PEJMAN RASTI

Analysis of Remote Sensing Image
Super Resolution using
Fluid Lenses

TARTU ÜLIKOOL
UNIVERSITAS TARTUENSIS
1632

DISSERTATIONES SCIENTIAE MATERIALIS UNIVERSITATIS TARTUENSIS

**18**

# PEJMAN RASTI

## Analysis of Remote Sensing Image Super Resolution using Fluid Lenses

Institute of Technology, Faculty of Science and Technology, University of Tartu, Estonia

Supervisors:

Assoc. Prof.　　　　Gholamreza Anbarjafari
University of Tartu
Tartu, Estonia

Prof.　　　　Rudolf Keifer
Ton Duc Thang University
Ho Chi Minh, Vietnam

Reviewer:

Assoc. Prof.　　　　Vitaly Skachek
University of Tartu
Tartu, Estonia

Opponent:

Prof.　　　　Enis Cetin
Bilkent University
Ankara, Turkey

The public defense will take place on May 23, 2017 at Auditorium 121 in Nooruse 1, Tartu, Estonia.

# Contents

# List of Tables

# LIST OF ABBREVIATIONS

**BoW** . . . . . . . . . . .     **B**ag **o**f **W**ords

**CR** . . . . . . . . . . . . .     **C**omputerized **T**omography

**CMOS** . . . . . . . . . .     **C**omplementary **M**etal-**O**xide **S**emiconductor

**CNN** . . . . . . . . . . .     **C**onvolutional **N**eural **N**etworks

**CS** . . . . . . . . . . . . .     **C**ycle **S**panning

**DARPA** . . . . . . . .     **D**efence **A**dvanced **R**esearch **P**roducts **A**gency

**DEA** . . . . . . . . . . .     **D**ielectric **E**lastomer **A**ctuator

**DESA** . . . . . . . . . .     **D**ielectric **E**lastomer **S**tack **A**ctuator

**DT-CWT** . . . . . . . .     **D**ual **T**ree **C**omplex **W**avelet **T**ransform

**DWT** . . . . . . . . . . .     **D**iscrete **W**avelet **T**ransform

**Eqn** . . . . . . . . . . . . .     **E**quation

**GWN** . . . . . . . . . . .     **G**aussian **W**hite **N**oise

**HH** . . . . . . . . . . . . . High-High

**HL** . . . . . . . . . . . . . High-Low

**HMM** . . . . . . . . . . . Hidden Markov Model

**Hof** . . . . . . . . . . . . . Histogram of Flow

**HoG** . . . . . . . . . . . Histogram of Gradient

**HP** . . . . . . . . . . . . . Head Pose

**HR** . . . . . . . . . . . . . High Resolution

**IBP** . . . . . . . . . . . . . Iterative Back Projection

**iCV-F** . . . . . . . . . . . intelligent Computer Vision Face database

**IDWT** . . . . . . . . . . Inverse Discrete Wavelet Transform

**IIBP** . . . . . . . . . . . Iterative Interpolation Back Projection

**LA-BSF** . . . . . . . . Local Adaptive Bivariate Shrinkage Function

**LFW** . . . . . . . . . . . Labeled Faces in the Wild

**LH** . . . . . . . . . . . . . Low-High

**LL** . . . . . . . . . . . . . Low-Low

**LR** . . . . . . . . . . . . . Low Resolution

**LSTM** . . . . . . . . . Long Short-Term Memory

**MBH** . . . . . . . . . . Motion Boundary Histogram

**MMSE** . . . . . . . . . Minimum Mean Square Error

**MRI** . . . . . . . . . . . Magnet Resonance Imaging

**MSE** . . . . . . . . . . . Mean Square Error

**NQM** . . . . . . . . . . Noise Quality Measure

**OCR** . . . . . . . . . . . Optical Character Recognition

**PDMS** . . . . . . . . . PolyDiMenthylSiloxane

**PSF** . . . . . . . . . . . . Point Spread Function

**PSNR** . . . . . . . . . . Peak Signal to Noise Ratio

**RNN** . . . . . . . . . . . Recurrent Neural Networks

**SR** . . . . . . . . . . . . . Super Resolution

**SRCNN** . . . . . . . . Convolutional Neural Network Super Resolution

**SSIM** . . . . . . . . . . Structural Similarity Index

**STD** . . . . . . . . . . . **ST**andard **D**eviation

**SVD** . . . . . . . . . . . **S**ingular **V**alue **D**ecomposition

**SVM** . . . . . . . . . . . **S**upport **V**ector **M**achine

**TRT** . . . . . . . . . . . **T**rajectory

**UQI** . . . . . . . . . . . **U**niversal **Q**uality **I**ndex

**VIFP** . . . . . . . . . . . **V**isual **I**nformation **F**idelity in **P**ixel domain

**WSNR** . . . . . . . . . . **W**eight **S**ignal to **N**oise **R**atio

**WT** . . . . . . . . . . . . **W**avelet **T**ransform

**WZP** . . . . . . . . . . **W**avelet **Z**ero **P**adding

# LIST OF SYMBOLS

$B$ .............. Bicubic Interpolation

$c$ ............... Cell

$d$ .............. Distance

$D_h$ ............ Dictionary of HR Patches

$D_l$ ............. Dictionary of LR Patches

$E$ ............. Feature Extraction Operator

$F$ ............. Forgert gate

$f$ ............. Map Function

$g$ .............. Cell output

$h$ .............. Thickness Distribution

$H$ ............. Blur Function

$H^{BP}$ .......... Back Projection Kernel

$k$ ............... Index of a patch in dictionaries

$l$ ............... Norm

$n$ ............... Additive noise

$o$ ............... Output gate

$p$ ............... Image patch

$p_h$ ............... HR image patches

$p_l$ ............... LR image patches

$Q$ ............... Decimation Operator

$R$ ............... Linear operator

$S$ ............... Sigmoid function

$s_\lambda$ ............... Even function

$t$ ............... Time

$u$ ............... Dummy Variable

$W$ ............... Weight matrix

$w$ ............... Rotation Speed

$y$ .............. Pixel of LR image

$\alpha$ .............. Alpha blending

$\beta$ .............. Tradeoff controller

$\gamma$ .............. Material Properties as Density

$\eta$ .............. Viscosity

$\lambda$ .............. Window function

$\mu$ .............. Mean

$\xi$ .............. Scale-up index

$\sigma$ .............. variance

$\Psi_h$ .............. HR image

$\Psi_l$ .............. LR image

$\Psi_s$ .............. SR image

# PUBLICATIONS INCLUDED IN THIS THESIS

**Directly Related**

Journal papers (1.1)

1. Rasti, P., Hous, H., Schlaak, H.F., Kiefer, R., Anbarjafari, G.: Dielectric elastomer stack actuator-based autofocus fluid lens. Applied optics 54(33), 9976–9980 (2015).

2. Rasti, P., Nasrollahi, K., Orlova, O., Tamberg, G., Moeslund, T.B., Anbarjafari, G.: Reducible dictionaries for single image super-resolution based on patch matching and mean shifting. Journal of Electronic Imaging pp. 1–22 (Accepted) (2017)

3. Rasti, P., Taşmaz, H., Daneshmand, M., Kiefer, R., Ozcinar, C., Anbarjafari, G.: Satellite image enhancement: systematic approach for denoising and resolution enhancement. DYNA-Ingeniería e Industria 91(3) (2016).

Proceedings (3.1)

4. Rasti, P., Orlova, O., Tamberg, G., Ozcinar, C., Nasrollahi, K., Moeslund, T.B., Anbarjafari, G.: Improved interpolation kernels for super-resolution algorithms. In: IEEE International Conference on Image Processing Theory, Tools and Applications. IEEE (2016).

5. Rasti, P., Kiefer, R., Anbarjafari, G.: Autofocus liquid lens by using sharpness measurement. In: 2015 23nd Signal Processing and Communications Applications Conference (SIU). pp. 608–611. IEEE (2015).

6. Rasti, P., Uiboupin, T., Escalera, S., Anbarjafari, G.: Convolutional neural network super resolution for face recognition in surveillance monitoring. In: International Conference on Articulated Motion and Deformable Objects. pp. 175–184. Springer (2016).

7. Rasti, P., Daneshmand, M., Alisinanoglu, F., Ozcinar, C., Anbarjafari, G.: Medical image illumination enhancement and sharpening by using stationary wavelet transform. In: Signal Processing and Communication Application Conference (SIU), 2016 24th. pp. 153–156. IEEE (2016).

8. Rasti, P., Kesküla, A., Haus, H., Schlaak, H.F., Anbarjafari, G., Aabloo, A., Kiefer, R.: A passive autofocus system by using standard deviation of the image on a liquid lens. In: SPIE Smart Structures and Materials+ Nondestructive Evaluation and Health Monitoring. pp. 94301Q–94301Q. International Society for Optics and Photonics (2015)

9. Nasrollahi, K., Escalera, S., Rasti, P., Anbarjafari, G., Baro, X., Escalante, H.J., Moeslund, T.B.: Deep learning based super-resolution for improved action recognition. In: Image Processing Theory, Tools and Applications (IPTA), 2015 International Conference on. pp. 67–72. IEEE (2015).

10. Bellantonio, M., Haque, M.A., Rodriguez, P., Nasrollahi, K., Telve, T., Escarela, S., Gonzalez, J., Moeslund, T.B., Rasti, P., Anbarjafari, G.: Spatio-temporal pain recognition in cnn-based super-resolved facial images. In: 23rd International Conference on Pattern Recognition (ICPR) (2016)

Book Chapter (3.2)

1. Anbarjafari, G., Rasti, P., Daneshmand, M., Ozcinar, C.: Resolution enhancement based image compression technique using singular value decomposition and wavelet transforms. In: Baleanu, D. (ed.) Wavelet Transform and Some of Its Real-World Applications, chap. 2. InTech (2015)

Book (6.4)

1. Anbarjafari, G., Rasti, P.: Illumination Enhancement: Image and Video, vol. 1. LAP LAMBERT Academic Publishing (2016)

**Indirectly Related**

Journal papers (1.1)

1. Rasti, P., Samiei, S., Agoyi, M., Escalera, S., Anbarjafari, G.: Robust non-blind color video watermarking using QR decomposition and entropy analysis. Journal of Visual Communication and Image Representation 38, 838–847 (2016).

2. Rasti, P., Daneshmand, M., Anbarjafari, G.: Statistical approach based iris recognition using local binary pattern. DYNA 92, 76–81 (2017).

3. Anbarjafari, G., Agoyi, M., Laur, L., Rasti, P.: A robust color image watermarking scheme using entropy and QR decomposition. Radioengineering (2015).

4. Kivilo, A., Zondaka, Z., Kesküla, A., Rasti, P., Tamm, T., Kiefer, R.: Electro-chemo-mechanical deformation properties of polypyrrole/dodecylbenzene-sulfate linear actuators in aqueous and organic electrolyte. RSC Advances 6(99), 96484–96489 (2016)

Proceedings (3.1)

1. Bolotnikova, A., Rasti, P., Traumann, A., Lusi, I., Daneshmand, M., Noroozi, F., Samuel, K., Sarkar, S., Anbarjafari, G.: Block based image compression technique using rank reduction and wavelet difference reduction. In: Seventh International Conference on Graphic and Image Processing. pp. 981702–981702. International Society for Optics and Photonics (2015).

2. Pjatkin, K., Daneshmand, M., Rasti, P., Anbarjafari, G.: Probability distribution function based iris recognition boosted by the mean rule. In: Intelligent Computing and Internet of Things (ICIT), 2014 International Conference on. pp. 47–50. IEEE (2015).

Keep your dreams alive. Understand to achieve anything requires faith and belief in yourself, vision, hard work, determination, and dedication. Remember all things are possible for those who believe.

Gail Devers

# ABSTRACT

In the field of image processing, image resolution enhancement plays a foremost role. Resolution enhancement provides a clearer image with a higher resolution. So far, the enhancement technique is widely applied in various photogrammetric images. However, due to the restriction of the CCD sensors, the number of pixels in the sensor is not sufficient in some cases. The image quality is affected and restricted. To solve this problem, the enhancement techniques are expended mainly in two categories: one is a hardware solution; another is a software solution. In this thesis, an investigation has been done on both hardware and software solutions.

First, an adaptive liquid lens is proposed [98] as well as two passive autofocusing systems [99, 100] for the hardware side, due to a lack of sufficient room in cell phones or small cameras that allow users to move a rigid lens in a range of focal lengths. For this purpose, a new Dielectric Elastomer Stack Actuator autofocus liquid lens is designed and fabricated that shows more advantages in comparison to the available dielectric elastomer base liquid lens.

Second, in the software approach, a new algorithm for the enhancement of satellite images [106], new interpolation kernels to improve resolution [104] and a new dictionary based Super resolution (SR) algorithm are proposed in order to enhance the resolution of images. The quantitative and qualitative analyses of the experimental results show the superiority of the proposed techniques over the conventional and state-of-the-art methods.

Third, an investigation has been done on the advantages of using SR algorithms in other areas of computer vision, such as automatic action recognition, face recognition, and pain detection.

# CHAPTER 1

# INTRODUCTION

The resolution of any image is an important property of the image which directly describes the expected amount of details that can be represented through the image. This is, however, limited to the capabilities of the image acquisition device, mainly the imaging sensor [74]. The size of this sensor defines the available spatial resolution of produced images. Thus, for increasing the spatial resolution of an image, one could increase the sensor density by reducing the spaces allocated to each pixel on the sensor, or simply by increasing the size of the sensor. The latter solution is expensive and slows down the imaging process, while the former solution decreases the amount of light incident on each sensor, which in turn increases the shot noise [74].

Applying various signal processing techniques is another approach for enhancing the resolution of an image. One of the famous techniques for this purpose is super-resolution (SR) [79, 24]. The basic idea behind SR methods is to recover/estimate one or more high-resolution (HR) image(s) from one or more low-resolution (LR) images [128, 95, 102]. Huang and Tsai [43] as pioneers of SR proposed a method for improving the spatial resolution of satellite images of earth, where a large set of translated images of the same scene is available. They showed that SR, using multiple offset images of the same scene and a proper registration, can produce better HR images compared to spline interpolation. Since then, SR methods have become common practice for many applications in different fields, such as remote sensing [66, 101]; surveillance video [20, 67]; medical imaging, such as ultrasound, magnetic resonance imaging (MRI), and computerized tomography (CT) scan [47, 70, 71, 82]; optical character recognition (OCR) problems or face recognition [16, 38, 114, 36, 77].

Different techniques have been developed for performing SR [79]. One such technique is based on sparse representation.

On the other hand, there is not enough room in cell phones or small cameras that allow users to move a rigid lens in a range of focal lengths. To solve the prob-

lem, an adaptive liquid lens [60] is proposed that enables small cameras to focus without needing any extra room. However, traditional approaches, which use an electric current to change the surface shape of a liquid, require a lot of power. Auto-focus liquid lenses are divided into two primary types [61, 41]: the reflective type that is used as a variable mirror, as applied in reflector telescopes, and the transmissive type that is based on an active change of the convex/concave lens shape over immiscible fluids with a different refractive index.

One or more fluids are used to make an infinitely variable liquid lens without any moving parts by controlling the meniscus shape, i.e. the shape of the liquid/liquid or liquid/air interface. The fluids could be manipulated mechanically and electrically, which takes advantage of the surface tension of the liquid [126]. In the mechanical method, a mechanical displacement of surfaces is used in order to adjust the shape of the lens [33, 84], while electrowetting [13, 109] is used to change the surface tension in the electric method. Liquid lenses are suitable for some applications, such as endoscopic medical imaging [19], fiber-optic telecommunication systems or micro cameras, due to the boundary between the two fluids which forms an extremely regular and smooth surface [61]. In this work, we develop a Dielectric Elastomer Actuator (DEA) liquid lens with a passive auto-focus system.

Overall, we are contributing to the field of optics and image SR by introducing the following new steps:

- Design and fabrication of a novel dielectric elastomer stack actuator auto-focus liquid lens with a new design that shows more advantages compared to available dielectric elastomer base liquid lenses.

- A new algorithm for the denoising and enhancement of satellite images.

- Introducing new interpolation kernels of $Lanczos$ and $sicexp$ and a $\alpha$ blending algorithm to improve resolution in SR algorithms. These kernels are tested on two state-of-the-art SR algorithms and show a better performance compare with using normal bicubic kernel.

- A novel low complexity dictionary based SR algorithm which shows a better performance when compared with the state-of-the-art SR algorithms.

- Demonstration of advantages of using SR algorithms in other areas such as action recognition and face recognition and pain detection.

The rest of this thesis is organized as follows. A literature review on available liquid lenses, as well as a detailed overview of the proposed auto-focused liquid lens, are presented in Chapter 2. Chapter 3 contains a comprehensive literature review of SR algorithms. All proposed SR algorithms, as well as the effects of

SR algorithms on different computer vision applications (in particular, biometric recognition), are provided and discussed in Chapter 3. In Chapter 4, the experimental results on the auto-focused liquid lens and the proposed SR techniques are reported, and detailed discussion of the results is provided. Finally, the conclusion section concludes the thesis.

# CHAPTER 2

# LIQUID LENSES

## 2.1  Autofocus Fluid Lenses

Autofocus fluid lenses can be divided into two primary types: transmissive and reflective. The reflective type works as a variable mirror applied in reflector telescopes. The principal of the reflective liquid lens is based on the variable mirror (for example, by using mercury and applying centripetal forces to create a smooth reflective concavity), that reduces the cost of such lenses by a factor of ten when compared with the traditional ways of fixed curved glass. The telescope based on the liquid reflective mirror (mercury) resides at the University of British Columbia.

In contrast to the reflective fluid lens, the transmissive fluid lens is based on an active change of the convex/concave lens shape over immiscible fluids with a different refractive index. The change of the liquids can be obtained mechanically or electrically. The advantage of such transmissive fluid lenses is the high optical quality in the range of 10 m.

Figure 2.1 shows a design of a fluid lens of Varioptic [3]. Varioptics uses manipulation of two fluids electrically based on electrowetting techniques.

The liquid lens uses two isodensity liquids, an insulator, and a conductor. The variation of voltage leads to a change of curvature of the liquid-liquid interface, which in turn leads to a change of the focal length of the lens [13]. A similar design of the liquid lens cell has been made by Philips using electrowetting techniques. The difference can be found in the coating of the tube with one part hydrophobic and the other with a hydrophilic material forming a hemispherical lens-shaped mass at the end of the tube. Both electrical induced shape lens devices require a high voltage between 2-4 kV to obtain focus changes. The mechanical manipulation of the surface tension between the two liquids is introduced over functional elements, such as pumping liquids, forcing the lens shape based on the Singapore type [35].
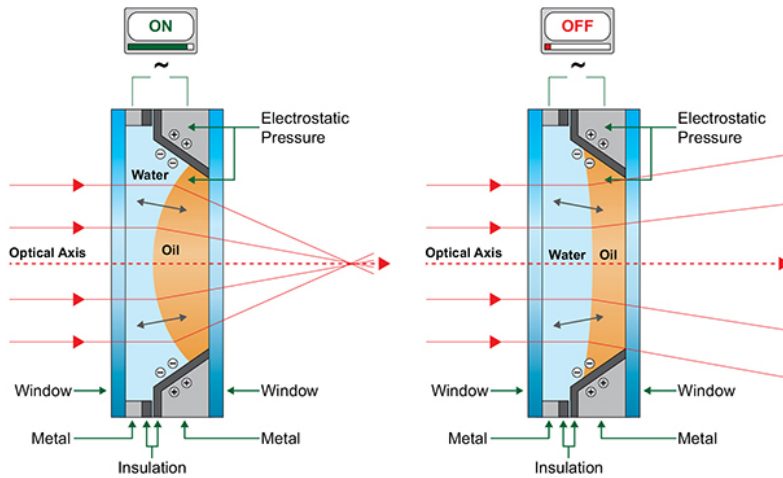
Figure 2.1: Design of the Varioptic fluid lens. The image is taken from [3].

Many extremely small cameras and cell phones simply do not have enough room for moving a rigid lens in the distance that is required for a range of focal lengths. An adaptive liquid lens, however, enables small cameras to focus without requiring any extra room. However, traditional approaches, which use an electric current to change the surface shape of a liquid, require a lot of power. The fluid lens technology needs to be adapted on the conducting polymer membrane actuator and where the changes of the shape of the drop can thereby be achieved by low energy actuator forces.

### 2.1.1 Related Work

Due to some problems of the liquid lenses which limit their practical use, such as membrane puncture, high stress, and a high driving voltage requirement, Lau et al. [62] separated dielectric elastomer actuator (DEA) from lens membrane. In fact, they used a liquid-immersed DEA as a diaphragm pump to inflate or deflate the liquid lens by hydraulic pressure. They generated an 8 mm lens by using VHB 4910 as an actuator material, graphite powder with oil immersion as an electrode material and a driving voltage of up to 1.8 kV. Figure 2.2 shows the lens and the concepts of its principal working area.

In [115] the authors presented a liquid lens system which makes use of an in-line, transparent electroactive polymer actuator. Their lens was constructed from a transparent liquid of fixed volume, a stiff frame, a passive elastomer membrane and an electroactive elastomer membrane. In their work, VHB 4910 as an actuator material and a single-walled carbon nanotube as an electrode material are used for
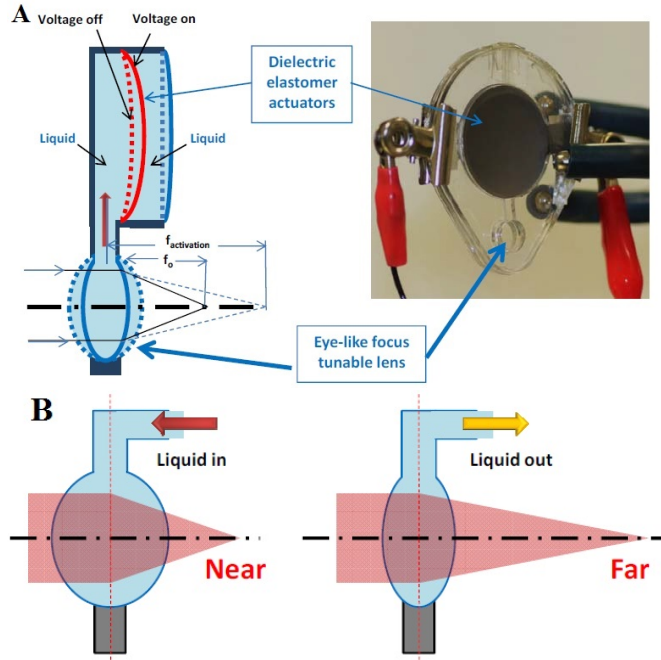
Figure 2.2: schematic (A) and principal concepts of the lens (B) [62].

the fabrication of the lens. The driving voltage of their lens is up to 5 kV. Figure 2.3 shows the structure of the lens.
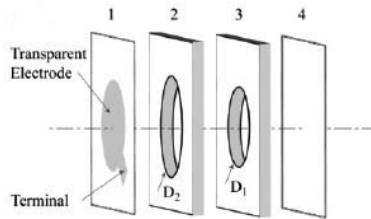


Figure 2.3: Construction of the tunable lens, consisting of 3 main parts: dielectric elastomer actuator membrane (part 1), frame (parts 2 and 3), and passive elastomer membrane (part 4). The aperture of the lens is defined by the smallest diameter of the frame cavity or D1 [115].

Carpi et al. [18] made a liquid lens that consists of a fluid-filled elastomeric lens integrated with an annular elastomeric actuator working as an artificial muscle. In order to generate an 8 mm lens, they used VHB 4905 as an actuator material, Carbon grease as an electrode material and a driving voltage of up to 4 kV. A schematic of the lens is shown in figure 2.4.
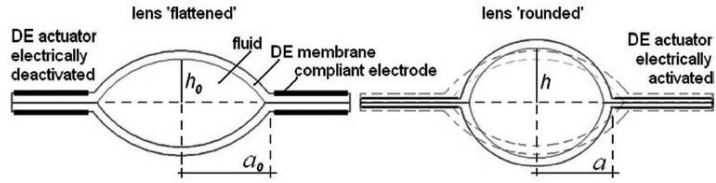
Figure 2.4: The schematic of the lens [18].

## 2.2 Design and Fabrication of the Proposed Liquid Lens

In order to fabricate the proposed liquid lens, two insoluble liquids and a membrane actuator are needed. In this case, distilled water and silicon oil are used, which are separated with a membrane actuator. The membrane has a hole with a diameter of 4 mm in the middle. A meniscus is formed by the liquids in correspondence to the central hole of the actuator as shown in figure 2.5.
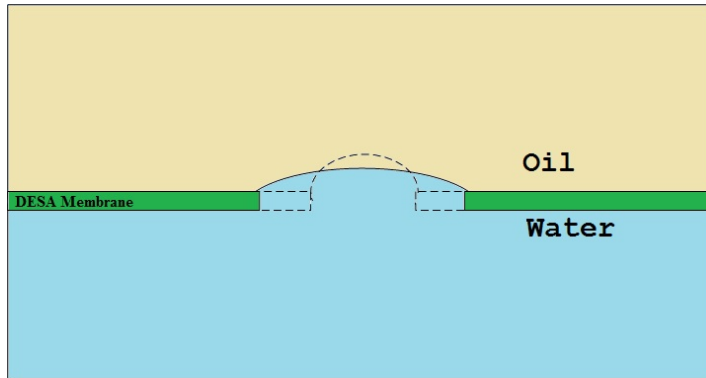


Figure 2.5: The working principal of an autofocus liquid lens based on mechanical lens change in applying membrane actuators [55].

