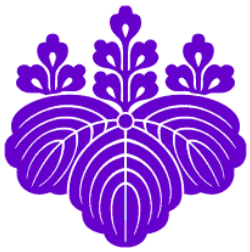
Study on Algorithms for Image Super-resolution based on Filtering and Learning Methods



Muhammad Haris

Graduate School of Systems and Information Engineering

Department of Intelligent Interaction Technologies

University of Tsukuba

A thesis submitted for the degree of

Doctor of Philosophy in Engineering

March 2017

”Bismillahi r-rahmani r-rahim”
In the name of Allah , the Entirely Merciful, the Especially Merciful.

”Laa ilaha illa-llah, muhammadur-rasulu-llah”
There is no God but Allah, and Muhammad is his messenger

“...Not even the weight of an atom that is within the Heavens nor in the Earth is hidden from Him...” (Surah Saba, Verse 3)

Acknowledgements

Spending 5 years living and studying in Tsukuba University is undeniably one of life episodes that I will cherish the most. During these years I met truly impactful people that broaden my perspective, and without their support (in any respect), this study would never been completed. Especially, I must express my deepest gratitude to my academic advisor, Professor Hajime Nobuhara for his endless support and kind assistance during this past 6 years. From him I learn how to see the problem, and derive the optimal solution.

I am very much indebted to Prof. Tsutomu Maruyama and Prof. Itaru Kitahara for tremendous support and suggestion during research process. I also thank my thesis committee members: Prof. Yasunori Endo and Prof. Shibuya Sensei for review and comments. I would also like to extend my heartfelt gratitude to CMU Lab Members for the warmth and supportive environment during my study, especially Hashikami Hidenobu who introduce me to image super-resolution.

This sincere gratitude also goes to my favorite brothers, Abdul Karim and Mahmoud Ben Othman thank you for becoming more than good friends and may Allah pour you with his blessings.

I gratefully acknowledge the scholarship I received during doctoral's degree as this journey would not have been accomplished without financial support from Indonesia Endowment fund (LPDP).

Most important, I want to thank the core supporter of my study. My parents, sisters and brother, thank you for showering me with abundant cheering and unending support. For my wife, Dita, who has been through thick and thin for these past three years, thank you for the continuous encouragement. For Shofiyyah Kyoko, our first baby daughter, thank you for becoming your parents' source of happiness and strength. This thesis is dedicated for both of you.

Abstract

This thesis focuses on developing theory and algorithms for the single-image super-resolution problem based on filtering and learning methods. Our proposed methods are divided into three categories.

First part, First-order Derivatives- based Super-resolution is filtering based method. A single fast super-resolution method based on first-order derivatives from neighbor pixels is proposed. The basic idea of the proposed method is to exploit a first-order derivatives component of six edge directions around a missing pixel; followed by back projection to reduce noise estimated by the difference between simulated and observed images. Using first-order derivatives as a feature, the proposed method is expected to have low computational complexity, and it can theoretically reduce blur, blocking, and ringing artifacts in edge areas compared to previous methods. Experiments were conducted using 900 natural grayscale images from the USC-SIPI Database. We evaluated the proposed and previous methods using peak signal-to-noise ratio, structural similarity, feature similarity, and computation time. Experimental results indicate that the proposed method clearly outperforms other state-of-the-art algorithms such as fast curvature based interpolation.

Second part, Super-Resolution via Adaptive Multiple Sparse Representation is learning based method. We propose a super-resolution algorithm based on adaptive sparse representation via multiple dictionaries for images taken by Unmanned Aerial Vehicles (UAVs). The super-resolution attainable through the proposed algorithm can increase the precision of 3D reconstruction from UAV images, enabling the production of high-resolution images for constructing high-frequency time series and for high-precision digital mapping in agriculture. The basic idea of the proposed method is to use a field server or ground-based camera to take training images and then construct multiple pairs of dictionaries based on selective sparse representations to reduce instability during the sparse coding process. The dictionaries are classified on the basis of the edge orientation into five clusters: 0, 45, 90, 135, and non-direction.

The proposed method is expected to reduce blurring, blocking, and ringing artifacts especially in edge areas. We evaluated the proposed and previous methods using peak signal-to-noise ratio, structural similarity, feature similarity, and computation time. Our experimental results indicate that the proposed method clearly outperforms other state-of-the-art algorithms based on qualitative and quantitative analysis. In the end, we demonstrate the effectiveness of our proposed method to increase the precision of 3D reconstruction from UAV images.

Last part, Deep Residual Learning Super-resolution is learning based method. The light and efficient residual network for super-resolution is proposed. We adopt inception module from GoogLeNet to exploit the features from the low-resolution images and residual learning to have fast training steps. The proposed network called Deep Residual Learning Super-resolution (DRLSR). The network is proven to have fast convergence and low computational time. It is divided into three parts: feature extraction, mapping, and reconstruction. In the feature extraction, we apply inception module followed by dimensional reduction. Then, we map the features using simple convolutional layer. Finally, we reconstruct the HR component using inception module and 1×1 convolutional layer. The experimental results show our proposed method can reduce more than half of computational time from the-state-of-the-art methods, while still having clean and sharp images.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

Computer vision is a multidisciplinary field that deals with how computers can be use to gain high-level understanding from digital images or videos. From the perspective of engineering, it seeks to automate tasks that the human visual system can do [26]. Computer vision tasks include methods for acquiring, processing, analyzing and understanding digital images. It deals with the extraction of high-dimensional data from the real world in order to represent it as numerical or symbolic information. There are many computer vision algorithms such as object recognition and object tracking. To get accurate result, the input images must be in acceptable quality and resolution. However, numerous object was taken in low-resolution (LR) due to several reason such as small charge-coupled device (CCD) sensors or image compression. Therefore, the ability of super-resolution (SR) to create high-resolution (HR) image and enhance the quality to get more accurate result is necessary as shown in Fig. 1.1.

