Mälardalen University
School of Innovation Design and Engineering
Västerås, Sweden

Thesis for the Degree of Master of Science in Engineering -
Dependable Systems 30.0 credits

TOWARDS RELIABLE COMPUTER VISION IN AVIATION: AN EVALUATION OF SENSOR FUSION AND QUALITY ASSESSMENT

Björklund Emil
contact@emilbjorklund.se
Hjorth Johan
johan@famhjorth.se

Examiner: Mikael Ekström
Mälardalen University, Västerås, Sweden

Supervisors: Masoud Daneshtalab
Mälardalen University, Västerås, Sweden

Company supervisor: Per-Olof Jacobson
SAAB AB, Järfälla, Sweden

June 10, 2020

Acknowledgment

We want to thank SAAB for letting us write our Master Thesis at their company and provide the appropriate equipment for our research. Thank you, Per-Olof Jacobsson, for your engagement in our work and research. It has been a pleasure to have you as our supervisor at SAAB, always providing us with great insights. We also want to express our sincerest gratitude to our supervisor at Mälardalens University, Docent Masoud Daneshtalab. Daneshtalab has provided us with knowledge and inspiration, allowing us to evolve in our work. Thank you, Doctor Håkan Forsberg, for providing us with feedback on the thesis it was greatly appreciated. We want to thank Doctor Martin Ekström, for enabling our academic research by providing us with insights and tools. Sincere thanks to Henrik Falk, for trying to relieve all forms of stress and giving essential insights. To the Dependable Systems class of 2020, at Mälardalen University, thank you for your support and showing interest at our countless presentations. To our friends Najda Vidimlic and Alexandra Levin, thank you for providing us with tools enabling the validation of our results regarding object detection, you are incredible. Furthermore, it has been a great pleasure to work closely with you during this project. A big thank you for all the great conversations and discussions during the days. It would not be the same without you guys. Lastly, to our families, Linnéa, Camilla, Astrid, parents, and siblings, without you, we would not have made it this far. Thank you for your invaluable support!

Abstract

Research conducted in the aviation industry includes two major areas, increased safety and a reduction of the environmental footprint. This thesis investigates the possibilities of increased situation awareness with computer vision in avionics systems. Image fusion methods are evaluated with appropriate pre-processing of three image sensors, one in the visual spectrum and two in the infra-red spectrum. The sensor setup is chosen to cope with the different weather and operational conditions of an aircraft, with a focus on the final approach and landing phases. Extensive image quality assessment metrics derived from a systematic review is applied to provide a precise evaluation of the image quality of the fusion methods. A total of four image fusion methods are evaluated, where two are convolutional network-based, using the networks for feature extraction in the detailed layers. Other approaches with visual saliency maps and sparse representation are also evaluated. With methods implemented in MATLAB, results show that a conventional method implementing a rolling guidance filter for layer separation and visual saliency map provides the best results. The results are further confirmed with a subjective ranking test, where the image quality of the fusion methods is evaluated further.

Keywords— Aviation, Image Fusion, Image Registration, Image Quality Assessment

What a beautiful puzzle

-Simon Anthony

Table of Contents

List of Figures

List of Tables

Acronyms

ADS-B Automatic Dependent Surveillance-Broadcast.

ATC Air Traffic Control.

AWGN Additive White Gaussian Noise.

BEMD Bi-dimensional Empirical Mode Decomposition.

BIMF Bi-dimensional Intrinsic Mode Function.

BRISQUE Blind/Referenceless Image Spatial Quality Evaluator.

CAT Clear Air Turbulence.

CNN Convolutional Neural Network.

CPU Central Processing Unit.

CSR Convolutional Sparse Representation.

DCT Discrete Cosine Transform.

DDR Double Data Rate.

EASA European Union Aviation Safety Agency.

EFVS Enhanced Flight Vision System.

EGPWS Enhanced Ground Proximity Warning System.

EVS Enhanced Vision System.

FAA Federal Aviation Administration.

FABEMD Fast and Adaptive Bi-dimensional Mode Decomposition.

Fast-FMI Fast Feature Mutual Information.

FMI Feature Mutal Information.

FPGA Field Programmable Gate Array.

FR-IQA Full Reference Image Quality Assessment.

GPWS Ground Proximity Warning System.

ICAO International Civil Aviation Organization.

IQA Image Quality Assessment.

IR Infrared.

JPDF Joint Probability Density Function.

LWIR Long-wavelength infrared.

MPDF Marginal Probability Density Functions.

MSD Multi-Scale Decomposition.

MSE Mean Squared Error.

MVG Multi-variate Gaussian.

NIQE Natural Image Quality Evaluator.

NR-IQA No Reference Image Quality Assessment.

NSS Natural Scene Statistics.

PMMW Passive Millimeter Wave.

PSNR Peak Signal-to-Noise Ratio.

Q Quality.

$\mathbf{Q^{AB/F}}$ Edge Preservation Quality Index.

RA Resolution Advisory.

ResNet Residual Network.

RGB Red Green Blue.

RGF Rolling Guidance Filter.

RR-IQA Reduced Reference Image Quality Assessment.

SA Situation Awareness.

SIDWT Shift Invariant Discrete Wavelet Transform.

SR Sparse Representation.

SURF Speeded-Up Robust Features.

SWIR Short-wavelength infrared.

TA Traffic Advisories.

TAWS Terrain Awareness and Warning System.

TCAS Traffic Alert and Collision Avoidance System.

VGG Visual Geometric Group.

VSM Visual Saliency Map.

WLS Weighted Least Squared Optimization.

ZCA Zero-Phase Component Analysis.

1 Introduction

Lack of visual representation of the surroundings due to poor visibility is a significant contributor to fatal accidents within the civil aviation industry [1]. A well known civil aviation disaster is the "Tenerife airport disaster" which took place on the Island of Tenerife, Spain, in 1977. Two Boeing 747 passenger jets collided on the runway resulting in 583 fatalities. According to the official Report (Subsecretaria de aviacion civil), the primary factor leading to the accident is that the captain of KLM Flight 4805 decided to take-off after he heard the Air Traffic Control (ATC)-Clearance on the radio despite the presence of Pan Am Flight 4805 still taxiing on the runway. It was speculated that his decision was a result of stress as the airplane was forced to land at Tenerife and not on Gran Canaria as scheduled. Earlier that day, a bomb detonated at the airport of Las Palmas, resulting in a higher amount of traffic to the Los Rodeos Airport, Tenerife. That day, there was severe fog present in the area, neither airplanes nor control tower had a visual perception of the runway. Hence, the pilots in each respective aircraft were not able to locate each other [2]. With today's computer vision technologies and systems, this accident could have been prevented or limited.

According to Boeing [3], between 2009 and 2018, 49% of all fatal accidents occurred during the final approach and landing phases compared to 12% during takeoff and initial climb. Sensors capable of sensing the environment in dense fog, rain, and reduced lighting conditions are examined within this thesis. The sensors are capable of capturing different wavelengths of the electromagnetic spectrum to provide a more complete understanding of the environment. Furthermore, the ability of combining data gathered from these sensors into one unified output is evaluated. The work performed in this thesis aims to increase Situation Awareness (SA) within aircraft operations according to Endsley's definition for SA [4] and focuses on the final approach and landing phases.

Four image fusion techniques are explored, through Image Quality Assessment (IQA) and to some extent performance. Image fusion aims to combine salient features from different sensors to increase information contained in the output. Fusion methods can be divided into subcategories depending on underlying theories. In this thesis, methods including neural network-, Sparse Representation (SR)-, multi-scale transform- and saliency-based methods are implemented and assessed. However, methods tend to combine different theories in various stages of the fusion process [5]. Difficulties occur trying to fuse multiple images such as differences in resolution, wavelengths of captured electromagnetic radiation and sensor placement [6]. Sensors evaluated in this thesis have different characteristics and are pre-processed with image registration before applying image fusion. Assessing the performance and quality of the systems mentioned above requires extensive metrics and evaluation. For this thesis, models regarding No Reference Image Quality Assessment (NR-IQA) are used in combination with subjective perception to compare and evaluate computer vision configurations and fusion methods.

1.1 Industry-Academia Collaboration

This thesis is a collaboration with SAAB Avionics Systems Business Unit which is part of SAAB's global aircraft and defense group [7]. The department develops electronic components, mechanical components, and software to be used in airplanes, helicopters, and other demanding applications. Therefore, the safety-critical applications of these systems require a high degree of reliability. Clean Sky 2 is a partnership between the European Commission and the aviation industry within the European Union. The Clean Sky 2 research programme aim to develop innovative, cutting-edge technology aimed at reducing CO2, gas emissions and noise levels produced by aircraft [8]. SAAB is part of the Clean Sky 2 program, contributing to the Large Passenger Aircraft Programme as well as the Systems Programme for Avionics Extended Cockpit. This work provides a knowledge base for SAABs work in Clean Sky 2 with the aims to reduce unnecessary time in air with increased SA.

2 Background

When maneuvering an airplane, pilots have several aids at their disposal [9]. These aids ensure that the pilot has a correct perception of the airplane's surroundings and the correctness of data fed to the pilot is of high importance. Aids available to pilots are used as a way to increase safety by informing pilots of possible collisions and includes information about other aircraft, the elevation of the terrain, or the position of aerial or ground-based objects. Automatic Dependent Surveillance-Broadcast (ADS-B), Traffic Alert and Collision Avoidance System (TCAS) and Terrain Awareness and Warning System (TAWS) are examples of aids available to pilots an example illustrating these systems can be seen in Figure 1.

2.1 Supporting Systems

One of the pilot aids available, ADS-B is a safety system used to improve the SA for Air Traffic Control (ATC) and avoid collisions. ADS-B is designed to operate in environments without radar coverage. ADS-B includes two major parts, ADS-B OUT and ADS-B IN. ADS-B OUT is the broadcasting part of ADS-B and is responsible for sending messages in a periodic manner. Messages sent by ADS-B OUT is received by ATC and other aircraft ADS-B IN systems. The messages contain information about the horizontal and vertical position of aircraft along with aircraft identification [10].