The diameter and the shape of the meniscus are changeable by field-induced actuator deformation. The meniscus will work as a lens if light passes through the hole. The focal length of the lens is adjustable by applying the different electrical potential to the membrane actuator, which modifies the meniscus shape. A rectangular shape is considered for all parts of the initial device. The fabricated device through this research work is assembled easily in a flexible design with actuator membranes of various sizes [55]. Figure 2.6 shows the prototype device.

DEA transforms electric energy into mechanical work which consists of an elastic dielectric sandwiched between two compliant electrodes. Maxwell stresses force a compression of the dielectric material sandwiched between the electrodes
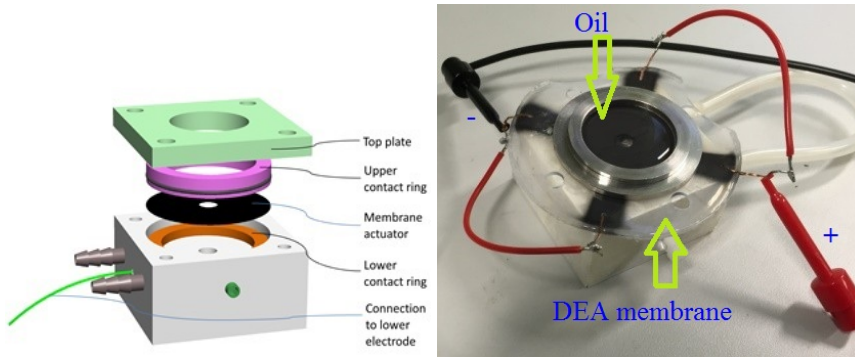
Figure 2.6: The prototype device which consists of a frame, a thin DESA membrane with a hole in the middle, silicon oil and water.

when a potential difference is applied to them. In this work, two DESA are fabricated by a fully automated fabrication process [68, 110]. Membranes are made of 30 layers that each layer has a thickness 45 $\mu$m but with a different active area. One of the membranes has an active area of 40 mm and another membrane has an active area of 20 mm. Figure 2.7 and figure 2.8 shows the principal design of a DESA and our proposed membranes, respectively.
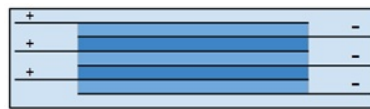


Figure 2.7: The principal design of a DESA.
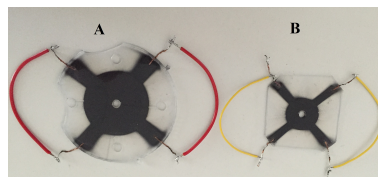


Figure 2.8: (A) membrane with an active area of 40 mm and (B) membrane with an active area of 20 mm.

According to [68], the polymer should fulfill three main requirements in order to produce a thin elastomer film. These requirements are fast cross-linking, cross-linking by polyaddition to avoid scission products and uncured components with a low viscosity. Due to these reasons, the Elastosil P7670 (Wacker Silicones) was used as an elastomeric film that meets these demands. The fabrication process of

membranes consists of three steps. In the first steps, two components of uncured polydimethylsiloxane (PDMS) are mixed in order to make an elastomer film and applied onto the disk of a spin coater. The thickness of the dielectric is directly dependent on the rotational speed of the spin coater that varies between $5\mu$m to $100\mu$m. In the second steps, thermal heating is used in order to speed up the curing process of the elastomer. Finally, in the last step, graphite powder is sprayed onto the dielectric through a shadow mask. These steps are repeated consecutively 30 times to fabricate the membranes. It is important to notice that in order to ensure encapsulation, a PDMS layer covers the last sprayed electrode of the DESA membrane. This process takes approximately 150 min for each membrane. A schematic of the fabrication process is shown in figure 2.9.
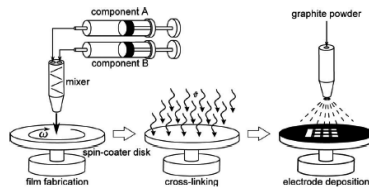


Figure 2.9: Process steps of fabrication of DESA. The image is taken from [68].

The thickness of the resulting film can be predicted by Eqn. (2.1) [68].

$$h(r) = \frac{h_0(r)}{\sqrt{1 + 4K h_0(r)^2 t}} \quad , \quad K = \frac{\gamma w^2}{3\eta} \tag{2.1}$$

where $h(r)$ is thickness distribution, $\gamma$ is material properties as density, $w$ is rotation speed, $\eta$ is viscosity, $t$ is time and $h_0(r)$ is initial thickness of the film.
A significant decrease in driving voltage shows the superiority of the proposed work in comparison with previous works. A comparison of the proposed work and some of the state-of-the-art methods is presented in section 4.1.

# CHAPTER 3

# IMAGE SUPER-RESOLUTION

## 3.1 Introduction

In this section, first, we present some notation. The LR and HR images are represented as matrices $\Psi_l \in \mathbb{R}^{N_l \times M_l}$ and $\Psi_h \in \mathbb{R}^{N_h \times M_h}$, where $N_h = \xi N_l$ and $M_h = \xi M_l$ and $\xi > 1$ is a integer scale-up factor. The blur operator is denoted by $H : \mathbb{R}^{N_h \times M_h} \longrightarrow \mathbb{R}^{N_h \times M_h}$, and the decimation operator for a factor $\xi$ in each axis is denoted by $Q : \mathbb{R}^{N_h \times M_h} \longrightarrow \mathbb{R}^{N_l \times M_l}$, which discards rows/-columns from the input image (the nearest neighbor interpolation function is used as a decimation function in this work). A sigmoid function is denote by $S$ is real-valued and differentiable, having a non-negative or non-positive first derivative, one local minimum, and one local maximum. The sigmoid function, also called the sigmoidal curve [123] or logistic function is defined by the equation $S(x) = \frac{1}{1+e^{-x}}$. A convolution operation between an image and a kernel is represented by $(*)$ which mathematically explains $A[x_a, y_a] * B[x_b, y_b] = C[x_c, y_c] = \sum_{i=1}^{x_a-1} \sum_{j=1}^{y_a-1} A[i,j] B[x-i, y-j]$ where $x_a$ and $y_a$ are a pixel location in image $A$. $x_b$ and $y_b$ are a pixel location in kernel $B$.

Image resolution refers to the number of pixels in an image. The resolution, called spatial resolution as well, is sometimes identified by the width and height of an image as well as the total number of pixels in the image. The resolution of an image is one of its important properties. However, this is limited to the capabilities of the image acquisition device, mainly the imaging sensor [74]. The size of this sensor defines the available resolution of produced images. Thus, for increasing the resolution of an image, one could increase the sensor density by reducing the space allocated to each pixel on the sensor, or simply by increasing the size of the sensor. The latter solution is expensive and slows down the imaging process, while the former solution decreases the amount of light incident on each sensor which in turn increases the shot noise [74].

Applying various signal processing techniques is another approach for en-

hancing the resolution of an image. One of the famous techniques for this purpose is SR [79, 24]. The basic idea behind SR methods is to recover/estimate one or more HR image(s) from one or more LR images [128, 95, 102]. HR image offers a high pixel density and thereby more details about the original scene. Huang and Tsai [43], as pioneers of SR, proposed a method for improving the resolution of satellite images of earth, where a large set of translated images of the same scene are available. They showed that SR, using multiple offset images of the same scene and a proper registration, can produce better HR images compared to spline interpolation. Since then, SR methods have become common practice for many applications in different fields, such as remote sensing [66, 101]; surveillance video [20, 67]; medical imaging, such as ultrasound, magnetic resonance imaging (MRI), and computerized tomography (CT) scans [47, 70, 71, 82]; optical character recognition (OCR) problems or face recognition [16, 38, 114, 36, 77]. SR algorithms are divided into two main categories of frequency domain algorithms and spatial domain algorithms [77]. These categories consist of several subcategories which are described in the sequel. Figure 3.1 shows a block diagram of these categories.
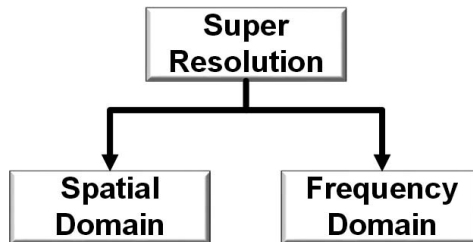


Figure 3.1: Main Categories of SR methods.

### 3.1.1 Frequency Domain SR Algorithms

SR algorithms of this group first transform the input LR image(s) to the frequency domain and then estimate the HR image in this domain. Finally, they transform back the reconstructed HR image to the spatial domain [31, 22, 121, 46, 75, 85]. Depending on the transformation employed for transforming the images to the frequency domain, these algorithms are generally divided into two groups: Fourier-transform based and wavelet-transform based methods. Figure 3.2 shows a block diagram of these algorithms.
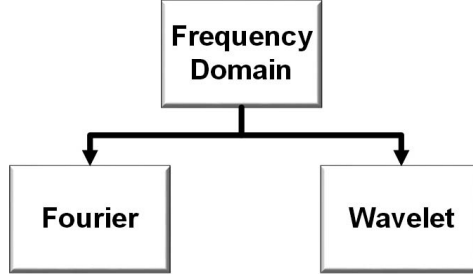
Figure 3.2: Subcategories of SR frequency domain.

### 3.1.2 Spatial Domain SR Algorithms

Based on the number of available LR observations, SR algorithms can be generally divided into two groups: single image based and multiple image-based algorithms. The algorithms included in these groups are explained in the following subsections.

#### Multiple Images based SR

Multiple-image SR algorithms receive several low-resolution images of the same scene as input and usually employ a registration algorithm to find the transformation between them. This transformation information is then used along with the estimated blurring parameters of the input low-resolution images to combine them into a higher scale framework to produce a super-resolved output image. For multiple-image SR algorithms to work properly, there should be sub-pixel displacements between input LR images. Furthermore, these sub-pixel displacements should be estimated properly by the registration algorithm, which is usually a challenging task, especially when a complicated motion of non-rigid objects, like the human body, needs to be modeled. These algorithms are guaranteed to produce proper higher resolution details; however, their improvement factors are usually limited by factors close to two [79]. Figure 3.3 shows subcategories of these types of SR algorithms.

Iterative back projection (IBP) methods are among the first methods developed for spatial-based SR. Some recent works are done based on the IBP method, such as works at [102, 95, 101, 97, 96]. The IBP model has two critical steps: the first step is construction of the model for the imaging process, and the second step is image registration. The first step can be described as

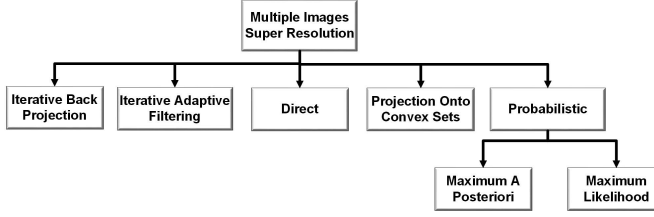$$\Psi_{l_k}(y) = QH \times \Psi_s(x) + n_k \qquad (3.1)$$

Figure 3.3: Subcategories of Multiple Images based SR.

where $\Psi_{l_k}$ are $k^{th}$ observed LR images, $y$ denote the pixel of LR images influenced by the area of $x$ of the SR image $\Psi_s$, $H$ is a blur kernel, $Q$ means decimating operator and $n_k$ is an additive noise term. More specifically, $Q$ performs a down-sampling with an integer factor of $\xi$, and $n_k$ is an independent identically distributed additive white Gaussian noise. In the first part of this method, a true (initial) SR image is assumed. Based on the imaging model given in Eqn. (3.2), different LR images are evaluated. Given the calculated LR images, a new SR image is obtained. Afterward, this new SR image is used to generate the new set of LR images. If this new set of LR images is the same as the earlier (previous) set, then the assumed SR image is the true SR image; otherwise, the error image obtained from the difference between the LR images is back projected to the assumed SR image. This process is repeated until no error image is left. IBP can be mathematically represented as

$$\Psi_s^{n+1}\left(x\right) = \Psi_s^n\left(x\right) + \sum_y \left(\Psi_{l_k}\left(y\right) - \Psi_{l_k}^n\left(y\right)\right) \times H^{BP} \qquad (3.2)$$

where $\Psi_s^n$ is the estimated SR image after $n$ iteration, $\Psi_{l_k}^n$ are calculated LR images from the imaging model of $\Psi_s^n$ after $n$ iteration, and $H^{BP}$ is the back projection kernel. Back projection kernel is a deblurring kernel which has the following relationship with the blurring kernel of the imaging model of Eqn. (3.1).

### Single Image based SR

As shown in figure 3.4, single image SR algorithms are divided into subcategories of learning based methods and reconstruction based methods.

Single-image SR algorithms do not have the possibility of utilizing sub-pixel displacements because they only have a single input. Instead, they employ a training step to learn the relations between a set of high-resolution images and their low-resolution counterparts. This learned relation is then used to predict the missing high-resolution details of the input low-resolution images. Depending on the relations between the training low- and high-resolution images, these algorithms can produce HR images that are far better than their inputs, by improvement factors
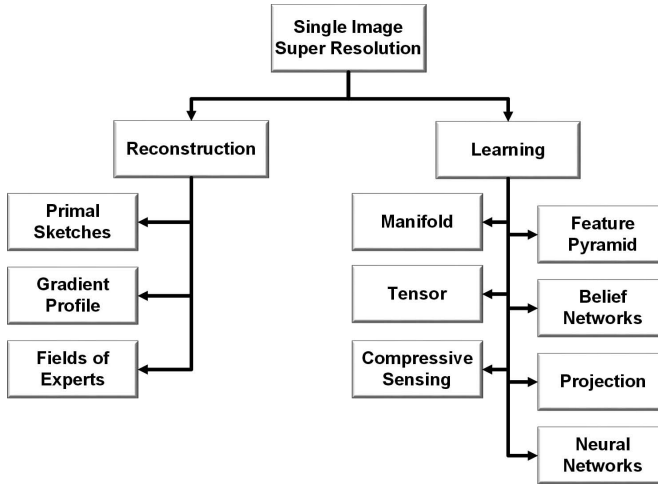
Figure 3.4: Subcategories of Single Image based SR.

that are much larger than two [79].

We focus on this type of algorithm in the rest of this thesis. The following sub-sections will present some approaches in order to find a good solution for image SR.

### 3.1.3 Background

Before talking about proposed algorithms, we introduce problems that we are facing as well as methods that can hep in implementing the proposed algorithms. There are various quality factors in satellite images. Two of the most important ones are noises and resolution issues. When a satellite image is being captured there are some noise added to the images and these noise have different sources, such as noise added during the data transmission from the capturing station to the research centers, or noises added by the image acquisition tools [116]. Interpretation of a noisy image is difficult for human observers. Such a noisy image needs to be denoised. Furthermore, the noisy images sent by satellites cannot be directly processed. A pre-processing stage is needed whereas image denoising is one of these pre-processing stages. The aim of image denoising is to remove the noise while keeping significant features of the image [117].

Interpolation in image processing is a method to increase the number of pixels in a digital image and is widely used in many image processing applications such as real-time actuator positioning [21], multiple description coding [108], and SR [102, 23]. There are three well-known interpolation techniques, namely nearest neighbour, bilinear and bicubic which use nearest neighbour, bilinear and bicubic

kernels for interpolation, respectively. It should be noticed that in sampling, kernel is a core function which is being defined to estimate the value of the created position in a sampling scenario. Bicubic interpolation is more sophisticated than the other two techniques and produces smoother edges. The bicubic kernel that is used for interpolation is defined in Eqn. 3.3.

$$\psi\left(x\right) = \begin{cases} \left(a+2\right)\left|x\right|^3 - \left(a+3\right)\left|x\right|^2 + 1 & \left|x\right| \leqslant 1 \\ a\left|x\right|^3 - 5a\left|x\right|^2 + 8a\left|x\right| - 4 & 1 \leqslant \left|x\right| \leqslant 2 \\ 0 & otherwise \end{cases} \qquad (3.3)$$

where $a = -0.5$ in the present implementation.

Wavelets are also play a significant role in many image processing applications. The wavelet analysis procedure for adoption of a wavelet prototype function, called an analyzing wavelet or mother wavelet. The two-dimensional wavelet decomposition of an image is performed by applying the one-dimensional Discrete wavelet Transform (DWT) along the rows of the image first, and then the results are decomposed along the columns. This operation results in four decomposed subband images referred to the approximate band (LL), vertical band (LH), horizontal band (HL), and diagonal detail band (HH). The frequency components of those subbands cover the full frequency spectrum of the original image.

Image resolution enhancement by using wavelets is a relatively new subject and recently many new algorithms have been proposed [25, 132, 118]. Carey et al. [17] have attempted to estimate the unknown details of wavelet coeffcients in an effort to improve the sharpness of the reconstructed images. Their estimation was carried out by investigating the evolution of wavelet transform extrema among the same type of subbands. Edges identifed by an edge detection algorithm in lower frequency subbands were used to prepare a model for estimating edges in higher frequency subbands and only the coeffcients with signifcant values were estimated as the evolution of the wavelet coeffcients. Another wavelet based image super resolution technique is Wavelet Domain Zero Padding and Cycle-Spinning (WZP and CS) [119]. The proposed method adopts the cycle-spinning methodology in the wavelet domain [119]. There are more state-of-the-art techniques which are using Dual Tree Complex Wavelet Transform (DT-CWT), Stationary Wavelet Transform (SWT), and also DWT [8, 23, 24].

A popular image denoising method is Local Adaptive Bivariate Shrinkage Function (LA-BSF) which requires priori knowledge of noise and marginal variances [112, 92]. This method uses non-Gaussian bivariate distributions and benefts from the dependencies between the wavelet coeffcients and their parents. It is reported that the models exploiting the dependency between wavelet coeffcients give better results than the ones using an independence assumption [112]. The performance is improved by estimating model parameters in a local neighbourhood. Satisfactory image denoising results have been obtained using DT-CWT based LA-BSF

in [112, 92].

The DT-CWT has been developed by N. G. Kingsbury [52, 53, 54] in order to overcome the disadvantages of classical DWT (i.e. lack of shift-invariance and poor directional selectivity). The DT-CWT uses specially designed real filters in order to provide the desired characteristics of the transform (i.e. approximately shift-invariance, good directional selectivity (for two or more dimensions), perfect reconstruction, limited redundancy, 2:1 for 1-D ($2m$ for $m$-D), efficient computation, only twice the simple DWT for 1-D ($2m$ times for $m$-D)) [54]. A one level DT-CWT decomposition results in two parallel trees (real and imaginary). The DT-CWT has the advantages of approximately shift-invariance and good directional selectivity (for two or more dimensions) over the classical DWT which are essential for many signal processing applications [52, 54].

## 3.2 Proposed Algorithms

### 3.2.1 Satellite Image Enhancement: Systematic Approach for Denoising and Resolution Enhancement

There are two significant parts, namely, denoising [116] and resolution enhancement in this proposed method [106]. Firstly, the noisy image is decomposed into subbands by using six-level DT-CWT. A one level 2D DT-CWT results in two complex-valued low frequency subband coefficients and six complex valued high frequency subband coefficients oriented at $+75°$, $+45°$, $+15°$, $-15°$, $-45°$, and $-75°$ [54]. The DT-CWT uses specially designed real filters which are different at the first-level and remaining levels of the transform. In this algorithm, a (9,7)-tap Antonini biorthogonal filter set is used at the first-level, and 6-tap Q-shift dual filters are used at the remaining levels of the DT-CWT [53]. The reason for using the DT-CWT in the denoising algorithm is that it has the properties of approximately shift-invariance and good directional selectivity lacking in the classical DWT. These properties are essential for many signal processing applications, including denoising [52, 54]. These noisy subband coefficients are denoised using the LA-BSF algorithm [112, 92]. The LA-BSF requires prior knowledge of the noise variance $\sigma n^2$ and marginal variance $\sigma^2$ for each wavelet coefficient. In the denoising algorithm, the marginal variance for the $k^{th}$ coefficient is estimated using neighboring coefficients in a region $N(k)$. Where $N(k)$ is defined as all the coefficients within a square shaped window that are centered at the $k^{th}$ coefficient. The LA-BSF is applied to the magnitude of complex coefficients since the real and imaginary parts are not shifted invariant. It is assumed that the images are corrupted by the Gaussian noise. A $7 \times 7$ window size $N(k)$ is used for the best denoising results [113].

In image resolution enhancement by using interpolation the main loss is on its high frequency components (i.e., edges), which is due to the smoothing caused by interpolation. Hence, in order to increase the quality of the super-resolved image, preserving the edges is essential. As it was previously mentioned, in this work DWT [72] has been employed in order to preserve the high-frequency components of the image. In the resolution enhancement stage, DWT is used to decompose an input image into subband images. LH, HL, and HH subband images contain the high-frequency components of the input satellite image which are obtained, respectively, as follows:

$$L_{j+1}[row][y] = \sum_{i=0}^{N-1} w[i] \times LL_j[row][y-i] \qquad (3.4)$$

$$H_{j+1}[row][y] = \sum_{i=0}^{N-1} h[i] \times LL_j[row][y-1-i] \qquad (3.5)$$

$$LL_{j+1}[x][col] = \sum_{i=0}^{N-1} w[i] \times L_j[x-i][col] \qquad (3.6)$$

$$LH_{j+1}[x][col] = \sum_{i=0}^{N-1} h[i] \times L_j[x-1-i][col] \qquad (3.7)$$

$$HL_{j+1}[x][col] = \sum_{i=0}^{N-1} w[i] \times H_j[x-i][col] \qquad (3.8)$$

$$HH_{j+1}[x][col] = \sum_{i=0}^{N-1} h[i] \times H_j[x-1-i][col] \qquad (3.9)$$

where $j = \{0,1,...,L-1\}$, $x = \{0,1,...,X/2^{j+1}-1\}$, $y = \{0,1,...,Y/2^{j+1}-1\}$, $row = \{0,1,...,X/2^j-1\}$, $col = \{0,1,...,Y/2^{j+1}-1\}$, $LL_0[x][y] = \psi[x][y]$ and $\psi[x][y] \in \mathbb{R}^{N \times N}$

In the resolution enhancement step, the bicubic interpolation is applied to high-frequency subband images. Here, instead of using LL, which contains less high frequency information compare to the input image, we are using the input satellite image for re-sampling by interpolation which increases the quality of the super-resolved image. By applying interpolation to the input satellite image and high-frequency subbands and then by using the inverse DWT (IDWT), the output image will contain sharper edges than the image obtained by re-sampling of the satellite image directly. This is due to the fact that the interpolation of isolated high-frequency components in HH, HL, and LH will preserve more high-frequency components after the interpolation of the respective subbands separately than interpolating the satellite image directly. In all steps of the proposed noisy satellite image resolution enhancement technique, the db.9/7 wavelet as mother wavelet function and a bicubic kernel as interpolation technique have been used. Figure 3.5 and algorithm 1 show the proposed noisy satellite image resolution enhancement system. Experimental results and comparison of the algorithm with state-of-the-art algorithms are described in section of 4.2
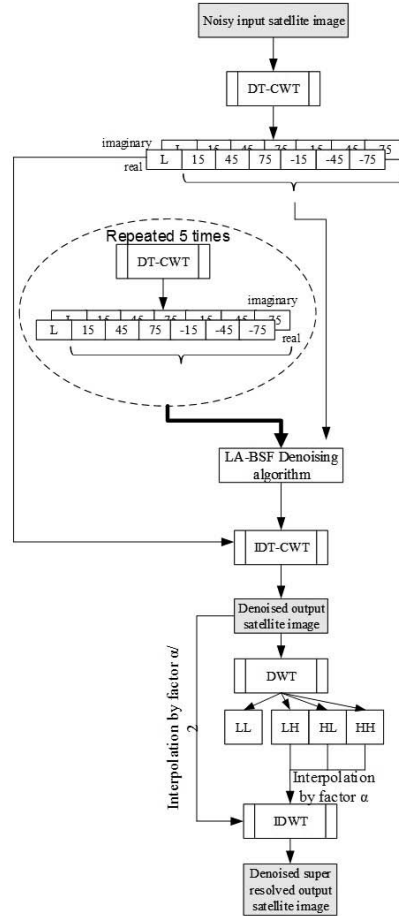
Figure 3.5: The proposed noisy satellite image resolution enhancement technique. Denoising part of the figure is taken from [116].

## 3.2.2 New Two-Dimensional Sampling Kernel-based Resolution Enhancement

In this section, some new kernels and an $\alpha$ blending algorithm are introduced which help to improve image resolution in SR algorithms. The main contribution of this work [104] is deriving new sets of kernels which can be used in sampling. The term of $\alpha$ blending algorithm in this thesis refers to a process of combining two different images.

---
**Algorithm 1** Satellite image enhancement
---
**Require:** LR Noisy Satellite Image
**Ensure:** HR Image
    **Decomposition** of noisy image to subbands by using a six-level DT-CWT.
    **Keep** the first level output and repeat the process for five more times.
    **Applying** LA-BSF denoising algorithm on results of the first level and sixth level.
    **Inverse decomposition** by using IDT-CWT.
    **Applying DWT** on the output of IDT-CWT .
    **Applying interpolation** on LH, HL and HH subbands by a factor of $\xi$ in parallel.
    **Applying interpolation** on the output of IDT-CWT by a factor of $\frac{\xi}{2}$.
    **Replace** the interpolated image of the output of IDT-CWT to LL subband
    **Inverce DWT** on interpolated subbands.
---

### Problem

SR algorithms usually employ one of the common upsampling kernels, like bilinear or bicubic, as an initial guess for desired HR image, and then they differ from each other in the way they fine tune this initial HR image and/or converge to desired HR image [30, 32, 76]. Though these common kernels are simple and efficient, we show in this section that they can be replaced by better alternatives resulting in better super-resolved images quantitatively in terms of known assessment measures, like Peak Signal to Noise Ratio (PSNR) and Structural Similarity Measure (SSIM) and qualitatively in terms of the visual appearance of the generated super-resolved images.