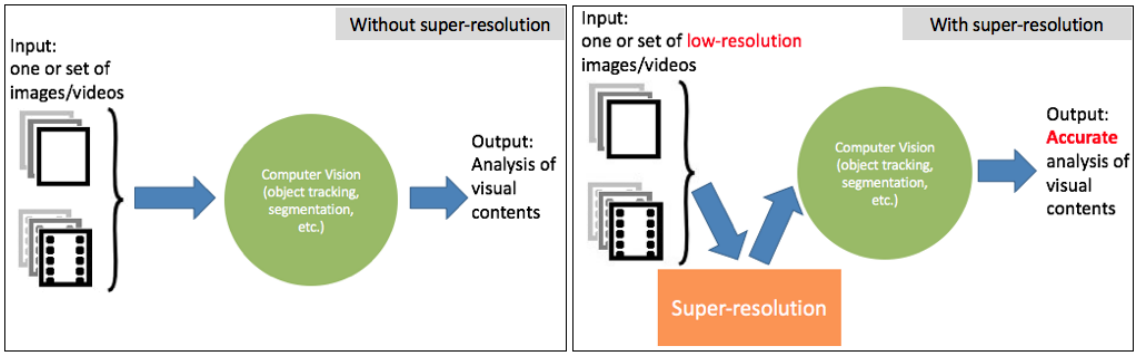


Figure 1.1: The use of super-resolution in computer vision task

SR algorithms were motivated to solve the problems caused by digital imaging devices [31]. The invention of digital scanners facilitate the conversion from paper-based docu-

ments into digital images. However, the image quality was poor, in low-resolution, and present some noise from COD sensors. With the goal of acquiring sharper and higher resolution image, SR algorithms were developed to combine multiple input LR from repeatedly scanning the same document with shifts and rotations.

Digital image data are unfortunately often at a lower quality than the desired one, because of several possible causes: spatial and temporal down-sampling due to noise degradation, high compression, etc. When we consider still images, the new sources of image contents, like the Internet or mobile devices, have generally a lower quality than high-definition display standard. Moreover, if we consider the past production, there is an enormous amount of images collected in the years, that are valuable but have a poor quality. The need of increasing the resolution of an image can also be required by the particular application context. Many applications, e.g. video surveillance and remote sensing, in fact, require the display of images at a desired resolution, for specific computer vision tasks like object recognition, zoom-in operations, or 3D reconstruction. For example, in Fig. 1.2, we can clearly see that after preprocessing using SR algorithm, the accuracy of 3D model increased. From these reasons, the urgency to improve the image quality is very important issue.

Figure 1.2: The example of super-resolution algorithm in 3D reconstruction

With the improvement of computational capability and mobile imaging devices, single SR has gained more attention with proven success. The fundamental difference is the number of input LR images required for SR to produce HR image. Since there is merely one input image, the formulation becomes an under-determined problem rather than an over-determined one as posited in the classical SR research. Because the problem is ill-posed and the available image data are limited, priors are exploited in the process to determine the generated pixel intensities. Numerous methods have been proposed based on different image properties and they can be roughly categorized into two approaches: filtering-based (non-learning) and learning-based.

Filtering methods include, among others, analytic interpolation methods, e.g. traditional bilinear and bicubic interpolation, which compute the missing intermediate pixels in the enlarged HR grid by averaging the original pixel of the LR grid with fixed filters. Edge-direction-based algorithms have been used to improve the limitation of traditional methods by exploiting local features such as edges by adapting each interpolating surface locally and assuming local regularity in a curvature. Once the input image has been upscaled to HR via interpolation, image sharpening methods can be applied. Sharpening methods aim at amplifying existing image details, by changing the spatial frequency amplitude spectrum of the image: in this way, the existing high frequencies in the image are enhanced, thus producing a more pleasant and richer output image.

Starting nineties, many powerful algorithms have been developed to solve different problems in a variety of scientific areas. Among single-image SR methods, the other important category is represented by algorithms that make use of machine learning techniques or learning-based approach. Although covering different meanings, machine learning can be generally referred to as that branch of artificial intelligence that concerns the construction and study of algorithms that can learn from data. In SR, learning method aims at estimating missing high-resolution detail that is not present in the original image, by adding new plausible high frequencies from the training data.

Several fundamental questions are still remained for single SR. In this thesis, we aim to address some of these important issues. For example, what are the important structures that can exploit and ensure for high-quality results? How to learn generating high-resolution image patches from low-resolution with and without learning process? In summary, single SR involves exploiting rich information contained in a single image. The challenges of single SR include recognizing important visual artifacts, refilling the HR details, and rendering them as faithfully and aesthetically pleasing as possible to be able to increase more accurate result on doing computer vision task. Addressing these challenges effectively and efficiently is the main motivation behind the research in this thesis.

1.2 Organization

Interested in the SR approach to the task of increasing the resolution of an image, and intrigued by the effectiveness of filtering- and learning-based techniques, during this doctorate we mostly investigated the SR problem and the application to it. On filtering based SR, we focus on reducing computational complexity by using only first-order derivative which involve only subtraction operator. In the other hand, learning-based SR procedures are patch-based procedures: the input image is partitioned into patches and from a single

LR input patch a single HR output patch is estimated via learning methods by learning the correspondences stored in the learned system. Finally, the whole set of estimated HR patches is then reconstruct to finally build the super-resolved image.