Another safety system is TCAS. This system collects information from other aircraft ATC transponders to identify potential threats and hazards, in essence, collision threats. The system creates a safe zone (volume), based on bearing, altitude and response times between aircraft, in the airspace.

TCAS provides two types of advisories with appropriate aural and visual warnings. Traffic Advisories (TA) indicates the relative position of an aircraft intruding the safe zone. This advisory activates when an aircraft is approximately 20 - 48 seconds from a collision. TA, requires an operational Mode S transponder, capable of transmitting aircraft addresses in a 24-bit format or an ATCRBS transponder to identify an aircraft. Resolution Advisory (RA) require Mode S or Mode C paired with mode A transponders to provide warnings. The Mode C transponder is capable of transmitting pressure altitude of the aircraft and when paired with Mode A transmits altitude and identification messages. RA activates when an intruder is approximately 15-35 seconds from a collision and provides computer calculated vertical maneuvers to increase separation to prevent the collision [11]. Both TA and RA is issued by the TCAS computer. When RA is activated, the pilot is required to immediately respond to the commands from RA and disregard commands from the ATC controller. Lastly, aircraft are often equipped with TAWS, a broad term in civil aviation including the Ground Proximity Warning System (GPWS) and the Enhanced Ground Proximity Warning System (EGPWS). The fundamental purpose of the system is to alert when the aircraft is close to terrain. GPWS collects direct measurements from sensors, commonly a radio-altimeter, to determine height above ground. With height data and aircraft speed, different warnings are triggered. EGPWS utilizes the GPWS sensors together with information from databases to determine risks with controlled flight into terrain along with advanced terrain mapping on visual displays [12]. The improvement in such supporting systems is an ongoing process where development is performed continuously. This paper will evaluate vision sensor technologies to be used as a method to increase pilot awareness.

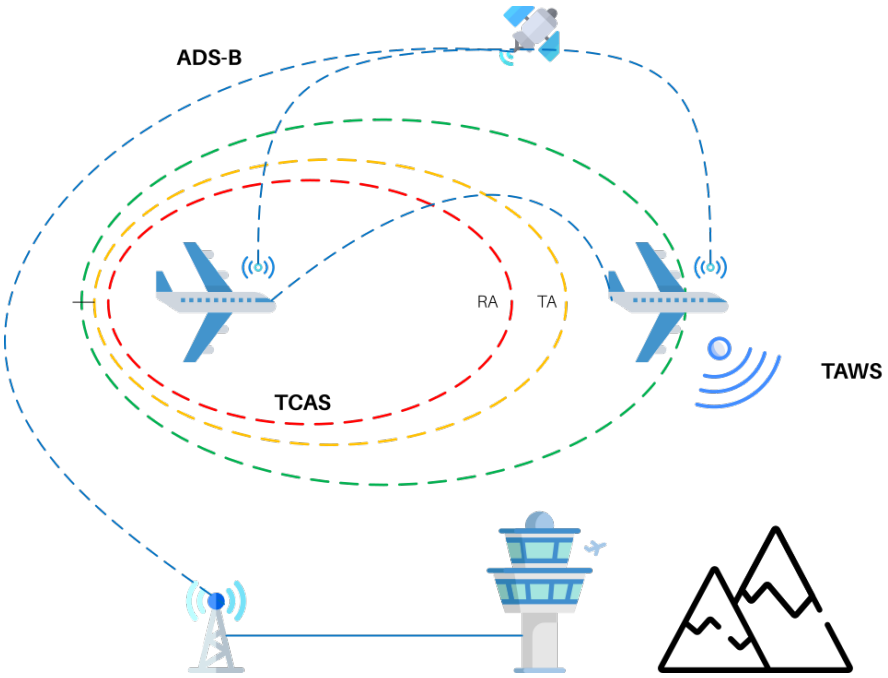


Figure 1: The figure illustrates a subset of the standard systems of large passenger aircraft. TAWS responsible for terrain awareness, TCAS accountable for checking the airspace for other aircraft as well as ADS-B sending informational messages.

2.2 Vision Technologies in Aviation

New technologies made available from other industries may be of interest to airplane manufacturers. However, a strict development process of avionics systems may prove that such technologies are impossible to implement. The stringent development process in the aircraft industry is imposed by the Federal Aviation Administration (FAA), European Union Aviation Safety Agency (EASA), and manufacturers. If computer vision technologies are to be implemented within aircraft, a rigorous process of testing and evaluation is needed to prove whether new technologies are suitable. Autonomous vehicles are an example of a growing market where new technologies, such as sensor fusion, are researched and developed [13].

By merging outputs from multiple vision sensors, an improved view of the surroundings may be created. According to the work of Luo and Kay [14], the primary purpose of fusing data from different sensors is to enable different systems operational applications in unregulated environments, without the need for complete human interaction. A system may not have complete knowledge of it's surrounding, due to the environment containing non-static objects, and fusing data from multiple sensors may improve the environmental information available. Before the fusion of data from multiple sensors is possible, registration between the different sensors is required, to ensure that all sensor data matches spatially. In an example with two vision sensors, images taken from both sources need to be captured at the same point in time, and the placement of features in one image needs to match features of the other image. When implementing multiple sensors, redundant information from the sensors will be captured, the redundant information refers to the same information captured by both sensors. The redundant information may be used to decrease the error of the multisensor system, or allow the system to operate in a degraded mode if one sensor fails. As multiple sensors capture the same information, the accuracy of the system may increase. Data captured from sensors that are not redundant is complementary data, which is data that one sensor can capture but not the other sensors in the system. When fusing data gathered from the sensors, the complementary data is often added directly to the corresponding part of the output. The redundant data is instead fused by adding all sensory data to depict a correct view of the surroundings. An example of a vision based fusion process can be seen in Figure 2 - 4. The images used in the example is based on a dataset provided by SAAB showing final approach and landing

of an aircraft.

(a) Sensor 1. (b) Sensor 2.

Figure 2: Input images for Figure 3 and 4 where 2a is captured using a SWIR sensor and 2b is captured using a sensor in the visual spectrum.

(a) Details from sensor 1. (b) Details from sensor 2. (c) Details merged.

Figure 3: An example of extracted details from Sensor 1 (2a) and Sensor 2 (2b), and all extracted details fused, to be used as input to 4.

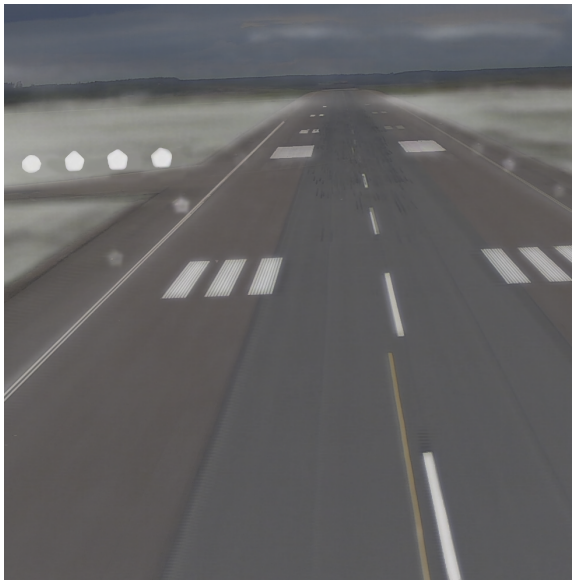


Figure 4: Illustration of a fused output image, using low frequency data from 2 and details seen in 3c, added together to create a fused output.

According to Drajic and Cvejic [15], there are two advantages to using vision sensor fusion.

Fused images are better suited for both human and machine perception, and the amount of transferred data is decreased compared to transferring all data from the source sensors. As an example Taehwan et al. [16], fuse an Infrared (IR) sensor and a radar to achieve object detection, at both day and night, as well as different weather conditions. Another example from Krotosky et al. [17] shows that fusing an IR-camera and a visual range camera may be used in surveillance systems, to detect the presence of moving persons. The benefit of implementing the IR sensor, is the sensors capability of providing structural information regardless of light conditions. Enhanced Vision System (EVS) is the terminology for vision-based systems in the aviation domain.

According to Spitzer et al. [18] EVS systems mitigates the following situations:

- *"Loss of vertical and lateral spatial awareness with respect to flight path"*
- *"Loss of terrain and traffic awareness during terminal area operations"*
- *"Unclear escape or go-around path even after recognition of problem"*
- *"Loss of attitude awareness in cases where there is no visible horizon"*
- *"Loss of situation awareness relation to the runway operations"*
- *"Unclear path guidance on the airport surface"* [18]

The image is typically displayed on a head-up display to enable monitoring of the system and preserve a direct visual perception of the situation. This configuration is approved by the FAA as "Enhanced Flight Vision System (EFVS), and are systems with the purpose of meeting requirements of enhanced flight visibility. Enhanced flight visibility is defined by FAA (14 CFR § 1.1) as: *"The average forward horizontal distance, from the cockpit of an aircraft in flight, at which prominent topographical objects may be clearly distinguished and identified by day or night by a pilot using an enhanced flight vision system"* [18]. Object detection requires feature-rich imagery data leading to demanding requirements of an EVS system. Using vision sensor fusion in EVS systems is explored in this thesis as a means to improve SA. If the output from a fusion process is of such low quality, that it's impossible to interpret, there would be no possibility of improving the SA.

2.3 Image Quality

As stated by Keelan [19], personal preferences may impact the quality grading of an image. An example of this would be a person grading the quality of an image depicting an old family member giving a high score, disregarding image degradations such as noise. It's further explained that personal preferences come into place as either first-party or second-party assessments. Where first-party-assessment implies the person who has taken the picture, and second-party-assessment implies the subject of the picture. If professional photographers perform the first-party assessment, their opinions may fit well into quantifiable image quality criteria. Whereas the second party, on the other hand, taking a more subjective route regarding the depiction of the subject in the image, may not fit well at all.

Several challenges exist to measure image quality. The quality is often measured by human visual perception. However, models for objective image quality measurements exist and are an ongoing topic in the research community. Evaluation of different sensors and fusion methods performance concerning object detection is difficult to quantify due to the lack of a clearly defined baseline regarding parameters that impact the object detection capacity.