More specifically, we introduce two kernels of *Lanczos* and *Sincexp* for upsampling, instead of bicubic interpolation, in the context of SR. The resulted super-resolved images by these two kernels, alone, are not always better than the super-resolved images generated by SR algorithms that are built upon bicubic interpolation, but the combined results of these two kernels are always better than the results obtained by SR algorithms when they use bicubic interpolation for their initial guess. To show this, we have employed the introduced kernels in two known SR algorithms of [83] and [50].

In order to show the efficiency of the proposed kernels, we have chosen the resolution enhancement problem. The general algorithm of the proposed resolution enhancement technique is shown in figure 3.6. Firstly the LR image is super-resolved by using [83] and [50] techniques in which Lancoz and Sincexp kernels are adopted. The final HR image is achieved by combing the output of each kernel. The best value of $\alpha$ is chosen by monitoring the performance of the system for

over 1000 images and observing the $\alpha$ value which results in the best quantitative output during the training stage.
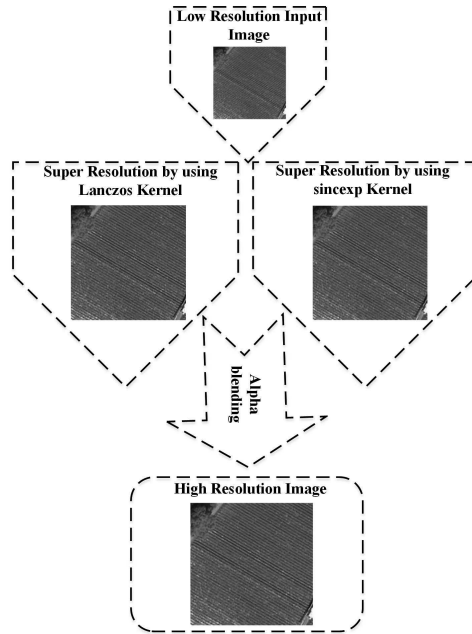


Figure 3.6: The block diagram of the proposed method.

## Sampling Kernels

An image is a combination of contours and smooth areas. For smooth areas, a kernel which has good approximation properties for smooth functions is needed (which is guaranteed if we have an estimate of the order of approximation via high order of modulus of smoothness). A good candidate is the sinc function, but it is not in $L^1$ where $L^1$ is the 1-norm which mathematically deffined $\|x\|_1 = \sum_{i=1}^{n} |x_i|$. Therefore, another kernel is needed with good properties. The Lanczos kernel [59] can be a good candidate due to its simple structure (the kernel will be described later). At the same time, Lanczos kernel is a reproducing kernel of a space which coincides with a corresponding $L^p$ $(p \geqslant 1)$ Bernstein space [80] (like the sinc kernel for $1 < p < \infty$) and this Bernstein space are dense in $L^p$ $(p \geqslant 1)$ where $L^p$ is the $p$-norm which mathematically defied $\|x\|_p = (\sum_{i=1}^{n} |x_i|^p)$. This gives an estimate of the order of approximation via an arbitrary order of modulus of smoothness. For the contour parts of an image, a kernel whose derivatives are rapidly decreasing with a good energy concentration is needed, like *Sincexp*, which has exponential decay.

For sampling, we usually use the interpolation methods that fit into a general scheme of interpolation with convolution where their spectral properties are described at [48, 40]. In the convolution, we can use an even kernel $s_\lambda$ defined via the Fourier cosine transform of an even window function $\lambda$ and given by

$$s_\lambda(x) = \int_0^\infty \lambda(u) \cos(\pi x u) \, du. \tag{3.10}$$

In the above equation, $x$ is a space variable, $u$ is a dummy variable and $\lambda$ is an arbitrary continuous even function which satisfies $\lambda(0) = 1$, $\lambda(2k) = 0$ $(k \in \mathbb{Z})$.

With the window function

$$\lambda_{L,n}(u) := \begin{cases} 1 & 0 \leqslant u \leqslant \frac{n-1}{2n}, \\ \frac{1}{2}(1 + n(1 - 2u)) & \frac{n-1}{2n} \leqslant u \leqslant \frac{n+1}{2n}, \\ 0 & u \geqslant \frac{n+1}{2n} \end{cases} \tag{3.11}$$

the Eqn. (3.10) defines the Lanczos $n$-kernel [59]

$$s_{L,n}(x) := \operatorname{sinc} \frac{x}{n} \operatorname{sinc} x, \tag{3.12}$$

in which $n \in \mathbb{N}$ is a parameter. Figures 3.7 and 3.8 show the Lanczos kernel in spatial and frequency domain respectively.
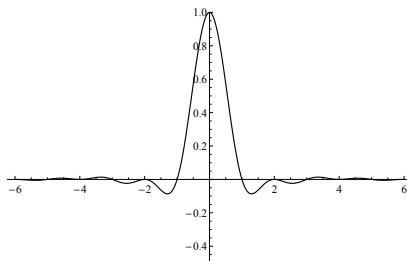


Figure 3.7: The Lanczos function in the spatial domain.

Some other kernels are defined, by using the window functions with the support $[-1, 1]$, i.e.

$$\lambda(u) = \begin{cases} l(u) & |u \leqslant 1| \\ 0 & elsewhere \end{cases} \tag{3.13}$$

where $l(u)$ is some window function with conditions as in (3.10). For example, the cosine based function $l_H(u) := \frac{1}{2}(1 + \cos(\pi u))$ defines the Hann kernel [57]

$$s_H(x) := \frac{1}{2} \frac{\operatorname{sinc} x}{1 - x^2} \tag{3.14}$$
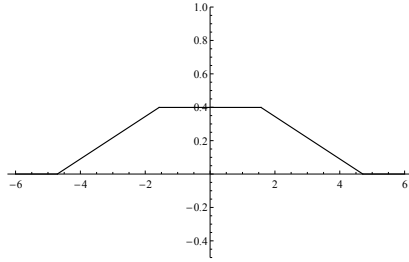
Figure 3.8: The Lanczos function in frequency domain.
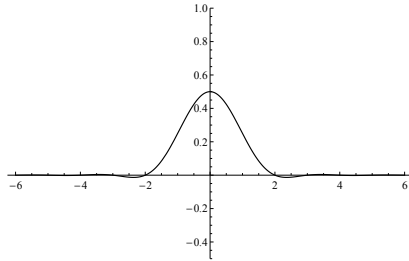


Figure 3.9: The Hann function in the spatial domain.
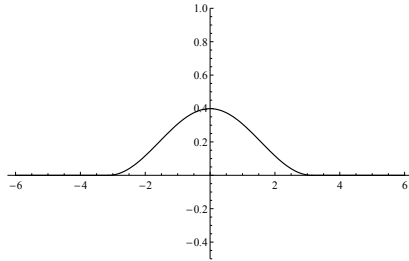


Figure 3.10: The Hann function in the frequency domain.

Figures 3.9 and 3.10 show the Hann kernel in spatial and frequency domain respectively. The general cosine-sum $l_{B,\mathbf{a}}(u) := \sum_{k=0}^{m} a_k \cos(k\pi u)$ with parameter vector $\mathbf{a} = \{a_k\}$ defines the Blackman-Harris kernel [58]

$$s_{B,\mathbf{a}}(x) := \frac{1}{2} \sum_{k=0}^{m} a_k \Big( \operatorname{sinc}(x - k) + \operatorname{sinc}(x + k) \Big) \qquad (3.15)$$

provided the sums of odd and even coefficients are both equal to $1/2$, i.e. ($\lfloor x \rfloor$ is

the largest integer less than or equal to $x \in \Re$)

$$\sum_{k=0}^{\left\lfloor \frac{m}{2} \right\rfloor} a_{2k} = \sum_{k=1}^{\left\lfloor \frac{m+1}{2} \right\rfloor} a_{2k-1} = \frac{1}{2}. \tag{3.16}$$

In fact, kernel in Eqn. (3.14) is a special case of the Blackman-Harris kernel and can be obtained by taking $m = 1$ in 3.15. Figures 3.11 and 3.12 show the Blackman-Harris kernel in spatial and frequency domain respectively. The
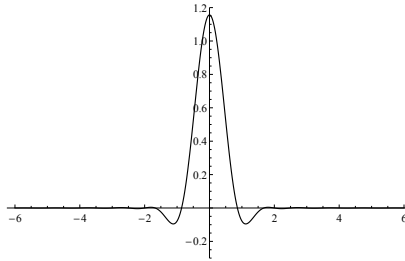


Figure 3.11: The Blackman-Harris function in the spatial domain.
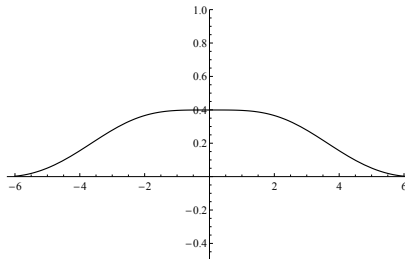


Figure 3.12: The Blackman-Harris function in the frequency domain.

function $l_{G,\delta}(u) := \frac{1}{2}\left[\operatorname{erf}\left(\frac{u+\pi}{2\delta}\right) - \operatorname{erf}\left(\frac{u-\pi}{2\delta}\right)\right]$, where $\operatorname{erf}$ is an error function $\operatorname{erf}(x) = \frac{2}{\sqrt{x}}\int_0^x e^{-t^2}\,dt$ and $\delta > 0$ is a parameter, which defines the sinc kernel with a Gaussian multiplier (the corresponding interpolation technique is further referred to as *Sincexp*) [111]

$$s_{G,\delta}(x) := e^{-\delta^2 x^2} \operatorname{sinc} x. \tag{3.17}$$

Figures 3.13 and 3.14 show the Sincexp kernel in spatial and frequency domain respectively.

The described kernels are more sophisticated than the common linear and cubic spline kernels as they are closer approximations to the theoretically optimal
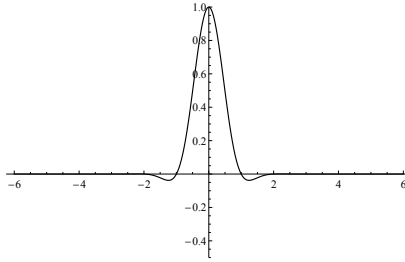
Figure 3.13: The Sincexp function in the spatial domain.
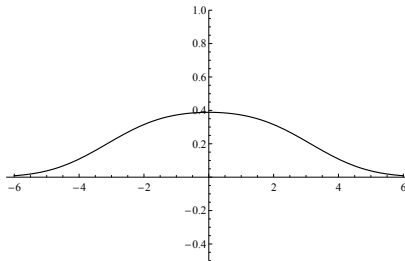


Figure 3.14: The Sincexp function in the frequency domain.

$sinc$ function, and thus often yielding a sharper image. The quality of the kernels can be described in terms of the $k$-order modulus of smoothness [15]. The higher order it is, the better interpolation results we have for the smooth functions (images) or smooth regions of functions (images).

The cubic spline is a positive kernel. Such kernels have many advantages, but also some disadvantages such as the impossibility of obtaining a modulus of smoothness of an order higher than 2. On the contrary, kernels defined in (3.12), (3.15), (3.17) are not positive kernels, and this fact yields a high-order modulus of smoothness [111].

Among all kernels, the Blackman-Harris kernel has the most possibilities to set its free parameters; therefore, using it can be more complicated. If the particular Blackman-Harris kernel corresponds to the $k$-order modulus of smoothness, then we have good interpolation results for the polynomial function of degree less than or equal to $k$ (in terms of images, the higher degree of the polynomial corresponds to a smoother image). Some choices of parameters may result in the fact that at the sample points, the evaluated pixel value does not equal the initial value. To fix this, $a_0 = 1/2$ must be taken in ( 3.15) [59].

It may be noticed that Eqn. (3.10) can also be used to obtain common interpolation kernels, such as cubic and spline (in this case, we must take the window function $\lambda_{S,3}(u) := \text{sinc}^3(\frac{u}{2})$). Note that for a given scale factor $w$ which determines the length (hereinafter support-length) of the truncated kernel sup-

44

port, which is $\left[\frac{-w}{2}; \frac{w}{2}\right]$ in the case of upsampling and $\left[\frac{-w}{(2\xi)}; \frac{w}{(2\xi)}\right]$ in the case of downsampling, the number of support points involved in the computation is $n =$ support-length$+2$. Hence, there is no connection between $n$ in the kernel (3.12) and the number of support points (this $n$ merely defines the shape of the kernel function). In order to increase the number of support points, the support-length should be increased (i.e. the support should be less truncated). Computations show that increasing the number of support point leads to a slightly better performance (provided the number of points is chosen wisely, i.e. the symmetry of the kernel is preserved; in our case, an even number should work well). It is also possible to increase the support-length and therefore the number of points not by taking less truncated support, but by scaling the kernel function. However, this will not give good results because of the smoothing effect. We will use these kernels for initial enlargement in SR algorithms and help them to show better performances.

### Adopted SR Techniques

In general, there are two models of image decimation in the literature. In the first one, a blur operator and a decimation operation plus an additional noise are applied to the input image in order to decimate it to an LR image, as shown in Eqn. (3.18):

$$\Psi_l = QH\Psi_h + n \tag{3.18}$$

where $\Psi_l$ is the LR image, $\Psi_h$ is defined as the input HR image, $H \in \mathbb{R}^{N_h \times N_h}$ is a blur operator, $Q \in \mathbb{R}^{N_l \times N_h}$ is a decimation operator for a factor $\xi$ in each axis, which discards rows/columns from the input image where $\xi > 1$, and $n$ is additional noise.

In the second approach, a decimation operation is used for the decimation, as shown in Eqn. (3.19):

$$\Psi_l = Q\Psi_h \tag{3.19}$$

In this research work, the second approach has been used in order to represent the effectiveness of the proposed kernels. Two state-of-the-art techniques, namely a statistical prediction model based on sparse representations [83] and example-based learning [50] are adopted by using the proposed kernels in order to generate an HR image.

**Statistical Prediction Model based on Sparse Representation Resolution Enhancement**

The technique proposed in [83] benefited sparse representation for generating an HR image. Here, we briefly introduce this algorithm.

The LR image can be seen as a blurred and down-sampled version of the HR

image. This observation model can be formulated as follows:

$$\Psi_l = QH\Psi_h + n \tag{3.20}$$

Since there are many HR images satisfying the reconstruction constraint for a given LR image, the process of recovering $\Psi_h$ from $\Psi_l$ is ill-posed. An effective way to deal with this problem is sparse representation.

Sparse representation has become an important tool for single image SR. The single image SR problem via sparse representation consists of the coding stage and the linear combination stage. The coding stage is formulated as follows:

$$\overset{Min}{\varphi}\left\{\left\|\tilde{\Psi}_l - \tilde{D}\varphi\right\|_2^2 + \lambda\left\|\varphi\right\|_1\right\} \tag{3.21}$$

where $\tilde{\Psi}_l = \begin{bmatrix} E\Psi_l \\ \beta z \end{bmatrix}$ and $\tilde{D} = \begin{bmatrix} ED_l \\ \beta PD_h \end{bmatrix}$, $D_h$ and $D_l$ are HR dictionary and LR dictionary respectively, $E$ is a feature extraction operator, $\Psi_l$ is the input LR image patch, $P$ extracts the region of overlap between the current target patch and previously reconstructed image patch, $\varphi$ is the sparse coefficients matrix and $z$ contains the values of the previously reconstructed HR image on the overlap. The parameter $\beta$ controls the tradeoff between matching the LR input and finding a HR patch that is compatible with its neighbors.

In the linear combination stage, the $i^{th}$ image patch $p_{h_i} \in \mathbb{R}_n$, which is extracted from the HR image $\Psi_h$, can be represented as a sparse linear combination in the HR dictionary $D_h \in \mathbb{R}^{n \times K}$:

$$p_{h_i} \approx D_h\varphi_i, \quad \varphi_i \in \mathbb{R}^k, \|\varphi_i\|_0 \ll K \tag{3.22}$$

where $\varphi_i$ is the sparse coefficient vector which is computed by the input LR image patch and the LR dictionary.

**Example-based learning for single-image resolution enhancement**

The work in [50] is an extension of [49], which uses kernel ridge regression in order to estimate the high-frequency details of the underlying HR image. Also, a combination of gradient descent and kernel matching pursuit is considered and allows time-complexity to be kept to a moderate level. Actually, the proposed method improves the SR method presented in [30]. In this algorithm, for a given set of training data points $(x_1, y_1), ..., (x_l, y_l) \subset$ $mathbbR^M \times$ $mathbbR^N$, the following regularized cost functional is minimized.

$$O\left(\{f^1, ..., f^N\}\right) = \sum_{i=1,...,N}\left(\frac{1}{2}\sum_{j=1,...,N}\left(f^i(x_j) - y_j^i\right)^2 + \frac{1}{2}\lambda\left\|f^i\right\|_H^2\right) \tag{3.23}$$

where $y_j = \left[ y_j^1, ..., y_j^N \right]$ and $H$ is a reproducing kernel Hilbert space. Due to the reproducing property, the minimizer of Eqn. (3.23) is expanded in kernel functions:

$$f^i (\cdot) = \sum_{j=1,...,l} a_j^i k (x_j, \cdot), \quad i = 1, ..., N \tag{3.24}$$

where $k$ is the generating kernel for $H$, which is chosen as a Gaussian kernel $\left( k (x, y) = \exp \left( -\|x - y\|^2 / \sigma_k \right) \right)$. Eqn. (3.23) is the sum of individual convex cost functionals for each scalar-valued regressor and can be minimized separately. The final estimation of pixel value for an image location $(x, y)$ is then obtained as the convex combination of candidates given in the form of a softmax:

$$\tilde{\Psi}_H (x, y) = \sum_{i=1,..N} w_i (x, y) Z (x, y, i) \tag{3.25}$$

where $w_i (x, y) = \exp \left( -\frac{|d_i(x,y)|}{\sigma_C} \right) / \left[ \sum_{j=1,...,N} \exp \left( -\frac{|d_j(x,y)|}{\sigma_C} \right) \right]$ and $Z$ is the initial SR image that is generated by a bicubic interpolation.
Actually, the proposed kernels improves the SR method presented in [30].

### $\alpha$-blending

Each super-resolved image includes some special details of the image. In order to incorporate all these details, we need to combine the super-resolved images in a single image. For this purpose, the following $\alpha$ blending technique by using an $\alpha$ is adopted:

$$\Psi_H = \alpha \times \Psi_{H_L} + (1 - \alpha) \times \Psi_{H_S}, \tag{3.26}$$

in which $\Psi_{H_L}$ and $\Psi_{H_S}$ are the HR images produced by the *Lanczos* (we set in Eqn. (3.12) $n = 2$) and *Sincexp* (we set in Eqn. (3.17) $\delta = 1$) based SR algorithms, respectively. The experimental results over 1025 non-facial images and 1013 facial image sample images randomly selected from some databases, showed that the considered forms of *Lanczos* and *Sincexp* are more effective than the considered form of *Blackman-Harris* (we set in Formula 3.15 $m = 3$ and parameter vector **a** corresponding to $4^{th}$ order modulus of smoothness and minimal first absolute moment); hence, in $\alpha$ blending, only the HR images produced by *Lanczos* and *Sincexp* were used. The reason for the weaker performance of *Blackman-Harris* could be that in this particular case at the sample points the evaluated pixel value did not equal the initial value.

The reason why such a $\alpha$-blending of kernels works well is that *Lanczos* is good for smooth parts but generates Gibbs phenomenon on contours, while *Sinexp* follows the image more precisely on contours but is not so good on smooth areas (more sensitive to noise). The parameter $\alpha$ characterizes the proportions of smoothness and contour parts for given types of images.

### 3.2.3 Reducible Dictionaries for Single Image Super-Resolution based on Patch Matching and Mean Shifting

In the proposed algorithm, we are contributing to the field of image SR by introducing the following new approach:

- We do not employ complex techniques such as sparse representation[129] that make the system slow. We just utilize separate dictionaries for LR and HR patches.

- The minimum distance between input LR patches and LR patches in the LR dictionary is used for choosing a corresponding HR patch in the HR dictionary.

- A mean shift method is used for illumination enhancement in order to avoid blocking effects in the super-resolved image.

Two acquisition models are commonly used in the literature [83] to describe how an LR image is generated from an HR image, and each of them has a different rationale. The first assumes that prior to decimation, a known low pass filter is applied on the image,

$$\Psi_l = Q\{H\{\Psi_h\}\} + v \tag{3.27}$$

where $v$ is an additive noise in the acquisition process. The corresponding problem of reconstructing $\Psi_h$ from $\Psi_l$ is also referred to in the literature as zooming deblurring [131]. The second acquisition model assumes [83, 129, 28] that there is no blur prior to decimation, namely $\Psi_l = Q\{\Psi_h\} + v$, so that image reconstruction is cast as a pure interpolation (zooming) problem. In other words, the problem is only filling out the missing pixels between the original pixels in the input LR image, which remain unaltered in the recovered HR image. In this work, the second model is considered, and also the images are assumed to be noise free, i.e. $v = 0$.

Let $P^k = R_n^k \Psi$ be an image patch of size $n \times n$ centered at location $k$ and extracted from the image $\Psi$ by the linear operator $R_n^k$ in spatial domain with a range of $n$ in two dimensions. Hence the LR and HR patches are extracted as

$$
\begin{aligned}
P_l^k &= R_n^k \Psi_l \\
P_h^k &= R_{\xi n}^k \Psi_h
\end{aligned}
\tag{3.28}
$$

where $\xi$ is the scale-up factor, $h$ refers to HR and $l$ refers to low resolution. Extracted patches of HR images and their corresponding LR patches are saved in two dictionaries of HR patches $D_h$ and LR patches $D_l$, respectively. Hence, every

$P_l^k \in D_l$ has a correspondence in $D_h$. The mapping between these two pairs can be expressed by $f$ as shown below:

$$P_h^k = f\left(P_l^k\right) \tag{3.29}$$

where $f$ is a mapping function that shows the location of $P_h^k$ in $D_h$ which is correspondence with $P_l^k$ in $D_l$. The main motivation for the proposed model is the desire to predict for each LR patch a missing HR detail via a pair in the created $D_l$ and $D_h$ dictionaries. Following the block diagram of the proposed system, shown in figure 3.15, we first find all the patches of the input LR image, $\widetilde{\Psi}_l$, using $R_n^q$ centered at location $q$:

$$\widetilde{P}_l^q = R_l^q \widetilde{\Psi}_l \tag{3.30}$$
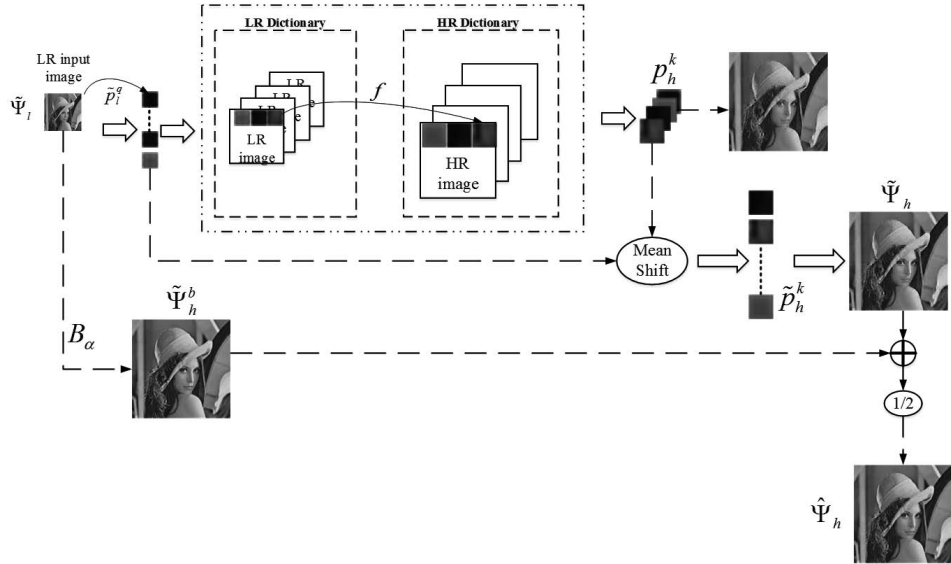


Figure 3.15: The block diagram of the proposed system.

Then the minimum distance between each patch, $\widetilde{P}_l^q$, and all the patches in $D_l$ will be calculated by

$$d_\kappa = \min_k \left(d(P_l^k, \widetilde{P}_l^q)\right)$$
$$\text{where} \quad d\left(P_l^k, \widetilde{P}_l^q\right) = \sqrt{\sum_i (\widetilde{P}_{l_i}^q - P_{l_i}^k)^2} \tag{3.31}$$

where $\kappa$ refers to the index of the patch in $D_l$, which has the minimum distance from the $\widetilde{P}_l^q$ of the input LR image.

Having found a patch in the LR dictionary with a minimum distance to $P_l^q$, its HR corresponding patch is found in $D_h$ and replaced in the HR image by

$$\widetilde{P}_h^q = f\left(P_l^\kappa\right) \tag{3.32}$$

In order to avoid a sudden change of illumination [7], a simple illumination enhancement will be applied to $\widetilde{P}_h^q$ by moving its mean, $\mu_{\widetilde{P}_h^q}$, towards the mean of $\widetilde{P}_l^q$, $\mu_{\widetilde{P}_l^q}$, using

$$\widetilde{P}_h^q = \widetilde{P}_l^q \times \frac{\mu_{\widetilde{P}_l^q}}{\mu_{\widetilde{P}_h^q}} \tag{3.33}$$

where $\mu_{\widetilde{P}_l^q}$ and $\mu_{\widetilde{P}_h^q}$ are a mean of LR and HR patches respectively.