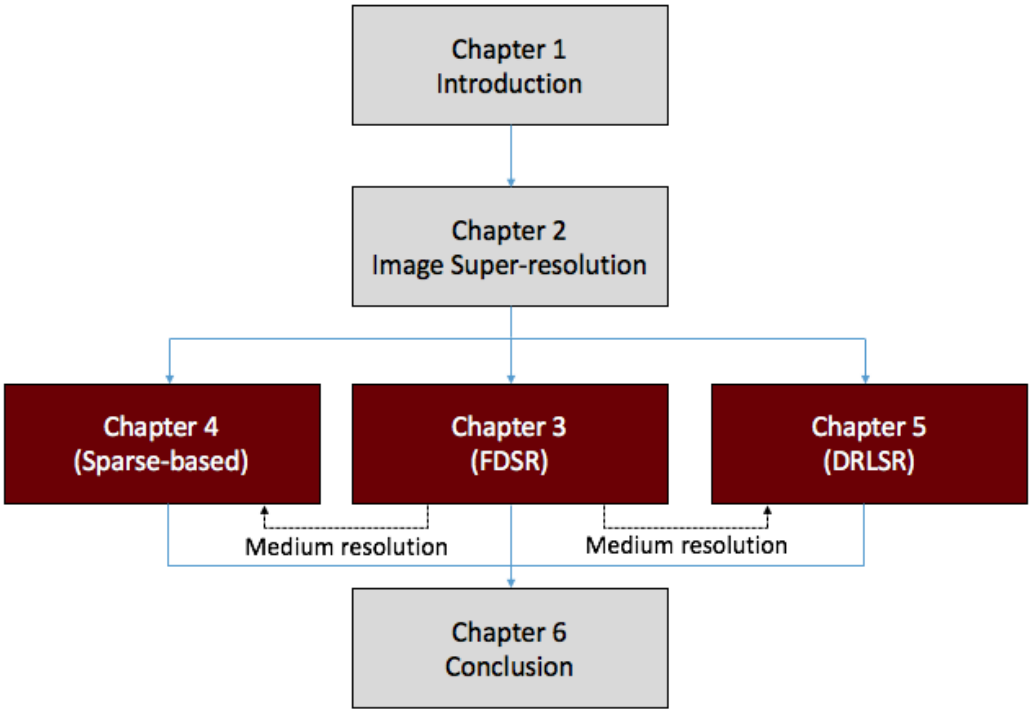


Figure 1.3: Research flowchart

The rest of this manuscript is structured as illustrated in Fig. 1.3. We start with Chapter 1 by explaining the introduction and motivation of our thesis. In Chapter 2, we give a general overview of SR and going deeper into the classification. The novel filtering based methods presented in Chapter 3 are a single fast SR method based on first-order derivatives from neighbor pixels which exploit a first-order derivatives component of six edge directions around a missing pixel; followed by back projection to reduce noise estimated by the difference between simulated and observed images. In Chapter 4, we presented an SR algorithm based on adaptive sparse representation via multiple dictionaries for images taken by Unmanned Aerial Vehicles (UAVs) which construct multiple pairs of dictionaries based on selective sparse representations to reduce instability during the sparse coding process. Then, to deal with very high non-linear relation between high- and low resolution images, we exploit the deep learning capability to propose efficient and fast architecture of convolutional neural networks based SR in Chapter 5. Finally, in Chapter 6 we end the thesis by summarizing our accomplishments, drawing conclusions from them and discussing about future directions.

Chapter 2

Image Super-resolution

2.1 Introduction

Super-resolution (SR) is the process of obtaining high-resolution (HR) image from one or more input low-resolution (LR). Numerous SR algorithms have been proposed and attracted many researchers to investigate the theory and application of SR [22]. It is found that SR can be applied in many practical applications such as image and video enhancement, medical images analysis, text analysis, satellite imaging, facial recognition. They are mainly divided based on the input and output image assumptions which can be categorized into two different types: spatial or temporal. In the spatial domain, SR aims to create an image with higher resolution and sharper image. While in the temporal domain, SR aims to insert extra frames in the video. Spatial SR or image SR has many applications and is the focus of this thesis. In the following, the term SR refers to algorithms in the spatial domain unless mentioned otherwise.

Depending on the input image, SR is mainly divided into two types: single- and multi-image SRs. Multi-image SR requires multiple images to acquire intrinsic characteristics. It then combines the information to construct a higher resolution image. Multi-image SR is highly suitable for video enlargement. It can exploit intrinsic characteristics that may differ from one sequence to another as illustrated in Fig. 2.1. For example, Liu et al. [19] proposed a Bayesian approach to adaptive video SR that involved simultaneous estimation of underlying motion, blur kernel, and noise level to reconstruct original HR frames; however, this approach has high computational complexity. Furthermore, the accuracy of multi-image SR is highly dependent to the variation of input LR images which is unnatural to obtain multiple images using common camera with different and complex motion, and known parameters.

The other method, single-image SR, requires only a single image to construct a higher resolution image. Single-image SR typically exploits the characteristics of the input image

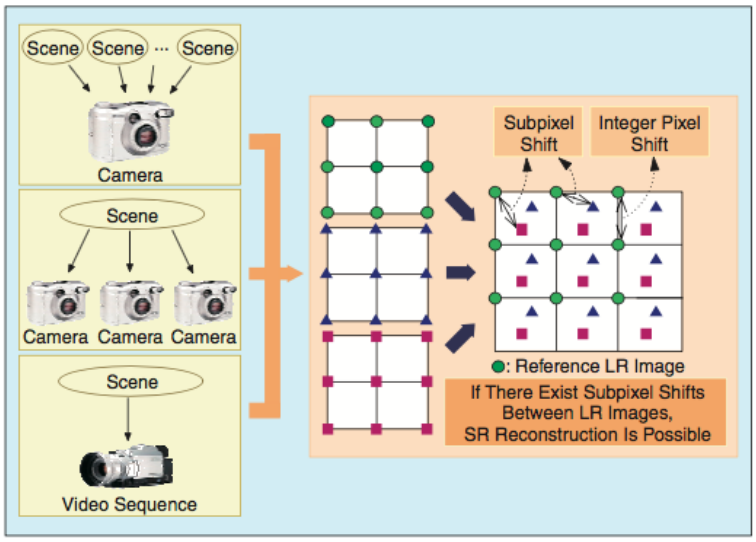


Figure 2.1: Basic premise for multi-image super-resolution [24]

and uses prior knowledge to learn the relationship between the LR and HR images. Single-image SR filled the missing pixels by observed the input LR or training data as illustrated in Fig. 2.2. Therefore, in this thesis, we focus on single-image SR which is highly applicable to the real world.

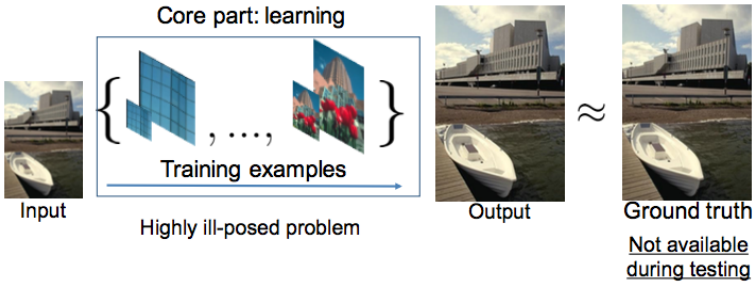


Figure 2.2: Basic premise for single-image super-resolution

Based on the approaches [22], single-image SR can be divided into three approaches: filtering-based, learning-based (non-direct examples), and reconstruction-based (direct examples). Each approach has published many research papers and designed for both specific and general purpose. However, the reconstruction-based method is eliminated from this dissertation because it requires high computational load for searching adequate instances in the exemplar set. If the exemplar set is large, the load for searching adequate exemplars will be high. Moreover, reconstruction approaches did not require training phase which make direct learning to the examples and produce more noise and instability during enlargement process.

The filtering methods were proven to have short computational time. However, it is hard to achieve the optimal result. Furthermore, the learning-based methods was able to accurately estimate the HR information by using training data but require long computational time. The more detail description of these two approaches are explained in the following sections.