3 Related Work

According to ICAO [20], visibility is defined as:

a) *"the greatest distance at which a black object of suitable dimensions, situated near the ground, can be seen and recognized when observed against a bright background;"*

b) *"the greatest distance at which lights in the vicinity of 1 000 candelas can be seen and identified against an unlit background."*

Airspace visibility depends on the opacity and illumination of the atmosphere. At the right conditions, the environment can be observed within the 400-700 nanometer electromagnetic spectrum (visual range). As pointed out in the introduction section 1, accidents may occur if aircraft operates in bad visibility. Fog decreases the ability to observe the environment within the visual range and occurs when the air's relative humidity reaches 100% resulting in water vapor [21]. Fog negatively impacts electromagnetic radiation in the atmosphere for waves with a wavelength of less than 1 cm. In conditions where fog is present, scattering occurs due to micro-physical structures (aerosols). Beier and Gemperlein [21] conducts an experiment to improve visibility in fog conditions, by simulating IR-cameras within the spectrum of 3-5 μm and 8-12 μm. According to the simulation IR-cameras improve the range of visibility for all types of aerosols in conditions for Clear Air Turbulence (CAT) I and CAT II. However, at extreme conditions with dense fog, there is no improvement utilizing IR cameras with the 3-5 μm and 8-12 μm range [21].

3.1 Image Fusion Theories

This section introduces the fundamental concepts and state-of-the-art practices regarding image fusion evaluated in this thesis. Amongst the field of research, the steps conducted in an image fusion process can be divided and categorized into some fundamental theories. The theories include but are not limited to, decomposition of input source images, where details are extracted from the source image. Representation of the source image is conducted in SR to enhance the performance of the fusion method. A set of rules on how the fusion of separated data shall be reconstructed by selecting what feature or sensor is dominant in some parts of the output. Feature extraction and fusion may also be conducted using Neural networks, and such solutions may increase the performance of the fusion method.

3.1.1 Transform Based Multi-scale Decomposition

Multi-Scale Decomposition (MSD) is a method for separating the source images into layers often represented with a pyramid structure. The most common pyramid is the Laplacian pyramid. This method is proven successful and can be divided into four stages: Low-pass filtering, sub-sampling, interpolation, and differencing [22]. Another approach to MSD is the wavelet transform. For example, the discrete wavelet transform decomposes the source image with filtering to obtain high- and low-frequency sub-images. The drawback of a discrete wavelet transform is the occurrence of oscillations, aliasing, and shift variance. In addition to discrete wavelet transform, several other wavelet transform-based methods for decomposition exists, e.g., lifting wavelet transform and spectral graph wavelet transform. Edge preserving filter is a MSD method that separates the source image to a base layer and a detailed layer. The significant contribution of this method is the spatial preservation capabilities and reduction of artifacts around edges. The base layer composes of a smoothed image and the detailed layer containing several sub-layers of different scales. This method is widely adopted in other methods, e.g. bilateral filter, mean filter, and weighted least square filter [5].

3.1.2 Transform Based Sparse representation

Sparse Representation (SR) methods are similar to MSD based methods in the sense that SR methods also belong to the transform domain-based approaches. However, some key differences exist between the techniques, one being that SR represents the source images from a dictionary, based on training images. The dictionary tends to be domain-independent, providing a reasonable

interpretation of the source images. Second, SR operates over patches distinctive to MSD operating over different decomposition levels, where patches refer to the source image divided into blocks (patches). Patches, in combination with a sliding window approach, provide a more reliable result, when compared to MSD to image misregistration. The theory behind SR is that image signals can be represented as a linear combination of some bits of a dictionary. First, the source image is segmented as patches represented as a vector. SR is performed on the patches with a dictionary. The next step is to combine the representations with a fusion rule. Finally, the fused image is reconstructed from the sparse coefficients. The dictionary and quality of such a dictionary are crucial for a satisfying output. There are several methods for constructing a dictionary [23]. For example, Yang and Li [24] creates a dictionary based on a set of functions containing Discrete Cosine Transform (DCT). Other approaches exist as well e.g. short-time Fourier transform CVT. A dictionary contains prototype signals named atoms. For each image signal, there is a linear combination of atoms in the dictionary that approximates to the real signal. This approach is an NP-hard problem; however, there exist attempts to minimize the computational overhead with sliding window approaches.

3.1.3 Fusion Rules

There are several fusion rules, with the coefficient combination method being the most common one. The coefficient combination method includes two major strategies, choose-max and weighted average. Authors adopting the choose-max strategy have different coefficients of interest depending on the implementation. For example, Chai et al. [25] implement the choose-max strategy with coefficients based on contrast and energy for applications in the medical domain. Furthermore, the weighted strategy combines the image layers based on a weight map. The weight maps can be generated using several approaches, with saliency analysis being a state-of-art practice [5]. Apart from coefficient combination methods operating on pixel-level additional region-based fusion rules exist. For example, the salient region rule implemented by Li et al. [26] constructs a saliency map based on a 31 by 31 window applied with laplacian filtering and local average.

3.1.4 Neural Network Based Fusion

The field of image fusion has adopted methods based on neural networks, the most common one being Convolutional Neural Network (CNN). CNN's, in combination with other techniques such as MSD, has been proven successful. Research shows that conventional methods for image fusion mentioned above have difficulties in pursuing state-of-the-art results compared to deep-learning methods. However, deep-learning methods also encounter challenges in the field. Mainly the absence of a large and specific dataset for training, and the challenge of constructing a network to handle a specific fusion task. A popular method for deep-learning image fusion is supervised aimed to learn a multistage feature representation [27].

3.2 Computer Vision

In a paper by Vygolov [1], an implementation that combines three optical-electronic sensors to capture the visual range of light, short-wave IR, and long-wave IR is presented. The purpose of the short-wave IR sensor is to provide visibility of essential features and light at night as well as bad weather, whereas the long-wave IR sensor increase sensitivity when fog is present. Enhancement of the image is performed with Multiscale Retinex before image fusion to obtain multiscale brightness. The approach for image fusion in the paper is Pytiev's morphological approach. Histogram segmentation is used to extract morphological shapes from the short-wave IR sensor. The visual sensor and long-wave IR sensor are projected to the short-wave IR picture by calculating the mean of a corresponding area and weigh the sum of the projections [1].

In the millimeter-regime the result produced by the sensor depends on the operating frequency. Transmission losses occur when the radio-wave frequency matches the resonant frequencies of molecules in the atmosphere. At 35, 94, 140, and 220GHz, the attenuation is relatively modest and objects reflecting the down-well radiation will provide high contrast to, for example, a human body. The reflection and emission of an object in the millimeter-regime are determined by the emissivity ϵ. For an ideal radiator (absorber) $\epsilon = 1$ and an ideal reflector (nonabsorbent) $\epsilon = 0$.

The emissivity can be expressed as a function of surface roughness, angle of observation and the materials dielectric properties. The image of a PMMW sensor consists of the observed radiometric temperature of a scene. The observations are based on the emissions of objects, the reflection of the sky's radiation, and atmospheric emissions between object and sensor [28].

As Song et al. describe [29], PMMW imaging systems are used for the detection of metallic objects. As the sensing system is entirely passive, no emission occurs, while the detection of the environment is still possible because of the high reflection of background radiation on metallic objects. The authors further explain that PMMW in itself is not good at spatial resolution or details, and therefore propose a method fusing PMMW images with images in the visible range. The fusion of the two technologies is proposed as a solution to the PMMW sensor's inability to capture details by adding details from the visual range. By finding an ultimate fusion of the two captured images, a "true scene" is expected.

As different vision-sensors are developed for specific wavelengths, the use-cases for each vision-sensor differ slightly. There are multiple ways to achieve a fusion of collected data. As explained by Xia et al. [30], data fusion may occur at the signal, pixel, or feature level. Where signal level fusion is based on row data, pixel fusion is based on pixel-to-pixel matching, and feature level is based on features extracted from sensors.

A significant challenge regarding image fusion is the concept of image registration. In essence, the image alignment, differences in resolution, the field of view, and distortion significantly complicates the image fusion process, particularly in real-time applications. According to Zitova et al. [31], the image registration process consists of four steps. The first step is feature detection, where features are detected in both images. The features consist of lines, corners, and other distinct points of interest. The second step is matching the detected features in both images. Both steps may be conducted either manually or automatically. The third step is to estimate the transform model and aligning the moving and reference images. The final step is where the actual transformation of the image occurs. Putz et al. [32] test both fusion and image registration in a multi-modal configuration. The test is conducted on a Field Programmable Gate Array (FPGA) for real-time applications, and data shows that the Laplacian pyramid and Fast and Adaptive Bi-dimensional Mode Decomposition (FABEMD)) methods provide better results when compared to Shift Invariant Discrete Wavelet Transform (SIDWT) and simple mean method.

The FABEMD algorithm is based on image decomposition into oscillatory sub-signals and a series of zero-mean Bi-dimensional Intrinsic Mode Function (BIMF), a simplified version of Bi-dimensional Empirical Mode Decomposition (BEMD). The algorithms start with a decomposition of both initial images, followed by combining the values of two BIMFs for each decomposition level. The third step is to combine two residues, and lastly, sum all combined components to a fused image. Laplacian pyramid algorithm is based on pyramid generation, the algorithm process the image through a low-pass filter and apply sub-sampling by a factor of two. The filter is often a 5x5 window with Gaussian coefficients. The result is a pyramid of sub-sampled images with a reduced spectral band. The fused image is calculated from the input pyramids with the selection of pixels with a higher intensity. This method preserves the contrast ratio of the final image [33]. In the work of Antoniewicz [33], the above algorithms are implemented in an FPGA to meet critical time requirements. Implementing image fusion on an FPGA reduces the processing time significantly due to the parallel and pipelined capabilities compared to a traditional Central Processing Unit (CPU). The images are sent from an image codec to the Double Data Rate (DDR) memory, the FPGA then processes one frame at the time with implemented fusion algorithms.

