Analysis of Thermal Super Resolution and Thermal-RGB Stereo Matching Algorithms

A Project Report

submitted by

JYOTHIS C G

in partial fulfilment of the requirements
for the award of the degree of

MASTER OF TECHNOLOGY

DEPARTMENT OF ELECTRICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY MADRAS.

MAY 2019

THESIS CERTIFICATE

This is to certify that the thesis titled **Analysis of Thermal Super Resolution and Thermal-RGB Stereo Matching Algorithms**, submitted by **Jyothis C G**, to the Indian Institute of Technology, Madras, for the award of the degree of **Master of Technology**, is a bona fide record of the research work done by him under our supervision. The contents of this thesis, in full or in parts, have not been submitted to any other Institute or University for the award of any degree or diploma.

Dr. Kaushik Mitra
Research Guide
Assistant Professor
Dept. of Electrical Engineering
IIT-Madras, 600 036

Place: Chennai

Date: 9^{th} May 2019

ACKNOWLEDGEMENTS

I would like to start by thanking my project guide Dr. Kaushik Mitra for guiding and supporting me through the entire duration of my Mtech project. His inputs have been invaluable to the completion of this project. Secondly I would like to thank research scholar Honey Gupta who helped me in solving some of the technical difficulties I faced during my project. I would also like to extend my thanks to my Faculty advisor and HoD Prof. Devendra Jalihal who offered his advice during my entire M.Tech. Last but not the least I would also like to thank my family and friends who offered their advice and support during this project. Finally special thanks to God Almighty for being with me during every single step.

ABSTRACT

With the thermal cameras becoming more affordable there is an increase in the number of potential use cases such as pedestrian detection, visual odometry, leakage detection etc. But these commercial cameras suffer from low resolution, poor signal-to-noise ratio and halo effects. Hence there is a need to use software techniques to improve the quality of thermal camera output. This thesis examines the effectiveness of state-of-the-art RGB SR techniques on thermal image SR resolution. We also try to propose a method to estimate the disparity map from thermal and RGB stereo images. Such a disparity map can be potentially used for all-day vision since the presene of thermal image allows operation in challenging environments such as fog, rain and varying illumination conditions.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

LR	Low resolution
HR	High resolution
SR	Super resolution
TIR	Thermal Infrared
NIR	NIR
GT	Ground Truth
RCAN	Residual Channel Attention Network
IITM	Indian Institute of Technology, Madras

CHAPTER 1

Introduction

Accurate and high quality thermal images are of utmost importance in several applications such as pedestrian detection, surveilance, military , fire detection, gas detection etc. These applications cannot utlise the conventional RGB camera as they cannot be used in challenging conditions such as in the presence of smoke or fog and in dark environments. Thermal cameras can overcome these problems as they capture the infrared radiation emitted by the object as compared to the RGB camera which captures the light reflected off an object. Thermal cameras are manufactured to capture temperature information either in the mid-wavelength region (3 - 8 μm) or in the long-wavelength region (8 - 15 μm).

Although thermal imaging has all these advantages the detector arrays used are really increase the manufacturing cost and thereby the cost of the camera. This severely limits the use of high quality thermal camera to specific use cases such as military etc. Commercially available cameras which are low cost compared to the higher end ones suffer from low signal-to-noise ratios, halo effects etc. Hence the images from these low cost detectors cannot be supplied directly as input to the other algorithms. One possible solution is to use software based solutions to improve the resolution of these low resolution images. With the advent of deep learning, several solutions have been proposed which needs only a single image to achieve the same result.

Most of the works on thermal image super resolution uses the same architecture as the RGB super resolution techniques. But the amount of literature in the field of thermal imaging is smaller compared to that of RGB SR. In this study we will try to use the state-of-the-art RGB SR algorithms for thermal image SR and to determine the optimal training domain. We also examine the efectiveness of this algorithm to improve the output of low cost thermal detectors such as FLIR AX8.

Thermal images also proved to be important in estimating disparity maps and in 3D reconstruction of environments with varying illumination and night time. Although ordinary stereo matching algorithms are robust enough to give good results in most

conditions it cannot be used in the special cases mentioned above. Most of the stereo matching algorithms consider a rectified pair of images as input. But cross-modal stereo matching is difficult due to the different appearance of the objects in the two cases and the low mutual information between them. This eliminates the possibility of using patch by patch matching to estimate disparity and restricts the methods to those based on feature based matching. Here we will try to use deep learning based approaches to estimate disparity between rectified thermal and RGB images and how the RGB based stereo matching techniques can be adapted to this particular problem.

The organization of the thesis is as follows. The second chapter gives a brief overview on the infrared sensor. It also explains the operation differences between the low and high cost thermal detectors. Third chapter explains the need for thermal image super resolution and some of the earlier works in this field. It also gives a overview on the current state-of-the-art SR method and its architecture. Fourth chapter explains the need for cross-modal stereo matching and some of the methods which have been proposed to perform Thermal-RGB stereo matching. Fifth chapter shows the experimental results obtained and sixth chapter lists the conclusions which are drawn on the basis of the results.

CHAPTER 2

Thermal Imaging

2.1 Infrared Sensors: Theory

A thermographic camera (also called an infrared camera or thermal imaging camera or infrared thermography) is a device that forms a heat zone image using infrared radiation. These cameras can capture wavelengths as long as 14 μm. There use is called thermography.

Although IR radiations are invisible to the human eye, IR cameras are able to detect them. Their operation differs from a digital camera operating on VS in that the classical Charge Coupled Device (CCD) is replaced by a Focal Plane Array (FPA), which is made of materials and alloys sensitive to IR wavelengths.