2.2 Filtering-Based Approaches

Filter-based approaches focus on obtaining reasonably good result with short computational time. The focus of this approach is to be able to minimize the use of computational resource and mainly works on spatial domain. The first conventional methods utilize low complexity and easy implementation. The classic nearest neighbor, bilinear, and bicubic interpolation methods have been widely applied for real-time processing in image viewers and image-processing tools [23]. However, these methods produce unnatural images due to excessive blurring and jagged artifacts [1]. Such conventional methods do not use a prior model between HR and LR images, which plays a strong role in algorithm performance relative to quality improvement.

Edge-direction-based algorithms, often called edge-adaptive algorithm, have been used to overcome that limitation by exploiting local features such as edges [11, 18, 14, 15, 30]. For example, new edge directed interpolation (NEDI) [18] provides good results by adapting each interpolating surface locally and assuming local regularity in a curvature. Fast curvature based interpolation (FCBI) [11], inspired by NEDI, obtains interpolated pixels by averaging two pixels determined by second-order directional derivatives of image intensity. An improved version of the FCBI algorithm, i.e., iterative curvature based interpolation (ICBI), which optimizes interpolated pixels using iterative correction has been introduced [11]. Haris et al. [15] proposed the improvement of FCBI algorithm by introducing single-image SR that extends from two to six directions and accommodates a wide range of the interpolating directions of the missing pixels, then improve the result by back-projection algorithm.

Many researcher explore on this approach because it is highly suitable for real-time application due to its low computation and insensitivity to training data. Moreover, this type of SR is very easy to implement. However, the result cannot produce sharper and clearer HR image compare to learning-based approach.

2.3 Learning-Based Approaches

Learning-based SR were first introduced in 1985 by [21] which used neural-network to improve the resolution of fingerprint images. This approach can be divided into two types of input domain: spatial- and frequency-based. In the spatial-based approaches, the SR algorithm directly extracts the features or high frequency components from the pixel values. However, in frequency-based approach, the input image first transforms to the frequency domain, such as wavelet transform and fourier transform, then transforms back to spatial domain.

Takeda et al. [28] generalized the use of spatially adaptive (steering) kernel regression, which produces results that preserve and restore details with minimal assumptions about local signal and noise models. An improvement of previous algorithm also proposed using adaptive enhancement and spatiotemporal up-scaling of videos without explicit motion estimation [29]. However, this method is not robust and is sensitive to parameters such as smoothing.

Danielyan et al. [4] proposed spatially adaptive filtering in the image domain and projection in a wavelet domain. Mallat et al. [20] introduced a class of inverse problem estimators computed by adaptively mixing a family of linear estimators corresponding to different priors computed over a wavelet frame. Demirel et al. [5] investigated discrete wavelet transform to decompose the input image into different sub-bands. Celik et al. [2] exploit a forward and inverse dual-tree complex wavelet transform to construct an HR image from the given LR image. However, these methods are computationally very complex.

SR using sparse representation has become popular because of its ability to naturally encode the semantic information of images [8]. By collecting representative samples in order to create an over-completed dictionary, it is possible to discover the correct basis for correctly encoding an input image. The studies by Yang et al. [32] and Zeyde et al. [33] focused on using a single pair of dictionaries; intuitively, however, using a single pair of dictionaries can produce many redundancies, which may cause instability during the image reconstruction process.

The latest convolutional neural networks (CNNs) is used in many image processing algorithm with large improvement in accuracy. On SR algorithm, Cao dong et al.[6] has demonstrated a CNNs' ability mapping LR to HR patches called Super-resolution Convolutional Neural Networks (SRCNN). The method is constructed by a very simple and a lightweight structure CNNs using two hidden layers and 3×3 filter size. Jiwon Kim et al. [17] introduces Very Deep Convolutional Networks (VDSR), a very deep CNN with residual learning, which proven have accurate result but have critical issues on convergence

speed. VDSR includes 20 layer of CNN using 3×3 filter size. The recent improvement has been published. FSRCNN [7] demonstrated superior performance than previous SRCNN. They focused on improving the current SRCNN and proposed faster and more accurate algorithm. FSRCNN redesign the network using three main principal: deconvolution, dimension shrinking, and smaller filter.

2.4 Our Contributions

The SR algorithm is the core algorithm to support computer vision tasks, such as pattern recognition and 3D reconstruction. It has the ability to transform the input image/video to acceptable resolution for improving the accuracy of computer vision tasks. However, in terms of the application, the requirements of each task are different and unique. For example, in video streaming application, the SR algorithm has to offer low computation algorithm without the use of training data to avoid the bottleneck during data transfer in the network. In the application for satellite images, the training data is limited, the proposed SR algorithm should be insensitive to training data. Moreover, in 3D reconstruction, the details and quality of input images are necessary, we should use many training data to improve the proposed SR algorithm.

The existing SR problems solved by varieties solutions offered by researchers. The same with our research, we aim to offer various solutions which suitable for many applications depend on the requirements. Nowadays, the researchers focus on dividing SR based on the theoretical approach as mentioned in the previous section. However, in the application problems, the author found three main problems existed during SR algorithm implementation: computational time, sensitivity to training data, and quality improvement. Therefore, in this dissertation, we deeply investigate the SR based on the application problems which divided into three categories: non training data, limited training data, and unlimited training data.

On non training data approaches which is low computational process, we proposed filtering based methods using first-order derivatives from neighbor pixels on six edge directions around a missing pixel, then followed by back projection to reduce noise estimated by the difference between simulated and observed images. The next proposed method is insensitive to training images. We develop an adaptive sparse representation via multiple dictionaries based on selective sparse representations to reduce instability during the sparse coding process using limited training data. Last, we propose a method where training data is unlimited. In this case, we propose to use convolutional networks which has been proven to construct the best image quality.

In details, we also show the importance of feature variation in developing SR algorithms. In our proposed methods, we focus to use multiple features, such as multiple edge direction and convolution filter, to extract the contextual information from the input images or videos. Moreover, we show that multiple feature extractions are not only able to increase the quality of SR result, but also deliver efficient and low computation algorithm if treated correctly based on the nature of the images.

In summary, we offer the solution for different problems based on the main implementation problem. We aim to develop SR algorithm as a service where the end user can easily choose the required SR algorithm for each application. With many application requirements, the end user can use our proposed methods easily and produce the expected result.

