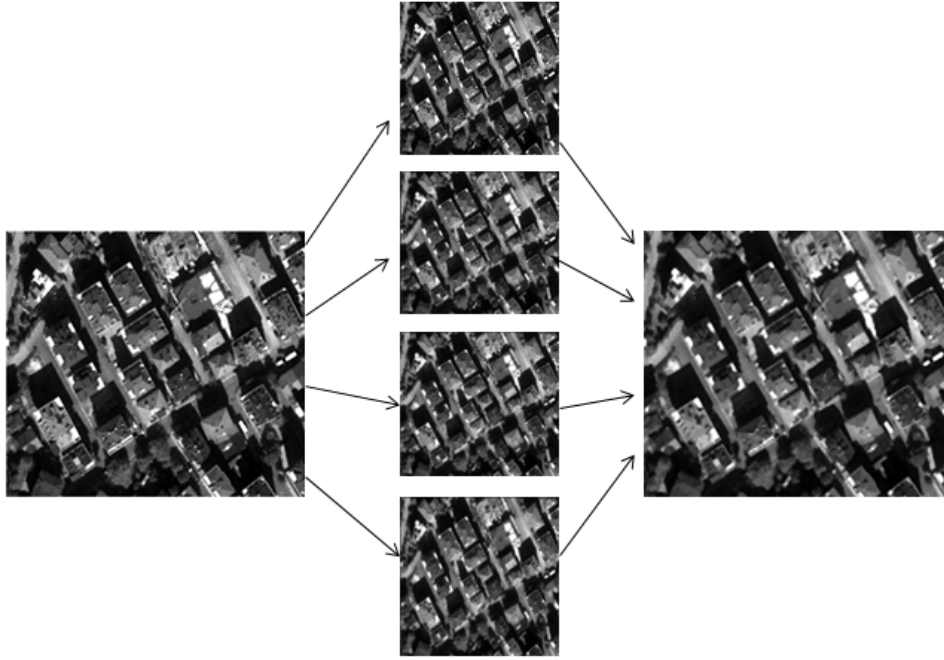


Figure 5.1 : First synthetic low resolution images (LR) are created from original image on the left. The HR image reconstructed from the LR images (middle) is shown on the right

To implement synthetic data scenario, we create four low resolution images by first decimating the high resolution image as it can be seen in Figure 5.1. Next, we choose a reference image and estimate other images according to this reference. At the same time, since we know warping parameters, we could control sub pixel shift calculating method accuracy. Lastly, the proposed super resolution methods are applied to these four low resolution images and results are compared with first original image.

5.2.1 Experiments on Synthetically Generated Images

Table 5.1 shows that except L1+Laplacian method, all test results give better results than those of the interpolated image. Both PSNR and SSIM metrics are confirmed the methods succes.

Table 5.1 : Comparison of results with respect to two image similarity measures PSNR and SSIM.

Method	PSNR	SSIM
Bicubic	23.9846	0.8177
L1+Laplacian	23.7745	0.8068
L1+BTM	24.0620	0.8280
L2+Laplacian	24.0428	0.8325
L2+BTM	24.1459	0.8480

5.2.2 Experiments on Real Data

The visual enhancements on the obtained HR images are observed. Some sample results are given in Figure 5.2, 5.3, 5.4 and 5.5.