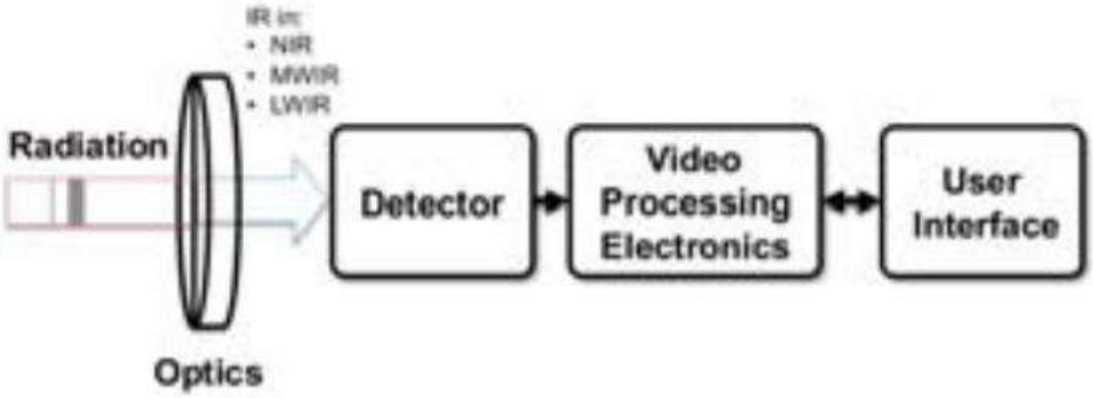


Figure 2.1: Simplified block diagram of Thermal camera

The main components of a thermal camera are shown in Fig 2.1 and it consists of a lens which focuses the incoming radiation onto the detector, an electronic element which converts the opto-electronic signals into images and a software part that interfaces between the came and the user.

According to the working principle the FPA can be classified into the following two categories: (i) photon detectors and (ii) thermal detectors. The former class corresponds to detectors in which radiation absorption happens through interaction with photons. The changed electrical distribution of the material gives rise to electrical signals. These electrical variations are used to estimate the amount of incident optical power.

These detectors require cryogenic cooling. Hence semi-conductor based infrared sensors are heavy, expensive and inconvenient for applications like ADAS.

In thermal detectors the incident radiation causes a change in the temperature or other physical property of the semiconductor material. This variation is used to generate an electrical output proportional to the incident radiation. A typical low cost thermal detector can make use of bolometer which can convert the incoming photon flux into heat thereby changing the electric resistance of the detector element. Currently thermal detectors are available for commercial use as opposed to photon detectors which are restricted to military applications.

These thermal detectors do not require cooling. Despite these advantages photon detectors were popularly believed to be have more speed and higher wavelength selectivity as compared to the thermal counterparts. Advances in the micro miniaturization in the 90's allowed arrays of bolometers or thermal detectors. This compensated for the moderate sensitivity and low frame rate of the thermal detectors. Large arrays also offered high quality imagery and better response time and eventually the manufacturing cost also dropped.

Spectral band	Spectral wavelength (μm)
Visible	0.4 - 0.7
Near Infrared (NIR)	0.78 - 1.0
Short-Wave Infrared (SIR)	1 - 3
Mid-Wave Infrared (MIR)	3 - 5
Long-Wave Infrared (LWIR)	8 - 12

Table 2.1: General spectral bands based on atmospheric transmission and sensor technology

2.2 Applications

This sections give an overview about some of the major application areas of thermography.

- **Inspections of Mechanical Components:**
  Infrared cameras can safely inspect mechanical systems from various industries to detect issues before they become major problems. Some of these include:
  - Finding air leaks and clogged condenser tubes in refrigeration systems
  - Locate and identify overheating bearings, excessive oil temperatures in pumps, compressors, fans, and blowers

- **Aerospace Industry:**
  Aerospace sets the greatest demands on Infrared camera systems due to the high safety and material requirements presented. Aerospace firms can use thermography to test active heat flows on new composite materials to ensure the next generation of lighter, more fuel-efficient aircraft remain as safe as today's models.

- **Electronics and Electrical Industry:**
  Not only does it prevent humans from having direct contact with these systems and circuits, testing and detection can be conducted without interrupting the flow of power. Manufacturing industries can also benefit from electrical thermography to monitor possible overheating, keep a close eye on tank levels, process line inspections, and even assess the condition of circuit boards.

2.3 Thermal Cameras

Two thermal cameras were used for collecting the data used in this project. This section gives a brief explanation about these cameras:

- **FLIR E95:**
  E95 cameras offer superior sensitivity and a true 42° field of view in a user-friendly, handheld platform. These cameras can detect even subtle indications of building deficiencies and moisture intrusion.
  Specifications:
  - Focal length: 17 mm
  - IR Resolution: 464 x 348 pixels
  - Spectral Range: 7.5 - 14 μm

- **FLIR AX8:**
  FLIR AX8 is a thermal sensor with imaging capabilities. Combining thermal and visual cameras in a small, affordable package, the AX8 provides continuous temperature monitoring and alarming for of critical electrical and mechanical equipment.
  Specifications:
  - IR Resolution: 80x60 pixels
  - Spectral range: 7.5 - 13 μm
  - Thermal Sensitivity: < 0.10°C @ +30 ° C (+86 °F) / 100 mK
  - Object Temperature Range: -10°C to 150 °C

(a) FLIR AX8 (b) FLIR E95

Figure 2.2: FLIR Thermal cameras

CHAPTER 3

Single Image Thermal Super resolution

3.1 Need for Thermal Super resolution

All objects emit IR radiation depending on their temperature and emissivity. Since thermal cameras are designed to capture these IR radiations if offers serious advantages over visible cameras in difficult environments such as rain, smoke and fog.

Still IR cameras have its own set of disadvantages. Some of these are listed below:

- Climate, air currents and the presence of other radiation sources can produce a confusing picture
- High quality thermal cameras are expensive due to the high cost of the sensor element.
- Detector arrays are expensive and its difficult to manufacture in large sizes. Hence the low cost detectors produce images with pixel spacing significantly lower than the underlying image resolution.