Chapter 3

First-order Derivatives- based Super-resolution

3.1 Introduction

The need for a fast super-resolution (SR) method has become increasingly necessary due to increased availability of SR hardware such as televisions and smartphones, which have low computational capacity. Mobile devices have limited ability to enlarge images and videos, which are still available in lower-resolution formats (such as older videos on the Internet). The primary problem of an enlarging process is to predict missing areas using existing pixels. Therefore, developing an algorithm to predict the most suitable pixel value in the missing area effectively is extremely challenging.

Depending on the input image, SR is primarily divided into two types, i.e., single- and multi-image SRs. Multi-image SR requires multiple images to acquire intrinsic characteristics. It then combines the information to construct a higher resolution image. However, in daily applications, it is unnatural to obtain multiple images using common camera with known parameters. Single-image SR requires only a single image to construct a higher resolution image. Single-image super-resolution typically exploits the characteristics of the input image and uses prior knowledge to learn the relationship between the low- (LR) and high-resolution (HR) image. Therefore, our proposed method uses single-image SR which is highly applicable to the real world.

Utilizing their low complexity and easy implementation, classic nearest neighbor, bilinear, and bicubic interpolation methods have been widely applied for real-time processing in image viewers and image-processing tools [23]. However, these methods produce unnatural images due to excessive blurring and jagged artifacts [1]. Such conventional methods do not use a prior model between HR and LR images, which plays a strong role in algorithm performance relative to quality improvement.

Multi-image SR is highly suitable for video enlargement. It can exploit intrinsic characteristics that may differ from one sequence to another. Liu et al. [19] proposed a Bayesian approach to adaptive video SR that involved simultaneous estimation of underlying motion, blur kernel, and noise level to reconstruct original HR frames; however, this approach has high computational complexity.

Takeda et al. [28] generalized the use of these techniques to spatially adaptive (steering) kernel regression, which produces results that preserve and restore details with minimal assumptions about local signal and noise models. An improvement that uses adaptive enhancement and spatiotemporal up-scaling of videos without explicit motion estimation has been proposed [29]. However, this method is not robust and is sensitive to parameters such as smoothing.

Danielyan et al. [4] proposed spatially adaptive filtering in the image domain and projection in a wavelet domain. Yang et al. [32] designed a pair of sparse to construct an HR image. Mallat et al. [20] introduced a class of inverse problem estimators computed by adaptively mixing a family of linear estimators corresponding to different priors computed over a wavelet frame. Demirel et al. [5] used discrete wavelet transform to decompose the input image into different sub-bands. Celik et al. [2] used a forward and inverse dual-tree complex wavelet transform to construct an HR image from the given LR image. However, these methods are computationally very complex.

Edge-direction-based algorithms, often called edge-adaptive algorithm, have been used to overcome that limitation by exploiting local features such as edges [11, 18, 14]. For example, new edge directed interpolation (NEDI) [18] provides good results by adapting each interpolating surface locally and assuming local regularity in a curvature. Fast curvature based interpolation (FCBI) [11], inspired by NEDI, obtains interpolated pixels by averaging two pixels determined by second-order directional derivatives of image intensity. An improved version of the FCBI algorithm, i.e., iterative curvature based interpolation (ICBI), which optimizes interpolated pixels using iterative correction has been introduced [11].

Learning from the FCBI algorithm, we propose single-image SR that extends from two to six directions and accommodates a wide range of the interpolating directions of the missing pixels. The use of first-order derivatives can reduce computational complexity because the main process uses only a subtraction operator. As mentioned before, previous interpolation methods have several drawbacks, including (1) blurring, blocking, and ringing artifacts in edge areas; (2) less smoothness along edges; (3) discontinuity along edges; and (4) high computational complexity. Therefore, a simple and fast mechanism to interpolate edges based on the largest first-order derivatives is proposed to solve these problems. Ta-

ble 3.1 shows a comparison of the proposed method and previous methods based on our experiment results.

Table 3.1: Comparison between proposed algorithm and previous methods (◯ = good, △ = normal, × = not good)

Method	Computation Time	Image Quality
Nearest neighbor	◯	×
Bilinear	◯	×
Bicubic	△	△
KR[29]	△	△
SpR[32]	×	◯
SME[20]	×	◯
FCBI[11]	◯	△
ICBI[11]	△	◯
NEDI [18]	△	△
DFDF[35]	△	△
Proposed	◯	◯

Chapter 4

Super-Resolution via Adaptive Multiple Sparse Representation

4.1 Introduction

The use of unmanned aerial vehicles (UAVs) in agriculture has increased in recent years [25, 9, 12]. The use of UAVs offers alternatives to manual breeding methods in agriculture, which are laborious, time-consuming, unreliable, and often impossible to implement. For example, high-frequency time series data are almost impossible to obtain without the use of a UAV. Moreover, large-scale, hilly landscapes make it impractical to manually analyze each tree individually using hand-held or ground-based devices. The use of UAVs can overcome such limitations, and UAV imaging offers advantages in terms of high-resolution data and precise 3D imaging.

Table 4.1: Comparison of agricultural monitoring systems (◯ = superior, △ = average, × = poor).

Method	Hand-held device	Ground-based device	UAV	Aircraft	Satellite
Frequency	×	△	◯	△	×
Coverage	×	×	△	◯	◯
Cost	◯	△	△	×	×
User friendly	◯	△	◯	×	×
Resolution	◯	◯	△	△	×

Examples of some of the advantages offered by the use of UAVs over traditional field-based monitoring methods are listed in Table 4.1. UAV imaging can efficiently provide high-frequency time series data, whereas aircraft and satellite systems are very complicated and their use requires arrangements be made in advance. Hand-held and ground-based

devices have short preparation times but require long execution times. In terms of coverage, aircraft and satellites perform well because they can rapidly image several hectares in area, but they produce low-resolution images. By contrast, UAVs can provide better resolution as they have adjustable flight altitudes. Although hand-held and ground-based devices can provide the best resolution because they can observe parts of plants in detail, they cannot be used for large area and coverage or to produce high-frequency time series data. UAVs also require lower expenditures than aircraft or satellite as UAV sensors are much cheaper. As a UAV can be operated autonomously, control by the end user is much simpler. These advantages make UAV utilization in agricultural monitoring quite useful by offering a new perspective from which to monitor the ground with high precision [34].