This process is repeated until the last patch of the input LR image. Finally, all founded HR patches are merged together according to their location in order to get the HR image, $\widetilde{\Psi}_h$. The generated image has some blocking effect, which is not desired. In order to remove this effect, the LR input image, $\widetilde{\Psi}_l$, is also interpolated using bicubic interpolation with the same scaling factor, using

$$\widetilde{\Psi}_h^b = B_\xi\left\{\widetilde{\Psi}_l\right\} \tag{3.34}$$

where $B$ is the bicubic interpolation operator with a scaling-up factor of $\xi$. Finally, the HR image, $\widehat{\Psi}_h$, is calculated by averaging the HR image obtained by merging patches, $\widetilde{\Psi}_h$, and the HR image obtained by bicubic interpolation, $\widetilde{\Psi}_h^b$, as shown in

$$\widehat{\Psi}_h = \frac{\widetilde{\Psi}_h + \widetilde{\Psi}_h^b}{2} \tag{3.35}$$

The general steps of the proposed single image SR method are summarized and shown in Algorithm 2 and are illustrated also in figure 3.15.

One of the main constraints of dictionary based SR algorithms is the huge size of the dictionaries. There exist many patches within the dictionaries which are very similar to each other. As one contribution to this work, in order to reduce the number of patches in the dictionary, we have tried to find the set of patches which span the vector space, which includes all the patches as well as the structural similarity between patches. For the first approach, the selected patches are independent or almost independent from each other. In order to find the independent and almost independent patches, we have used SVD. Then the Euclidean distance between the vector of singular values of patches has been calculated. The patches that are discriminant enough from the others (i.e., the calculated cross distance is bigger than the threshold value, $\tau$) are kept in the dictionary. The threshold, $\tau$, can be selected based on the application. In this work, we have reduced the size of the dictionary by 30 percent, while the visual quality has dropped only 0.07 dB (discussed more in the next section).

---
**Algorithm 2** Single image SR scheme
---
**Require:** LR Image and scale-up factor

**Ensure:** HR Image

    **Image interpolation** by using bicubic interpolation with the scale-up factor to generate a scaled-up image from $\widetilde{\Psi}_l$.

    **Extract LR patches** centered at locations $q$ from the LR image, described at Eqn. (3.30).

    **for** $q$ **do**

        **Compute the minimum distance between the LR patch** and the LR patches in $D_l$, described at Eqn. (3.31).

        **Find the corresponding HR patch** with the LR patches that has minimum distance, described at Eqn. (3.32).

        **Enhance the illumination** of HR patch by using a mean shift technique, described at Eqn. (3.33) .

        **Replace the HR patch** in $\widetilde{\Psi}_h$.

    **end for**

    **Find the average between** $\widetilde{\Psi}_h$ and $\widetilde{\Psi}_h^b$.

    **Generate HR image**, described at Eqn. (3.35).
---

Another approach of dictionary reduction proposed in this work is based on the structural similarity index (SSIM). In order to remove patches which are structurally similar to other patches, a similarity between all patches is calculated by using SSIM; then patches with a similarity index more than threshold $\tau$ are removed from the dictionaries. Similar to the previous approach, the threshold, $\tau$, can be selected based on the application. In this approach, we have reduced the size of the dictionary by 45 percent, while the visual quality has dropped only 0.08 dB (discussed more in the next chapter).

## 3.3   Super Resolution Applications

### 3.3.1   Deep Learning Convolutional Network Super Resolution for Improving Face Recognition in Surveillance Monitoring

In this work, first the Viola-Jones [122] algorithm was used to extract only faces from each image to reduce the effects of background and clothing to face recognition results. Also, all images are downsampled by a factor of 4 in order to simulate the situation of images taken by using surveillance cameras.

In this proposed system, we use the SR method of [27], which is based on a CNN with three layers. Given an image $\Psi_h$, a low-resolution upscaled image $\Psi_l$ is created using bicubic interpolation. The first layer extracts overlapping patches from the image $\Psi_l$ and represents each patch as a high-dimensional vector. Instead of using pre-trained bases, such as PCA, DCT or Haar, in this case, these bases are optimized during the network optimization. This first layer can be expressed as an operation $F_1$:

$$F_1(\Psi_l) = \max(0, W_1 * \Psi_l + B_1) \qquad (3.36)$$

where $W_1$ and $B_1$ represent the filters and biases, respectively. In our case, $W_1$ is of a size $9 \times 9 \times 64$, where $9 \times 9$ is the spatial size of the filters and $64$ is the number of the filters. $B_1$ is a $64$ dimensional vector, where each element is associated with a filter.

The second layer performs a non-linear mapping of each of the $64$-dimensional vectors from the first layer to a $32$ dimensional vector. The operation performed by this second layer can be formulated as

$$F_2(\Psi_l) = \max(0, W_2 * F_1(\Psi_l) + B_2) \qquad (3.37)$$

where $W_2$ is of the size $64 \times 5 \times 5 \times 32$ and $B_2$ is $32$-dimensional. Each of those output $32$-dimensional vectors is conceptually a representation of a high-resolution patch that will be used for reconstruction in the last layer. Although more convolutional layers can be added to increase the non-linearity, we use the configuration of [27] to avoid computational complexity of the method.

The output from the second layer is then passed to the last layer, where the reconstruction is performed. This layer emulates the classical method that averages the predicted overlapping high-resolution patches to create the final full image. This process can be formulated as

$$F(\Psi_l) = W_3 * F_2(\Psi_l) + B_3 \qquad (3.38)$$

where $W_3$ is of size $32 \times 5 \times 5$ and $B_3$ is a 1D vector (since we use single channel images). The logic after this formulation is that if the high-resolution patches are

in the image domain, the filters act as an averaging filter, while if the representations are in some other domains, $W_3$ behaves like first projecting the coefficients onto the image domain and then averaging. In both cases, $W_3$ is a set of linear filters.

Learning the end-to-end mapping function $F$ requires the estimation of parameters $\Theta = \{W_1, W_2, W_3, B_1, B_2, B_3\}$. Following the work in [26], this is achieved through minimizing the loss between the reconstructed images $F(\Psi_l; \Theta)$ and the corresponding ground truth high-resolution images $\Psi_h$. Given a set of high-resolution images $\{\Psi_h^i\}$ and their corresponding low-resolution images $\{\Psi_l^i\}$, Mean Squared Error (MSE) is used as the loss function. This loss is minimized using stochastic gradient descent with the standard backpropagation. In particular, the weight matrices are updated as

$$\Delta_{i+1} = 0.9 \cdot \Delta_i + \eta \cdot \frac{\partial L}{\partial W_i^l}, \quad W_{i+1}^l = W_i^l + \Delta_{i+1} \qquad (3.39)$$

where $l \in \{1, 2, 3\}$ and $i$ are the indices of layers and iterations, $\eta$ the learning rate ($10^{-4}$ for $l \in \{1, 2\}$ and $10^{-5}$ for $l = 3$) and $\frac{\partial L}{\partial W^l}$ is the $l^{th}$ derivative. The filter weights of each layer are initialized by drawing randomly from a Gaussian distribution with zero mean and standard deviation 0.001 (and 0 for biases). Figure 3.16 shows an overview of the SR method that is used in the proposed method. Each database used in our work has 10 poses per person. Five of each pose of per-
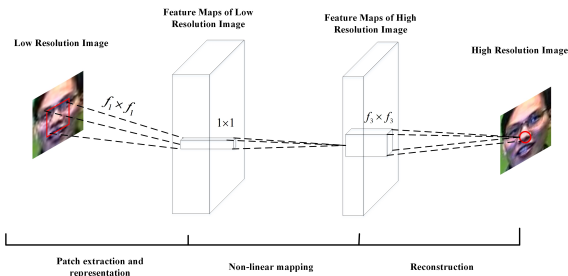


Figure 3.16: The flowchart of the SR method. Image is taken from [28].

sons are used to train HMM [73], and the remaining 5 are used for testing. Both training and test images go through a face recognition process which is divided into seven steps as will be explained in detail here. These steps are: filtering, generating observation vectors, feature extraction, feature selection, quantization, training and face recognition. In filtering, a $3\times3$ minimum filter is applied to the face image to remove unwanted artifacts, such as highlights in subjects' eyes due to flash, and salt noise. After the filtering is done, the face image is converted into a one-dimensional sequence. This is done by sliding a $L \times W$ window from the top to the bottom of the image, which creates a sequence of overlapping blocks of

width $W$ and height $L$ of each face image of width $W$ and height $H$. Next, the features are extracted. Here, instead of using gray values of pixels in sampling windows, SVD coefficients are used as features. Once the features are extracted, a subset of features that lead to the smallest classification error and computational cost are extracted from SVD which contains three matrices ($\mathbf{U}$, $\Sigma$ and $\mathbf{V}$): two first coefficients of $\Sigma$ ($\Sigma_{11}$ and $\Sigma_{22}$) and one first coefficient of $\mathbf{U}$ ($U_{11}$) are used to associate each block. This significantly decreases the length of the observation vectors and also the computational complexity and sensitivity to noise, changes in illumination, shift and rotation. Since SVD coefficients have innately continuous values, which can lead to an infinite number of possible observation vectors that can't be modeled by discrete HMM, the features need to be quantized. Considering vector $X = (x_1, x_2, ..., x_n)$ the quantized value of $x_i$ is computed as below:

$$x_{iquantized} = \left[ \frac{x_i - x_{imin}}{(x_{imax} - x_{imin})/D_i} \right] \tag{3.40}$$

($x_{imax}$) and ($x_{imin}$) are the maximum and minimum that ($x_i$) can get in all possible observation vectors, respectively, and ($D_i$) is the number of quantization levels. Here the first feature ($\Sigma_{11}$) is quantized into 10, second ($\Sigma_{22}$) into 7 and third ($U_{11}$) into 18 levels. After each face image is represented by observation vectors, they are modeled by seven-state HMM. The Baum-Welch algorithm [91] is used to train the HMM model for each person in the database. Finally, for each test image, the probability of the observation vector ($O$) in respect to each HMM face model ($\theta$) is calculated for classification. A person on a test image ($m$) is classified as person ($a$) if

$$P\left(O^{(m)} \mid \theta_a\right) = \max_n P\left(O^{(m)} \mid \theta_n\right) \tag{3.41}$$

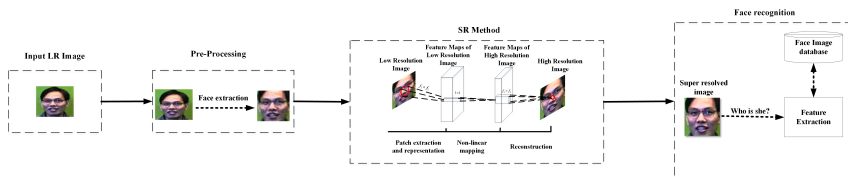Figure 3.17 shows an overview of the proposed method.



Figure 3.17: The flowchart of the proposed system.

### 3.3.2 Deep Learning based Super-Resolution for Improved Action Recognition

In this section, we investigate the importance of the state-of-the-art deep learning-based CNN super-resolution algorithm of [28] in improving recognition accuracies of the state-of-the-art activity recognition algorithm of [124] for working with low-resolution images. To the extent of our knowledge, activity recognition in different resolutions has not been studied that much, except in Ahad et al. [5] in which appearance-based directional motion history image has been used to recognize various levels of video resolutions. However, our proposed system is the first one in which super-resolution algorithms have been employed for improving quality of low-resolution input images before action recognition. We show in this section, that such superresolution algorithms produce high-resolution details that are not necessarily recovered by simple upscaling algorithms, like bicubic interpolation. It is shown in this paper, that combining the results of the deep learning-based super-resolution and the bicubic interpolation, through an alpha blending approach, produces images that are of better quality compared to the input low-resolution images. We show that employing such higher resolution images improves the recognition accuracy of a state-of-the-art action recognition algorithm. Figure 3.18 shows the importance and need for an upsampling algorithm in low-resolution images for action recognition algorithms. It is shown in this figure that regardless of the targeted action, the recognition accuracy of action recognition algorithms drops considerably as the resolution of the images drops.

The block diagram of the proposed system is shown in figure 3.19. Having a low-resolution input video, the proposed system upsamples it by a bicubic interpolation and the deep learning SR algorithm of [27] in parallel. Then these two upsampled videos are combined using a simple $\alpha$ blending technique to produce a high-resolution video. Next, the dense trajectories of [125] are used to perform action recognition on the high-resolution video. These algorithms are briefly revisited in the following subsections.

### $\alpha$ blending of different upsampled videos

Following the block diagram of the proposed system shown in figure 3.19, having upsampled low-resolution videos by the bicubic algorithm and the above mentioned deep learning-based SR algorithm, we need to combine them. To do so, we use the following $\alpha$ blending technique:

$$\bar{\Psi}_H = \alpha\Psi_s + (1 - \alpha)\Psi_b \tag{3.42}$$

| Down-sampling factor | | |
| --- | --- | --- |
| Original resolution | 4 | 8 |
|  |  |  |
|  |  |  |
|  |  |  |
| Recognition rates using bag of words (BoW) | | |
| 62.2% | 30.73% | 23.7% |

Figure 3.18: The importance of resolution in action recognition algorithms: recognition accuracy of the state-of-the-art action recognition algorithm of [125] drops as the resolution of images of different actions of the Hollywood2 dataset drops.
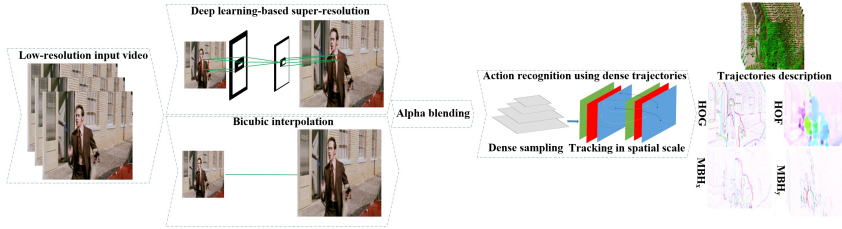


Figure 3.19: The block diagram of the proposed system.

in which, $\Psi_s$ and $\Psi_b$ are the high-resolution images produced by the SR and bicubic algorithms, respectively. The reason for using these two different upsampling algorithms is that they produce different high-resolution details. This can be seen from the difference images (fourth column) of figure 3.20. These images have been obtained by subtracting the two upsampled images from each other after mean filtering the super-resolved image by a kernel of size 3×3. It can be seen from these different images that these two upsampling algorithms produce high-resolution images that are not necessarily the same. These high-resolution details mostly focus around the subject of interest and thus are important for action recognition. Having produced the high-resolution image, hereafter shown by $\bar{\Psi}_H$, it is fed to the action recognition algorithm described in the next section.
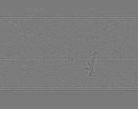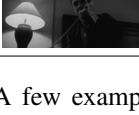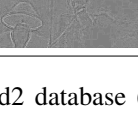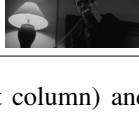
| input | 3x interpolation (bicubic) | 3x super-resolved | difference | combined by $\alpha = 0.2$ |
|---|---|---|---|---|
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |

Figure 3.20: A few examples from the Hollywood2 database (first column) and their upsampling counterparts by bicubic interpolation (second column) and deep learning-based SR (third column). Difference between the two upsampling algorithms is shown in Difference image (fourth column). It can be seen from Difference images that the two upsampling algorithms do not necessarily produce similar high-resolution images, and hence complement each other (last column) for recognition purposes. This is further verified in the experimental results section.

### Dense trajectories-based action recognition

This algorithm starts with a feature extraction step, in which the dense trajectories are considered as the criteria, which help determine the type of action perceived. To do that, first, feature points are densely sampled on a grid spaced by $K$ pixels, and then tracked in each spatial scale separately. A median filter is used to track each point $\bar{\Psi}_H^p = (x_p, y_p)$ from the $p^{\text{th}}$ frame to the $p + 1^{\text{th}}$, in a dense optical flow field $W_t = (h_t, v_t)$, where $v_t$ denotes the vertical component, and $h_t$ stands for the horizontal component of the optical flow:

$$\tilde{\Psi}_H^{p+1} = (x_{p+1}, y_{p+1}) = (x_p, y_p) + (M * W)|_{(x_p, y_p)}, \qquad (3.43)$$

where $M$ is the median filter kernel of size $3 \times 3$. It is worth mentioning that median filter is more robust compared to bilinear interpolation, and improves trajectories for points at motion boundaries. Afterward, the points associated with subsequent frames are concatenated to form trajectories in the form $(\bar{\Psi}_H^p, \bar{\Psi}_H^{p+1}, \dots)$, for each of which, five descriptors are extracted: the trajectory itself, in other words, the concatenation of normalized displacement vectors, along with HOG, HOF, and MBH [125].

Dense trajectories and the corresponding descriptors are extracted from training and test videos, and the descriptors are normalized by means of the so-called RootSIFT approach, i.e. square root after $L_1$ normalization.

For the aim of this work, following the same strategy as the one utilized in [125], $100,000$ training descriptors were randomly selected for clustering, whereby training and test videos were represented through the aforementioned bag-of-features representation. Then feature spaces were combined by the classification model, which is tantamount to taking an SVM with an $\chi^2$ kernel into account [125].

Finally, for the classification procedure, Fisher Vectors from the latter descriptors were concatenated so as to represent each video on the basis of a linear SVM.

## 3.4   Spatio-Temporal Pain Recognition in CNN-based Super-Resolved Facial Images

In this section, we first describe the facial pain-expression database to be used in our investigation. We then describe the procedure of generating facial images with different resolutions and, finally, the deep learning-based classification framework for the experiment.

### 3.4.1   The database

We use the UNBC-McMaster Shoulder Pain database collected by the researchers at McMaster University and the University of Northern British Columbia [69]. The database contains facial video sequences of participants who had been suffering from shoulder pain and were performing a series of active and passive range of motion tests to their affected and unaffected limbs on multiple occasions. The database also contains FACS information of the video frames, self-reported pain scores in sequence level and facial landmark points obtained by an appearance model. The database was originally created by capturing facial videos from 129 participants (63 males and 66 females). The participants had a wide variety of occupations and ages. During data capturing, the participants underwent eight standard range-of-motion tests: abduction, flexion, and internal and external rotation of each arm separately. Participants' self-reported pain score, along with offline independent observers' rated pain intensity, were recorded. At present, the UNBC-McMaster database contains 200 video sequences with 48398 FACS coded frames of 25 subjects.

### 3.4.2   Obtaining Pain-Expression Data with Varying Face Resolution

We created multiple datasets by obtaining the original images from the UNBC-McMaster database and then varying the resolutions by down-up sampling or SR algorithms. The down-up sampling was accomplished by simply down-sampling the original images and then up-sampling the down-sampled images to the same resolution of the original images by employing a cubic-interpolation.
In order to generate SR images, a state-of-the-art technique, namely example-based learning [50], is adopted. We use the down-sampled images as input to the SR algorithm and obtain the super-resolved images.

### 3.4.3 Deep Hybrid Classification Framework

We use a hybrid framework that is a combination of CNN and RNN to exploit both spatial and temporal information of facial pain expressions for pain detection. The hybrid pain detection framework is depicted in figure 3.21. In order to extract discriminative facial features, we fine-tune `VGG_Faces` [81], a 16-layer pre-trained CNN with 2.6M facial images of 2.6K people. Concretely, we replace the last layer of the CNN by a randomly initialized fully-connected layer with the three pain levels to recognize and set its learning rate as ten times the learning rate of the rest of the CNN.

Once fine-tuned, we extract the features of the `fc7` layer of the fine-tuned model and use them as input to a Long-Short Term Memory (LSTM) Recurrent Neural Network (RNN) [39]. LSTMs are particular implementations of RNN that make use of the forget ($F$), input ($i$), and output ($o$) gates so as to solve the vanishing or exploding gradient problems, making them suitable for learning long-term time dependencies. These gates control the flow of information through the model by using point-wise multiplications and sigmoid functions $S$, which bound the information flow between zero and one:

$$i(t) = S(W_{(x \to i)}x(t) + W_{(g \to i)}g(t-1) + b_{(1 \to i)}) \tag{3.44}$$

$$F(t) = S(W_{(x \to F)}x(t) + W(g \to F)g(t-1) + b_{(1 \to F)}) \tag{3.45}$$

$$z(t) = \tanh(W_{(x \to c)}x(t)) + W_{(g \to c)}g(t-1) + b_{(1 \to c)}) \tag{3.46}$$

$$c(t) = F(t)c(t-1) + i(t)z(t), \tag{3.47}$$

$$o(t) = S(W_{(x \to o)}x(t) + W_{(g \to o)}g(t-1) + b_{(1 \to o)}) \tag{3.48}$$

$$g(t) = o(t)\tanh(c(t)), \tag{3.49}$$

where $z(t)$ is the input to the cell at time $t$, $c$ is the cell, $W$ is a weight matrix, $b$ represent a bias vector and $g$ is the output of the cell. $W_{(x \to y)}$ are the weights from $x$ to $y$. More detail can be found in the original implementation [65].

Labels are predicted sequence-wise, *i.e.* given a sequence of $n$ frames $f_i \in \{f_1, ..., f_n\}$, the target prediction is the pain level of the $f_n$ frame. Thus, training is set so that the information contained in the past frames is used in order to predict the current pain level. We optimize the LSTM with *Adam* [51] with an initial learning rate of 0.001 so as to alleviate the hyper-parameter tuning problem.
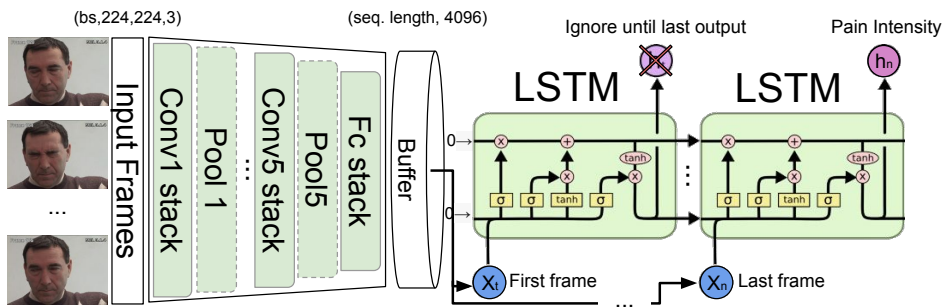
Figure 3.21: The block diagram of the deep hybrid classification framework based on a combination of CNN and RNN.

# CHAPTER 4

# EXPERIMENTAL RESULTS AND DISCUSSIONS

In this chapter, experimental results of our proposed algorithms are shown and comparison between our results with several state-of-the-art algorithms have been done. It should be noticed that all algorithms which are used for comparison are re-built with the same parameters used in the literature (in some cases original codes of the algorithm were downloaded from the website of corresponding authors) except properties of liquid lenses mentioned in section 4.1 which is taken directly from the literature.

## 4.1  Auto-focus Fluid Lenses

The shape of the meniscus is changed by applying different voltage. A laser beam is guided through the lens in order to show the change of the meniscus and to measure it. Figure 4.1 shows the measurement setup of this research.

As is shown in figure 4.3, changes in the size of the laser beam image depend on the applied voltage on the actuator membrane. When the voltage is increased, the hole dimension is decreased, which makes a change in the meniscus. Due to this reason, the size of laser beam image is increased. The red dots show the size of the laser beam image versus applied voltage for an actuator membrane with an active area of 40mm, and blue dots show the size of the laser beam image verses applied voltage for an actuator membrane with an active area of 20mm.

The hole diameter variations (strain $(L - L_0)/L_0$ in percent) also are optically measured with a CMOS camera, as shown in figure 4.2.

As was observed, the focal length of the lens is decreased by applying more voltage. Although the maximum applied voltage on the membrane varies between 1150V to 2250V, depending on the thickness of the membrane actuator, the
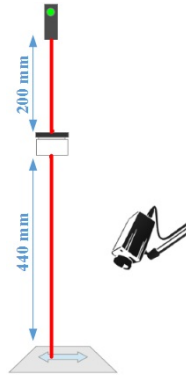
Figure 4.1: Setup for lens experiment.
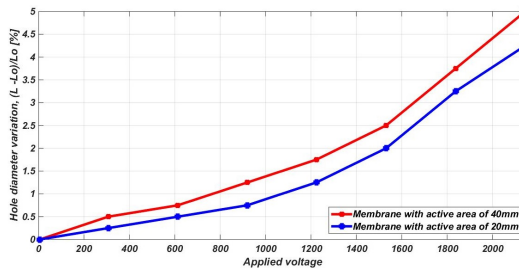


Figure 4.2: Variation of hole diameters versus applied voltage.
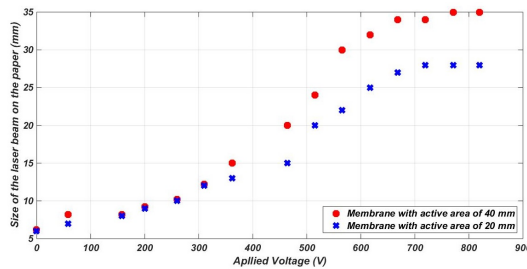


Figure 4.3: Change of laser beam image size on the plane versus applied voltage to membrane actuator.

change in the size of the central hole and shape of meniscus starts from 50V to approximately 750V. Figure 4.4 illustrates the laser beam image on the plane.