To summarise the thermal cameras available for commercial purposes suffers from low signal-to-noise ratios, blurring and halo effects. This severely limits their applicability in practical situations.

One possible solution to overcome the resolution problems posed by the low cost thermal detector is to super resolve the images prior to processing. These methods are significantly cheaper than the high resolution cameras. Furthermore the existing LR techniques can be used.

3.2 Related works

Single image super-resolution algorithms focuses on generating a high resolution image from a single low resolution image. These algorithms can be broadly classified into 3 major categories (i) interpolation based (ii) reconstruction based and (iii) learning based methods.The interpolation based technique is the simplest of the three but it fails to produce a good quality output and the details are often vague. The reconstruction based method utilises prior information and several existing degradation models such as down-sampling, motion blurring and optical distortion to reconstruct the high resolution image. The third method which focuses on learning based techniques utilises a large data-set of low quality - high quality image pairs to achieve super resolution.

With the advent of deep learning there have significant improvements in the field of single image super resolution. Convolutional Neural networks based networks [[3], [4], [9], [12]] have achieved significant improvements over single image super resolution techniques. Don et al. [3] proposed one of the earliest works named as SRCNN which used a 3 layer CNN to achieve SR. Later Kim et al. increased the depth of the network to 20 with VDSR [9] which showed considerable over SRCNN. This showed that the network depth is an important factor in visual recognition tasks. Ledig et al. [10] introduced ResNet [6] to construct a deeper network with perceptual losses [8] and generative adversarial network (GAN) [5] for photo-realistic SR. Later Lim et al. [12] built a very wide network EDSR and a very deep one MDSR by using simplified residual blocks. The performance improvements of EDSR and MDSR indicates that the depth of the network is an important factor in single image SR tasks.

However the amount of literature in the field of single image thermal super resolution have been fairly sparse. Choi et al. [2] proposed one of the earliest works in thermal super resolution. The paper used a simple CNN network with bicubic interpolated version of the LR image as input. It consisted of 3 convolutional layers with spatial padding to preserve the resolution of the image. His work was followed by Lee et al. [11] in 2017 which proposed a Res-Net based architecture to achieve the result. The network is composed of multiple layers for feature extraction and matching followed by a single layer for reconstruction. Compared to the previous works the focus of the network was on predicting the high frequency details in the form of residuals. This was based on the idea that the low frequency information in the LR and GT HR

image was highly correlated thereby eliminating the need for computing low frequency information. The paper also performed a performed a comparison to determine the best training domain. Comparsions were done between thermal infrared, NIR and the four RGB domains namely gray, lightness, intensity and brightness. Based on their analysis brightness domain was observed to give better results.

Most of these papers were trained on publicly available pedestrain datasets like KAIST which severly affected the generalisation capability of the networks. One possible solution to this problem was the idea of transfer learning. Sun et al. [15] in 2018 used the idea of transfer learning to achieve better results. The network is initially trained on a large database of natural RGB images and then fine-tuned by a relatively smaller dataset of thermal images which was found to yield better outputs. Moreover it eliminated the need for upscaling the LR image prior to giving it as input by the use of deconvolutional layers thereby bringing the computational cost as well.

To summarise the super resolution algorithms which have developed for RGB images can be used without much modification to thermal images. The current state-of-the art algorithm in RGB single image super resolution was proposed by Zhang et al. [19] in 2018 which will be explained in more detail in the next section.

3.3 Image Super-Resolution Using Very Deep Residual Channel Attention Networks: A Brief Overview

This section gives a brief overview about the super resolution technique proposed by Zhang et al. in [19]. Although this method is intended for use with RGB images as input recent works in thermal SR literature indicates that the RGB SR algorithms can be extended to thermal SR without any modification in the architecture.

Most recent CNN-based methods deals treat channel-wise features with the same importance which limits the flexibility of the network to deal with different features. For instance the low frequency information can be directly bypassed to the output without any loss of optimality thereby saving precious computational time. Also depth of the network is an important factor in the case of SR. But as the depth increases it becomes more difficult to train and with diminishing returns only. This method overcomes the

above problems with the following major contributions:

- A residual in residual structure which helps to train very deep networks. The presence of long and short skip connection allows the network to bypass low frequency information.
- Channel attention mechanism to rescale features taking into consideration the interdependence between features

The Residual Channel Attention Network (RCAN) consists of four parts: shallow feature extraction layer, residual-in-residual deep feature extraction layer, upscale module and reconstruction part.