3.3 Image Fusion Quality Assessment

Complications arise when assessing fused images. In an example by Qu et al. [34] there are applications where no ideal fused image exists, thus making it impossible to implement an error based assessment. As a means to objectively measure image quality on images processed by image fusion, several methods exist. For instance, the method by Qu et al. [34] compares a fused output image with input images by calculating the mutual information contained in the fused image. The amount of information carried over in the fusion process is determined and used as a measure of image fusion performance. However, the method doesn't take into consideration what information is considered essential. In order to increase speed-performance Haghighat and Razian

[35] propose another method, Fast Feature Mutual Information (Fast-FMI) which show similar results to other Feature Mutal Information (FMI) methods. The proposed Fast-FMI method divides the full image into smaller squares, which helps reduce complexity. Furthermore, the method calculates an average of all the mutual information obtained from the smaller windows, which is summed and seen as the mutual information of the entire image. In a method proposed by Xydeas and Petrović [36], edge information is seen as valuable information. Edge information still present in the fused image is calculated as an indication of the fusion process. It should be noted that this method only pertains to pixel-level fusion methods.

3.4 Image Restoration

In recent years, deep learning has emerged as a solution for image denoising. For instance Liu et al. propose several methods in [37, 38, 39]. According to a survey by Tian et al. [40], the most significant disadvantage for conventional methods of denoising [41, 42, 43] is manually tuning of parameters and complex optimization problem, resulting in high computational costs. Current deep learning models face challenges of noise that are deviating from Additive White Gaussian Noise (AWGN), resulting in problems with real noise e.g., low light.

D. Park and H. Ko. propose a method for the restoration of fog-degraded images. With an atmospheric scattering model together with a depth estimation, the Red Green Blue (RGB) channels and contrast can be restored. The method reads RGB values for each pixel and estimates the depth d, an opening and closing reconstruction is performed and a β is estimated to find maximum entropy [44].

The movement of a sensor when capturing an image, cause image blur. Image blur is a phenomenon that depends on the motion of the sensor relative to the scene during the exposure time of the sensor. In a paper by Li et al. [45], several methods are compared for image restoration of blurred images. The results show that the Wiener filter and Blind restoration provides the most accurate restoration. However, it is stated that the image quality is degraded after restoration.

3.5 Summary Of Related Work

Concerning the extreme environments of aircraft operations, sensors suitable in other domains may prove unsuitable in the aircraft domain. If similar systems are implemented in an aircraft, the reliability and robustness of the system's operation are of high importance.

As an aircraft operates in a broad spectrum of weather conditions, sensors used are required to deliver an adequate perception of the environment in all phases of operation. Research shows that fog drastically decreases the performance of a vision sensor in the visual range and that IR sensors can sense the environment in moderate fog conditions [32]. Other research projects experiment with sensor fusion in a variety of techniques and algorithms. The experiments are proven successful, however, not tested in the aircraft domain [33].

4 Problem Formulation

Decision-making systems are critical systems demanding high reliability. The importance of correct data gathered from sensors is crucial. As an example, the automotive industry has implemented vision-based sensors, fusing images obtained by cameras and other sensors [13]. The aviation industry is moving towards an automated future and the ability to identify hazards in crucial operations, for instance, landings are of importance. Today, well-proven systems can detect other aircraft with the utilization of transponders and radars together with human visual perception. As an example, aircraft operating unmanned airports require the ability to detect obstacles not equipped with transponders or not discovered by radar. To increase detection capacity, at differing weather conditions, vision sensor fusion will be evaluated using appropriate sensors. Both CNN based methods and more traditional methods will be tested.

This thesis aims to explore the possibilities of utilizing vision sensor technologies in the aircraft domain to enable object detection and increase safety. As sensors are to be evaluated for a specific purpose, quality criteria need to be determined. The assessment of image quality is a broad topic, where both subjective and objective measurements may be implemented in multiple ways. Furthermore, assessment techniques of fused images are problematic as no perfect fused image exists in this specific sensor configuration. Therefore an adequate fusion performance method needs to be obtained and used as a measure of this specific implementation.

4.1 Hypotheses

The work in this thesis is based on the following hypothesis. The hypothesis is acquired from research conducted in the related literature.

With consistent vision sensor acquisition the imagery output can be used for object detection in the aviation domain, regardless of environmental conditions.

4.2 Research Questions (RQ):

To test the hypothesis stated in section 4.1, the following research questions have been formulated.

RQ1) What image quality assessment techniques are required to determine the output quality of evaluated image processing methods?

RQ2) What are the similarities and differences between state-of-the-art vision based fusion methods?

RQ3) What sensor fusion technique provides adequate results with respect to given quality metrics from RQ1 in an aircraft environment?

RQ4) What correlation exists between the detection capacity of objects and image quality?

RQ5) What is the most correct way to assure that output from one sensor matches that of other sensors in a sensor setup of two IR sensors and one visual spectrum sensor?

4.3 Overview

Table 1 aims to provide easier navigation in the thesis, displaying the corresponding sections to each research question.

Table 1: Section Overview.

RQ	Method	Result	Discussion	Conclusion
1	5.2, 5.6.1, 5.7	7.1	8.1	9.1
2	5.6	7.3	8.3, 8.6, 8.7	9.2
3	5.6.1, 5.7	7.3, 7.4	8.3	9.3
4	5.8	7.5	8.5	9.4
5	5.5	7.2	8.2	9.5

5 Method

A systematic review is conducted to answer RQ1 and provides a knowledge basis for the following research. The systematic review follows a structured approach, where relations, patterns, and identifications are identified in publications.

Proceeding the systematic review, RQ2, RQ3, RQ4 and RQ5 are answered with experimental research. The experimental setup consists of two IR sensors and one sensor in the visual range. During the registration process, the output from the IR sensors is treated as moving images and fitted against the visual range sensor. After both IR images have been registered, all three images are fused using one of the multiple fusion methods. The output from all of the proposed methods is evaluated by applying multiple Image Quality Assessment (IQA) methods and subjective evaluation. The subjective evaluation of the fused images is performed by people of varying expertise, ranging from pilots or experienced in image processing to novice. The validity of the research is

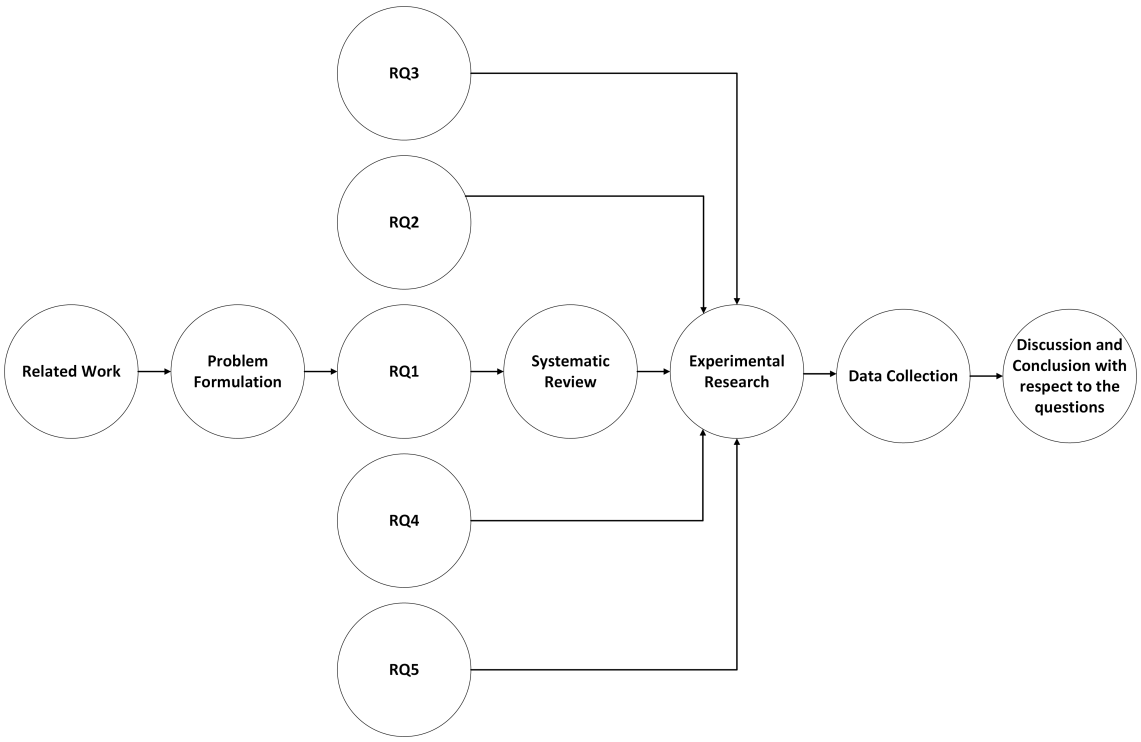


Figure 5: Research flowchart.

discussed in the discussion section. However, the measures to increase validity is a continuous process throughout the project. Validity can be divided into four main categories. Construct validity, content validity, face validity, and criterion validity. For example," Do the constructed tools and experiment represent the measure of variables intended?", "Do the review and experiment cover all aspects of the subject?" and "What is the correlation between these papers results and other literature?" [46]. According to Keelan [19], personal preference takes place while grading image quality. To mitigate personal preference outside the scope of sensor evaluation, specific image qualities are stated as image quality criteria in the subjective ranking of images. The work aims to evaluate vision-based sensors in an aircraft operating environment, where the conclusions and experiments made are not optimized for general purpose use.