Table 4.1 shows a comparison of the proposed work and some of the state of the art methods. A maximum driving voltage of 750V is required to achieve far focus (at 44cm) for the prototype. A significant decrease in driving voltage shows the superiority of the proposed work compared with previous works.
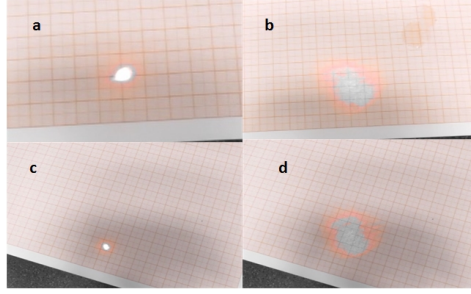
Figure 4.4: Image of laser beam (a) 6v on the actuator membrane with active area of 20mm, (b) 700V on the actuator membrane with active area of 20mm, (c) 6v on the actuator membrane with active area of 40mm, (d) 700V on the actuator membrane with active area of 40mm.

Table 4.1: Performance comparison among various designs of tunable liquid lens using DEA.

|  | Carpi et al [18] | Shian et al [115] | Lau et al [62] | Proposed Work |
|---|---|---|---|---|
| Actuator material | VHB 4905 | VHB 4910 | VHB 4910 | PDMS |
| Electrode material | Carbon grease | Single-walled carbon nanotube | Graphite powder with oil immersion | Graphite powder |
| Lens diameter | 7.6 mm | — | 8 mm | 4 mm |
| Driving voltage | Up to 4 kV | Up to 5 kV | Up to 1.8 kV | Up to 750 V |
| Accommodation or tuning range | 3 cm to 10 cm | 16 cm to 770 cm | 15 cm to 50 cm | 19 cm to 44 cm |

## 4.2   Satellite Image Enhancement: Systematic Approach for Denoising and Resolution Enhancement

In this section, the proposed algorithm in section 3.2.1 is tested and compared with some state-of-the-art algorithms. Figures 4.5 and 4.6 illustrate LR images taken from several aerospace and geosciences resources. We assume that Gaussian white noise (GWN) of zero mean and standard deviation (STD) $\sigma = 25$ is added to the images as shown in (b). The denoised images via the DT-CWT based LA-BSF algorithm are given in (c). Then the denoised images are super-resolved by using bicubic interpolation (d), WZP (e) and DWT (proposed) (f). Objective test (PSNR) results of the proposed image enhancement system are given in Table 4.2.

It should be notice that the term 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. Because many signals have a very wide dynamic range, (ratio between the largest and smallest possible values of a changeable quantity) the PSNR is usually expressed in terms of the logarithmic

decibel scale. PNSR is a standard quality metric of a reconstructed image. If we can show that an algorithm or set of algorithms can enhance a degraded known image to more closely resemble the original, then we can more accurately conclude that it is a better algorithm.

The quantitative test results and visual qualities of the final results indicate that the proposed technique achieves sharper images than the ones achieved by direct interpolation and WZP.
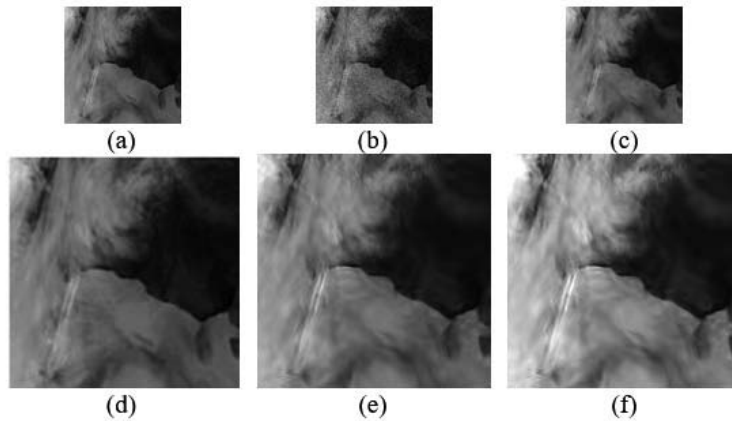


Figure 4.5: Original low contrast image from Antarctic Meteorological Research Centre (a), the noisy image (PSNR=20.41 dB at STD=25) (b), denoised image using DT-CWT based LA-BSF algorithm (PSNR=29.58 dB at STD=25) (c), resolved images (after denoising) by using: bicubic interpolation (d), WZP (e), and DWT (proposed) (PSNR=31.78 dB at STD=25) (f).
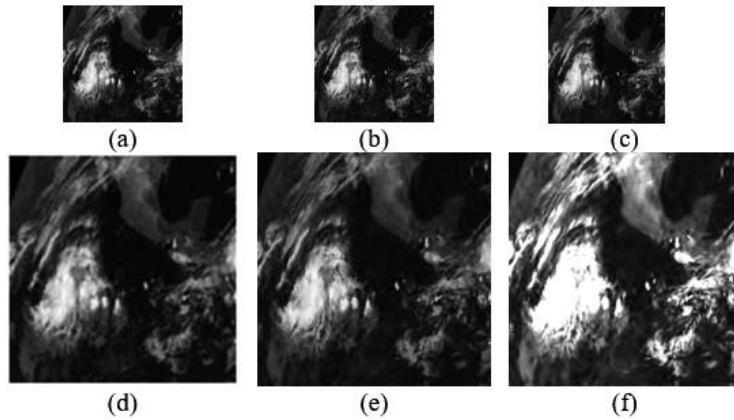
Figure 4.6: Original low contrast image from Antarctic Meteorological Research Centre (a), the noisy image (PSNR=20.79 dB at STD=25) (b), denoised image using DT-CWT based LA-BSF algorithm (PSNR=26.12 dB at STD=25) (c), resolved images (after denoising) by using: bicubic interpolation (d), WZP (e), and DWT (proposed) (PSNR=28.74 dB at STD=25) (f).

Table 4.2: Objective test (PSNR) results of proposed image enhancement system (denoising and resolution enhancement).

| | PSNR (dB) of enhanced image | |
| --- | --- | --- |
| Method | Fig. 4.5 | Fig. 4.6 |
| Gaussian filter + WZP ($\sigma = 15$) | 28.04 | 26.07 |
| Gaussian filter + WZP ($\sigma = 25$) | 23.67 | 22.59 |
| LA-BSF + WZP ($\sigma = 15$) | 28.26 | 26.78 |
| LA-BSF + WZP ($\sigma = 25$) | 23.08 | 23.33 |
| Gaussian filter + (WZP and CS) ($\sigma = 15$) | 29.11 | 27.48 |
| Gaussian filter + (WZP and CS) ($\sigma = 25$) | 24.97 | 24.25 |
| LA-BSF + (WZP and CS) ($\sigma = 15$) | 31.49 | 30.11 |
| LA-BSF + (WZP and CS) ($\sigma = 25$) | 25.15 | 25.89 |
| Gaussian filter + DWT ($\sigma = 15$) | 30.03 | 29.92 |
| Gaussian filter + DWT ($\sigma = 25$) | 25.45 | 26.01 |
| Proposed Technique ($\sigma = 15$) | 33.09 | 31.78 |
| Proposed Technique ($\sigma = 25$) | 30.82 | 28.74 |

## 4.3 New Two-Dimensional Sampling Kernel based Resolution Enhancement

To show that replacing the bicubic interpolation kernel of an SR algorithm with the introduced kernels and then combining the results using the $\alpha$ blending technique, discussed in the previous section, improves the results of the used SR algorithm, we have chosen two state-of-the-art SR algorithms, namely Sparse-SR [83] and Example-SR [50]. In addition to the proposed kernel and bicubic kernel, a Bicubic-spline (B-spline) kernel is used in SR algorithms. The B-spline preserves the shape of the data by inserting one or two additional knots in the subinterval where the interpolants do not attain the desired shape characteristics of data [4, 29].

The Sparse-SR technique proposed in [83] avoids any invariance assumption, which is a common practice in sparsity-based approaches treating this task. A parametric model which captures the statistical dependencies between the sparsity patterns of the LR and HR coefficients and between the corresponding nonzero coefficients is used. The HR patches are predicted by using minimum mean square error estimation. The work in [50] which is an extension of [49] uses kernel ridge regression in order to estimate the high-frequency details of the underlying HR image. Also, a combination of gradient descent and kernel matching pursuit is considered, which allows time-complexity to be kept to a moderate level.

To conduct the experiments, we have created a database of 1025 non-facial images and 1013 facial images from multiple databases, such as LFW [42], the Helen database [63] and Van Hateren's Natural Image Database [37], as shown in figure 4.7. Then the following experiments have been run:

- Have applied the Sparse-SR algorithm of [83] to the dataset without changing the original bicubic interpolation kernel.

- Have applied the Example-SR algorithm of [50] to the dataset without changing the original bicubic interpolation kernel.

- Have applied the Sparse-SR algorithm of [83] to the dataset when the original bicubic kernel of the algorithm has been replaced by the B-spline Kernel [64, 11].

- Have applied the Example-SR algorithm of [50] to the dataset when the original bicubic kernel of the algorithm has been replaced by the B-spline Kernel [64, 11].

- Have applied the Sparse-SR algorithm of [83] to the dataset when the original bicubic kernel of the algorithm has been replaced by the kernels in-

troduced in this work and the results are combined using the mentioned $\alpha$ blending.

- Have applied the Example-SR algorithm of [50] to the dataset when the original bicubic kernel of the algorithm has been replaced by the kernels introduced in this work and the results are combined using the mentioned $\alpha$ blending.

Figure 4.7: Some facial and non-facial images.



Quantitative comparison of the obtained results for the above-mentioned experiments can be found in Tables 4.4 and 4.6 for an enlargement size of 2 and in Tables 4.5 and 4.7 for an enlargement size of 4. These tables report the results using the common factors that are used for assessing the quality of SR algorithms, namely PSNR and SSIM, respectively. The results in these tables are given individually for some known benchmark images and also for the total dataset (in the last row of the tables). Figure 4.8 also shows the results generated using the above-mentioned experiments for some of the known benchmark images.

It can be seen from both Tables 4.4 and 4.6 and figure 4.8 that the results generated by both SR algorithms using the proposed kernels (the last two columns of the tables and also the figure) are both quantitatively and visually (qualitatively) better than the results of the SR algorithms when their original bicubic interpolation kernel has been used. Figures 4.9, 4.10 and 4.11, as well as Table 4.3,

Figure 4.8: Visual comparison of adopted algorithms where bicubic and proposed kernels are used. From left to right, the first column shows LR images, the second and third columns show the super resolved-images by using Example-SR with proposed kernels and bicubic kernel respectively. The fourth and fifth columns show super-resolved images by using Sparse-SR with proposed kernels and bicubic kernel respectively.

show the same comparison on several remote sensing images. Table 4.3 shows the average PSNR results of 21 classes of remote sensing images, in which each category consists of 100 images [130]. The magnification used is 2 in all of the
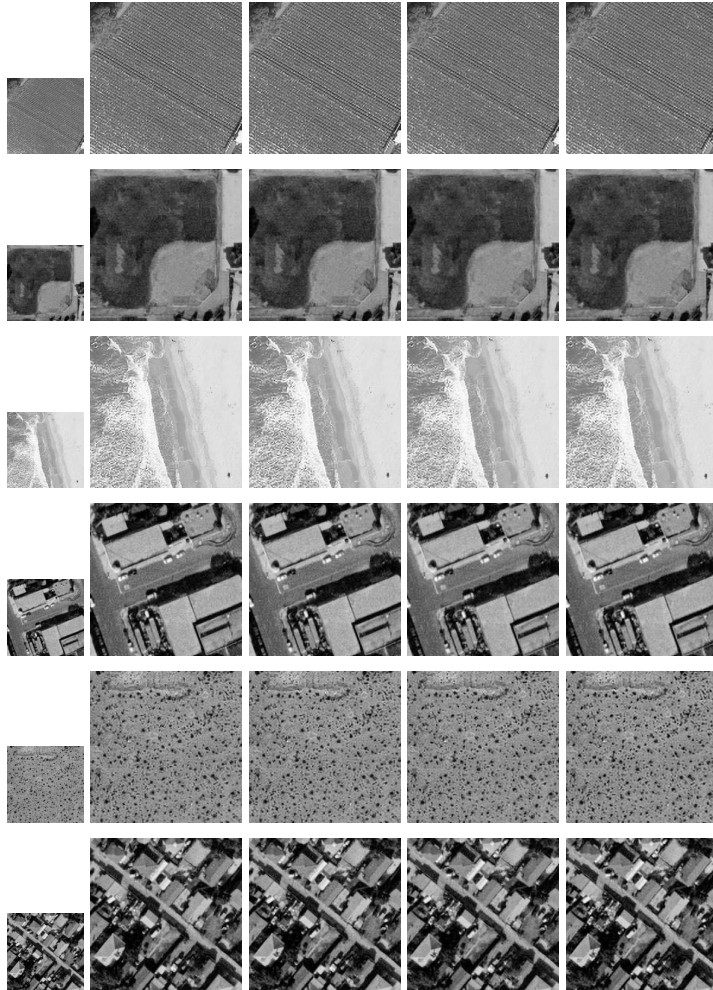
Figure 4.9: Visual comparison of adopted algorithms where bicubic and proposed kernels are used. From left to right, the first column shows LR images, the second and third columns show the super-resolved images by using Example-SR with proposed kernels and bicubic kernel respectively. The fourth and fifth columns show super-resolved images by using Sparse-SR with proposed kernels and bicubic kernel respectively.

experiments.

In order to show the effect of changing $\alpha$ values in the $alpha$ blending technique, the outputs of the Sparse-SR [83] through using each of the proposed kernels have been combined by using different alpha values, and accordingly, the PSNR values have been calculated for facial and non-facial images separately. Figure 4.12 illustrates the change in the quality of the output based on different alpha values.
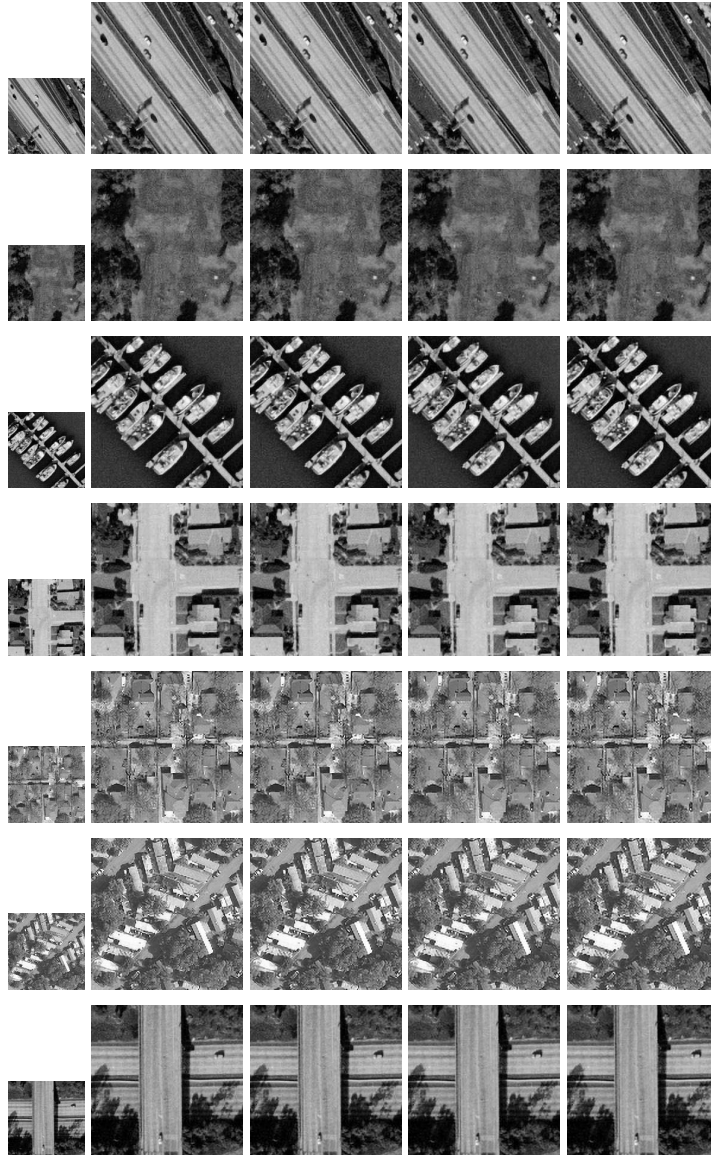
Figure 4.10: Visual comparison of adopted algorithms where bicubic and proposed kernels are used. From left to right, the first column shows LR images, the second and third columns show the super-resolved images by using Example-SR with proposed kernels and bicubic kernel respectively. The fourth and fifth columns show super-resolved images by using Sparse-SR with proposed kernels and bicubic kernel respectively.

By picking the best value of $\alpha$ for facial images, which is $\alpha = 0.8$, 2330 faces in the Helen database [63] have been super-resolved by a factor of 2. As
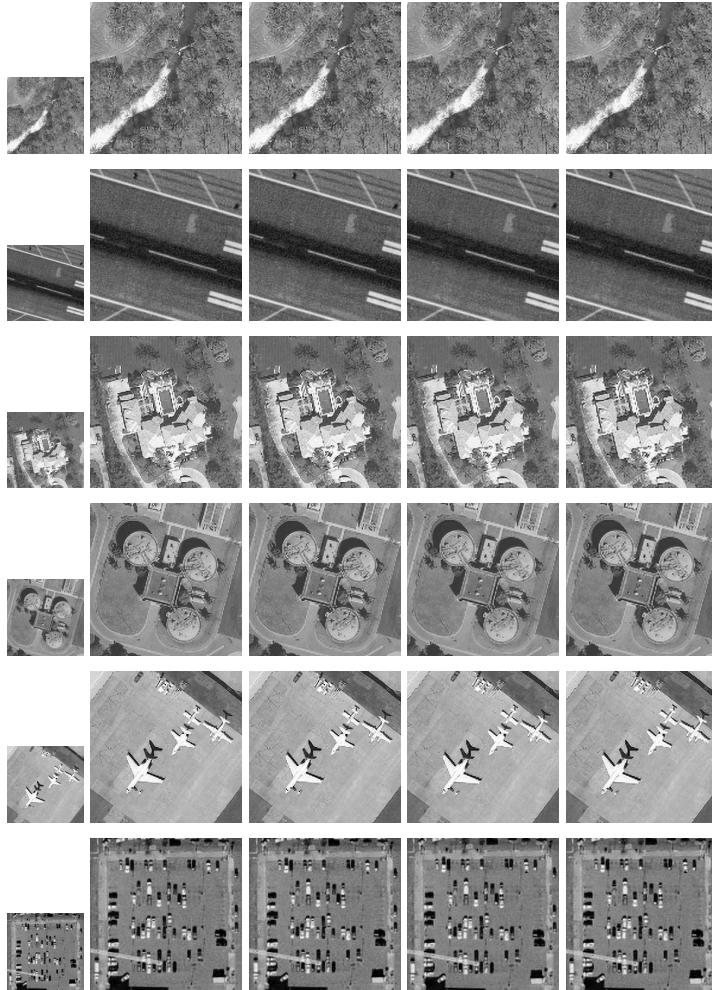
Figure 4.11: Visual comparison of adopted algorithms where bicubic and proposed kernels are used. From left to right, the first column shows LR images, the second and third columns show the super-resolved images by using Example-SR with proposed kernels and bicubic kernel respectively. The fourth and fifth columns show super-resolved images by using Sparse-SR with proposed kernels and bicubic kernel respectively.

shown in Table 4.8, the quantitative results by means of PSNR show that the proposed method performs 1.41dB, on average, better than the bicubic. This can be observed from SSIM results as well.
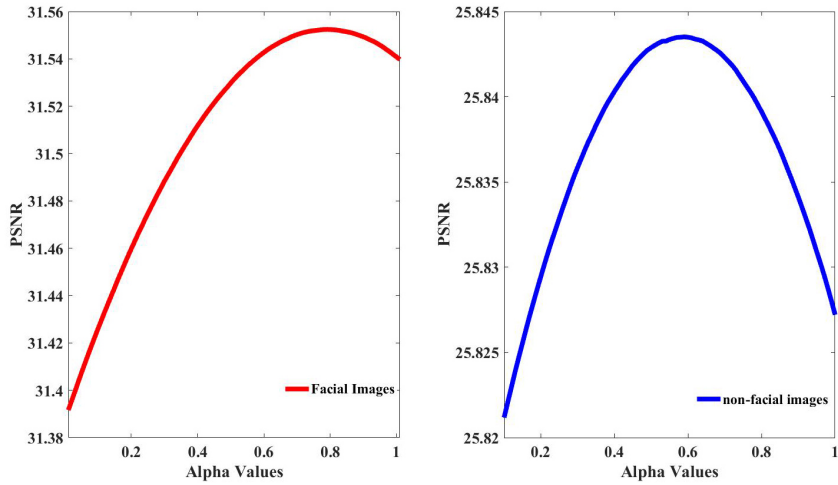
Figure 4.12: Average PSNR results of facial and non-facial images with Sparse-SR algorithm of [83] on different value of alpha.

Table 4.3: The Average PSNR values in dB for 21 categories of remote sensing images adopting Sparse-SR [83] and Example-SR [50] with bicubic and the proposed kernels based resolution enhancement for enlargement factor of 2.

| | PSNR Value | | | |
| | Bicubic Kernel | | Proposed kernels with $\alpha$ blending | |
| | Sparse-SR | Example-SR | Sparse-SR | Example-SR |
|---|---|---|---|---|
| Agricultural | 25.49 | 24.98 | 24.98 | 25.10 |
| Airplane | 25.86 | 25.53 | 25.53 | 25.53 |
| Baseballdiamond | 30.56 | 30.00 | 30.00 | 30.03 |
| Beach | 33.63 | 33.05 | 33.05 | 33.11 |
| Building | 23.90 | 23.61 | 23.61 | 23.61 |
| Chaparral | 24.03 | 23.82 | 23.82 | 23.88 |
| Denseresidential | 24.01 | 23.81 | 23.81 | 23.79 |
| Forest | 26.02 | 25.58 | 25.58 | 25.65 |
| Freeway | 26.88 | 26.47 | 26.47 | 26.57 |
| Golfcourse | 30.61 | 30.06 | 30.06 | 30.09 |
| Harbor | 21.86 | 21.52 | 21.52 | 21.52 |
| Intersection | 25.24 | 24.93 | 24.93 | 24.93 |
| Mediumresidential | 23.55 | 23.27 | 23.27 | 23.27 |
| Mobilehomepark | 22.14 | 21.88 | 21.88 | 21.86 |
| Overpass | 23.92 | 23.51 | 23.51 | 23.52 |
| Parkinglot | 21.14 | 20.71 | 20.71 | 20.72 |
| River | 26.02 | 25.53 | 25.53 | 25.60 |
| Runway | 27.21 | 26.57 | 26.57 | 26.77 |
| Sparseresidential | 26.20 | 25.86 | 25.86 | 25.9 |
| Storagetanks | 25.67 | 25.30 | 25.30 | 25.30 |
| Tenniscourt | 26.74 | 26.38 | 26.38 | 26.38 |

Table 4.4: The PSNR values in dB for the used images adopting Sparse-SR [83] and Example-SR [50] with B-spline, bicubic and the proposed kernels based resolution enhancement for an enlargement factor of 2.

| Images | PSNR values | | | | | |
|---|---|---|---|---|---|---|
| | B-spline Kernel | | Bicubic Kernel | | Proposed kernels with $\alpha$ blending | |
| | Sparse-SR [83] | Example-SR [50] | Sparse-SR [83] | Example-SR [50] | Sparse-SR [83] | Example-SR [50] |
| Anna | 29.12 | 28.73 | 31.93 | 30.82 | 32.53 | 31.53 |
| Butterfly | 19.57 | 19.48 | 21.89 | 21.52 | 23.20 | 21.54 |
| Comic | 19.01 | 18.90 | 20.59 | 20.47 | 22.15 | 20.51 |
| Flowers | 23.67 | 23.12 | 24.58 | 24.23 | 26.12 | 24.25 |
| Foreman | 22.49 | 27.44 | 30.52 | 30.37 | 32.36 | 31.39 |
| Francoise | 19.86 | 20.19 | 22.79 | 22.32 | 24.38 | 22.36 |
| Girl | 23.73 | 27.78 | 29.02 | 28.49 | 30.62 | 28.53 |
| Lena | 24.39 | 26.39 | 28.49 | 27.46 | 30.09 | 28.05 |
| Man | 22.59 | 21.95 | 23.61 | 23.21 | 25.24 | 23.23 |
| Pepper | 25.09 | 25.89 | 27.06 | 26.79 | 27.85 | 26.83 |
| Starfish | 21.77 | 22.09 | 23.99 | 22.71 | 25.69 | 23.73 |
| Zebra | 22.16 | 22.68 | 23.94 | 23.17 | 25.07 | 23.79 |
| Average of 2000 images | 25.29 | 25.48 | 27.39 | 26.10 | 28.64 | 27.85 |

Table 4.5: The PSNR values in dB for the used images adopting Sparse-SR [83] and Example-SR [50] with B-spline, bicubic and the proposed kernels based resolution enhancement for an enlargement factor of 4.

| Images | PSNR values | | | | | |
|---|---|---|---|---|---|---|
| | B-spline kernel | | Bicubic Kernel | | Proposed kernels with $\alpha$ blending | |
| | Sparse-SR [83] | Example-SR [50] | Sparse-SR [83] | Example-SR [50] | Sparse-SR [83] | Example-SR [50] |
| Anna | 26.05 | 24.92 | 25.50 | 27.95 | 25.91 | 28.16 |
| Butterfly | 16.87 | 16.44 | 16.33 | 18.16 | 16.54 | 18.68 |
| Comic | 16.91 | 17.24 | 15.88 | 17.25 | 15.98 | 17.68 |
| Flowers | 21.25 | 21.18 | 20.77 | 21.48 | 20.80 | 21.94 |
| Foreman | 23.96 | 24.68 | 25.52 | 27.86 | 25.49 | 28.47 |
| Girl | 27.27 | 27.37 | 27.42 | 26.64 | 27.45 | 27.12 |
| Lena | 22.69 | 24.38 | 21.84 | 25.62 | 22.05 | 26.24 |
| Man | 20.94 | 21.27 | 20.41 | 21.23 | 20.51 | 21.67 |
| Pepper | 23.72 | 24.25 | 22.74 | 25.28 | 23.13 | 25.69 |
| Starfish | 20.55 | 19.80 | 20.49 | 20.77 | 20.61 | 21.24 |
| Zebra | 18.16 | 18.20 | 17.05 | 20.02 | 17.19 | 20.53 |

## 4.4 Reducible Dictionaries for Single Image Super-Resolution based on Patch Matching and Mean Shifting

In this work, 581 HR images from different standard test image databases [2] are used to make the HR dictionary, $D_h$, and its corresponding LR dictionary, $D_l$.