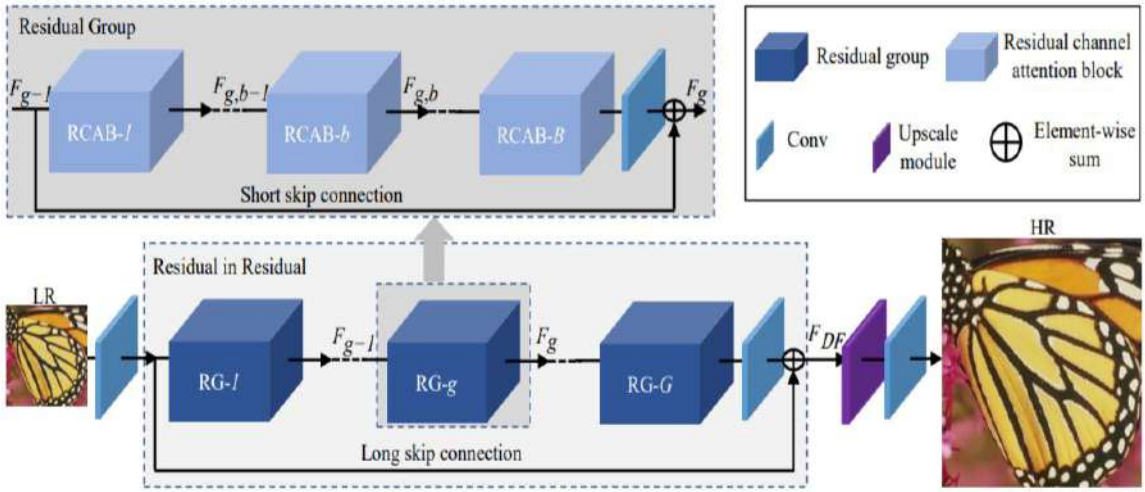


Figure 3.1: RCAN Architecture

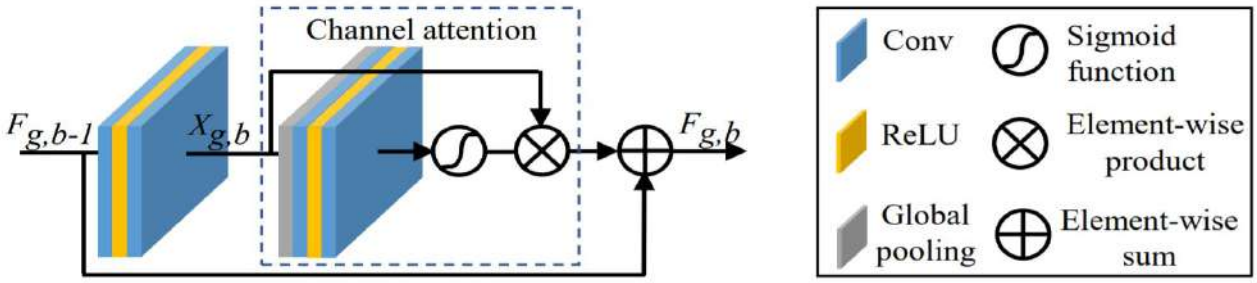


Figure 3.2: Residual Channel Attention Block

Given a training set $\{I^i_{LR}, H^i_{LR}\}^N_{i=1}$ which contains N LR inputs and their HR counterparts. The objective is to minimise the L_1 loss function between the RCAN reconstruction and the ground truth.

$$L(\Theta) = \frac{1}{N}\sum_{i=1}^{N} ||H_{RCAN}(I^i_{LR}) - I^i_{HR}||_1$$

where Θ denotes the parameter set of the network.

CHAPTER 4

Thermal - RGB Stereo Matching

4.1 Motivation

Recent works in the field of robotics, material classification and pedestrian detection have shown that multiple modalities can perform much better than single sensor based systems. Moreover there has been an increase in the number of commercial thermal cameras (FLIR One Pro which can be used with a smartphone) which offer cheaper IR imaging solutions at the cost of resolution and signal-to-noise ratio. In most of the cases the thermal sensors are also coupled with an RGB camera which are of high resolution as compared to their thermal counterparts. This Thermal - RGB images form a stereo pair which can be used to reconstruct disparity maps as well as depths in challenging environments such as during night-time or in the presence of fog. Hence they have wide applications in the field of self-driving cars, fire and rescue, military etc.

This dual-stereo set up can also help overcome some of the common stereo reconstruction failure cases such as reflections - which will be more pronounced in the case of visible images but not in thermal images as they are sensitive only to the IR radiation coming from the object. However thermal - rgb stereo matching is a very challenging problem owing to the lack of texture in thermal images and the low mutual information between the two modalities. Here we will try to leverage some of the newer deep learning based stereo matching techniques to perform cross-modal stereo matching.

4.2 Related works

Most of the works on cross-modal stereo matching has been related to thermal images in the near infrared region. This is less challenging as compared to thermal images in the LWIR region. The wavelength of NIR is close enough to red colour and hence therma images captured in the NIR range posess a high degree of correlation with the

red channel of their RGB stereo pair. Most of the exising technique make use of this idea to achieve results. However the same cannot be extended directly to the LWIR range since the wavelength is too large and hence the appearance of the images are totally different.

This section will cover some of the latest works in NIR stereo matching as well as RGB stereo matching since there were no papers in the RGB-LWIR category. Some of the ideas from these papers will be utilised in the process to come up with an acceptable solution for this problem.

In stereo estimation, [18] used CNN to compute the matching cost between two image patches. They utilised a Siamese network which takes the same sized patches from the left and right images with a few fully-connected layers to predict the matching cost. The model was trained to minimize a binary cross-entropy loss. In similar spirit to [18], [17] investigated different CNN based architectures for comparing image patches. They found concatenating left and right image patches as different channels to be better.

Luo et al. in [13] proposed a deep learning architecture for stereo matching between RGB images which utilises a Siamese network to estimate the disparity. The major contribution of this paper was the reduced computational time obtained by replacing the concatenation layer in the previous architectures with a dot product layer which was used to compute the correlation between the output of the Siamese representations.

Ryu et al.in [14] defined a new image transform called 'Local Area Transform' which was consistent even with non-linearly deformed image pairs induced by different modality conditions. This involved transforming the image into a local domain which is invariant under nonlinear intensity deformations such as radiometric, photometric and spectral deformations. The method also showed improvement in cross-spectral template matching and cross-spectral feature matching with RGB-NIR data.

Zhi et al.in [20] proposed an architecture for simulatenously transforming images across spectral bands and to estimate disparity between RGB-NIR stereo pairs. A material aware loss term is also introduced into the loss function to account for regions with unreliable matching such as light sources, glass windshields and glossy surfaces.

4.3 Proposed methods

This section explains about the three different approaches which have been adopted to tackle the problem of Thermal-RGB stereo matching. Most of these algorithms are in the preliminary stage and hence the outputs are not optimal. However they do exhibit promise and the methods can be adapted to the specific constraints posed by the different modalities with appropriate modification.

4.3.1 Naive approach using Siamese network

In this method a deep learning based stereo matching approached proposed in [13] is used. This method is intended to be used with RGB stereo pairs. But in this case the dataset is replaced by the rectified thermal-RGB stereo pairs from the CATS dataset. The network utilises a Siamese architecture with shared weights and a dot product layer with the two representations of the Siamese network as input.