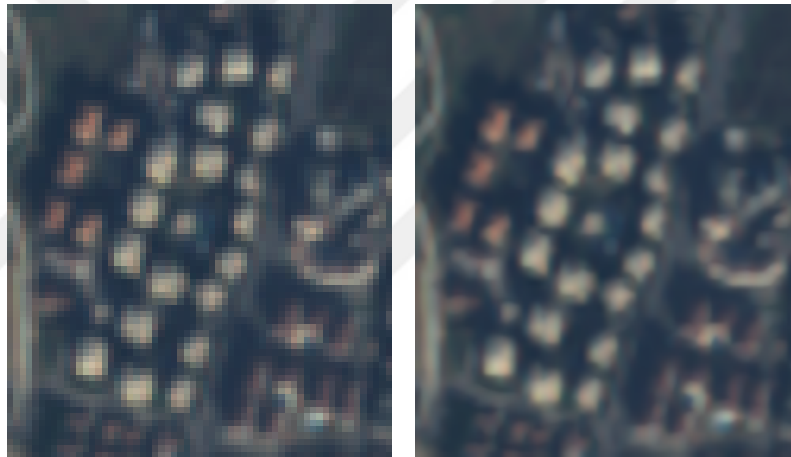


Figure 5.2 : SPOT multispectral and its VDSR result

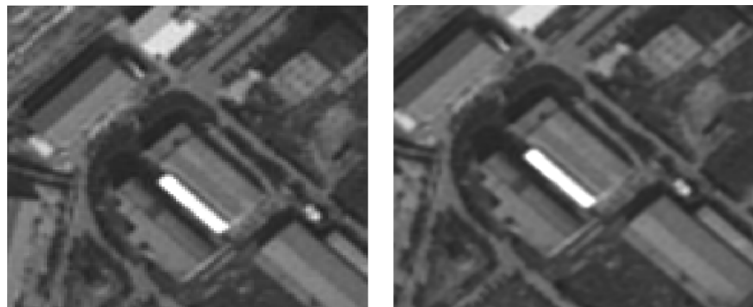


Figure 5.3 : SPOT original PAN image (left) and SR result (right).



Figure 5.4 : Pléiades original MS image (left) and SR result (right).

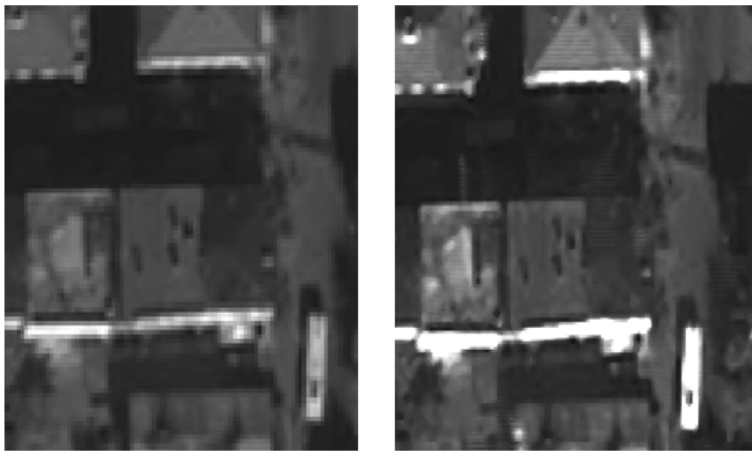


Figure 5.5 : Pléiades original PAN image (left) and SR result (right).

5.3 Single Frame Super Resolution Experiments

The flowchart in figure 4.4 illustrates the proposed approach graphically. As an input image, original RS data can be given. Reconstruction of Intensity image by VDSR/SRCNN is the first result of this method. Then inverse IHS transform is applied. The proposed method is realized for panchromatic and multispectral images separately. For panchromatic images, IHS transform is not required because those data have only a single band of gray image. To quantify the SR quality, whole pipeline that is presented in figure 4.4 is used for a validation scenario. For a real application, the first downsampling phase is removed and the input is given directly IHS transform phase.

5.3.1 Experiments on Synthetically Generated Images

In table 5.2, 5.3, 5.4 and 5.5, SSIM, PSNR, SAM and ERGAS results can be seen for bicubic, SRCNN and VDSR results for Pleiades Multispectral, Spot Multipsectral, Pleiades Panchromatic and SPOT Panchromatic images respectively. The best results

Table 5.2 : Results for Pleiades Multispectral

	Scale	Bicubic	SRCNN Result	VDSR Result
PSNR	2	23.2104	15.1006	24.6881
	3	18.9716	13.9766	19.4993
	4	16.6274	12.9718	17.3835
SSIM	2	0.9687	0.9678	0.9745
	3	0.9177	0.9117	0.9233
	4	0.8650	0.8583	0.8759
SAM	2	0.5102	0.6264	0.4980
	3	0.7520	0.8222	0.7307
	4	0.9528	0.9936	0.9155
ERGAS	2	0.9269	2.5274	0.7925
	3	1.4980	2.8742	1.4158
	4	1.9551	3.2245	1.8002