Table 4.6: The SSIM values for the used images adopting Sparse-SR [83] and Example-SR [50] with B-spline, bicubic and the proposed kernels based resolution enhancement for an enlargement factor of 2.

| Images | SSIM Values | | | | | |
| | B-spline kernel | | bicubic kernel | | Proposed kernels with $\alpha$ blending | |
| | Sparse-SR [83] | Example-SR [50] | Sparse-SR [83] | Example-SR [50] | Sparse-SR [83] | Example-SR [50] |
|---|---|---|---|---|---|---|
| Anna | 0.8409 | 0.8355 | 0.9254 | 0.9097 | 0.9349 | 0.9168 |
| Butterfly | 0.7423 | 0.7535 | 0.8323 | 0.8349 | 0.8701 | 0.8395 |
| Comic | 0.6667 | 0.6623 | 0.7427 | 0.7409 | 0.7788 | 0.7450 |
| Flowers | 0.7634 | 0.7366 | 0.8014 | 0.7885 | 0.8359 | 0.7911 |
| Foreman | 0.8668 | 0.8489 | 0.8987 | 0.8920 | 0.9218 | 0.8997 |
| Francoise | 0.7951 | 0.7529 | 0.813 | 0.7872 | 0.8701 | 0.7877 |
| Girl | 0.7541 | 0.6891 | 0.7347 | 0.7137 | 0.7799 | 0.7150 |
| Lena | 0.8181 | 0.9117 | 0.9243 | 0.9059 | 0.9475 | 0.9189 |
| Man | 0.7030 | 0.8571 | 0.8603 | 0.8409 | 0.9059 | 0.8417 |
| Pepper | 0.8344 | 0.909 | 0.9187 | 0.9006 | 0.9410 | 0.9093 |
| Starfish | 0.7600 | 0.7499 | 0.8124 | 0.8080 | 0.8480 | 0.8204 |
| Zebra | 0.7835 | 0.904 | 0.9023 | 0.8918 | 0.9313 | 0.8987 |
| Average of 2000 images | 0.7415 | 0.8493 | 0.855 | 0.8126 | 0.8750 | 0.8405 |

Table 4.7: The SSIM values for the used images adopting Sparse-SR [83] and Example-SR [50] with B-spline, bicubic and the proposed kernels based resolution enhancement for an enlargement factor of 4.

| Images | SSIM values | | | | | |
| | B-spline kernel | | Bicubic Kernel | | Proposed kernels with $\alpha$ blending | |
| | Sparse-SR [83] | Example-SR [50] | Sparse-SR [83] | Example-SR [50] | Sparse-SR [83] | Example-SR [50] |
|---|---|---|---|---|---|---|
| Anna | 0.6952 | 0.7729 | 0.6010 | 0.7973 | 0.6174 | 0.8127 |
| Butterfly | 0.5481 | 0.6881 | 0.5213 | 0.6890 | 0.5241 | 0.7091 |
| Comic | 0.4255 | 0.5472 | 0.4008 | 0.5352 | 0.4044 | 0.5483 |
| Flowers | 0.6051 | 0.6822 | 0.6023 | 0.6570 | 0.6035 | 0.6706 |
| Foreman | 0.7965 | 0.8289 | 0.8195 | 0.8391 | 0.8209 | 0.8495 |
| Girl | 0.6998 | 0.7377 | 0.6905 | 0.6284 | 0.6909 | 0.6427 |
| Lena | 0.7313 | 0.7589 | 0.7068 | 0.8509 | 0.7072 | 0.8605 |
| Man | 0.5682 | 0.6348 | 0.5578 | 0.7181 | 0.5584 | 0.7323 |
| Pepper | 0.7665 | 0.7782 | 0.7415 | 0.8511 | 0.7417 | 0.8631 |
| Starfish | 0.5955 | 0.6582 | 0.5982 | 0.6692 | 0.6014 | 0.6814 |
| Zebra | 0.5545 | 0.598 | 0.5436 | 0.7482 | 0.5471 | 0.7596 |

These images are selected from different categories, such as face images, natural images, and texture images. Some of these images are shown in figure 4.13.

Table 4.8: The average result with the best $alpha$ value for the Helen Database [63].

|  | PSNR(dB) | SSIM |
|---|---|---|
| Bicubic | 27.44 | 0.8774 |
| SR algorithm Sparse-SR [83] with new kernel | 28.85 | 0.9002 |

In order to make HR and LR dictionaries, the nearest neighbor interpolation is used to decimate HR images to LR images by a factor of 2, i.e. in our experimental results we are conducting SR with a scaling factor of $\xi = 2$. Then patch sizes of $8 \times 8$ and $4 \times 4$ are chosen for making HR and LR dictionaries, respectively. Also, the input LR test images to all SR techniques used for the experimental results have been obtained by downsampling their HR original counterparts using the nearest neighbor kernel. It should be mentioned that the test images are not used in constructing the dictionaries.

In the experimental results, many well-known benchmark images such as Butterfly, Comic, Flowers, Foreman, Girl, Lena, Man, Pepper, Starfish, and Zebra, as well as the LFW face database [42] with 13259 images, are used. PSNR, SSIM index, Visual Information Fidelity in Pixel Domain (VIFP), universal quality index (UQI), noise quality measure (NQM), Weight signal to noise ratio (WSNR) and mean square error (MSE) are used in order to evaluate imperceptibility characteristics quality measurement. Table 4.11 shows the PSNR values, SSIM values, VIFP values, UQI values, NQM values, WSNR values and MSE values for the convolutional neural network SR method [28], Peleg and Elad [83], the Yang method [127] that is called KK, the A+ method [120] and the proposed SR techniques for the aforementioned images.

For Lena's image, the PSNR of the proposed method is 3.50 dB and 3.01 dB higher than those of the convolutional neural network SR method (SRCNN) [28] and the SR method proposed by Peleg et al. [83], respectively.

Table 4.11 illustrates that the greatest difference is for the pepper image that is about 4.06 dB and 4.48 dB more than the Peleg and Elad method in [83] and SRCNN in [28], respectively. The average of the PSNR results of our proposed method are approximately 1.4 dB and 0.7 dB more than the SRCNN method in [28] and Peleg and Elad's method in [83], respectively. Figure 4.14 shows the visual comparison of the SRCNN [28], the proposed method in [83] and the proposed method for some of the images used in the comparison. The results show that the proposed method performs better than the conventional and the state-of-the-art methods. The proposed method also shows a good performance in the area of remote sensing images. Tables 4.9 and 4.10 show the average PSNR and SSIM results of the proposed algorithm on 21 classes of remote sensing images,

Figure 4.13: Some images that are used for making dictionaries which are employed with the experimental results of this work.

in which each class consists of 100 images.

Table 4.9: The Average PSNR values in dB for 21 categories of remote sensing images.

|  | Sparse-SR | SRCNN | Proposed Method |
|---|---|---|---|
| Agricultural | 25.49 | 26.43 | 27.42 |
| Airplane | 25.86 | 26.53 | 27.63 |
| Baseballdiamond | 30.56 | 31.33 | 32.09 |
| Beach | 33.63 | 34.28 | 35.09 |
| Building | 23.90 | 24.64 | 25.43 |
| Chaparral | 24.03 | 25.02 | 25.90 |
| Denseresidential | 24.01 | 24.86 | 25.57 |
| Forest | 26.02 | 26.80 | 28.08 |
| Freeway | 26.88 | 27.59 | 28.40 |
| Golfcourse | 30.61 | 31.36 | 32.29 |
| Harbor | 21.86 | 22.46 | 23.37 |
| Intersection | 25.24 | 26.06 | 26.78 |
| Mediumresidential | 23.55 | 24.35 | 25.44 |
| Mobilehomepark | 22.14 | 22.91 | 23.78 |
| Overpass | 23.92 | 24.63 | 25.94 |
| Parkinglot | 21.14 | 21.63 | 22.48 |
| River | 26.02 | 26.77 | 28.11 |
| Runway | 27.21 | 27.81 | 29.23 |
| Sparseresidential | 26.20 | 27.05 | 27.99 |
| Storagetanks | 25.67 | 26.44 | 27.38 |
| Tenniscourt | 26.74 | 27.60 | 28.42 |
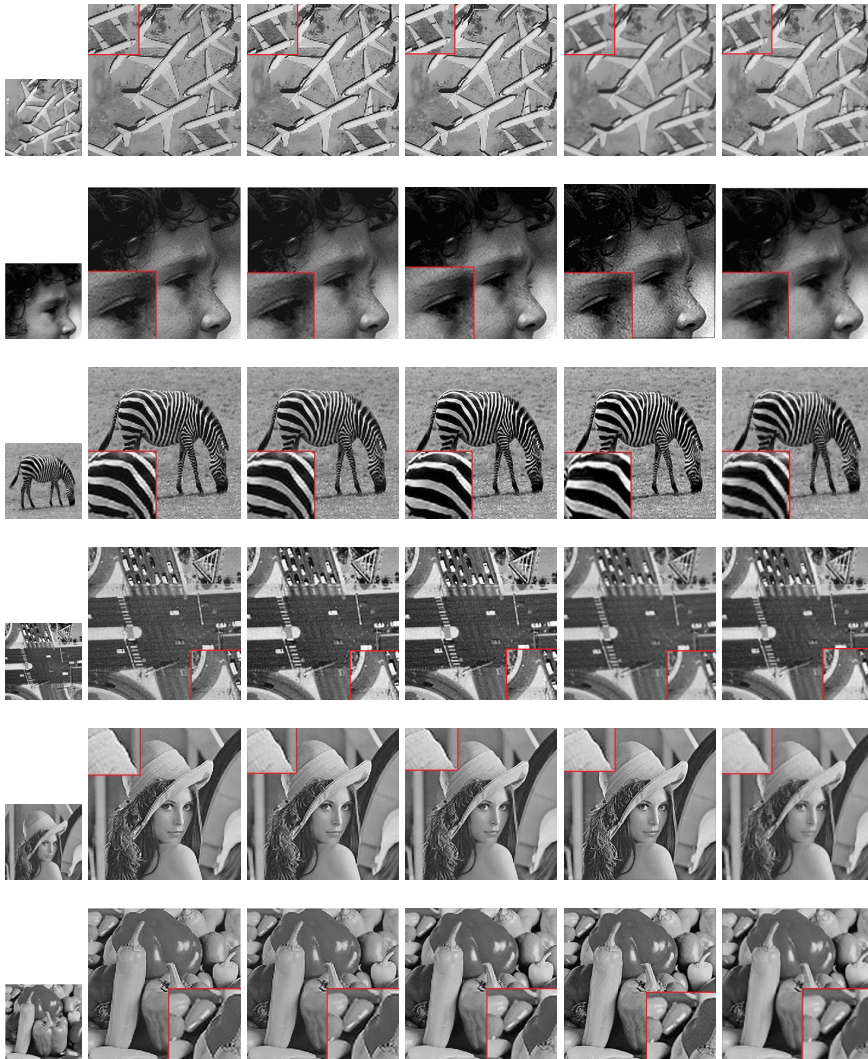
Figure 4.14: Visual comparison of state-of-the-art algorithms. From left to right, the first column shows LR images, the second and third columns show the super-resolved images by using SRCNN [28], Sparse-SR [83] respectively. The fourth and fifth columns show super-resolved images by using the KK method [127] and A+ [120] respectively. The last column shows the proposed method with mean shift.

## 4.5 Deep Learning Convolutional Network Super Resolution for Improving Face Recognition in Surveillance Monitoring

Four face databases are used in the experiments. These databases are the facial recognition technology (FERET) database [87, 86], the Essex Faces facial images

Table 4.10: The Average SSIM values in dB for 21 categories of remote sensing images.

|  | Sparse-SR | SRCNN | Proposed Method |
|---|---|---|---|
| Agricultural | 0.6876 | 0.6922 | 0.7280 |
| Airplane | 0.8219 | 0.8188 | 0.8450 |
| Baseballdiamond | 0.8265 | 0.8253 | 0.8522 |
| Beach | 0.8943 | 0.8932 | 0.9060 |
| Building | 0.8187 | 0.8158 | 0.8326 |
| Chaparral | 0.7509 | 0.7473 | 0.7863 |
| Denseresidential | 0.8095 | 0.8085 | 0.8277 |
| Forest | 0.7274 | 0.7223 | 0.7702 |
| Freeway | 0.8072 | 0.8036 | 0.8267 |
| Golfcourse | 0.8216 | 0.8198 | 0.8499 |
| Harbor | 0.8334 | 0.8263 | 0.8458 |
| Intersection | 0.8130 | 0.8108 | 0.8316 |
| Mediumresidential | 0.7575 | 0.7543 | 0.7921 |
| Mobilehomepark | 0.7710 | 0.7653 | 0.7956 |
| Overpass | 0.7581 | 0.7559 | 0.7943 |
| Parkinglot | 0.7892 | 0.7782 | 0.8050 |
| River | 0.7505 | 0.7443 | 0.7880 |
| Runway | 0.8052 | 0.7987 | 0.8394 |
| Sparseresidential | 0.7503 | 0.7490 | 0.7877 |
| Storagetanks | 0.8127 | 0.8092 | 0.8330 |
| Tenniscourt | 0.8160 | 0.8151 | 0.8370 |

collection [1], the Head Pose Image database (HP) [34] and our recently introduced iCV Face Recognition database (iCV-F). The iCV-F database consists of face images of 31 subjects with 10 images for each subject. The database includes people wearing glasses or not and with varying skin color. Models were asked to make different facial expressions while the photos were taken. Figure 4.15 shows some images of the iCV-F database.

The facial images first have been passed through the Viola-Jones face detector,



Figure 4.15: Figures of our own database.

and the segmented faces are resized to $60 \times 60$ pixels. These images are then downsampled by a factor of 4 in order to achieve LR input images. Figure 4.16 shows the LR and super-resolved images. The LR images have the size of $15 \times$

Table 4.11: Comparison of results of the proposed method with SRCNN [28], Peleg & Elad method [83], KK [127] and A+ [120], where bold numbers show the best performance.

| | | Butterfly | Comic | Flowers | Foreman | Girl | Lena | Man | Pepper | Starfish | Zebra | 1000 facial images |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SRCNN [28] | PSNR | 22.45 | 21.03 | 23.58 | 30.33 | 30.95 | 24.93 | 23.68 | 27.21 | 24.52 | 20.75 | 31.95 |
| | SSIM | 0.8282 | 0.7360 | 0.7679 | 0.8898 | 0.8012 | 0.8277 | 0.7423 | 0.8814 | 0.8045 | 0.7275 | 0.9452 |
| | VIFP | 0.3752 | 0.2893 | 0.3069 | 0.5110 | 0.3883 | 0.4021 | 0.3104 | 0.4486 | 0.3538 | 0.2727 | 0.5842 |
| | UQI | 0.7542 | 0.7426 | 0.6821 | 0.7202 | 0.6738 | 0.7581 | 0.6995 | 0.8045 | 0.7699 | 0.7180 | 0.8561 |
| | NQM | 16.36 | 14.77 | 15.83 | 20.89 | 21.95 | 16.65 | 16.13 | 17.49 | 16.62 | 11.94 | 20.49 |
| | WSNR | 24.92 | 26.02 | 24.58 | 34.61 | 30.48 | 27.56 | 24.78 | 27.71 | 27.78 | 24.23 | 29.25 |
| | MSE | 369.61 | 497.12 | 284.85 | 60.30 | 52.24 | 208.97 | 278.43 | 123.51 | 229.47 | 546.64 | 49.52 |
| Peleg [83] | PSNR | 22.95 | 21.68 | 24.13 | 30.55 | 31.69 | 25.42 | 24.32 | 27.63 | 25.12 | 21.67 | 32.33 |
| | SSIM | 0.8415 | 0.7499 | 0.7857 | 0.9011 | 0.8291 | 0.8416 | 0.7638 | 0.8941 | 0.8205 | 0.7502 | 0.9516 |
| | VIFP | 0.3858 | 0.2973 | 0.3176 | 0.5260 | 0.4111 | 0.4141 | 0.3229 | 0.4650 | 0.3661 | 0.2845 | 0.5997 |
| | UQI | 0.7756 | 0.7587 | 0.7056 | 0.7409 | 0.6775 | 0.7792 | 0.7221 | 0.8252 | 0.7894 | 0.7407 | 0.8549 |
| | NQM | 16.36 | 14.80 | 15.84 | 20.79 | 22.02 | 16.58 | 16.14 | 17.44 | 16.55 | 12.10 | 20.46 |
| | WSNR | 24.94 | 26.09 | 24.60 | 34.60 | 30.54 | 27.51 | 24.79 | 27.72 | 27.76 | 24.41 | 29.42 |
| | MSE | 329.52 | 428.53 | 251.08 | 57.30 | 44.02 | 186.70 | 240.65 | 112.26 | 200.17 | 442.50 | 45.08 |
| KK [127] | PSNR | 21.64 | 20.25 | 22.76 | 30.34 | 29.76 | 25.08 | 22.83 | 25.39 | 24.08 | 19.95 | 31.01 |
| | SSIM | 0.8410 | 0.7420 | 0.7700 | 0.8899 | 0.7785 | 0.8336 | 0.7380 | 0.8789 | 0.8133 | 0.7206 | 0.9342 |
| | VIFP | 0.3705 | 0.2808 | 0.2985 | 0.5084 | 0.3724 | 0.4063 | 0.3027 | 0.4366 | 0.3507 | 0.2634 | 0.5762 |
| | UQI | 0.7782 | 0.7551 | 0.6914 | 0.7238 | 0.6788 | 0.7636 | 0.7090 | 0.8149 | 0.7905 | 0.7205 | 0.8381 |
| | NQM | 16.72 | 15.00 | 16.06 | 20.95 | 22.26 | 16.75 | 16.38 | 17.04 | 16.95 | 12.15 | 20.49 |
| | WSNR | 24.07 | 25.07 | 23.32 | 34.60 | 29.18 | 27.55 | 23.54 | 25.88 | 26.84 | 23.28 | 27.96 |
| | MSE | 445.29 | 614.01 | 344.68 | 60.13 | 68.77 | 202.04 | 338.53 | 188.13 | 254.40 | 658.52 | 168.24 |
| A+ [120] | PSNR | 18.79 | 17.37 | 21.19 | 25.59 | 24.66 | 24.66 | 20.69 | 24.31 | 20.41 | 20.10 | 31.82 |
| | SSIM | 0.7015 | 0.6359 | 0.6842 | 0.8071 | 0.6254 | 0.7367 | 0.6095 | 0.6881 | 0.7051 | 0.6769 | 0.9186 |
| | VIFP | 0.3140 | 0.2366 | 0.2790 | 0.4292 | 0.2740 | 0.3452 | 0.2464 | 0.3115 | 0.2960 | 0.2901 | 0.5788 |
| | UQI | 0.6594 | 0.6423 | 0.5799 | 0.6168 | 0.5060 | 0.5706 | 0.5492 | 0.5203 | 0.6764 | 0.6719 | 0.7917 |
| | NQM | 12.33 | 12.57 | 18.36 | 16.65 | 17.26 | 21.01 | 20.44 | 21.50 | 12.52 | 15.66 | 23.40 |
| | WSNR | 20.49 | 19.88 | 22.27 | 29.90 | 23.37 | 26.30 | 21.23 | 25.55 | 23.13 | 22.58 | 27.18 |
| | MSE | 858.68 | 1190.32 | 494.01 | 179.65 | 222.24 | 222.34 | 554.82 | 241.22 | 591.87 | 635.56 | 122.98 |
| Proposed method | PSNR | **23.82** | **22.75** | **25.11** | **32.83** | **32.49** | **28.43** | **25.50** | **31.69** | **26.21** | **22.44** | 31.82 |
| | SSIM | **0.8675** | **0.7882** | **0.8394** | **0.9462** | **0.8706** | **0.8974** | **0.8106** | **0.9836** | **0.8695** | **0.7848** | **0.9412** |
| | VIFP | **0.4046** | **0.3192** | **0.3419** | **0.5436** | **0.4289** | **0.4707** | **0.3565** | **0.6372** | **0.3932** | **0.3021** | **0.5893** |
| | UQI | **0.7916** | **0.7801** | **0.7307** | **0.7502** | **0.7193** | **0.8082** | **0.7553** | **0.9336** | **0.8133** | **0.7526** | 0.8495 |
| | NQM | **17.29** | **16.38** | **17.69** | **22.02** | **24.30** | **19.70** | **18.33** | 21.85 | **18.41** | **14.38** | 22.71 |
| | WSNR | 24.48 | 26.00 | 24.51 | **35.49** | **30.89** | **30.06** | **25.31** | **30.21** | **27.93** | **24.77** | 29.96 |
| | MSE | **283.52** | **355.22** | **206.09** | **34.25** | **37.71** | **95.43** | **187.54** | **45.10** | **159.32** | **379.07** | 40.44 |

Table 4.12: Comparison of results of the proposed method with SRCNN [28], Peleg & Elad method [83], KK [127] and A+ [120] on 13259 images of LFW database [42], where bold numbers show the best performance.

| | Bicubic | SRCNN [28] | Peleg [83] | A+ [120] | KK [127] | Proposed method |
|---|---|---|---|---|---|---|
| PSNR | 31.97 | 31.95 | 32.33 | 25.16 | 30.76 | **33.07** |
| SSIM | 0.9449 | 0.9439 | 0.9504 | 0.8528 | 0.9400 | **0.9524** |
| VIFP | 0.5965 | 0.5828 | 0.5984 | 0.4648 | 0.5684 | **0.6058** |
| UQI | 0.8639 | 0.8585 | 0.8550 | 0.7478 | 0.8703 | **0.8885** |
| NQM | 21.28 | 20.59 | 20.56 | 16.33 | 20.86 | **23.02** |
| WSNR | 28.4 | 29.32 | 29.31 | 23.96 | 28.13 | **30.23** |
| MSE | 47.83 | 49.54 | 45.12 | 220.25 | 64.72 | **37.84** |

15 and super-resolved images $60 \times 60$.
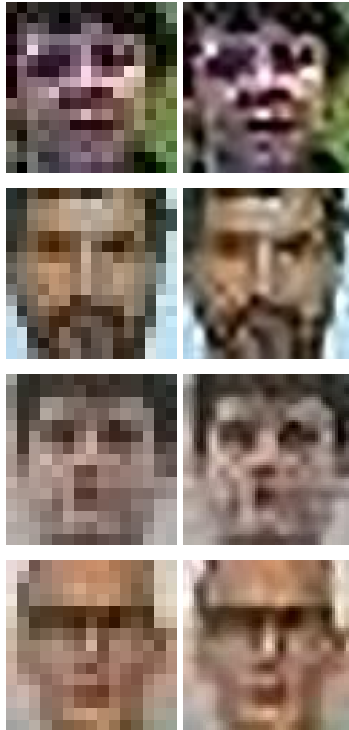
Figure 4.16: LR (15 × 15 at left) and SR (60 × 60 at right) version of images of different databases. The first row belongs to the Essex database, the second row is for the FERET database, and the third and forth rows belong to the HP and iCV-F databases, respectively.

## Databases

The FERET program was sponsored by the Department of Defense Counter-drug Technology Development Program through the Defense Advanced Research Products Agency (DARPA). It ran from 1993 through 1997, and its primary mission was to develop an automatic face detection system to assist security, intelligence, and law enforcement. The FERET database was collected to support testing and evaluation of face recognition algorithms. The photos were taken in a semi-controlled environment. The same physical setup was used in each photography session to maintain consistency throughout the whole database. There are some minor differences in images gathered on different dates due to reassembling the equipment for each session. The final corpus consists of 14051 eight-bit grayscale images with views of model heads ranging from frontal to both left and right profiles. [87, 86]. In 2003 a color version of the database was released by DARPA. It included 2413 high-resolution, 24-bit color still facial images of 856 individuals. The selected subset of FERET images database consists of 500 color

images.

The selected subset of Essex Faces94 facial images collection consists of 1500 images. The subjects of Essex database sat at a fixed distance from the camera and were asked to speak during the photoshoot. Speaking is for introducing variations of facial expressions. The original images of the database are $180 \times 200$ pixels. The background of all photos is plain green. No head scale is used. Minor changes in head turn, tilt and slant and in a position of the face appear in the images. No lighting variation occurs in the photos. Also, there is no individual hairstyle variation since the photos were taken in a single session [1].