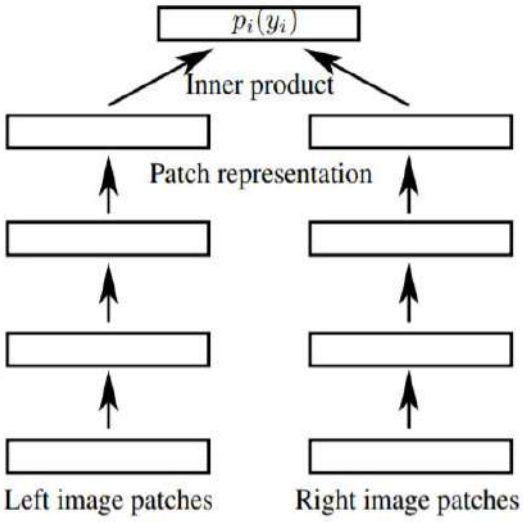


Figure 4.1: Stereo Matching Network [18]

The network learns a probability distribution over all disparities with a smooth target distribution. During training binary cross entropy loss is minimised. All the parameters of the network are the same as that proposed in the paper.

4.3.2 Preprocessing the images with Local Area Transform

This method uses the same architecture as in the previous section. The only difference is that the image pairs are converted into a local area domain which is invariant under non-linear intensity deformations especially photometric, spectral deformations etc.

The pseudo code for Local area transform given in [14] is as follows.

Algorithm Local area Transform

Input: input image **I**, half-window size l

Output: local area transformed image **Y**

Notation: integral histogram **H**, local histogram **h**

/* integral histogram computation */

for each pixel **p** = (x, y) **do**

H'(x, y) $\leftarrow$ **H**'(x, y - 1) + **I**(x, y)

end

for each pixel **p** = (x, y) **do**

H(x, y) $\leftarrow$ **H**(x - 1, y) + **H**'(x, y)

end

/* local histogram computation */

for each pixel **p** = (x, y) **do**

h(x,y) $\leftarrow$ **H**(x + l,y + l) + **H**(x - l,y - l) - **H**(x - l,y + l) - **H**(x - l,y + l)

end

/* local area computation */

R_k = **I(p)** - r

R_l = **I(p)** + r **for each** pixel **p** = (x, y) **do**

Y(x,y) $\leftarrow \sum_{b\epsilon(R_k,R_l)} \omega(b)$ x **h**(x,y,b)

end

Parameters r and σ control the interval of integration and the degree of Gaussian smoothing of the histogram, respectively.

4.3.3 Converting the thermal image into visible spectrum

Berg et al. in [1] proposed a network for spectral transformation of thermal images into the visible spectrum. This architecture will be used to convert the thermal image in each Thermal-RGB stereo pair. These converted thermal images and RGB image pairs will be used as the training data for the stereo matching method in 3.3.1

CHAPTER 5

Experiments and Results

This section gives a brief explanation about the different datasets used in this project

5.1 Datasets used

- **KAIST Multispectral Pedestrian dataset** This dataset [7] contains aligned thermal and RGB image captured in both in day and light environments and is used commonly used as a benchmark for many thermal imaging related computer vision tasks. We will be using this dataset to compare the performance of different thermal SR algorithms.

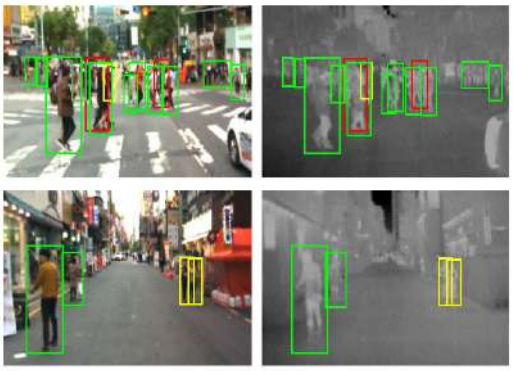


Figure 5.1: Images from KAIST dataset

- **FLIR ADAS dataset** This dataset features an initial set of more than 14,000 annotated summer driving thermal images captured at day and night and their corresponding RGB imagery for reference. FLIR ADAS was used for training data for the thermal RCAN model which was trained from the scratch exclusively on theral images.

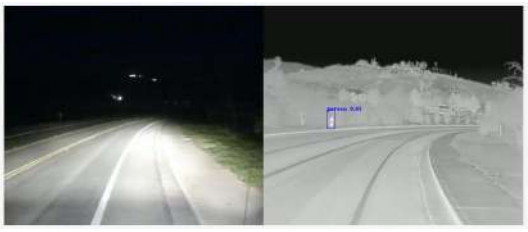


Figure 5.2: Image from FLIR ADAS dataset

- **High resolution thermal dataset** This dataset consists of generator set images captured using the FLIR E95 high resolution thermal camera and was used for fine-tuning the pre-trained RCAN model.
- **Color and Thermal Stereo datasaset** Color and Thermal Stereo (CATS) dataset [16] contains rectified RGB and thermal stereo pairs along with the depth information captured using LIDAR. This dataset was used as the primary training data for Thermal-RGB stereo matching problem.