5.1 Thesis Limitations

Regarding the limitations of this thesis, there are several factors to take into consideration. For instance, the possibility of implementing Full Reference Image Quality Assessment (FR-IQA) or Reduced Reference Image Quality Assessment (RR-IQA), as no known right image of the scene

depicted exists for this sensor setup. NR-IQA is instead what is conducted, to assess the different fusion methods. The personal preference impacting image grading is somewhat limited during subjective tests by selecting subjects with some prior knowledge in associated areas. Another example would be in what environmental conditions these methods are feasible to implement. Possibility to enhance visibility in fog or low light, as well as right weather conditions such as daylight and clear visibility, is explored. All other possible environmental conditions are explored and are therefore placed outside the scope of this thesis. The idea of implementing object detection in avionic systems is impressive, and this thesis can be seen as a step of improving the object detection system's ability to create a correct world view. This thesis tries to improve visibility for these types of systems, or as a possible pilot aid. However, one way to show an increase in object detection capabilities would be to implement real object detection, which is outside the scope of this work. The fused output is tested using a pre-trained network, but it's important to note that the network is not trained on the output from this thesis fused images. Therefore an improvement to image quality is instead used as an argument of object detection capabilities.

5.2 Systematic Review

A systematic review is an appropriate tool used to make conclusions based on a consensus in the literature. As such, a systematic review is a sufficient tool to identify methods of IQA. During the review, universal patterns and themes are identified, and publications are chosen based on a number of criteria:

- The Main focus of the paper should include Objective IQA.
- Literature shall position their work to other research, by evaluating their method in comparison to other methods.
- Literature should be published in an established journal or similar.

At the beginning of the review, publications with a large number of citations are selected to ensure the quality of publications further. Analyzing patterns and themes in a set of literature requires a systematic approach. A concept matrix provides an overview of the relationship between articles and concepts, helps to identify common patterns and themes in the literature as well as gaps. The review implements the following steps proposed by Webster and Watson [47]:

1) Review leading journals and conference proceedings with a reputation of high quality.
2) Review the citations of literature found in step 1. This helps to identify concepts leading to the state-of-the-art.
3) Identify important literature, referencing the findings in step 2. This provides a broad view of the research.

The primary source for references and data for the literature review is scientific articles and papers. The literature is collected and searched for at the following databases:

- Google Scholar
- IEEE Explore
- Research Gate

5.2.1 Image Quality Assessment

A common theme amongst the literature reviewed is the difficulties concerning NR-IQA. Furthermore, it is stated that the human eye is an expert in this area [48, 49, 50, 51, 52]. Wang and Bovik [48] describes the problem as *"mission impossible"* to quantify the quality of an image without a reference base-line. However, there exist good models for FR-IQA and RR-IQA as Wang and Bovik states [48], not applicable in the scope of this thesis.

Figure 6 illustrates the progress of literature made by notable authors in this domain. This map is the baseline for the literature review, achieving a systematic approach.

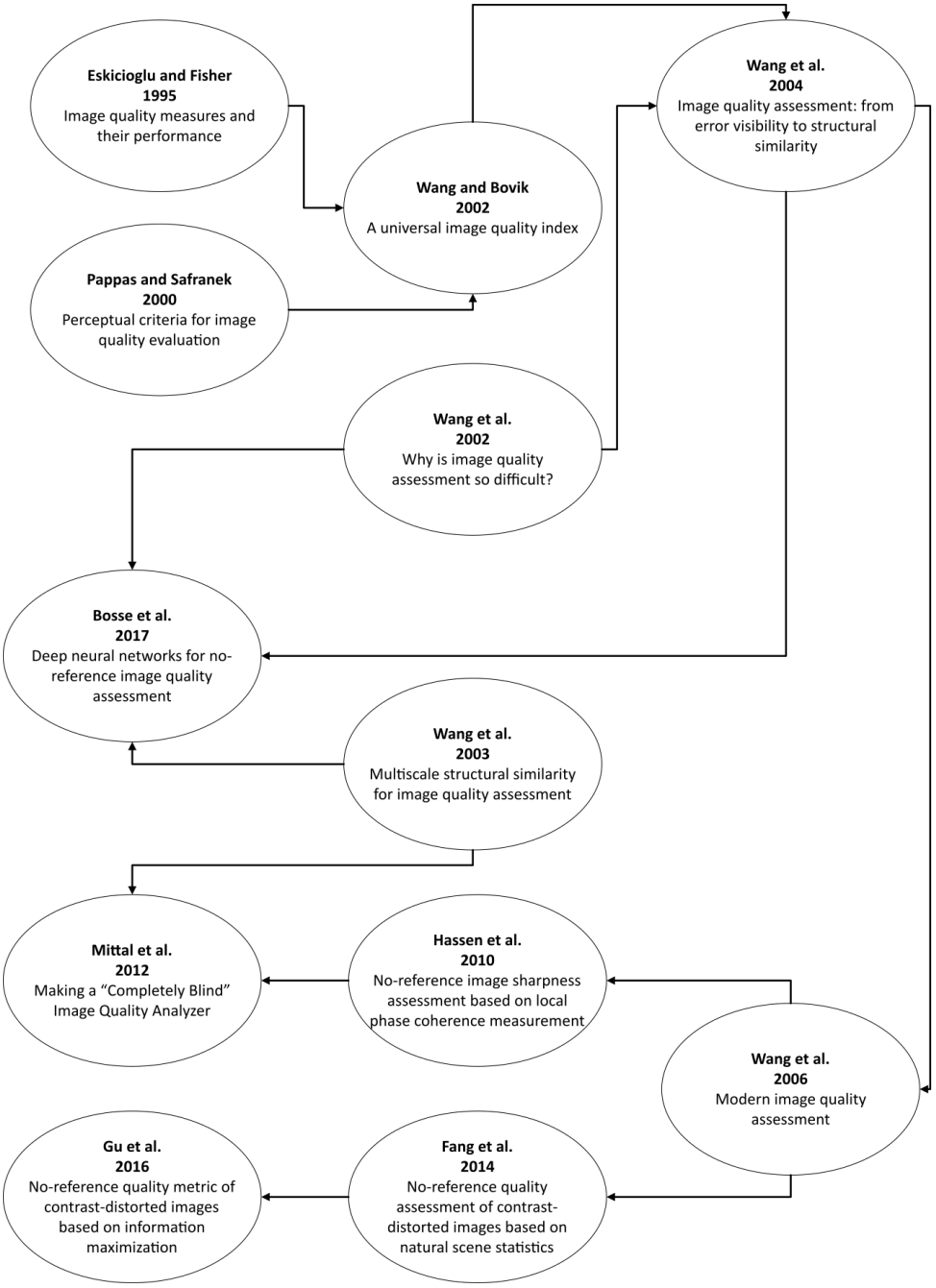


Figure 6: Reference tree of the structured literature review.

During the earlier years of objective IQA, there were some setbacks, due to the fact of objective image assessment not correlating well with subjective image quality assessment. This is a phenomenon that Eskicioglu and Fisher [53] tried to demonstrate by evaluating different objective quality measures in a gray-scale. In order to evaluate their technique of evaluation, test subjects with some prior image distortion knowledge were chosen. Correlation of the test subject's perceived image quality and IQA are later showed and varying results for different test images can be seen. However, they concluded that evaluating techniques with differing implementations, needed more parameters during evaluations.

Pappas et al. [54] evaluate objective criteria based on human perception of image quality and compares the objective techniques to Mean Squared Error (MSE). MSE technique, is a technique where a reference image is used in order to calculate the error. The objective techniques, on

the other hand, have an advantage as it determines quality based on different image distortions. Compared to the MSE technique, which suffers while calculating errors induced by different types of artifacts.

In order to mitigate the shortcomings of previous image quality measurement techniques such as Peak Signal-to-Noise Ratio (PSNR) and MSE, Wang and Bovik [49] introduces a new index for IQA. The new quality index value Quality (Q) is dependent on a combination of three parameters. The first parameter measures loss of linear correlation between a point in two images. The second parameter measures luminance distortion by comparing luminance intensity in points of the two images. The last parameter measures similarities of contrast between points of the two images. The index presented is supposed to be used as an independent method, applicable in multiple types of image processing implementations. Results show that the quality index Q detects quality degradation where MSE remains constant. Both Pappas et al. [54] as well as Wang and Bovik [49] show that a new philosophy is needed to achieve an objective quality assessment, correlating with perceived image quality. As pointed out by Wang et al. [50]: *"The best way to assess the quality of an image is perhaps to look at it because human eyes are the ultimate receivers in most image processing environments"* which may explain why both Eskicioglu et al. [53] and Pappas et al. [54] use human subjects as a reference in order to evaluate the assessed IQA techniques. It is further explained that a subjective Mean Opinion Score is not a practical solution as it is time consuming, inconvenient, and expensive. Wang et al. [50] proposes a new assessment method where the structural information in an image should be used as an error estimation, as older error estimation techniques calculate all types of distortions and may not correlate with perceived image quality. The paper shows that the new quality estimation technique is useful, but states that more research is needed in the field of structural information.

Wang et al. further investigate the usage of structural similarity [51], and shows how structural similarity compares to MSE by developing a technique that compares local patterns of pixel intensities. Both MSE and the new methods are tested in a variety of image distortions. Results show that MSE doesn't perform well, as the results of the MSE technique differs while used in different types of distortions. However, the new method developed shows good results with better consistency while comparing both MSE and structural similarity to qualitative visual appearance.

In order to increase the flexibility of image assessment methods Wang et al. [55] also proposes a new multi-scale structural similarity approach. The new approach shows that improvements can be made regarding performance, and a comparison is made to both single-scale approaches as well as other state of the art IQA methods if correct parameters have been chosen.

According to Hassen et al. [52] sharpness is one of the more important factors during the visual objective assessment of image quality and therefore proposes a sharpness measuring method. It is shown that the proposed method has correlations with subjective quality assessment. Furthermore, it's showed that by redefining blur as *"...the degradation of sharpness is identified as the loss of local phase coherence."* image distortions other than sharpness may be evaluated by the method as well.

Mittal et al. [56] state that NR-IQA methods require prior knowledge of distortions correlating with human perception in order to asses image quality. As a means to create a new model of assessment implementing measurable deviations occurring in images. The new model does not require any training using human graded distortions. To evaluate the model, a correlation between human perception is tested. The correlation between the new blind IQA method and human perception is evaluated alongside other IQA models. Testing shows that the blind model outperforms FR-IQA models while also performing similarly to other NR-IQA models.