The HP database consists of 2790 face images of 15 individuals with variations of pan and tilt angles from -90 to +90 degrees. It has 2 series of 93 images, all in different poses, for each person. The reason for having 2 series is to have both known and unknown faces for training and algorithms. The database features people of various skin colors and with or without glasses. The background of the images is neutral and uncluttered for focusing on face operations [34]. The subset of HP database used in this work includes 150 images. Faces in the selected subset are turned only in the horizontal direction.

Our own faces database (iCV-F) consists of 310 images. Photos were taken in 2 sessions in similar conditions. The database includes people wearing glasses or not and with various skin colors. Subjects were asked to make different facial expressions for each photo.

## Results

The results of the face recognition are given in Table 4.13. By comparing the recognition performances, it can be observed that super-resolved images result in an increase in recognition rate for all the databases used in the experimental procedures. Also, one may note that the recognition rate for the FERET and HP face databases are relatively low and this is because the HMM-based face recognition algorithm [73] is not robust in the pose changes, where both aforementioned databases are well-known for including faces with varying poses.

All in all, SR gives better results in face recognition in all cases. In this work, we have shown that super resolving the LR facial images will boost the face recognition rate.

Table 4.13: The correct recognition rate (in %) for different face databases with different resolution.

|          | LR images | SR images | Original size |
|----------|-----------|-----------|---------------|
| **Essex**  | 75.82     | 86.8      | 99.20         |
| **HP**     | 26.67     | 37.33     | 50.67         |
| **FERET**  | 20        | 21.60     | 31.6          |
| **iCV-F**  | 72.83     | 78.1      | 98.06         |

## 4.6 Deep Learning based Super-Resolution for Improved Action Recognition

In order to show the effectiveness of the proposed system in improving the quality of low-resolution images, and hence increasing the recognition accuracy of the dense trajectories-based action recognition algorithm of [125], we performed the following two experiments on the Hollywood2 database. On one hand, we assess the action recognition performance when processing LR videos. The goal of this experiment is to show how performance degrades when decreasing the resolution of videos. On the other hand, we evaluate the recognition performance after applying the super resolution method described in Section 3.3.2. The goal is to assess the recognition rate improvement obtained when the quality of LR videos has been enhanced. In the following, we first explain the employed databases, and then the details of the experimental results are given.

The Hollywood2 database has been employed. The database includes 12 classes of human actions with 10 classes of scenes distributed over 3669 video clips. The actions are answer phone, drive a car, eat, fight the person, get out of a car, hand shake, hug person, kiss, run, sit down, sit up and stand up. The database contains various video clips from about 70 movies. The database introduces a comprehensive benchmark for human action recognition in realistic and challenging settings and is used in experimental results in many state-of-the-art action recognition systems.

Two experiments were performed for the evaluation of the benefits of the mentioned upsampling technique for improving the action recognition performance on low-quality videos. In the first experiment, we evaluated the recognition performance of the dense trajectories method using low-resolution videos of different resolutions. For this experiment, we down-sampled the employed databases by down-sampling factors of four and eight. Each of these down-sampling factors results in a new database in which we applied the action recognition method described in Section 3.3.2. Results in this experiment on the entire database of Hollywood2 are shown in Figure 3.18. It can be seen from this figure that the

recognition rate generally drops as the resolution of the input images drops. This verifies the need for upsampling techniques for action recognition algorithms to work with low-resolution images.

In the second experiment, we assess the benefits of using the alpha blending technique in combining the super-resolved videos obtained by the deep learning algorithm and those obtained by the bicubic interpolation. To do that, a subset of the Hollywood2 database has been chosen and down-sampled by a factor of two. This subset contains 53 videos for training and 59 videos for testing, covering all the actions of the database by at least eight to ten videos for each action of the database. The images in the down-sampled subset have been upsampled by a factor of two by both the bicubic interpolation and the deep learning-based super-resolution of [27], separately. The recognition rate of the dense trajectories-based super-resolution algorithm of [125] have been obtained for these two upsampling cases and are reported in Table 4.14. Then the images upsampled by these two upsampling techniques (bicubic interpolation and deep learning-based super-resolution) are combined using the mentioned alpha blending technique, with an (experimentally determined) alpha value of 0.2. Next, recognition rates of the dense trajectories-based super-resolution algorithm of [125] have been obtained for the images that are combined using the alpha blending technique. The results are shown in the last column of Table 4.14.

The recognition results in Table 4.14 are reported for different descriptors obtained by the dense trajectories of [125], using a bag of words (BoW) technique. These descriptors are the trajectory (TRT), HoG, HoF, MBH, and their combined version following the method of [125]. For the classification purpose, the code words for each of the descriptors are classified using a Neural Network (with one hidden layer of neurons, trained for 50 epochs, with a learning rate of 0.1) and a linear SVM. It can be seen from this table that the proposed upsampling techniques using alpha blending produce better results compared to the other two cases using the Neural Network classifier. This verifies that deep learning-based super-resolution algorithms and other upsampling techniques (like bicubic interpolation) can complement each other in enhancing the quality of videos and hence improve the recognition performance of the state-of-the-art action of [125]. It should be also noted that the best performance among the approaches under consideration has been obtained by using the Neural Network with the TRT, while in [125] the best recognition rate has been obtained by using SVM from the combined classifier. However, the system of [125] uses the original images, while we have first down-sampled and then upsampled them. This means that the proposed combined classifier of [125] is not robust against down-sampling compared to TRT.

Table 4.14: The recognition accuracy of the dense trajectories-based action recognition algorithm of [125] for a subset of images of the Hollywood2 database that are first down-sampled by a factor of two and then are upsampled by different upsampling algorithms, including the one proposed in this thesis.

| - | - | Upsampling method | | |
|---|---|---|---|---|
| Classification method | Descriptor | Interpolated by bicubic | Super-resolved by deep learning | Combined by alpha blending |
| Neural Network | TRT | **36.78** % | **36.18**% | **39.28**% |
| Neural Network | HoG | 24.92% | 24.02% | 26.98% |
| Neural Network | HoF | 23.62% | 23.16% | 20.22% |
| Neural Network | MBHx | 20.37% | 21.76% | 25.69% |
| Neural Network | MBHy | 20.51% | 20.45% | 19.80% |
| Neural Network | combined | 29.87% | 23.57% | 34.47% |
| Liner SVM | TRT | 13.74% | 11.26% | 13.10% |
| Liner SVM | HoG | 11.47% | 14.18% | 13.32% |
| Liner SVM | HoF | 14.19% | 11.97% | 11.32% |
| Liner SVM | MBHx | 12.78% | 13.06% | 11.21% |
| Liner SVM | MBHy | 10.33% | 12.28% | 12.92% |
| Liner SVM | combined | 14.04% | 13.73% | 14.50% |

## 4.7 Spatio-Temporal Pain Recognition in CNN-based Super-Resolved Facial Images

### 4.7.1 Experimental Environment

As stated in Section 3.4, we evaluated the performance of pain detection in varying face resolution by employing the hybrid deep learning framework on the UNBC-McMaster Shoulder Pain database [69]. The video frames of the database showed patients who were suffering from shoulder pain while they were performing a series of active and passive range-of-motion tests. The pain indexes were computed by following the Prkachin and Solomon Pain Intensity (PSPI) scale from [89], and the pain levels vary in the interval 0-16 based on the FACS codes. Following [45], we classified each pain index into three categories of no pain (pain index lower than 1), weak pain (pain index between 2 and 6) and strong pain (pain index greater than 6). The three categories have been balanced by dropping consecutive no-pain frames at the beginning and at the end of each video, or by discarding entire video sequences which do not contain pain.

We applied the down-up sampling and SR algorithm described in Section 3.4.2 to generate three experimental datasets. The first dataset was created by using the original images from the UNBC-McMaster database (also used by [44]). The second and third datasets were denoted by 'SR$\frac{1}{4}$'and 'SR$\frac{1}{2}$', and were created by employing down-up sampling with the values $\frac{1}{4}$ and $\frac{1}{2}$, respectively, on the first dataset. The fourth dataset was denoted by 'SR2', and was created by employing the SR algorithm from Section 3.4.2 on the down-sampled images with factor $\frac{1}{2}$. The LSTM network was configured with 3 hidden-layers of 64 hidden-units each and a temporal window of 16 consecutive video frames. For the purpose of comparison, the experimental setup of the LSTM was kept fixed for all the experiments against the three datasets. The performances were estimated with

leave-one-subject-out cross-validation protocol.

### 4.7.2   The Obtained Results

Table 4.15 shows the results of the proposed system against the three sets. Here we report the accuracy in percentage for each of the three categories, namely "No Pain", "Weak Pain" and "Strong Pain". From the experiments, we can claim that the proposed method applied to super-resolved images is crucial since it reaches better performance than using the plain down-sampled versions. The latter is denoted by the amount of improvement appearing in the pain detection rate using the super-resolved images as the subjects while being compared against that of the LR ones. In other words, when recognizing the pains using the super-resolved images, a more powerful SR method leads to recognition rates closer to the case of considering the original ones. From the results, we can see that pain detection

Table 4.15: Pain detection results for the four experimental datasets created from the UNBC-McMaster [69] database.

| Semantic Ground Truth | Pain Index | $SR\frac{1}{4}$ | $SR\frac{1}{2}$ | SR2 |
|:---:|:---:|:---:|:---:|:---:|
| No Pain | 0, 1 | 55.3% | 62.22% | 55.78% |
| Weak Pain | 2, 3, 4, 5 | 73.1% | 67.7% | 75.94% |
| Strong Pain | >6 | 18.36% | 5.86% | 39.45% |
| F1-score | | 0.67 | 0.66 | 0.69 |
| Total | 0-16 | 62.43% | 62.64% | 65.34% |

is much better in super-resolved images compared to down-sampled ones by a large margin in the case of strong pain, while for the other two levels, namely no pain and weak pain, the performances are slightly better. This is due to the fact that stronger pain (compared to the weak and no pain cases) imposes more changes on the face, and these changes are more pronounced on super-resolved images; hence, the detection accuracy improves more by far in the strong pain class in comparison to the other classes. In order to see how temporal information affects the final results, we provide the SR2 accuracy when using a linear classifier on the plain CNN features against the LSTM predictions, which aggregates the temporal information in Table 4.16. Here the results are reported for each subject in the considered data set. As can be seen, temporal information improves the predictions by a large margin, $16\%$ on average, meaning that spatial features are not enough for determining the pain level on facial images. Thus, the temporal variation of the frames allows for finding higher level facial features, like FACS, which are central for predicting the PSPI pain score [90].

From Table 4.16, we notice that in two cases, specifically subject number 7 and subject number 8, the LSTM failed to improve the accuracy of the CNN. After a

detailed study of the dataset, we notice that sometimes, for both subjects, the pain index changes very rapidly among consecutive frames. The same pattern occurs (in a lighter form) also for subject 6, where improvement in the accuracy is not as good as for the other subjects. In addition, subject 7 is the only one that features only one video for the validation set, while subject 8 features three videos, one of which is very noisy with only 20 frames. We think that the aforementioned differences could be the key problems that lead to such a different performance for the different subjects.

Table 4.16: Comparison between CNN and LSTM performances on the SR2 dataset (in accuracy %). The CNN relies on the information of a single frame, while the LSTM takes into account variations in the images on the temporal axis. As can be seen, the LSTM enhances the accuracy prediction for all subjects, reaching 16% on average.

| subj. | CNN | LSTM |
|---|---|---|
| 1 | 40.9 | **58** |
| 2 | 50.3 | **61.5** |
| 3 | 52 | **63** |
| 4 | 50.7 | **65.6** |
| 5 | 50.3 | **82.5** |
| 6 | 42.5 | **48** |
| 7 | **29.3** | 28 |
| 8 | **47.3** | 40 |
| 9 | 50.4 | **81** |
| 10 | 44 | **65.4** |
| 11 | 50 | **82** |
| 12 | 51.3 | **66** |
| 13 | 51.9 | **65.5** |
| 14 | 30.2 | **60.5** |
| AVG | 45.8 | **61.9** |

# CONCLUSIONS

In this thesis, a study of both hardware and software solutions for image enhancement has been done. On the hardware side, a new liquid lens design with a DESA membrane located directly in the optical path has been demonstrated. Two prototypes with two different DESA, which have a 40 and 20 mm active area in diameter, were developed. The lens performance was consistent with the mechanics of elastomer deformation and relative focal length changes. A laser beam was used to show the change in the meniscus and to measure the focal length of the lens. The experimental results demonstrate that voltage in the range of 50 to 750 V is required to create change in the meniscus.

On the software side, a new satellite image enhancement system was proposed. The proposed technique decomposed the noisy input image into various frequency subbands by using DT-CWT. After removing the noise by applying the LA-BSF technique, its resolution was enhanced by employing DWT and interpolating the high-frequency subband images. An original image was interpolated with half of the interpolation factor used for interpolating the high-frequency subband images, and the super-resolved image was reconstructed by using IDWT.

A novel single-image SR method based on a generating dictionary from pairs of HR and their corresponding LR images was proposed. Firstly, HR and LR pairs were divided into patches in order to make HR and LR dictionaries respectively. The initial HR representation of an input LR image was calculated by combining the HR patches. These HR patches are chosen from the HR dictionary corresponding to the LR patches that have the closest distance to the patches of the input LR image. Each selected HR patch was processed further by passing through an illumination enhancement processing order to reduce the noticeable change of illumination between neighbor patches in the super-resolved image. In order to reduce the blocking effect, the average of the obtained SR image and the bicubic interpolated image was calculated.

The new kernels for sampling have also been proposed. The kernels can improve the SR by resulting in a sharper image. In order to demonstrate the effectiveness of the proposed kernels, the techniques from [83] and [50] for resolution enhancement were adopted. The super-resolved image was achieved by combining the HR

images produced by each of the proposed kernels using the alpha blending technique.

The proposed techniques and kernels are compared with various conventional and state-of-the-art techniques, and the quantitative test results and visual results on the final image quality show the superiority of the proposed techniques and kernels over conventional and state-of-art techniques.

# Bibliography

[1] Collection of facial images: Faces94. `http://cswww.essex.ac.uk/mv/allfaces/faces94.html`, accessed: 2015-12-28

[2] Standard test image. `http://www.imageprocessingplace.com/root_files_V3/image_databases.htm`, accessed: 2016-04-05

[3] The first liquid lens using the concept of electrowetting: Arctic series. `www.varioptic.com/technology/liquid-lens-autofocus-af`, accessed: 2016-09-19

[4] Abbas, M., Majid, A.A., Ali, J.: Positivity-preserving c2 rational cubic spline interpolation. ScienceAsia 39(2), 208–213 (2013)

[5] Ahad, M.A.R., Tan, J., Kim, H., Ishikawa, S.: A simple approach for low-resolution activity recognition. Int. J. Comput. Vis. Biomech 3(1) (2010)

[6] Anbarjafari, G., Agoyi, M., Laur, L., Rasti, P.: A robust color image watermarking scheme using entropy and QR decomposition. Radioengineering (2015)

[7] Anbarjafari, G.: An objective no-reference measure of illumination assessment. Measurement Science Review 15(6), 319–322 (2015)

[8] Anbarjafari, G., Demirel, H.: Image super resolution based on interpolation of wavelet domain high frequency subbands and the spatial domain input image. ETRI journal 32(3), 390–394 (2010)

[9] Anbarjafari, G., Rasti, P.: Illumination Enhancement: Image and Video, vol. 1. LAP LAMBERT Academic Publishing (2016)

[10] Anbarjafari, G., Rasti, P., Daneshmand, M., Ozcinar, C.: Resolution enhancement based image compression technique using singular value decomposition and wavelet transforms. In: Baleanu, D. (ed.) Wavelet Transform and Some of Its Real-World Applications, chap. 2. InTech (2015)

[11] Beatty, J.C., Barsky, B.A.: An introduction to splines for use in computer graphics and geometric modeling. Morgan Kaufmann (1987)

[12] Bellantonio, M., Haque, M.A., Rodriguez, P., Nasrollahi, K., Telve, T., Escarela, S., Gonzalez, J., Moeslund, T.B., Rasti, P., Anbarjafari, G.: Spatio-temporal pain recognition in cnn-based super-resolved facial images. In: 23rd International Conference on Pattern Recognition (ICPR) (2016)

[13] Berge, B., Peseux, J.: Variable focal lens controlled by an external voltage: An application of electrowetting. The European Physical Journal E 3(2), 159–163 (2000)

[14] Bolotnikova, A., Rasti, P., Traumann, A., Lusi, I., Daneshmand, M., Noroozi, F., Samuel, K., Sarkar, S., Anbarjafari, G.: Block based image compression technique using rank reduction and wavelet difference reduction. In: Seventh International Conference on Graphic and Image Processing. pp. 981702–981702. International Society for Optics and Photonics (2015)

[15] Butzer, P.L., Nessel, R.J.: Fourier Analysis and Approximation, vol. 1. Birkhäuser Verlag, Basel, Stuttgart (1971)

[16] Capel, D., Zisserman, A.: Super-resolution enhancement of text image sequences. In: Pattern Recognition, 2000. Proceedings. 15th International Conference on. vol. 1, pp. 600–605. IEEE (2000)

[17] Carey, W.K., Chuang, D.B., Hemami, S.S.: Regularity-preserving image interpolation. IEEE transactions on image processing 8(9), 1293–1297 (1999)

[18] Carpi, F., Frediani, G., Turco, S., De Rossi, D.: Bioinspired tunable lens with muscle-like electroactive elastomers. Advanced Functional Materials 21(21), 4152–4158 (2011)

[19] Chen, H.S., Lin, Y.H.: An endoscopic system adopting a liquid crystal lens with an electrically tunable depth-of-field. Optics express 21(15), 18079–18088 (2013)

[20] Cristani, M., Cheng, D.S., Murino, V., Pannullo, D.: Distilling information with super-resolution for video surveillance. In: Proceedings of the ACM 2nd international workshop on Video surveillance & sensor networks. pp. 2–11. ACM (2004)

[21] Daneshmand, M., Aabloo, A., Anbarjafari, G.: Size-dictionary interpolation for robots adjustment. Frontiers in bioengineering and biotechnology 3 (2015)

[22] De Santis, P., Gori, F.: On an iterative method for super-resolution. Journal of Modern Optics 22(8), 691–695 (1975)

[23] Demirel, H., Anbarjafari, G.: Discrete wavelet transform-based satellite image resolution enhancement. IEEE transactions on geoscience and remote sensing 49(6), 1997–2004 (2011)

[24] Demirel, H., Anbarjafari, G.: Image resolution enhancement by using discrete and stationary wavelet decomposition. Image Processing, IEEE Transactions on 20(5), 1458–1460 (2011)

[25] Do, M.N., Vetterli, M.: Rotation invariant texture characterization and retrieval using steerable wavelet-domain hidden markov models. IEEE Transactions on Multimedia 4(4), 517–527 (2002)

[26] Dong, C., Loy, C.C., He, K., Tang, X.: Learning a deep convolutional network for image super-resolution. In: Computer Vision–ECCV 2014, pp. 184–199. Springer (2014)

[27] Dong, C., Loy, C.C., He, K., Tang, X.: Image super-resolution using deep convolutional networks. CoRR abs/1501.00092 (2015), http://arxiv.org/abs/1501.00092

[28] Dong, C., Loy, C.C., He, K., Tang, X.: Image super-resolution using deep convolutional networks. IEEE transactions on pattern analysis and machine intelligence 38(2), 295–307 (2016)

[29] Fiorot, J.C., Tabka, J.: Shape-preserving $c^2$ cubic polynomial interpolating splines. Mathematics of Computation 57(195), 291–298 (1991)

[30] Freeman, W.T., Jones, T.R., Pasztor, E.C.: Example-based super-resolution. Computer Graphics and Applications, IEEE 22(2), 56–65 (2002)

[31] Gerchberg, R.: Super-resolution through error energy reduction. Journal of Modern Optics 21(9), 709–720 (1974)

[32] Glasner, D., Bagon, S., Irani, M.: Super-resolution from a single image. In: Computer Vision, 2009 IEEE 12th International Conference on. pp. 349–356. IEEE (2009)

[33] Gorman, C.B., Biebuyck, H.A., Whitesides, G.M.: Control of the shape of liquid lenses on a modified gold surface using an applied electrical potential across a self-assembled monolayer. Langmuir 11(6), 2242–2246 (1995)

[34] Gourier, N., Hall, D., Crowley, J.L.: Estimating face orientation from robust detection of salient facial features. In: ICPR International Workshop on Visual Observation of Deictic Gestures. Citeseer (2004)

[35] Graham-Rowe, D.: Liquid lenses make a splash. Nat. Photonics 6, 2–4 (2006)

[36] Gunturk, B.K., Batur, A.U., Altunbasak, Y., Hayes, M.H., Mersereau, R.M.: Eigenface-domain super-resolution for face recognition. Image Processing, IEEE Transactions on 12(5), 597–606 (2003)

[37] van Hateren, J.H., van der Schaaf, A.: Independent component filters of natural images compared with simple cells in primary visual cortex. Proceedings of the Royal Society of London B: Biological Sciences 265(1394), 359–366 (1998)

[38] Hennings-Yeomans, P.H., Baker, S., Kumar, B.V.: Recognition of low-resolution faces using multiple still images and multiple cameras. In: Biometrics: Theory, Applications and Systems, 2008. BTAS 2008. 2nd IEEE International Conference on. pp. 1–6. IEEE (2008)

[39] Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural computation 9(8), 1735–1780 (1997)

[40] Hou, H.S., Andrews, H.: Cubic splines for image interpolation and digital filtering. Acoustics, Speech and Signal Processing, IEEE Transactions on 26(6), 508–517 (1978)

[41] Hsieh, H.T., Wei, H.C., Lin, M.H., Hsu, W.Y., Cheng, Y.C., Su, G.D.J.: Thin autofocus camera module by a large-stroke micromachined deformable mirror. Optics express 18(11), 11097–11104 (2010)

[42] Huang, G.B., Mattar, M., Lee, H., Learned-Miller, E.: Learning to align from scratch. In: NIPS (2012)

[43] Huang, T.S., Tsay, R.Y.: Multiple frame image restoration and registration. In: Advances in Computer Vision and Image Processing. pp. 317–339 (1984)

[44] Irani, R., Nasrollahi, K., Moeslund, T.B.: Pain recognition using spatiotemporal oriented energy of facial muscles. In: 2015 IEEE Conference on

Computer Vision and Pattern Recognition Workshops (CVPRW). pp. 80–87 (June 2015)

[45] Irani, R., Nasrollahi, K., Simon, M.O., Corneanu, C.A., Escalera, S., Bahnsen, C., Lundtoft, D.H., Moeslund, T.B., Pedersen, T.L., Klitgaard, M.L., Petrini, L.: Spatiotemporal analysis of rgb-d-t facial images for multimodal pain level recognition. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops (June 2015)

[46] Ito, I., Kiya, H.: A new technique of non-iterative super-resolution without boundary distortion. In: ICASSP. pp. 1273–1276 (2012)

[47] Kennedy, J.A., Israel, O., Frenkel, A., Bar-Shalom, R., Azhari, H.: Super-resolution in pet imaging. Medical Imaging, IEEE Transactions on 25(2), 137–147 (2006)

[48] Keys, R.G.: Cubic convolution interpolation for digital image processing. Acoustics, Speech and Signal Processing, IEEE Transactions on 29(6), 1153–1160 (1981)

[49] Kim, K.I., Kim, D., Kim, J.H.: Example-based learning for image super-resolution. In: The third Tsinghua-KAIST Joint Workshop on Pattern Recognition (2004)

[50] Kim, K.I., Kwon, Y.: Example-based learning for single-image super-resolution. In: Pattern Recognition, pp. 456–465. Springer (2008)

[51] Kingma, D., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)

[52] Kingsbury, N.: Image processing with complex wavelets. Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences 357(1760), 2543–2560 (1999)

[53] Kingsbury, N.: A dual-tree complex wavelet transform with improved orthogonality and symmetry properties. In: Image Processing, 2000. Proceedings. 2000 International Conference on. vol. 2, pp. 375–378. IEEE (2000)

[54] Kingsbury, N.: Complex wavelets for shift invariant analysis and filtering of signals. Applied and computational harmonic analysis 10(3), 234–253 (2001)

[55] Kiveste, H., et al.: Fliud lens device construction and actuator performance based on electroactive polymers. Ph.D. thesis, University of Tartu (2013)

[56] Kivilo, A., Zondaka, Z., Kesküla, A., Rasti, P., Tamm, T., Kiefer, R.: Electro-chemo-mechanical deformation properties of polypyrrole/dodecyl-benzenesulfate linear actuators in aqueous and organic electrolyte. RSC Advances 6(99), 96484–96489 (2016)

[57] Kivinukk, A., Tamberg, G.: On sampling operators defined by the hann window and some of their extensions. Sampling Theory in Signal and Image Processing 2, 235–258 (2003)

[58] Kivinukk, A., Tamberg, G.: On blackman-harris windows for shannon sampling series. Sampling Theory in Signal & Image Processing 6(1) (2007)

[59] Kivinukk, A., Tamberg, G.: Interpolating generalized shannon sampling operators, their norms and approximation properties. Sampling Theory in Signal and Image Processing 8, 77–95 (2009)

[60] Kuiper, S., Hendriks, B.H., Huijbregts, L.J., Hirschberg, A.M., Renders, C.A., van As, M.A.: Variable-focus liquid lens for portable applications. In: Optical Science and Technology, the SPIE 49th Annual Meeting. pp. 100–109. International Society for Optics and Photonics (2004)

[61] Kuiper, S., Hendriks, B.: Variable-focus liquid lens for miniature cameras. Applied physics letters 85(7), 1128–1130 (2004)

[62] Lau, G.K., La, T.G., Shiau, L.L., Tan, A.W.Y.: Challenges of using dielectric elastomer actuators to tune liquid lens. In: Proc. SPIE. vol. 9056, p. 90561J (2014)

[63] Le, V., Brandt, J., Lin, Z., Bourdev, L., Huang, T.S.: Interactive facial feature localization. In: Computer Vision–ECCV 2012, pp. 679–692. Springer (2012)

[64] Lee, S., Wolberg, G., Shin, S.Y.: Scattered data interpolation with multilevel b-splines. IEEE transactions on visualization and computer graphics 3(3), 228–244 (1997)

[65] Léonard, N., Waghmare, S., Wang, Y.: Rnn: Recurrent library for torch. arXiv preprint arXiv:1511.07889 (2015)

[66] Li, F., Jia, X., Fraser, D.: Universal hmt based super resolution for remote sensing images. In: Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on. pp. 333–336. IEEE (2008)

[67] Lin, F.C., Fookes, C.B., Chandran, V., Sridharan, S.: Investigation into optical flow super-resolution for surveillance applications. In: Lovell, B.C., Maeder, A.J. (eds.) APRS Workshop on Digital Image Computing: Pattern Recognition and Imaging for Medical Applications. pp. 73–78. The University of Queensland, Brisbane (2005), `http://eprints.qut.edu.au/17945/`, the contents of this conference can be freely accessed online via the conference's web page (see hypertext link).