Figure 5.3: Images from CATS dataset

5.2 Thermal Image Super resolution

5.2.1 Comparison with existing methods

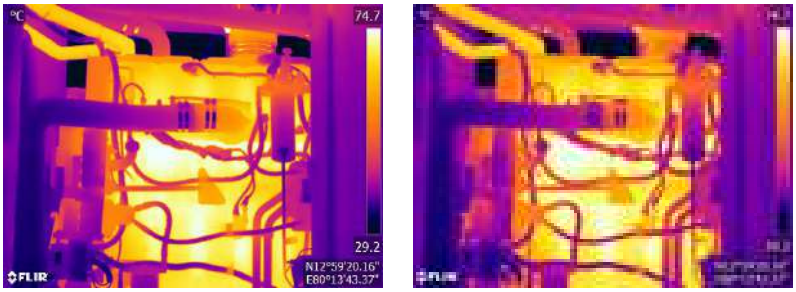
Here we will be comparing the performance of RCAN pre-trained model with the algorithm proposed in [2]. Since they have provided results only for x2 and x3 case we will be restricting our attention to those cases.

	x2	x2	x2	x3	x3	x3
	Bicubic	TEN	RCAN	Bicubic	TEN	RCAN
PSNR (dB)	42.97	44.212	46.234	39.67	40.4814	42.814

Table 5.1: Comparison of RCAN pre-trained with existing methods

5.2.2 Comparison between Thermal trained RCAN and RCAN pre-trained model

This section explains the experiments done to determine the optimal training domain for thermal image super resolution. The comparison was performed between RCAN trained



(a) Original HR Image

(b) Original image downscaled by x4

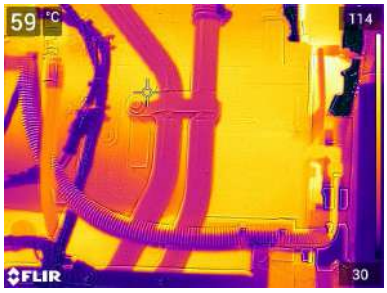
(c) Pre-trained RCAN model output

(d) Thermal trained RCAN output

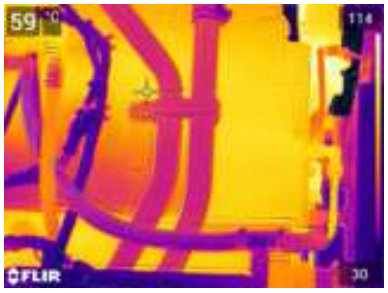
Figure 5.4: RCAN pre-trained and Thermal trained RCAN model comparison

5.2.3 Comparison between RCAN pre-trained and RCAN fine-tuned models

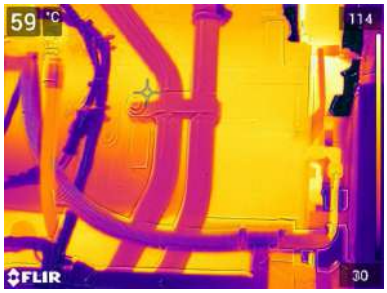
This section explains the experiments done to determine the optimal training domain for thermal image super resolution. The comparison was performed between RCAN pre-trained and RCAN fine-tuned using gen-set images.



(a) Original HR Image



(b) Original image downscaled by x4



(c) Pre-trained RCAN model output

(d) RCAN fine-tuned model output

Figure 5.5: RCAN pre-trained and RCAN fine-tuned model comparison: Result 1

(a) Original HR Image

(b) Original image downscaled by x4

(c) Pre-trained RCAN model output

(d) RCAN fine-tuned model output

Figure 5.6: RCAN pre-trained and RCAN fine-tuned model comparison: Result 2

The first result indicates that the fine-tuned RCAN model gives better output as compared to pre-trained model.

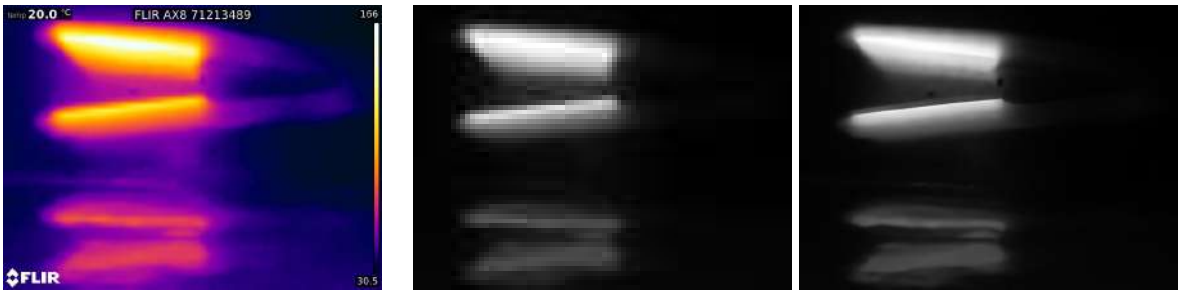
5.2.4 Results on FLIR AX8 Thermal camera

The native IR resolution of FLIR AX8 is 80x60 pixels and cannot be provided as input to other post processing algorithms. Hence it is internally upscaled by some interpolation algorithm which is not publicly available.

In this section we will compare how the RCAN pre-trained performs in comparison to FLIR's proprietary algorith, Since no GT is available we will be examining the perceptual quality rather than metrics like PSNR and SSIM in this case. The process involves extracting the 80x60 from the 640x480 size IR output of the camera and then super resolving using the pre-trained RCAN model.

(a) FLIR AX8 Output (640x480) (b) Raw Thermal Image (80x60) (c) Super resolved AX8

Figure 5.7: Genset image captured using FLIR AX8



(a) FLIR AX8 Output (640x480) (b) Raw Thermal Image (80x60) (c) Super resolved AX8