The main problems in constructing 3D high-resolution maps using UAV images are flight-time limitations and image quality from the target object. Taking aerial images of a large field will consume a large amount of time, and to reduce time consumption, it is necessary to set an optimum height for UAV flight. However, maximizing the height, which increase the viewing perspective of the UAV and thus potentially reduces the flight time, reduces the optical detail of a target object. Therefore, it is necessary to use a super-resolution (SR) technique to obtain higher-resolution, high-precision images of target objects [3].

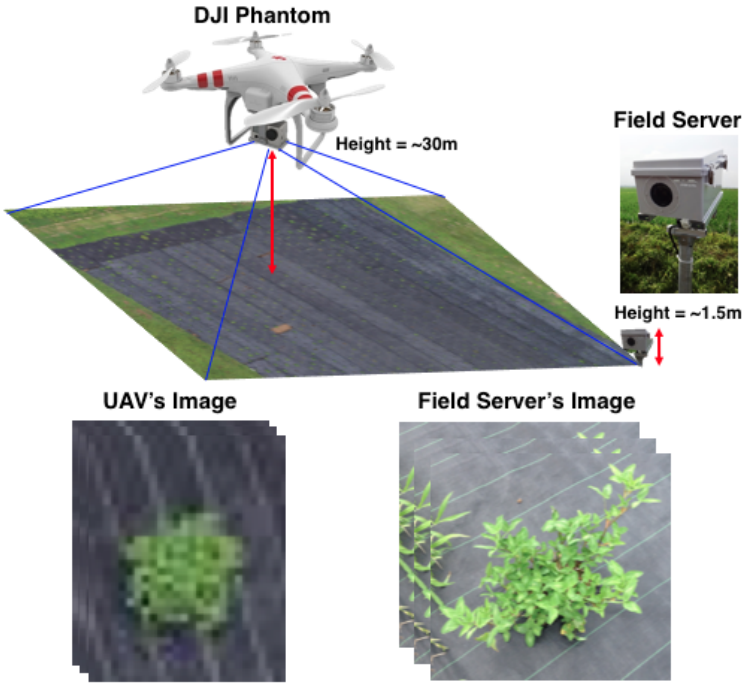


Figure 4.1: DJI Phantom and Field Server sample images.

Field Server (FS) systems [10, 13] can be used for ground-based monitoring via a series of small sensor nodes equipped with a Web server that can be accessed via the Internet

and communicate, unlike traditional sensor nodes, through a wireless LAN over a high-speed transmission network. An FS system can be easily installed for remotely monitoring field information anywhere. By including the functionality of a Web server in each module, an FS system can collectively manage each module over the Internet, producing high-resolution images that can be used as training images for an SR algorithm. A comparison of FS and UAV images is shown in Fig. 4.1.

Depending on the input image, SR imaging is primarily divided into two types, i.e., single- and multi-image SR imaging. Multi-image SR requires multiple images to acquire intrinsic characteristics; it combines the information from each image to construct a higher-resolution image. In day-to-day applications, however, it is unusual to obtain multiple images using a generic camera with known parameters. Single-image SR requires only a single image to construct a higher-resolution image - a much simpler task than multi-image SR. Single-image SR typically exploits the characteristics of the input image and uses prior knowledge to determine the relationship between a low- (LR) and high-resolution (HR) image. Our proposed method therefore uses single-image SR, which is highly suitable for the use real world applications. Furthermore, training based on SR can produce better prediction using a training model for enlarging images of phenotyping fields.

Owing to their low complexity and ease of implementation, classic nearest neighbor, bilinear, and bicubic interpolation methods have been widely applied in image processing [23]. However, such methods produce unnatural images due to excessive blurring and jagged artifacts [1].

Multi-image SR is highly suitable for video enlargement. It can exploit intrinsic characteristics that may differ from one sequence to another. Liu et al. [19] proposed a Bayesian approach to adaptive video SR that involved the simultaneous estimation of the underlying motion, blur kernel, and noise level to reconstruct original HR frames; however, this approach has high computational complexity.

Edge-direction-based algorithms, which are applied to single-image SR and often termed edge-adaptive algorithms, have been used to overcome computational complexity limitations by exploiting local features such as edges [11, 18, 14, 15]. For example, new edge directed interpolation (NEDI) [18] produces good imaging results by adapting each interpolating surface locally and assuming local regularity of curvature. Iterative curvature-based interpolation (ICBI), inspired by NEDI, produces interpolated pixels by averaging sets of two pixels using second-order directional derivatives of the image intensity [11].

SR using sparse representation has become popular because of its ability to naturally encode the semantic information of images [8]. By collecting representative samples in order to create an over-completed dictionary, it is possible to discover the correct basis for

correctly encoding an input image. The studies by Yang et al. [32] and Zeyde et al. [33] focused on using a single pair of dictionaries; intuitively, however, using a single pair of dictionaries can produce many redundancies, which may cause instability during the image reconstruction process.

In this paper, we propose adapting multiple pairs of dictionaries that classify by edge orientation in order to select the most suitable pair of dictionaries for a particular signal. These dictionaries are obtained by determining bases from HR images produced by FS. Following this, we discuss how input images from a UAV can be enlarged to obtain higher-resolution images. Finally, we demonstrate the effectiveness of the proposed method in reconstructing 3D images.

Chapter 5

Deep Residual Learning Super-resolution

5.1 Introduction

The availability of various types of images due to internet technologies provide big chance for learning algorithm to learn image characteristic deeply. This opportunity has been exploited by many researchers to develop robust super-resolution (SR) algorithms based on learning approaches. The main goal of SR is to recover high-frequency information from the input low-resolution (LR) image to be able to produce high-resolution (HR) one. Other goal of SR algorithm is to increase the accuracy of computer vision task. The SR algorithm is expected to reconstruct the LR input image in acceptable quality and resolution.

Currently, learning methods are widely used to map from LR to HR patches. Super-resolution using sparse representation shows its popularity because of the ability to naturally encode the semantic information of images [8]. By collecting representative samples in order to create an over-completed dictionary, it is possible to discover the correct basis for correctly encoding an input image. The studies by Yang et al. [32] and Zeyde et al. [33] focused on using a single pair of dictionaries; intuitively, however, using a single pair of dictionaries can produce many redundancies, which may cause instability during the image reconstruction process.

Lately, convolutional neural networks (CNN) is used in many image processing algorithm with large improvement in accuracy. On SR algorithm, Cao dong et al.[6] has demonstrated a CNNs' ability mapping LR to HR patches called Super-resolution Convolutional Neural Networks (SRCNN). The method is constructed by a very simple and a lightweight structure CNNs using two hidden layers and 3×3 filter size. Jiwon Kim et al. [17] introduces Very Deep Convolutional Networks (VDSR), a very deep CNN with residual learning, which proven have accurate result but have critical issues on convergence

speed. VDSR includes 20 layer of CNN using 3×3 filter size.