Table 5.3 : Results for Spot Multispectral

	Scale	Bicubic	SRCNN Result	VDSR Result
PSNR	2	24.7813	20.4221	25.2314
	3	19.9936	17.9177	20.1203
	4	17.6259	16.2477	17.7596
SSIM	2	0.9592	0.9665	0.9636
	3	0.8769	0.8798	0.8816
	4	0.7916	0.7905	0.7959
SAM	2	0.9289	1.1433	0.8927
	3	1.4989	1.5994	1.4682
	4	1.9688	2.0138	1.9400
ERGAS	2	1.5727	2.7336	1.5019
	3	2.6970	3.6783	2.6608
	4	3.5343	4.4787	3.4842

are indicated by boldface in the tables. 15 different images are implemented for every satellite images and for every scales. VDSR gave better results than bicubic method in all satellite images. However the same did not hold for SRCNN. One could argue that the deepness of VDSR model provides the desired performance in the SR.

Table 5.4 : Results for Pleiades Panchromatic

	Scale	Bicubic	SRCNN Result	VDSR Result
PSNR	2	23.7877	16.2573	24.6763
	3	19.3827	14.9937	19.9788
	4	16.8656	13.8396	17.4250
SSIM	2	0.8768	0.9094	0.8945
	3	0.6914	0.7072	0.7196
	4	0.5189	0.5144	0.5541

Table 5.5 : Results for SPOT Panchromatic

	Scale	Bicubic	SRCNN Result	VDSR Result
PSNR	2	25.4632	17.8374	26.1392
	3	21.4100	16.6753	21.5700
	4	19.4937	15.8567	19.6090
SSIM	2	0.8412	0.9009	0.8684
	3	0.5913	0.6286	0.6179
	4	0.3881	0.3937	0.4109

5.3.2 Experiments on Real Data

After quantitative validation found satiablely, visual enhancement is observed by applying CNN based super resolution to real remote sensing images. According to quantitative results, VDSR much better than SRCNN for remote sensing images. Therefore we applied only VDSR to real data to show visual enhancement results. Multispectral Pleiades and SPOT satellite images and their results with VDSR can be seen in figure 5.6, 5.7, 5.8, 5.9, 5.10, 5.11 and 5.12. Their improvement can be recognized both along line and edge features in images, as well as in homogeneous regions. The improvement is less visible in some parts for Pleiades Panchromatic images, as PHR is one of the highest resolution remote sensing satellites of the world. During these visual enhancement experiments, scale factor is chosen as 4 to see the best results. Secondly, to compare visual enhancements, original Pleiades panchromatic image and reconstructed Pleiades multispectral image can be seen in figure 5.13 and Original Pleiades panchromatic image and reconstructed Spot panchromatic image can be seen in figure 5.14 for same observation area.

Until this study, CNN based super resolution application to remote sensing images is limited. Therefore, our study has novelty in its first adaptation of CNN based super resolution to satellite imaging as well as incorporation of IHS transform domain for super resolution methods.