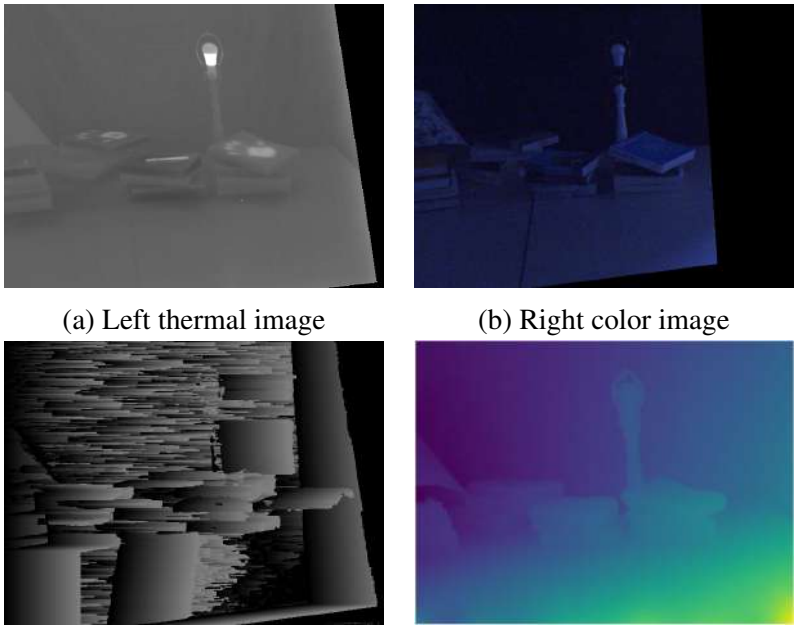
Figure 5.8: Image of hair dryer captured using FLIR AX8

These results indicate that the super resolved image obtained by using the pre-trained RCAN model achieves much sharper images as compared to FLIR's inbult SR technique.

5.3 Thermal - RGB Stereo Matching

5.3.1 Using Siamese network

Here we use the RGB stereo matching technique specified in [13] to estimate the disparity between the left thermal image and right color image. The results are not satisfactory due to the difference in image modallities.



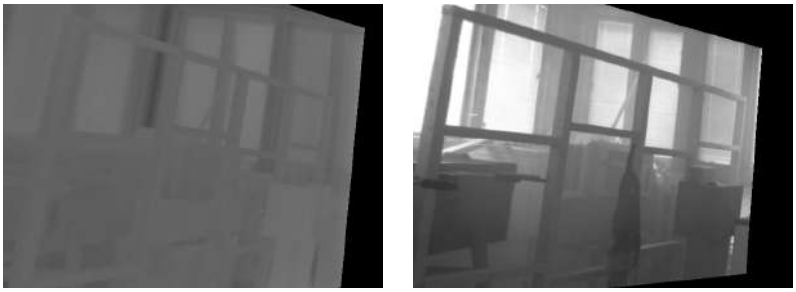
(a) Left thermal image (b) Right color image

(c) Output Disparity map (d) GT Disparity map

Figure 5.9: Disparity estimation using Siamese Network

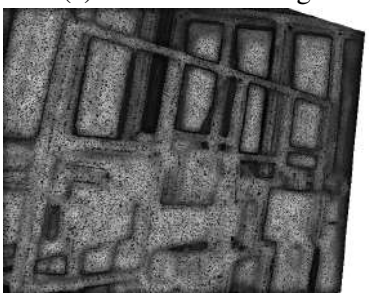
5.3.2 Preprocessing using Local Area Transform

In this technique the image pairs are converted using Local area transform prior to training the Siamese network.

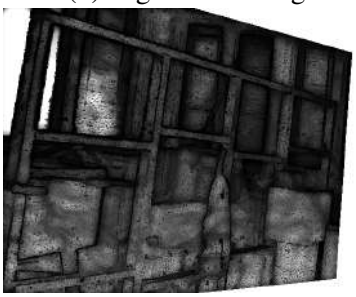


(a) Left thermal image

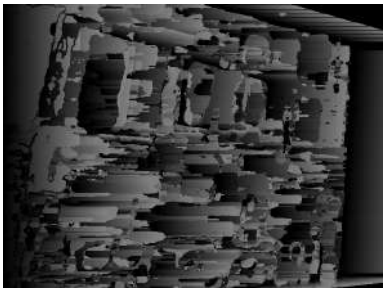
(b) Right color image



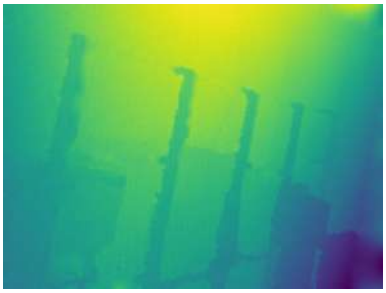
(c) LATS converted thermal image



(d) LATS converted color image



(e) Output Disparity map



(f) GT Disparity map

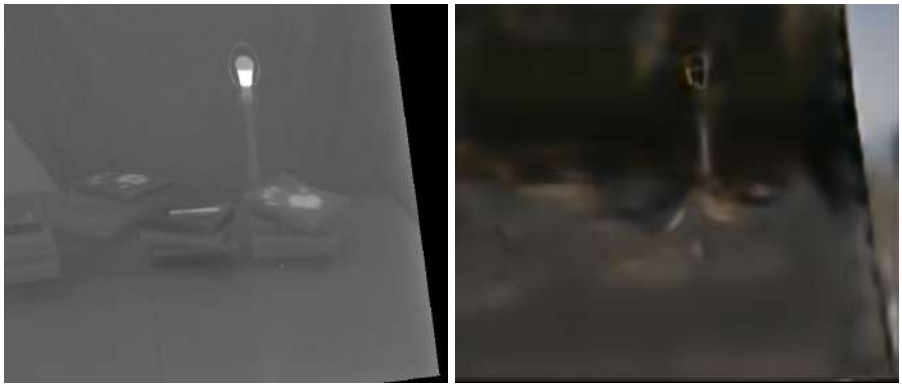
Figure 5.10: Disparity estimation on LAT preprocessed image

The images in the first row are the recitified image pair taken from the Color and Thrermal Stereo dataset. The second row corresponds to their LATS processed counterparts. The bottom row shows the ground truth disparity and the estimated disparity.

LAT shows improvement in the case of RGB-NIR stereo matching problems. However the same cannot be said regarding the LWIR images. It may be due to the large wavelength difference between NIR and LWIR.