The recent improvement has been published. FSRCNN [7] demonstrated superior performance than previous SRCNN. They focused on improving the current SRCNN and proposed faster and more accurate algorithm. FSRCNN redesign the network using three main principal: deconvolution, dimension shrinking, and smaller filter.

In this paper, we propose fast convergence and low-computation convolutional network for image super-resolution as shown in Fig. **??**. Our proposed network is inspired by inception module and residual learning. GoogleNet [27] introduces inception concept which use multiple type of filter size then combine it into one stream. This concept has been proven in the 2015 ILSVRC challenge. While, residual learning introduces by He et al. [16] to ease the training of networks and gain better accuracy.

Chapter 6

Conclusion and Future Works

6.1 Summary

The work of this thesis focused on the study of super-resolution (SR) as a technique to augment the spatial resolution of images, to a greater extent than conventional methods. In particular, we adopted the single-image SR approach based on filtering and learning methods. The filtering method predict the HR component based on curvature modeling using first-order derivatives. Then, the SR procedure based on machine learning paradigm, where the HR output image is predicted/estimated patch by patch: for each LR input patch we compute a model on the basis of local examples and we use this model to predict the related HR output patch.

In the first part, the main contribution is the extension of edge direction based on first-order derivatives for single-image SR. In the proposed method, we employ six edge directions and first-order derivatives as a feature to extract the interpolation direction. This is followed by a back-projection process to refine the image. The proposed method was implemented and evaluated. The results of our evaluations show that the our proposed method has the lowest computational complexity and demonstrates superior quality compared to other methods. The experiment results from both quantitative and qualitative analysis show that the proposed method outperforms previous method. Furthermore, the proposed method can preserve image details and reduce artifacts, such as blurring and ringing around edges.

In the second part, an SR based on adaptive multiple pairs of dictionaries for UAV images was proposed. The proposed method employs a classification based on edge orientation to obtain selective patches by creating five clusters, each of which obtains a pair of dictionaries A_l and A_h. The proposed method was implemented and out-performed other methods. The experimental results show the superiority of our proposed method for both quantitative and qualitative analysis by preserving detail and reducing artifacts such as

blurring and ringing around the edge. Our method was also proven effective for 3D reconstruction and produced an image superior to the original image from a 10m height. The use of a GPU application could further enhance our method by enabling opportunities to decrease its computational time.

In the third part, we proposed Deep Residual Learning Super-resolution (DRLSR). The network inspired by Inception module of GoogLeNet to produce multiple features during feature extraction and reconstruction process. Our strategies ensure the network having fast convergence and low computational time. The proposed network was assessed. The results show that our proposed network can cut half of computational time from the the-state-of-the-art network. Furthermore, our proposed network successfully exploit the Inception module and residual learning in the SR approach.

In summary, Fig. 6.1 shows the summary of our dissertation. We aim to solve the three main problems during SR implementation: computational time, sensitivity to training data, and quality improvement. In the beginning, we focus to create low computation SR algorithm which considered as filtering based method. Then, we investigate the ability of multiple sparse coding in the SR approach with insensitivity from training images. Finally, we develop efficient convolutional networks with superb quality compare to current-state-of-the-art methods. Moreover, the proposed method from Chapter 3 can be used as interpolated method to produce middle or medium resolution which is used in Chapter 4 and 5 to create training pairs.

6.2 Future works and perspectives

Apart from the results, we are aware that our work is far from finished. In the last section, we would highlight some questions as the future works.

First part is First-order Derivatives- based Super-resolution. The proposed method is very simple and light. However, the edge direction can be wrongly interpolated and cause some noise in the image. Currently, this noise can be polished by back-projection method. For the next step, we need to observe the edge and texture modeling using first-order derivatives. Furthermore, we can correctly interpolate the direction of the edge in the input image.

Second part is Super-Resolution via Adaptive Multiple Sparse Representation. Sparse-based method is notably one of the-state-of-the-art in the super-resolution methods. It has the ability to create basis to connect low-resolution image and high-resolution image. However, the sparse and dictionary initialization is crucial for this method. We can observe carefully the impact of the initialization. Moreover, the possibility of K-SVD giving the

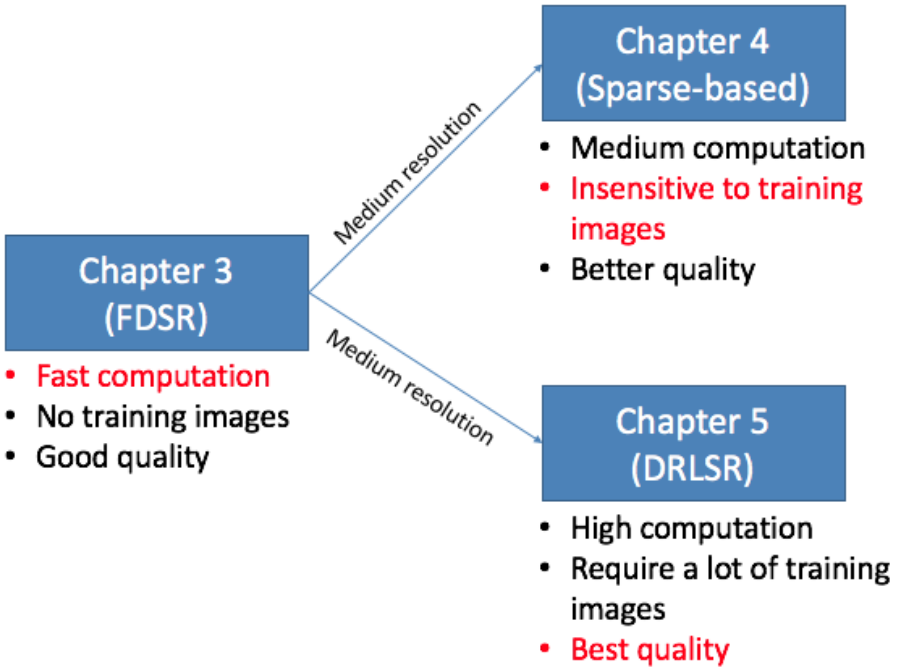


Figure 6.1: The summary of the proposed methods

local optimum solution is high especially using single dictionary. Therefore, the chance to improve the current methods is high.