Fang et al. [57] state in their paper, that contrast distortion often is a major contributor to perceived image quality. The results of the proposed method is evaluated against other IQA methods and comparisons are made based on the correlation of human visual perception. And promising results can be seen, however, additional development is needed in order to increase the performance.

According to Gu et al. [58] the main contributor to image quality, is the amount of information contained in the image. The statement is based on human perception, and how a person would determine quality. Therefore they propose an IQA, implementing information maximum by computing information contained, locally as well as globally. The reasoning behind both global and local information is explained as an image containing a large blue sky or area of green grass,

which may not locally contain much information yet globally be important to the perceived image quality. It is concluded that the developed NR-IQA method have better performance compared to other FR-IQA models as well as NR-IQA models. Further, the experiments show that the method has good capabilities at determining which image contains more contrast.

Recent studies as shown in Bosse et al. point towards a deep neural-network approach for IQA. In the work of Bosse et al. [59] a CNN is constructed for FR-IQA evaluated towards the LIVE, CISQ and TID2013 dataset. However, the authors claim that the proposed method can be used for NR-IQA with minor modifications. With a network of ten convolutional layers, five pooling layers, and two fully connected layers, the solution outperforms other state-of-the-art methods. On the other hand, the performance of CNN is heavily dependent on the dataset.

5.2.2 Summary of the Systematic Review

State-of-the-art solutions for IQA seems to be moving towards implementing CNN's to further increase performance. However, the majority of the solutions evaluate performance by comparing it to, among others, human subjective ranking. There also exist three major branches of IQA where this thesis focuses on the most difficult to implement, NR-IQA. Based on the findings a proposed method for evaluating image fusion is presented, see Figure 7.

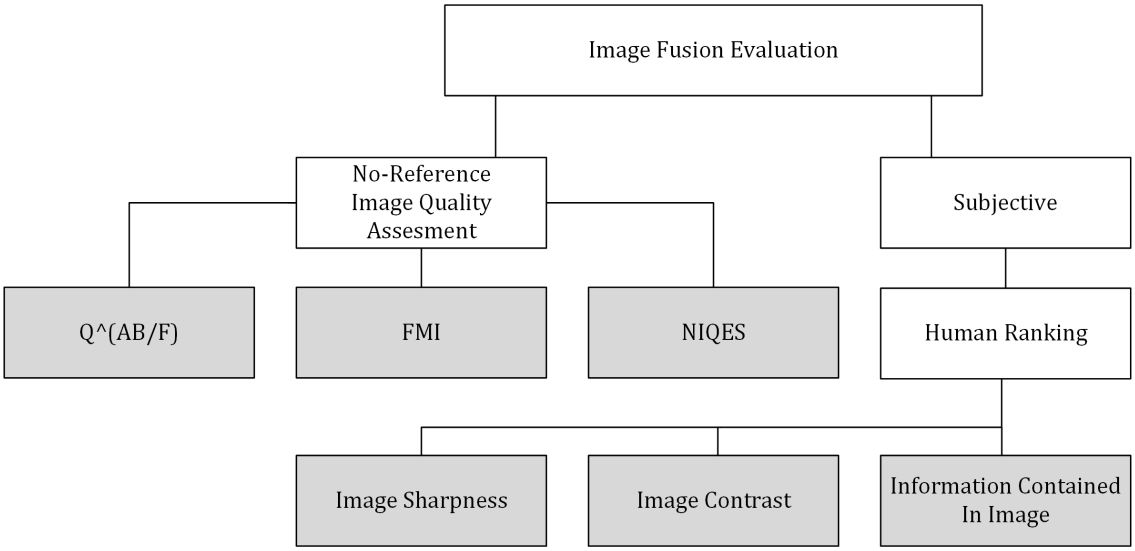


Figure 7: Proposed quality assessment method, with the Edge Preservation Quality Index ($Q^{\mathbf{AB/F}}$), Fast Feature Mutual Information, Natural Image Quality Evaluator (NIQE) and grading criteria for human ranking, according to the grey boxes.

5.3 Experimental Setup

Data from three sensors are used as input to the experiment, together with four different fusion techniques. Two of the three sensors is selected as moving images for the image registration part of the processing. The last sensor is used as a reference. As the visual sensor in this setup has a higher resolution than both IR sensors, the visual sensor is selected as the reference. Two transformation matrices are needed, one for each of the moving images. The two matrices are created by matching a set of points in the moving and reference pictures. The matching of points is done for both moving images, together with the reference visual spectrum image. The transformation of the two images results in empty areas when placed on top of the reference picture. Therefore cropping is necessary to ensure that the final image only consists of areas including all three sensors. Cropping coordinates are also pre-calculated and used together with the two previously constructed transformation matrices, and added to the experimental setup.

The process flow can be seen in Figure 8 and consists of two main parts, image registration, and image fusion. In the first step, image registration, the transform matrices transform the moving images and crops all three images. The output from the image registration part is then used as input to the image fusion, in the image fusion part, all three images are fused using one of the four suggested fusion methods.

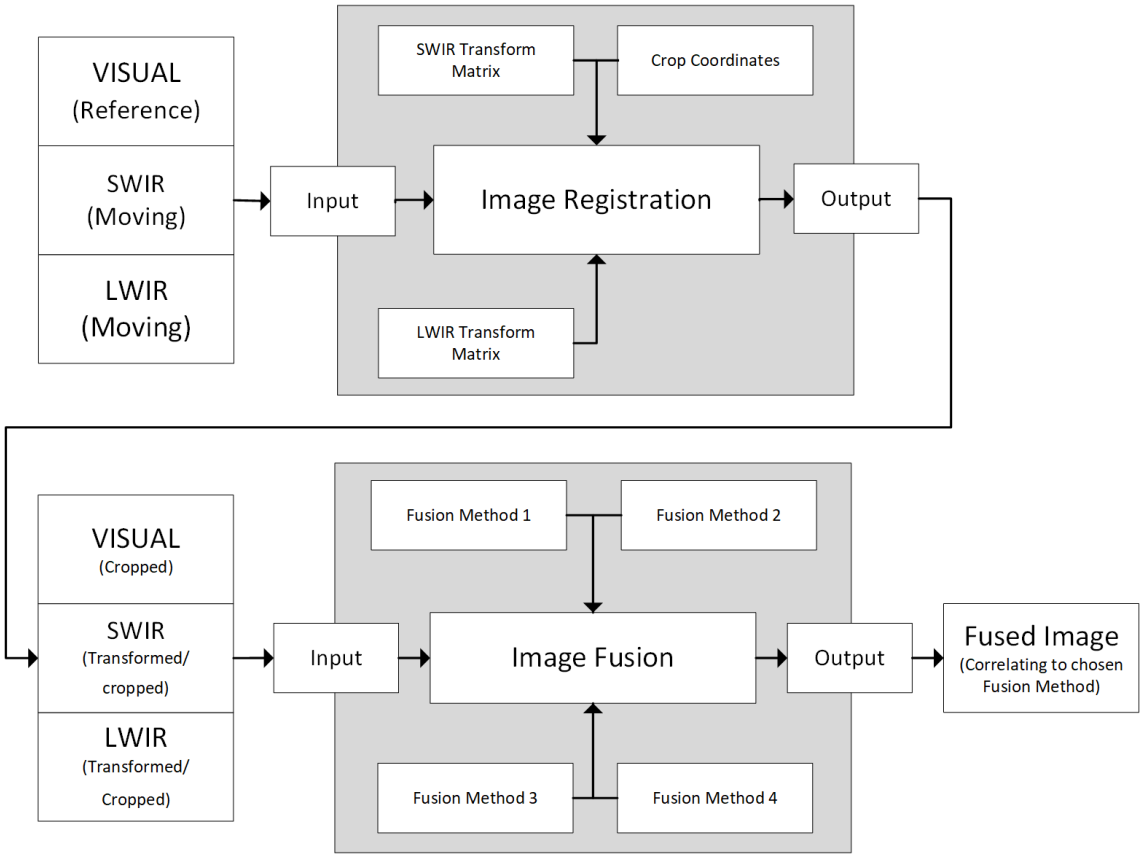


Figure 8: Image process flow, shown from captured sensor images to fused output of 3 sensors.

5.4 Sensor Configuration

This section explains the sensor setup used to collect imagery data for evaluation purposes in this thesis. The main setup is a concept rig with ”off-the-shelf” sensors provided by SAAB. All data were collected in a series of test flights conducted at a Swedish airfield in a controlled environment. The sensors were placed in a customized nose cone of the aircraft rigged to the frame, see Figure 9. During testing, a marshall CV 342-CSB [60] is used as a visual range sensor. The sensor uses both different resolution and framerate compared to the two other IR-sensors used. As the sensor can capture high detail images in the right weather conditions, it is a sufficient choice to use in the visible spectrum. The Rufus 640 Analog [61], on the other hand, is a Short-wavelength infrared (SWIR) sensor that has claims of working well in low-light and fog conditions. This thesis aims to enhance visibility in these conditions, and the Rufus 640 Analog is a good fit for this experiment. A complementary Long-wavelength infrared (LWIR), the Raven 640 Analog [62] is added to capture thermal properties regardless of weather and light conditions.

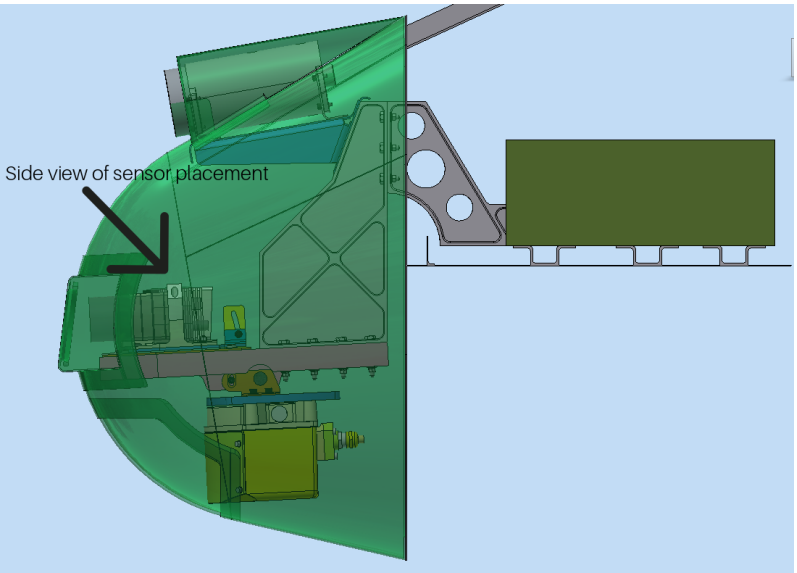


Figure 9: CAD drawing of the sensor setup used to collect imagery. As seen in the CAD drawing, sensors are rigged to the airframe.