5.3.3 Converting thermal images to visible spectrum

In this section we have used the pre-trained model of [1] in converting the left thermal image into the visible spectrum. The pre-trained model is trained using the aligned thermal-RGB pairs from the KAIST dataset.



(a) Thermal image (b) Converted visible spectrum image

Figure 5.11: Conversion of thermal to visible image

There is a loss of information in the converted thermal image and it cannotbe used for RGB stereo matching. This indicates the network trained on KAIST dataset is not a good choice for operating on the CATS dataset.

CHAPTER 6

Conclusion

From the experiments done it is observed that RGB super resolution algorithms can be used to solve the low resolution problem by the commercial thermal cameras. In most cases the optimal training domain is observed to be natural images. This may be attributed to the unavailability of high resolution thermal datasets with sufficiently diverse images. In cases where a high quality thermal dataset is available it maybe more advantageous to fine-tune the pre-trained model with the thermal images. The fine-tuned model is observed to give better results than the pre-trained model. So there is a trade-off between the cost incurred in creating a high resolution dataset and the improvement in results. But according to our study pre-trained would suffice in most scenarios.

Thermal-RGB stereo matching can improve the 3D construction in challenging environments such as in the presence of fog and rain. Unfortunately none of the methods proposed in the thesis were able to give satisfactory results. This may be due to the large difference in modalities and low mutual information between the images. The third method which involved converting thermal to the visible spectrum failed because the network was trained on pedestrian data and was unable to adapt to the images in the CATS dataset. This can be improved by creating a aligned thermal-RGB dataset with more diverse scenes. Adding an appropriate regularization term may also help to improve the result.

CHAPTER 7

Future work

In the case of thermal-RGB stereo matching none of the proposed methods could give satisfactory results. This may be due to the large difference in modality. This may be improved by including a graph-cut based regularization term in the network loss function. Also an unsupervised depth estimation which uses a combination of the left-right consistency idea and perceptual loss is another future work. Regarding the spectral transformation idea it is necessary to have an aligned thermal-RGB dataset which contains more diverse scenes. For thermal super-resolution can be improved by fine-tuning the pre-trained model with a larger dataset with more diverse scenes.

REFERENCES

[1] **Berg, A.**, **J. Ahlberg**, and **M. Felsberg**, Generating visible spectrum images from thermal infrared. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 2018.

[2] **Choi, Y.**, **N. Kim**, **S. Hwang**, and **I. S. Kweon**, Thermal image enhancement using convolutional neural network. *In 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2016.

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

[4] **Dong, C.**, **C. C. Loy**, and **X. Tang**, Accelerating the super-resolution convolutional neural network. *In European conference on computer vision*. Springer, 2016.

[5] **Goodfellow, I.**, **J. Pouget-Abadie**, **M. Mirza**, **B. Xu**, **D. Warde-Farley**, **S. Ozair**, **A. Courville**, and **Y. Bengio**, Generative adversarial nets. *In Advances in neural information processing systems*. 2014.

[6] **He, K.**, **X. Zhang**, **S. Ren**, and **J. Sun**, Deep residual learning for image recognition. *In Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.

[7] **Hwang, S.**, **J. Park**, **N. Kim**, **Y. Choi**, and **I. So Kweon**, Multispectral pedestrian detection: Benchmark dataset and baseline. *In Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015.

[8] **Johnson, J.**, **A. Alahi**, and **L. Fei-Fei**, Perceptual losses for real-time style transfer and super-resolution. *In European conference on computer vision*. Springer, 2016.

[9] **Kim, J.**, **J. Kwon Lee**, and **K. Mu Lee**, Accurate image super-resolution using very deep convolutional networks. *In Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.

[10] **Ledig, C.**, **L. Theis**, **F. Huszár**, **J. Caballero**, **A. Cunningham**, **A. Acosta**, **A. Aitken**, **A. Tejani**, **J. Totz**, **Z. Wang**, *et al.*, Photo-realistic single image super-resolution using a generative adversarial network. *In Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.

[11] **Lee, K.**, **J. Lee**, **J. Lee**, **S. Hwang**, and **S. Lee** (2017). Brightness-based convolutional neural network for thermal image enhancement. *IEEE Access*, **5**, 26867–26879.

[12] **Lim, B.**, **S. Son**, **H. Kim**, **S. Nah**, and **K. Mu Lee**, Enhanced deep residual networks for single image super-resolution. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 2017.

[13] **Luo, W.**, **A. G. Schwing**, and **R. Urtasun**, Efficient deep learning for stereo matching. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016.

[14] **Ryu, S.**, **S. Kim**, and **K. Sohn** (2017). Lat: Local area transform for cross modal correspondence matching. *Pattern Recognition*, **63**, 218–228.

[15] **Sun, C.**, **J. Lv**, **J. Li**, and **R. Qiu** (2018). A rapid and accurate infrared image super-resolution method based on zoom mechanism. *Infrared Physics & Technology*, **88**, 228–238.

[16] **Treible, W.**, **P. Saponaro**, **S. Sorensen**, **A. Kolagunda**, **M. O'Neal**, **B. Phelan**, **K. Sherbondy**, and **C. Kambhamettu**, Cats: A color and thermal stereo benchmark. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017.

[17] **Zagoruyko, S.** and **N. Komodakis**, Learning to compare image patches via convolutional neural networks. *In Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015.

[18] **Zbontar, J.** and **Y. LeCun**, Computing the stereo matching cost with a convolutional neural network. *In Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015.

[19] **Zhang, Y.**, **K. Li**, **K. Li**, **L. Wang**, **B. Zhong**, and **Y. Fu**, Image super-resolution using very deep residual channel attention networks. *In Proceedings of the European Conference on Computer Vision (ECCV)*. 2018.

[20] **Zhi, T.**, **B. R. Pires**, **M. Hebert**, and **S. G. Narasimhan**, Deep material-aware cross-spectral stereo matching. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018.