The last part is Deep Residual Learning Super-resolution. We aim to have light network yet constructing clear and sharp HR image. In the experiment, we have not observed and analyzed deeply regarding the advantages of Inception modules and various settings. The current network have high possibility to be trapped in local optimum solution. In the future, more efficient network is need to be designed to produce better quality of obtained HR image.

Bibliography

[1] Nicola Asuni and Andrea Giachetti. Accuracy improvements and artifacts removal in edge based image interpolation. *VISAPP (1)'08*, pages 58–65, 2008.

[2] Turgay Celik and Tardi Tjahjadi. Image resolution enhancement using dual-tree complex wavelet transform. *Geoscience and Remote Sensing Letters, IEEE*, 7(3):554–557, 2010.

[3] Dengxin Dai, Yujian Wang, Yuhua Chen, and Luc Van Gool. Is image super-resolution helpful for other vision tasks? In *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1–9. IEEE, 2016.

[4] Aram Danielyan, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Image upsampling via spatially adaptive block-matching filtering. In *Signal Processing Conference, 2008 16th European*, pages 1–5. IEEE, 2008.

[5] Hasan Demirel and Gholamreza Anbarjafari. Discrete wavelet transform-based satellite image resolution enhancement. *Geoscience and Remote Sensing, IEEE Transactions on*, 49(6):1997–2004, 2011.

[6] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 38(2):295–307, 2016.

[7] Chao Dong, Chen Change Loy, and Xiaoou Tang. Accelerating the super-resolution convolutional neural network. In *European Conference on Computer Vision*, pages 391–407. Springer, 2016.

[8] Michael Elad. Sparse and redundant representation modeling - what next? *Signal Processing Letters, IEEE*, 19(12):922–928, 2012.

[9] J Everaerts et al. The use of unmanned aerial vehicles (uavs) for remote sensing and mapping. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 37:1187–1192, 2008.

[10] Tokihiro Fukatsu and Masayuki Hirafuji. Field monitoring using sensor-nodes with a web server. *Journal of Robotics and Mechatronics*, 17(2):164–172, 2005.

[11] A. Giachetti and N. Asuni. Real time artifact-free image upscaling. *Image Processing, IEEE Transactions on*, 20(10):2760–2768, October 2011.

[12] GJ Grenzdörffer, A Engel, and B Teichert. The photogrammetric potential of low-cost uavs in forestry and agriculture. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 31(B3):1207–1214, 2008.

[13] Wei Guo, Tokihiro Fukatsu, and Seishi Ninomiya. Automated characterization of flowering dynamics in rice using field-acquired time-series rgb images. *Plant Methods*, 11(1):1–15, 2015.

[14] Muhammad Haris, Kazuhito Sawase, Muhammad Rahmat Widyanto, and Hajime Nobuhara. An efficient super resolution based on image dimensionality reduction using accumulative intensity gradient. *Journal of Advanced Computational Intelligence and Intelligent Informatics*, 18(4):518–528, 2014.

[15] Muhammad Haris, M Rahmat Widyanto, and Hajime Nobuhara. First-order derivative-based super-resolution. *Signal, Image and Video Processing*, 11(1):1–8, 2017.

[16] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *arXiv preprint arXiv:1512.03385*, 2015.

[17] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR Oral)*, June 2016.

[18] Xin Li and Michael T. Orchard. New edge-directed interpolation. *IEEE Transactions on Image Processing*, 10:1521–1527, 2001.

[19] Ce Liu and Deqing Sun. A bayesian approach to adaptive video super resolution. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 209–216. IEEE, 2011.

[20] Stéphane Mallat and Guoshen Yu. Super-resolution with sparse mixing estimators. *Image Processing, IEEE Transactions on*, 19(11):2889–2900, 2010.

[21] Eric Mjolsness. *Fingerprint Hallucination*. PhD thesis, California Institute of Technology, 1985.

[22] Kamal Nasrollahi and Thomas B Moeslund. Super-resolution: a comprehensive survey. *Machine vision and applications*, 25(6):1423–1468, 2014.

[23] M.A. Nuno-Maganda and M.O. Arias-Estrada. Real-time fpga-based architecture for bicubic interpolation: an application for digital image scaling. In *Reconfigurable Computing and FPGAs, 2005. ReConFig 2005. International Conference on*, pages 8 pp.–1, Sept 2005.

[24] Sung Cheol Park, Min Kyu Park, and Kang Moon Gi. Super-resolution image reconstruction: A technical overview. *IEEE Signal Processing Magazine*, 20:21–36, 2003.

[25] Santhosh K Seelan, Soizik Laguette, Grant M Casady, and George A Seielstad. Remote sensing applications for precision agriculture: A learning community approach. *Remote Sensing of Environment*, 88(1):157–169, 2003.

[26] Milan Sonka, Vaclav Hlavac, and Roger Boyle. *Image processing, analysis, and machine vision*. Cengage Learning, 2014.

[27] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–9, 2015.

[28] Hiroyuki Takeda, Sina Farsiu, and Peyman Milanfar. Kernel regression for image processing and reconstruction. *Image Processing, IEEE Transactions on*, 16(2):349–366, 2007.

[29] Hiroyuki Takeda, Peyman Milanfar, Matan Protter, and Michael Elad. Super-resolution without explicit subpixel motion estimation. *Image Processing, IEEE Transactions on*, 18(9):1958–1975, 2009.

[30] Qing Wang and Rabab Kreidieh Ward. A new orientation-adaptive interpolation method. *IEEE Transactions on Image Processing*, 16(4):889–900, 2007.

[31] Chih-Yuan Yang. *Example-Based Single-Image Super-Resolution*. PhD thesis, UNIVERSITY OF CALIFORNIA, MERCED, 2015.

[32] Jianchao Yang, John Wright, Thomas S Huang, and Yi Ma. Image super-resolution via sparse representation. *Image Processing, IEEE Transactions on*, 19(11):2861–2873, 2010.

[33] Roman Zeyde, Michael Elad, and Matan Protter. On single image scale-up using sparse-representations. In *Curves and Surfaces*, pages 711–730. Springer, 2012.

[34] Chunhua Zhang and John M Kovacs. The application of small unmanned aerial systems for precision agriculture: a review. *Precision agriculture*, 13(6):693–712, 2012.

[35] Lei Zhang and Xiaolin Wu. An edge-guided image interpolation algorithm via directional filtering and data fusion. *Image Processing, IEEE Transactions on*, 15(8):2226–2238, 2006.