5.5 Image Registration

As all data is pre-recorded, frames are collected and matched from each sensor by extracting frames at given time-intervals, resulting in an equal number of matching snapshots from each sensor.

5.5.1 Feature Detection and Matching

As seen in section 3.2, precise image registration is crucial for a successful fusion of multi-modal images. This process aligns the images acquired from different sources to a unified view, in essence, the alignment of one image to a fixed image [31]. In this thesis, the sensors implemented in the test rig have differences in resolution, lenses, and placement relative to each other. Feature selection is conducted both manually and automatically. In the manual selection method, points are placed by hand, in both the moving image and the reference image. An image with distinct features, both close and further away in the image, is selected to assure that the image matching is efficient across the entire image. The automatic selection and matching method tested is Speeded-Up Robust Features (SURF) [63]. The SURF method implements a detector and a descriptor. Where the detector is based on the Hessian matrix and the descriptor creates sub-sectors of the image by dividing the original image into 4x4 squares. Each sub-sector of the image is processed to calculate image intensity patterns.

5.5.2 Image Transformation

After the feature matching is completed, geometric transformation matrices are generated as either projective or affine. A geometric transformation refers to a set of operations that maps the source image to a new coordinate system. Affine transform preserves parallelism but manipulates translation see Equation (1), shear see Equation (2), rotation see Equation (3) and scale see Equation (4). Projective transform has the same capabilities as Affine transform but does not preserve parallelism see Equation (5) [64]. Translation Transform:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ t_x & t_y & 1 \end{bmatrix} \tag{1}$$

Shear Transform:

$$\begin{bmatrix} 1 & sh_y & 0 \\ sh_x & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{2}$$

Rotation Transform:

$$\begin{bmatrix} \cos(\theta) & \sin(\theta) & 0 \\ -\sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{3}$$

Scale Transform:

$$\begin{bmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{4}$$

Tilt Transform:

$$\begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{bmatrix} \tag{5}$$

(a) No transform. (b) Translation (c) Shear

(d) Rotation (e) Scale (f) Tilt

Figure 10: Example of transformed images.

5.5.3 Image Crop

As there are differences in resolution, and the sensor placement differs between all sensors, images were matched in a way where only one sensor was present in some parts of the images. As a consequence of areas in the image only containing data from one sensor cropping was performed on all images. Images were cropped in such a way that only areas where data from all sensors are present in the image were preserved.

5.6 Image Fusion

In this section, implementations for image fusion methods are presented. For the aims of this thesis, MATLAB [65] is used to implement fusion techniques and quality measures, see Appendix A. Furthermore, a test framework is developed to provide traceability of the conducted experiments with continuous logging of data, see Table 2.

Table 2: Example table of evaluation of fusion methods.

Time Stamp	**Method**	**Elapsed Time**	**Time/ Pixel**	**Fast-FMI**	**Q$\widehat{}$(AB/F)**	**NIQE**

5.6.1 Subjective Test for Fusion Evaluation

The subjective ranking of fusion algorithms follows Petrovic [66] method. Testing is conducted in a controlled environment, minimizing the risk of uncontrollable parameters affecting results. A custom graphical user interface is developed for subjective tests, see Appendix B. The user interface enables consistency when all subjects are faced with the same experience. The interface displays four images in a 2-by-2 matrix containing two source images on the top row and two fused images on the bottom row. The subject selects the preferred fused image by clicking on the image. The fused alternatives are randomly displayed on the bottom row, minimizing the risk of the subject being biased and repeatable selecting the same image in the matrix. The test generates a file containing data about the preferred fused algorithm and the time taken for the subject to choose. The subjects selected one of the preferred fused images based on the criteria shown in Figure 7. An option of not selecting any of the fused images is also possible. To avoid subject fatigue, the sample size is kept to 8 images. None of the subjects had reported degradation in eyesight or similar. However, corrected vision with glasses was allowed. See Table 3 for test setup.

Table 3: Table over ranking participants.

Subject	Background
1	Pilot/Avionics System Developer
2	Pilot/Avionics System Developer
3	Avionics System Developer
4	Avionics System Developer
5	Avionics System Developer
6	Avionics System Developer
7	Avionics System Developer
8	Avionics System Developer
9	Robotics Student
10	Robotics Student
11	Knowledge In Object Detection
12	Knowledge In Object Detection
13	Computer Science Student

5.6.2 Method 1: CNN with Visual Geometric Group (VGG)-19 model

The first method implemented is proposed by Li et al. [67]. The approach utilizes the VGG-19 [68] model trained on ImageNet [69]. The method decomposes the input images to detailed layers I_k^d and base layers I_k^b. The base layers are fused with a weighted-averaging strategy, and the detailed layers are fused using a multi-layer fusion strategy of features extracted with the VGG-19 model. Finally, the fused layers are added to one resulting image. Source images are taken as input, k denotes which input is used, and source images are denoted as I_k, where $k \in \{1, 2, 3\}$ in this example. The base layer is obtained by solving the Tikhonov Regularization problem, where the horizontal and vertical parameters are set to $g_x = [-1\ 1]$, $g_y = [-1\ 1]^T$ and $\lambda = 5$ in this method:

$$I_k^b = \arg\min_{I_k^b} ||I_k - I_k^b||_F^2 + \lambda(||g_x * I_k^b||_F^2 + ||g_y * I_k^b||_F^2) \tag{6}$$

After the base layers have been obtained, the detail layers I_k^d are extracted from the source image, by subtracting the base layer:

$$I_k^d = I - I_k^b \tag{7}$$

When both detail and base layers are separated, the fusion of the base layer is achieved using a weighted average strategy. Where x and y represents placement in the image and α denotes weight values of the pixel for each of the base layers, $\alpha_1 = \alpha_2 = \alpha_3 = \frac{1}{3}$:

$$F_b(x, y) = \alpha_1 I_1^b(x, y) + \alpha_2 I_2^b(x, y) + \alpha_3 I_3^b(x, y) \tag{8}$$

For the details of the image, a convolutional neural network VGG-19 [68] is used as a feature extractor of the detailed layers I_k^d. The extracted features are used to create weight maps and fused by multi-layer fusion. $\phi_k^{i,m}$ denotes the feature maps of $k'th$ sensor in the $i'th$ layer, where m is the channel for each layer ($m \in \{1, 2, ..., M\}, M = 64x2^{i-1}$). Where there's a Φ_i for each of the rectified unit layers (relu 1_1, relu 2_1, relu 3_1, and relu 4_1) in the network.

$$\phi_k^{i,m} = \Phi_i(I_k^d) \tag{9}$$

The contents for each $\phi_k^{i,m}$ at each position in the image is denoted as $\phi_k^{i,m}(x, y)$ After all features have been extracted from the source image, the initial pixel intensity map C_k^i is created by applying $l_1 - norm$, where $k \in \{1, 2, 3\}$ and $i \in \{1, 2, 3, 4\}$ [70]:

$$C_k^i(x, y) = ||\phi_k^{i,1:M}(x, y)||_1 \tag{10}$$

The created pixel intensity maps are then further processed by applying a block-based average operator to limit misregistration, where r determines block size. Using a more significant value of r can however lead to missing details:

$$\hat{C}_k^i = \frac{\sum_{\beta=-r}^{r} \sum_{\theta=-r}^{r} C_k^i(x + \beta, y + \theta)}{(2r + 1)^2} \tag{11}$$

Once the pixel intensity maps have been created, initial weight maps $W_k^i(x, y)$ for the pixels are created with K denoting the number of maps ($K = 3$) by normalizing the pixel intensity maps:

$$W_k^i(x, y) = \frac{\hat{C}_k^i(x, y)}{\sum_{n=1}^{K} \hat{C}_n^i(x, y)} \tag{12}$$

Using a VGG network solution, the pooling operator used is a subsampling operator, meaning that the size of created weight maps W_k^i differs from the size of I_k^d. Therefore an upscaling of W_k^i is needed. After the upscaling of the weight maps are completed, new weight maps: $\hat{W}_k^i$ all have matching sizes as I_k^d.Where $p, q \in \{0, 1, ..., (2^{i-1} - 1)\}$:

$$\hat{W}_k^i(x + p, y + q) = W_k^i(x, y) \tag{13}$$

The new weight maps $\hat{W}_k^i$ are used to extract features from source images detail layers:

$$F_d^i(x,y) = \sum_{n=1}^{K} \hat{W}_n^i(x,y) \times I_n^d(x,y),\ K = 3 \tag{14}$$

And, fusion of the details is completed by selecting the biggest value of each of the outputs acquired in Equation (14) at all possible positions (x, y)

$$F_d(x,y) = max[F_d^i(x,y)|i\ \in \{1,2,3,4\}] \tag{15}$$

Finally the entire image is fused together using the obtained detail and base layers:

$$F(x,y) = F_b(x,y) + F_d(x,y) \tag{16}$$

See Figure 11 for a visual representation of method 1.