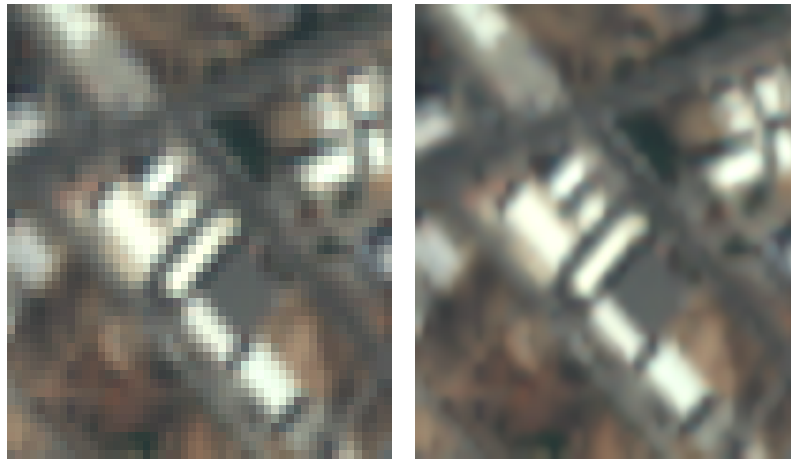


Figure 5.6 : SPOT multispectral and its VDSR result



Figure 5.7 : Pleiades multispectral and VDSR result

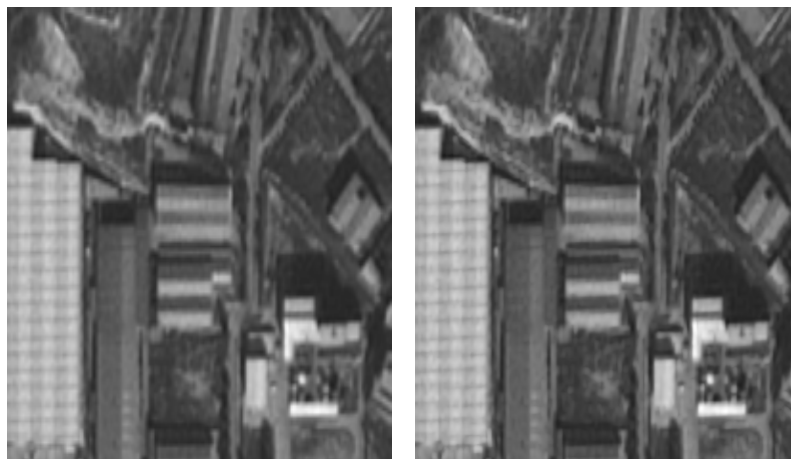


Figure 5.8 : SPOT panchromatic and its VDSR result



Figure 5.9 : Pleiades multispectral and its VDSR result

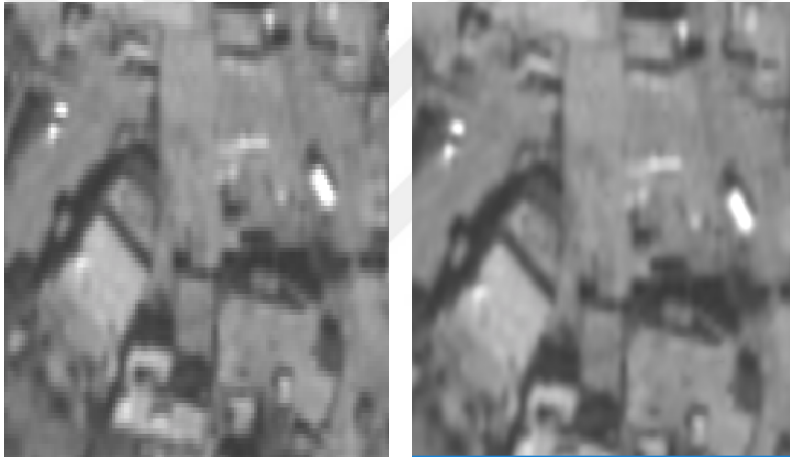


Figure 5.10 : Pleiades panchromatic and Its VDSR result

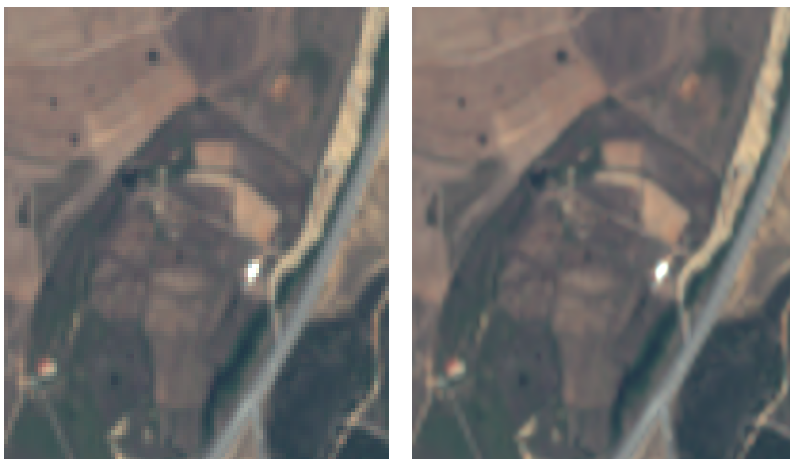


Figure 5.11 : SPOT multispectral and its VDSR result



Figure 5.12 : SPOT panchromatic and its VDSR result



Figure 5.13 : Original Pleiades panchromatic image (left) and VDSR result of multispectral Pleiades image (right) by a scale factor of 4.



Figure 5.14 : Original Pleiades panchromatic image (left) and VDSR result of multispectral Pleiades image (right) by a scale factor of 3.



6. CONCLUSIONS AND RECOMMENDATIONS

6.1 Summary and Contributions

To our knowledge, this is the first study that applies tri-stereo satellite images for multi-frame super resolution. Acquisition time is narrow comparing to other multi frame super resolution applications for remote sensing images. These results demonstrate the potential performance improvements in the super resolution of tri-stereo remote sensing images. On the other hand, problems in using tri-stereo images for multi-frame super resolution should be acknowledged. For instance, in mountainous regions, tri-stereo images have different radiometric values due to the impact of topography and sensor geometry, therefore, similarity optimization based methods are not ideal for those regions. Another disadvantage is for dynamic objects such as movement of vehicles on the roads and highways. Especially, in crowded cities, many cars exist and they are typically in motion. For Pléiades tri-stereo acquisition case, flight time of the satellite from one position to another is approximately 30 seconds in a sense that transposition is inevitable. To have three remote sensing images is more costly and less feasible than to have only one remote sensing images. Therefore, single frame super resolution is more advantageous if method is successful.

In this work, a state-of-art single frame super resolution method is presented for remote sensing images, based on deep learning based methods. VDSR showed high performance in all four quality metrics comparing to the other convolutional neural network model, SRCNN. PSNR, SSIM, ERGAS and SAM quality metrics are used for performance evaluations. Since remote sensing images have different features, generic methods from computer vision literature could not be directly applied, domain specific processing should be incorporated as we showed in this thesis. Almost all previous computer vision based related works, methods are applied to synthetic or basic camera images. In this thesis satellite images are used.