[68] Lotz, P., Matysek, M., Schlaak, H.F.: Fabrication and application of miniaturized dielectric elastomer stack actuators. Mechatronics, IEEE/ASME Transactions on 16(1), 58–66 (2011)

[69] Lucey, P., Cohn, J.F., Prkachin, K.M., Solomon, P.E., Matthews, I.: Painful data: The UNBC-McMaster shoulder pain expression archive database. In: 2011 IEEE International Conference on Automatic Face Gesture Recognition and Workshops (FG 2011). pp. 57–64 (Mar 2011)

[70] Maintz, J.A., Viergever, M.A.: A survey of medical image registration. Medical image analysis 2(1), 1–36 (1998)

[71] Malczewski, K., Stasinski, R.: Toeplitz-based iterative image fusion scheme for mri. In: Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on. pp. 341–344. IEEE (2008)

[72] Mallat, S.: A wavelet tour of signal processing. Academic press (1999)

[73] Miar-Naimi, H., Davari, P.: A new fast and efficient hmm-based face recognition system using a 7-state hmm along with svd coefficients. Iranian Journal of Electrical and Electronic Engineering 4(1), 46–57 (2008)

[74] Milanfar, P.: Super-resolution imaging. CRC Press (2010)

[75] Morin, R., Basarab, A., Kouamé, D.: Alternating direction method of multipliers framework for super-resolution in ultrasound imaging. In: 2012 9th IEEE International Symposium on Biomedical Imaging (ISBI). pp. 1595–1598. IEEE (2012)

[76] Nasir, H., Stanković, V., Marshall, S.: Singular value decomposition based fusion for super-resolution image reconstruction. Signal Processing: Image Communication 27(2), 180–191 (2012)

[77] Nasrollahi, K., Moeslund, T.B.: Finding and improving the key-frames of long video sequences for face recognition. In: Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on. pp. 1–6 (2010)

[78] Nasrollahi, K., Escalera, S., Rasti, P., Anbarjafari, G., Baro, X., Escalante, H.J., Moeslund, T.B.: Deep learning based super-resolution for improved action recognition. In: Image Processing Theory, Tools and Applications (IPTA), 2015 International Conference on. pp. 67–72. IEEE (2015)

[79] Nasrollahi, K., Moeslund, T.B.: Super-resolution: a comprehensive survey. Machine vision and applications 25(6), 1423–1468 (2014)

[80] Olevskii, A., Ulanovskii, A.: Interpolation in bernstein and paley–wiener spaces. Journal of Functional Analysis 256(10), 3257–3278 (2009)

[81] Parkhi, O.M., Vedaldi, A., Zisserman, A.: Deep face recognition. In: British Machine Vision Conference. vol. 1, p. 6 (2015)

[82] Peled, S., Yeshurun, Y.: Superresolution in mri: application to human white matter fiber tract visualization by diffusion tensor imaging. Magnetic resonance in medicine 45(1), 29–35 (2001)

[83] Peleg, T., Elad, M.: A statistical prediction model based on sparse representations for single image super-resolution. Image Processing, IEEE Transactions on 23(6), 2569–2582 (2014)

[84] Pelrine, R., Kornbluh, R.D., Pei, Q., Stanford, S., Oh, S., Eckerle, J., Full, R.J., Rosenthal, M.A., Meijer, K.: Dielectric elastomer artificial muscle actuators: toward biomimetic motion. In: SPIE's 9th Annual International Symposium on Smart Structures and Materials. pp. 126–137. International Society for Optics and Photonics (2002)

[85] Peng, Y., Yang, F., Dai, Q., Xu, W., Vetterli, M.: Super-resolution from unregistered aliased images with unknown scalings and shifts. In: 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 857–860. IEEE (2012)

[86] Phillips, P.J., Moon, H., Rizvi, S., Rauss, P.J., et al.: The feret evaluation methodology for face-recognition algorithms. Pattern Analysis and Machine Intelligence, IEEE Transactions on 22(10), 1090–1104 (2000)

[87] Phillips, P.J., Wechsler, H., Huang, J., Rauss, P.J.: The feret database and evaluation procedure for face-recognition algorithms. Image and vision computing 16(5), 295–306 (1998)

[88] Pjatkin, K., Daneshmand, M., Rasti, P., Anbarjafari, G.: Probability distribution function based iris recognition boosted by the mean rule. In: Intelligent Computing and Internet of Things (ICIT), 2014 International Conference on. pp. 47–50. IEEE (2015)

[89] Prkachin, K.M., Solomon, P.E.: The structure, reliability and validity of pain expression: Evidence from patients with shoulder pain. Pain 139(2), 267–274 (2008)

[90] Prkachin, K., Schultz, I., Berkowitz, J., Hughes, E., Hunt, D.: Assessing pain behaviour of low-back pain patients in real time: concurrent validity and examiner sensitivity. Behaviour Research and Therapy 40(5), 595 – 607 (2002)

[91] Rabiner, L.R.: A tutorial on hidden markov models and selected applications in speech recognition. Proceedings of the IEEE 77(2), 257–286 (1989)

[92] Rao, M.V.G., Vathsal, S.: Local adaptive bivariage shrinkage for medical image denoising. International Journal of Electronics Engineering (2009)

[93] Rasti, P., Daneshmand, M., Alisinanoglu, F., Ozcinar, C., Anbarjafari, G.: Medical image illumination enhancement and sharpening by using stationary wavelet transform. In: Signal Processing and Communication Application Conference (SIU), 2016 24th. pp. 153–156. IEEE (2016)

[94] Rasti, P., Daneshmand, M., Anbarjafari, G.: Statistical approach based iris recognition using local binary pattern. DYNA 92, 76–81 (2017)

[95] Rasti, P., Demirel, H., Anbarjafari, G.: Image resolution enhancement by using interpolation followed by iterative back projection. In: Signal Processing and Communications Applications Conference (SIU), 2013 21st. pp. 1–4. IEEE (2013)

[96] Rasti, P., Demirel, H., Anbarjafari, G.: Iterative back projection based image resolution enhancement. In: Machine Vision and Image Processing (MVIP), 2013 8th Iranian Conference on. pp. 237–240. IEEE (2013)

[97] Rasti, P., Demirel, H., Anbarjafari, G.: Improved iterative back projection for video super-resolution. In: 2014 22nd Signal Processing and Communications Applications Conference (SIU). pp. 552–555. IEEE (2014)

[98] Rasti, P., Hous, H., Schlaak, H.F., Kiefer, R., Anbarjafari, G.: Dielectric elastomer stack actuator-based autofocus fluid lens. Applied optics 54(33), 9976–9980 (2015)

[99] Rasti, P., Kesküla, A., Haus, H., Schlaak, H.F., Anbarjafari, G., Aabloo, A., Kiefer, R.: A passive autofocus system by using standard deviation of the image on a liquid lens. In: SPIE Smart Structures and Materials+ Nondestructive Evaluation and Health Monitoring. pp. 94301Q–94301Q. International Society for Optics and Photonics (2015)

[100] Rasti, P., Kiefer, R., Anbarjafari, G.: Autofocus liquid lens by using sharpness measurement. In: 2015 23nd Signal Processing and Communications Applications Conference (SIU). pp. 608–611. IEEE (2015)

[101] Rasti, P., Lusi, I., Demirel, H., Kiefer, R., Anbarjafari, G.: Wavelet transform based new interpolation technique for satellite image resolution enhancement. In: Aerospace Electronics and Remote Sensing Technology (ICARES), 2014 IEEE International Conference on. pp. 185–188. IEEE (2014)

[102] Rasti, P., Lüsi, I., Sahakyan, A., Traumann, A., Bolotnikova, A., Daneshmand, M., Kiefer, R., Aabloo, A., Anbarjafar, G., Demirel, H., et al.: Modified back projection kernel based image super resolution. In: Artificial Intelligence, Modelling and Simulation (AIMS), 2014 2nd International Conference on. pp. 161–165. IEEE (2014)

[103] Rasti, P., Nasrollahi, K., Orlova, O., Tamberg, G., Moeslund, T.B., Anbarjafari, G.: Reducible dictionaries for single image super-resolution based on patch matching and mean shifting. Journal of Electronic Imaging pp. 1–22 (Accepted) (2017)

[104] Rasti, P., Orlova, O., Tamberg, G., Ozcinar, C., Nasrollahi, K., Moeslund, T.B., Anbarjafari, G.: Improved interpolation kernels for super-resolution algorithms. In: IEEE International Conference on Image Processing Theory, Tools and Applications. IEEE (2016)

[105] Rasti, P., Samiei, S., Agoyi, M., Escalera, S., Anbarjafari, G.: Robust non-blind color video watermarking using QR decomposition and entropy analysis. Journal of Visual Communication and Image Representation 38, 838–847 (2016)

[106] Rasti, P., Taşmaz, H., Daneshmand, M., Kiefer, R., Ozcinar, C., Anbarjafari, G.: Satellite image enhancement: systematic approach for denoising and resolution enhancement. DYNA-Ingeniería e Industria 91(3) (2016)

[107] Rasti, P., Uiboupin, T., Escalera, S., Anbarjafari, G.: Convolutional neural network super resolution for face recognition in surveillance monitoring.

In: International Conference on Articulated Motion and Deformable Objects. pp. 175–184. Springer (2016)

[108] Rener, Y., Wei, J., Ken, C.: Downsample-based multiple description coding and post-processing of decoding. In: Control Conference, 2008. CCC 2008. 27th Chinese. pp. 253–256. IEEE (2008)

[109] Saurei, L., Mathieu, G., Berge, B.: Design of an autofocus lens for vga 1/4-in. ccd and cmos sensors. In: Optical Systems Design. pp. 288–296. International Society for Optics and Photonics (2004)

[110] Schlaak, H.F., Jungmann, M., Matysek, M., Lotz, P.: Novel multilayer electrostatic solid state actuators with elastic dielectric. In: Smart Structures and Materials. pp. 121–133. International Society for Optics and Photonics (2005)

[111] Schmeisser, G.: Interconnections between multiplier methods and window methods in generalized sampling. Sampling Theory in Signal and Image Processing 9, 1–24 (2010)

[112] Sendur, L., Selesnick, I.W.: Bivariate shrinkage functions for wavelet-based denoising exploiting interscale dependency. IEEE Transactions on signal processing 50(11), 2744–2756 (2002)

[113] Sendur, L., Selesnick, I.W.: Bivariate shrinkage with local variance estimation. IEEE Signal Processing Letters 9(12), 438–441 (2002)

[114] Sezer, O.G., Altunbasak, Y., Ercil, A.: Face recognition with independent component-based super-resolution. In: Electronic Imaging 2006. pp. 607705–607705. International Society for Optics and Photonics (2006)

[115] Shian, S., Diebold, R.M., Clarke, D.R.: Tunable lenses using transparent dielectric elastomer actuators. Optics express 21(7), 8669–8676 (2013)

[116] Taşmaz, H., Demirel, H., Anbarjafari, G.: Satellite image enhancement by using dual tree complex wavelet transform: Denoising and illumination enhancement. In: 2012 20th Signal Processing and Communications Applications Conference (SIU). pp. 1–4. IEEE (2012)

[117] Taşmaz, H., Erçelebi, E.: Image enhancement via space-adaptive lifting scheme exploiting subband dependency. Digital Signal Processing 20(6), 1645–1655 (2010)

[118] Temizel, A., Vlachos, T.: Image resolution upscaling in the wavelet domain using directional cycle spinning. Journal of Electronic Imaging 14(4), 040501–040501 (2005)

[119] Temizel, A., Vlachos, T.: Wavelet domain image resolution enhancement using cycle-spinning. Electronics Letters 41(3), 119–121 (2005)

[120] Timofte, R., Smet, V., Gool, L.: Anchored neighborhood regression for fast example-based super-resolution. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 1920–1927 (2013)

[121] Tsai, R., Huang, T.S.: Multiframe image restoration and registration. Advances in computer vision and Image Processing 1(2), 317–339 (1984)

[122] Viola, P., Jones, M.J.: Robust real-time face detection. International journal of computer vision 57(2), 137–154 (2004)

[123] Von Seggern, D.H.: CRC standard curves and surfaces with mathematica. CRC Press (2016)

[124] Wang, H., Schmid, C.: Action recognition with improved trajectories. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 3551–3558 (2013)

[125] Wang, H., Schmid, C.: Action recognition with improved trajectories. In: Computer Vision (ICCV), 2013 IEEE International Conference on. pp. 3551–3558. IEEE (2013)

[126] Wippermann, F., Schreiber, P., Bräuer, A., Craen, P.: Bifocal liquid lens zoom objective for mobile phone applications. In: Proc. SPIE. vol. 6501, pp. 650109–9 (2007)

[127] Yang, C.Y., Ma, C., Yang, M.H.: Single-image super-resolution: a benchmark. In: Computer Vision–ECCV 2014, pp. 372–386. Springer (2014)

[128] Yang, J., Huang, T.: Image super-resolution: Historical overview and future challenges. Super-resolution imaging pp. 20–34 (2010)

[129] Yang, J., Wright, J., Huang, T.S., Ma, Y.: Image super-resolution via sparse representation. Image Processing, IEEE Transactions on 19(11), 2861–2873 (2010)

[130] Yang, Y., Newsam, S.: Bag-of-visual-words and spatial extensions for land-use classification. In: Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems. pp. 270–279. ACM (2010)

[131] Yu, G., Sapiro, G., Mallat, S.: Solving inverse problems with piecewise linear estimators: From gaussian mixture models to structured sparsity. Image Processing, IEEE Transactions on 21(5), 2481–2499 (2012)

[132] Zhao, S., Han, H., Peng, S.: Wavelet-domain HMT-based image super-resolution. In: Image Processing, 2003. ICIP 2003. Proceedings. 2003 International Conference on. vol. 2, pp. II–953. IEEE (2003)

# ACKNOWLEDGMENTS

I would like to thank to my PhD advisors, Professors Gholamreza Anbarjafari (Shahab) and Rudolf Keifer, for supporting me during these past three years. Shahab is someone you will instantly love and never forget once you meet him. He is the funniest advisor and one of the smartest people I know. I hope that I could be as lively, enthusiastic, and energetic as Shahab and to someday be able to command an audience as well as he can. Rudolf has been supportive and has given me the freedom to pursue various projects without objection. He has also provided insightful discussions about the research.

I will forever be thankful to my former university advisor, Professor Hasan Demirel. Hasan has been helpful in providing advice many times during my graduate school career. He was and remains my best role model for a scientist, mentor, and teacher. I still think fondly of my time as a graduate student in his department. Hasan was the reason why I decided to pursue a career in research. His enthusiasm and love for teaching is contagious.

I am thankful to my colleagues, Olga Orlova, Prof. Gert Tamberg, Assoc. Prof. Kamal Nasrollahi, and Prof. Thomas B. Moeslund for their help on my research approach, especially on our works to develop new interpolation kernels. I also would like to thank Prof. Sergio Escalera for his new ideas on research purposes, which always open a new door where there were only walls.

Finally, and most importantly, I would like to thank my wife, Salma. Her support, encouragement, quiet patience and unwavering love were undeniably the bedrock upon which the past ten years of my life have been built. Her tolerance of my occasional vulgar moods is a testament in itself to her unyielding devotion and love. I thank my parents, Masoud and Shahla, for their faith in me and for allowing me to be as ambitious as I wanted. It was under their watchful eye that I gained so much drive and an ability to tackle challenges head on. Also, I thank Salma's parents, Naser and Batool. Your prayer for me was what sustained me thus far.

# KOKKUVÕTE
# (SUMMARY IN ESTONIAN)

# VEDEL-OBJEKTIIV ABIL SALVESTATUD KAUGSEIRE PILTIDE ANALÜÜS KASUTADES SUPER-RESOLUTSIOONI MEETODEID

Käesolevas doktoritöös uuriti nii riist- kui ka tarkvaralisi lahendusi piltide töötlemiseks. Riistvaralise poole pealt pakuti lahenduseks uudset vedelläätse, milles on dielektrilisest elastomeerist kihilise täituriga membraan otse optilisel teljel. Doktoritöö käigus arendati välja kaks prototüüpi kahe erineva dielektrilisest elastomeerist kihilise täituriga, mille aktiivne ala oli ühel juhul 40 ja teisel 20 mm. Läätse töö vastas elastomeeri deformatsiooni mehaanikale ja suhtelistele muutustele fookuskauguses. Muutuste demonstreerimiseks meniskis ja läätse fookuskauguse mõõtmiseks kasutati laserkiirt. Katseandmetest selgub, et muutuste tekitamiseks on vajalik pinge vahemikus 50 kuni 750 volti.

Tarkvaralise poole pealt pakuti uut satelliitpiltide parandamise süsteemi. Pakutud süsteem jagas mürase sisendpildi DT-CWT laineteisenduse abil mitmeteks sagedusalamribadeks. Pärast müra eemaldamist LA-BSF funktsiooni abil suurendati pildi resolutsiooni DWT-ga ja kõrgsagedusliku alamriba piltide interpoleerimisega. Interpoleerimise faktor algsele pildile oli pool sellest, mida kasutati kõrgsagedusliku alamriba piltide interpoleerimisel ning superresolutsiooniga pilt rekonstrueeriti IDWT abil.

Käesolevas doktoritöös pakuti tarkvaraliseks lahenduseks uudset sõnastiku baasil töötavat superresolutsiooni (SR) meetodit, milles luuakse paarid suure resolutsiooniga (HR) ja madala resolutsiooniga (LR) piltidest. Kõigepealt jagati vastava sõnastiku loomiseks HR ja LR paarid omakorda osadeks. Esialgse HR kujutise

saamiseks LR sisendpildist kombineeriti HR osi. HR osad valiti sõnastikust nii, et neile vastavad LR osad oleksid võimalikult lähedased sisendiks olevale LR pildile. Iga valitud HR osa heledust korrigeeriti, et vähendada kõrvuti asuvate osade heleduse erinevusi superresolutsiooniga pildil. Plokkide efekti vähendamiseks arvutati saadud SR pildi keskmine ning bikuupinterpolatsiooni pilt.

Lisaks pakuti käesolevas doktoritöös välja kernelid, mille tulemusel on võimalik saadud SR pilte teravamaks muuta. Pakutud kernelite tõhususe tõestamiseks kasutati [83] ja [50] poolt pakutud resolutsiooni parandamise meetodeid. Superresolutsiooniga pilt saadi iga kerneli tehtud HR pildi kombineerimise teel alpha blendingu meetodit kasutades.

Pakutud meetodeid ja kerneleid võrreldi erinevate tavaliste ja kaasaegsete meetoditega. Kvantitatiivsetest katseandmetest ja saadud piltide kvaliteedi visuaalsest hindamisest selgus, et pakutud meetodid on tavaliste kaasaegsete meetoditega võrreldes paremad.

# PUBLICATIONS

# CURRICULUM VITAE

**Personal data**

| | |
|---|---|
| Name | Pejman Rasti |
| Birth | September 18th, 1985 |
| | Esfahan, Iran |
| Citizenship | Iranian |
| Marital Status | Married |
| Languages | Persian, English |
| Address | Nooruse 1, Tartu |
| | 50411 Tartumaa |
| | Estonia |
| Contact | +372 588 46434 |
| | pejman.rasti@ut.ee |
| | pejman@icv.tuit.ut.ee |

**Education**

| | |
|---|---|
| 2014– | University of Tartu, Ph.D. candidate in Material science, Major in Image Processing and Computer Vision |
| 2012–2014 | Eastern Mediterranean University, M.Sc. in Electrical and Electronic Engineering |
| 2010–2012 | Azad University, B.Sc. in Electrical and Electronic Engineering, Major in Electronics. |

**Experience**

| | |
|---|---|
| 2014– | Intelligent Computer Vision (iCV) Research Group, Research Assistant |
| 2013–2014 | Eastern Mediterranean University, Technical Assistant in Cisco Laboratory |

| 2012–2013 | Eastern Mediterranean University, Student Assistant |
| 2010–2012 | Benco Company, Network Manager |

## Expertise

| 1 | Super Resolution |
| 2 | Image Acquisition & Enhancement |
| 3 | Biometric Recognition and Security |
| 4 | Video Processing |
| 5 | Deep Learning |

## Teaching Experience (teaching assistant)

| 1 | Digital Image Processing |
| 2 | Data Analysis and Calculation with Matlab |

## Professional Duties

Reviewer of   The following journals and conferences
Optics Letters
Optics Express
Signal, Image and Video Processing
Microscopy Research and Technique
Optics & Laser Technology
IETE Journal
The International Journal of Acoustics and Vibration (IJAV)
ICGIP 2016 conference
IPTA 2016 conference

IEEE Member since 2011

Former Vice Chair of IEEE Student Branch of University of Tartu

## Publication

Please have a look at my google scholar profile at

https://scholar.google.com/citations?user=hl8Jyr4AAAAJ&hl=en

# ELULOOKIRJELDUS

**Isikuandmed**

| | |
|---|---|
| Nimi | Pejman Rasti |
| Sünniaeg ja -koht | 18. september 1985<br>Esfahan, Iraan |
| Kodakondsus | iraanlane |
| Perekonnaseis | abielus |
| Keelteoskus | Pärsia, inglise |
| Aadress | Nooruse 1, Tartu<br>50411 Tartumaa<br>Estonia |
| Kontaktandmed | +372 588 46434<br>pejman.rasti@ut.ee<br>pejman@icv.tuit.ut.ee |

**Haridustee**

| | |
|---|---|
| 2014– | Tartu Ülikool, |
| 2012–2014 | Eastern Mediterranean Ülikool, MSc, põhja-Küpros |
| 2010–2012 | Azad Ülikool, BSc, Iraan, |

# DISSERTATIONES SCIENTIAE MATERIALIS
# UNIVERSITATIS TARTUENSIS

1. **Martin Järvekülg.** Tubular microstructures by Hf-, Zr- and Ti-butoxide gel sheet rolling. Tartu, 2011, 112 p.
2. **Sergei Vlassov.** Investigation of nanoscale interactions and mechanical properties of nanostructures using quartz tuning fork based real-time measurements. Tartu, 2011, 171 p.
3. **Margus Kodu.** Pulsed Laser Deposition of Magnesium Oxide and Barium Ternary Oxides for Plasma Display Protective Layers. Tartu, 2011, 89 p.
4. **Rainer Pärna.** Surface studies of some oxide films. Tartu, 2011, 129 p.
5. **Jevgeni Šulga**. Self-assembly and interaction of nanostructures. Tartu, 2011, 114 p.
6. **Wojciech Kuznik.** Quantum-chemical computer simulations of the linear and non-linear optical properties of pyrazoloquinoline and dicyanopyrazine derivatives. Tartu, 2012, 89 p.
7. **Leonid Dorogin.** Structural and tribological properties of zero- and one-dimensional nanocrystals. Tartu, 2012, 193 p.
8. **Viljar Palmre.** Fabrication and characterization of microporous carbon-based electroactive polymer actuators. Tartu, 2012, 99 p.
9. **Madis Paalo**. Synthesis of CNT-metal oxide nanocomposites: sol-gel process, rheology, structural and functional properties. Tartu, 2014, 175 p.
10. **Raul Välbe.** Developement of ionic liquid composites by sol-gel method for elaboration of industrial nano- and microstructures. Tartu, 2014, 97 p.
11. **Urmas Joost.** Impurity and preparation dependent properties of titania thin films. Tartu, 2014, 155 p.
12. **Madis Umalas**. Application of sol-gel technology for production of ceramic nanocomposites and functional coatings. Tartu, 2015, 136 p.
13. **Elena Samsonova.** Energy transfer probe as a tool to study morphological and structural origins of fluorescence quenching in rare-earth doped nano-phosphors. Tartu, 2015, 114 p.
14. **Jay Mondal.** Novel Corrosion Protective Nanostructured Composite Coatings. Tartu, 2016, 104 p.
15. **Kathriin Utt.** Metal oxide mesostructures for optical applications. Tartu, 2016, 88 p.
16. **Kaido Siimon.** Electrospun gelatin cross-linked by glucose. Tartu, 2016, 84 p.
17. **Marko Part.** Combined three-dimensional sol-gel structures and atomic layer deposited thin films. Tartu, 2017, 162 p.