6.2 Future Work

As a future work, a dedicated training model of remote sensing image super resolution should be performed by training with deep network and GPU usage will be necessary due to huge matrix sizes of remote sensing images. Available deep learning libraries can be utilized to implement data training with Python and Matlab such as Caffe, Matconvnet, Theano, Torch to train data. We foresee that the training networks should be deep as the results in this thesis indicate that deeper networks perform better than in the remote sensing application of this thesis.



REFERENCES

- [1] **Park, S.C., Park, M.K. and Kang, M.G.** (2003). Super-resolution image reconstruction: a technical overview, *IEEE Signal Processing Magazine*, 20(3), 21–36.
- [2] **Polh, C. and Van Genderen, J.** (1998). Multisensor image fusion in remote sensing: concepts, methods and applications, *International Journal of Remote Sensing*, 19(5), 823–854.
- [3] **Zavorin, I. and Le Moigne, J.** (2005). Use of multiresolution wavelet feature pyramids for automatic registration of multisensor imagery, *IEEE Transactions on Image Processing*, 14(6), 770–782.
- [4] **Schatzberg, A. and Devaney, A.J.** (1992). Super-resolution in diffraction tomography, *Inverse Problems*, 8(1), 149.
- [5] **Park, J., Kwon, Y. and Kim, J.H.** (2005). An example-based prior model for text image super-resolution, *Eighth International Conference on Document Analysis and Recognition (ICDAR'05)*, IEEE, pp.374–378.
- [6] **Shen, M., Xue, P. and Wang, C.** (2011). Down-sampling based video coding using super-resolution technique, *IEEE Transactions on Circuits and Systems for Video Technology*, 21(6), 755–765.
- [7] **Baker, S. and Kanade, T.** (2000). Hallucinating faces, *Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on*, IEEE, pp.83–88.
- [8] **Zhang, H., Zhang, L. and Shen, H.** (2012). A super-resolution reconstruction algorithm for hyperspectral images, *Signal Processing*, 92(9), 2082–2096.
- [9] **Kim, S., Bose, N.K. and Valenzuela, H.** (1990). Recursive reconstruction of high resolution image from noisy undersampled multiframes, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 38(6), 1013–1027.
- [10] **Farsiu, S., Robinson, M.D., Elad, M. and Milanfar, P.** (2004). Fast and robust multiframe super resolution, *IEEE Transactions on Image Processing*, 13(10), 1327–1344.
- [11] **Nguyen, N., Milanfar, P. and Golub, G.** (2001). A computationally efficient superresolution image reconstruction algorithm, *IEEE transactions on image processing*, 10(4), 573–583.

- [12] **Freeman, W.T., Jones, T.R. and Pasztor, E.C.** (2002). Example-based super-resolution, *IEEE Computer graphics and Applications*, 22(2), 56–65.
- [13] **Yang, J., Wright, J., Huang, T.S. and Ma, Y.** (2010). Image super-resolution via sparse representation, *IEEE transactions on image processing*, 19(11), 2861–2873.
- [14] **Dong, C., Loy, C.C., He, K. and Tang, X.** (2016). Image super-resolution using deep convolutional networks, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(2), 295–307.
- [15] **Jiji, C., Chaudhuri, S. and Chatterjee, P.** (2007). Single frame image super-resolution: should we process locally or globally?, *Multidimensional Systems and Signal Processing*, 18(2-3), 123–152.
- [16] **Yue, L., Shen, H., Li, J., Yuan, Q., Zhang, H. and Zhang, L.** (2016). Image super-resolution: The techniques, applications, and future, *Signal Processing*, 128, 389–408.
- [17] **Shen, H., Ng, M.K., Li, P. and Zhang, L.** (2009). Super-resolution reconstruction algorithm to MODIS remote sensing images, *The Computer Journal*, 52(1), 90–100.
- [18] **Ma, J., Chan, J.C.W. and Canters, F.** (2012). An operational superresolution approach for multi-temporal and multi-angle remotely sensed imagery, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 5(1), 110–124.
- [19] **Chan, J.C.W., Ma, J. and Canters, F.** (2008). A comparison of superresolution reconstruction methods for multi-angle CHRIS/Proba images, *SPIE Remote Sensing*, International Society for Optics and Photonics, pp.710904–710904.
- [20] **Zhang, H., Yang, Z., Zhang, L. and Shen, H.** (2014). Super-resolution reconstruction for multi-angle remote sensing images considering resolution differences, *Remote Sensing*, 6(1), 637–657.
- [21] **Milanfar, P.** (2010). *Super-resolution imaging*, CRC press.
- [22] **Zhao, Y. and Wu, A.** (2015). Super-Resolution Image Reconstruction Based on Wavelet Transform and Edge-Directed Interpolation, *Journal of Communication and Computer*, 12, 73–78.
- [23] **Keys, R.** (1981). Cubic convolution interpolation for digital image processing, *IEEE transactions on acoustics, speech, and signal processing*, 29(6), 1153–1160.
- [24] **Esteve, E.**, <https://www.semiwiki.com>, date: 14.11.2016.
- [25] **Maiseli, B.J., Elisha, O.A. and Gao, H.** (2015). A multi-frame super-resolution method based on the variable-exponent nonlinear diffusion regularizer, *EURASIP Journal on Image and Video Processing*, 2015(1), 1–16.

- [26] **Guide, E.U.** ENVI on-line software user's manual.
- [27] **Supergeo**, http://www.supergeotek.com/library_1_200904.aspx, date: 22.11.2016.
- [28] **Vandewalle, P., Süsstrunk, S. and Vetterli, M.** (2006). A frequency domain approach to registration of aliased images with application to super-resolution, *EURASIP Journal on Applied Signal Processing*, 2006, 233–233.
- [29] **Li, F., Jia, X., Fraser, D. and Lambert, A.** (2010). Super resolution for remote sensing images based on a universal hidden Markov tree model, *IEEE Transactions on Geoscience and Remote Sensing*, 48(3), 1270–1278.
- [30] **CNES**, <https://directory.eoportal.org/web/eoportal/satellite-missions/p/pleiades>, date: 14.11.2016.
- [31] **Zhao, S., Jin, R., Xu, X., Song, E. and Hung, C.C.** (2015). A Variational Bayesian Superresolution Approach Using Adaptive Image Prior Model, *Mathematical Problems in Engineering*, 2015.
- [32] **Babacan, S.D., Molina, R. and Katsaggelos, A.K.** (2011). Variational Bayesian super resolution, *IEEE Transactions on Image Processing*, 20(4), 984–999.
- [33] **Chibani, Y. and Houacine, A.** (2002). The joint use of IHS transform and redundant wavelet decomposition for fusing multispectral and panchromatic images, *International Journal of Remote Sensing*, 23(18), 3821–3833.
- [34] **LeCun, Y., Bengio, Y. and Hinton, G.** (2015). Deep learning, *Nature*, 521(7553), 436–444.
- [35] **Romero, A., Gatta, C. and Camps-Valls, G.** (2016). Unsupervised deep feature extraction for remote sensing image classification, *IEEE Transactions on Geoscience and Remote Sensing*, 54(3), 1349–1362.
- [36] **Karpathy, A.**, <http://cs231n.stanford.edu/>, date: 17.12.2016.
- [37] **Krizhevsky, A., Sutskever, I. and Hinton, G.E.** (2012). Imagenet classification with deep convolutional neural networks, *Advances in neural information processing systems*, pp.1097–1105.
- [38] **Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. and Rabinovich, A.** (2015). Going deeper with convolutions, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.1–9.
- [39] **Kim, J., Kwon Lee, J. and Mu Lee, K.** (2016). Accurate Image Super-Resolution Using Very Deep Convolutional Networks, *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR Oral)*.

- [40] **Dahl, G.E., Sainath, T.N. and Hinton, G.E.** (2013). Improving deep neural networks for LVCSR using rectified linear units and dropout, *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, IEEE, pp.8609–8613.
- [41] **Kim, J., Lee, J.K. and Lee, K.M.** (2016). Accurate image super-resolution using very deep convolutional networks, *CVPR*.
- [42] **Ling, Y., Ehlers, M., Usery, E.L. and Madden, M.** (2007). FFT-enhanced IHS transform method for fusing high-resolution satellite images, *ISPRS Journal of photogrammetry and Remote Sensing*, 61(6), 381–392.
- [43] **Wang, Z., Simoncelli, E.P. and Bovik, A.C.** (2003). Multiscale structural similarity for image quality assessment, volume 2, IEEE, pp.1398–1402.
- [44] **Wald, L.** (2000). Quality of high resolution synthesised images: Is there a simple criterion?, *Third conference" Fusion of Earth data: merging point measurements, raster maps and remotely sensed images"*, SEE/URISCA, pp.99–103.

CURRICULUM VITAE

Name Surname:Çağlayan Tuna

Place and Date of Birth:Gaziantep 27.05.1991

E-Mail:caglayantun@gmail.com

EDUCATION:

- **B.Sc.:** 2015, Istanbul Technical University, Telecommunication Engineering
- **M.Sc.:** 2017, Istanbul Technical University, Satellite Communication and Remote Sensing

PROFESSIONAL EXPERIENCE:

- 2015- Project Student at Istanbul Technical University Research and Application Center for Satellite Communications and Remote Sensing (ITU-CSCRS)

PUBLICATIONS, PRESENTATIONS:

- Caglayan Tuna ; Alper Akoguz ; Gozde Unal ; Elif Sertel; Resolution enhancement of tri-stereo remote sensing images by super resolution methods . Proc. SPIE 10004, Image and Signal Processing for Remote Sensing XXII, 1000409 (October 18, 2016); doi:10.1117/12.2241176.