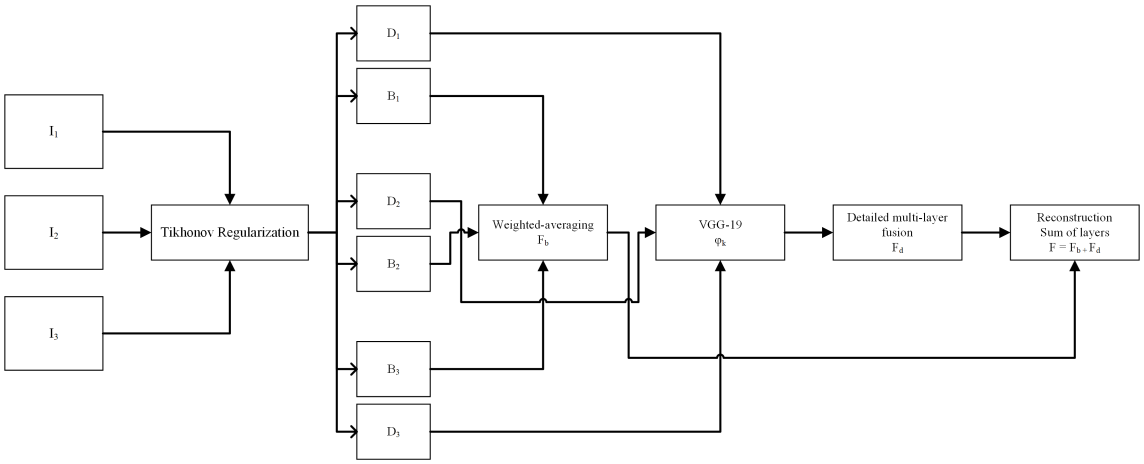


Figure 11: Illustration of Method 1, showing three images as input. The input images are optimized with Thikonov Regularization, and divided into base and detail layers. All detail layers are then fused using a multi-layer fusion strategy of features extracted with the VGG-19 model and the base layers are fused using a weighted averaging strategy.

5.6.3 Method 2: CNN with Residual Network (ResNet)50 model

Like the previous method described in section 5.6.2 Li et al. [71] utilizes a CNN, trained on ImageNet. Unlike previous authors who extracted detailed features using VGG-19 [68], this method implements the ResNet50 model. The model is a residual network based on ResNet [72] and is 50 layers deep with 5 convolutional layers. ResNet50 is used to extract details from the source images I_k (where $k \in \{1,2,3\}$) in blocks from the input source images, and processed using Zero-Phase Component Analysis (ZCA) operations. Features for each block is represented by $I_k^{i,1:C}$ containing C channels ($i \in \{1,2,...,5\}$). Each iteration of i, contains j number of convolutional channels, $j \in \{1,2,...C\}$.

For each of the input sources, the covariance matrix in ZCA is given by:

$$Co_k^{i,j} = I_k^{i,j} \times (I_k^{i,j})^T$$

And the decomposition as:

$$Co_k^{i,j} = U\Sigma V^T$$

The features $I_k^{i,1:C}$ is represented in a ZCA subspace as $\hat{I}_k^{i,1:C}$ and are obtained by Equation (17) together with the extracted features for each channel:

$$\hat{I}_k^{i,1:C} = (U(\Sigma + \epsilon I)^{-\frac{1}{2}} U^T) \times (I_k^{i,j}),\ j \in \{1,...,C\} \tag{17}$$

$l1$-norm of average values is used to create weight maps $I_k^{i,*}$ where $t = 2$:

$$I_k^{i,*} = \frac{\sum_{p=x-t}^{x+t} \sum_{q=y-t}^{y+t} ||\hat{I}_k^{i,1:C}(p,q)||_1}{(2t+1) \times (2t+1)}$$

The weight maps are then transformed into the size of the source images with bicubic interpolation [73]. Where bicubic interpolation increases the size of the smaller weight maps, by estimating pixel values and adding pixels. An example of bicubic interpolation can be seen in Equation (18) where the matrix represents an input image:

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \tag{18}$$

If the image is to be extended into a 4×4 matrix, the method starts by extending the image in one direction and after the initial extension, repeats the process in the next direction. First a padding of two zeroes is added at the edges of each row, to the matrix in Equation (18). The padding is added so that calculations of the edge values in the new image is possible, where each pixel requires two values on each side of the value to be calculated:

$$\begin{bmatrix} 0 & 0 & 1 & 2 & 3 & 0 & 0 \\ 0 & 0 & 4 & 5 & 6 & 0 & 0 \\ 0 & 0 & 7 & 8 & 9 & 0 & 0 \end{bmatrix} \tag{19}$$

The matrix in Equation (19) is then extended and shifted in the first direction, in this case the rightward direction:

$$\begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 2 & 2 & 3 & 3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 & 4 & 5 & 5 & 6 & 6 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 7 & 7 & 8 & 8 & 9 & 9 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{20}$$

Each pixel value in the matrix from Equation (20) is then averaged using weights, by the two values on either side of the pixel:

$$\begin{bmatrix} 0.9453 & 1.5811 & 2.4189 & 3.0547 \\ 3.9453 & 4.5811 & 5.4189 & 6.0547 \\ 6.9453 & 7.5811 & 8.4189 & 9.0547 \end{bmatrix}$$

The same method is then repeated in the downward direction and starts by adding two zeroes at the edge of each column instead of row, as the extension is now conducted downward:

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0.9453 & 1.58118 & 2.4189 & 3.0547 \\ 3.9453 & 4.5811 & 5.4189 & 6.0547 \\ 6.9453 & 7.5811 & 8.4189 & 9.0547 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{21}$$

And the matrix from Equation (21) is extended and shifted:

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0.9453 & 1.5811 & 2.4189 & 3.0547 \\ 0.9453 & 1.5811 & 2.4189 & 3.0547 \\ 3.9453 & 4.5811 & 5.4189 & 6.0547 \\ 3.9453 & 4.5811 & 5.4189 & 6.0547 \\ 6.9453 & 7.5811 & 8.4189 & 9.0547 \\ 6.9453 & 7.5811 & 8.4189 & 9.0547 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

And the final resized image is calculated by implementing weighted average, and the output of the interpolation is obtained as:

$$\begin{bmatrix} 0.7813 & 1.4170 & 2.2549 & 2.8906 \\ 2.6885 & 3.3242 & 4.1621 & 4.7979 \\ 5.2021 & 5.8379 & 6.6758 & 7.3115 \\ 7.1094 & 7.7451 & 8.5830 & 9.2188 \end{bmatrix}$$

In the next step of the fusion method soft-max is used to obtain the final weight maps.

$$w_k^i(x,y) = \frac{I_k^{i,*}(x,y)}{I_1^{i,*}(x,y) + I_2^{i,*}(x,y) + I_3^{i,*}(x,y)}$$

After the weight maps have been transformed into correct size, equal to input images, they are multiplied with each corresponding source image and fused with weighted averages:

$$F(x,y) = \sum_{k=1}^{3} w_k^i(x,y) * I_k(x,y) \tag{22}$$

An overview of method 2 can be seen in Figure 12.

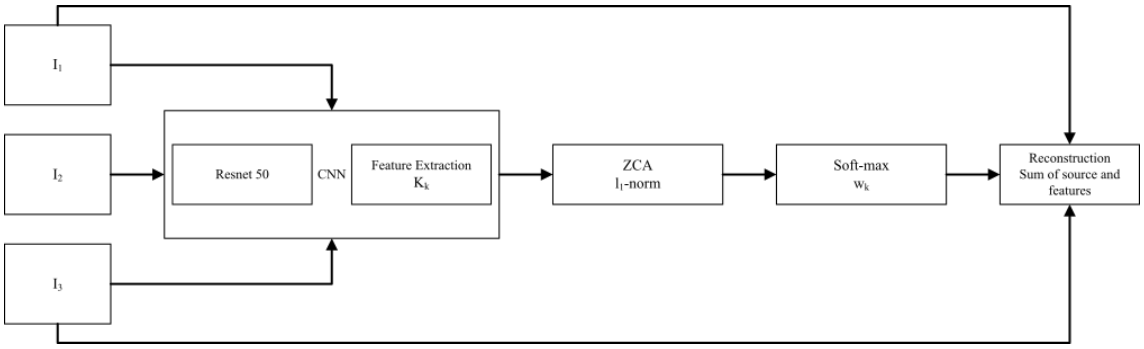


Figure 12: Illustration of Method 2 showing three input images. Features from the input source images are extracted using a CNN, and the features are furthered processed using ZCA and Soft-max to select which sensor have the most prominent features. Final output is then reconstructed by fusing selected features with base informations from all sensors.

5.6.4 Method 3: Convolutional Sparse Representation (CSR)

CSR by Liu et al. [6] is suggested as the third method to be evaluated. The method implements a dictionary learning method from Wohlberg [74]. This method takes K number of source images (denoted as $I_k, k \in \{1, ..., K\}$), in this thesis $K = 3$ and the dictionary filters are denoted as $d_m, m \in \{1, ..., M\}$ All source images are divided into two layers, a base layer I_k^b and a detail layer I_k^d. The layers are separated by Equations (6) and (7) explained in section 5.6.2. The detailed layers are then further processed to calculate sparse coefficient maps. Where the coefficient maps are denoted as $C_{k,m}$ and d_m represents a set of filters from the dictionary and $m \in \{1, ..., M\}$ from Wohlberg's method [74] method:

$$\arg\min_{C_{k,m}} \frac{1}{2} \left|\left| \sum_{m=1}^{M} d_m * C_{k,m} - I_d^k \right|\right|_2^2 + \alpha \sum_{m=1}^{M} ||C_{k,m}||_1 \ \alpha = 0.01 \tag{23}$$

$C_{k,1:m}(x,y)$ represents sparse coefficients at a position of x, y with M dimensions. $l_1 - norm$ is then used for each of the calculated $C_{k,1:M}$ values to create a pixel level measure map for each of the pixels in the source images and $A_k(x,y)$ is obtained:

$$A_k(x,y) = ||C_{k,1:m}(x,y)||_1 \tag{24}$$

To decrease the possibility of misregistration of $A_k(x, y)$ an averaging strategy of all positions is used to calculate $\bar{A}_k(x, y)$. Where r represents window size and should be as big as possible to avoid misregistration. Yet small enough that details still appear, for multimodal fusion that often contains small features a small value of r is preferred:

$$\bar{A}_k(x, y) = \frac{\sum_{p=-r}^{r} \sum_{q=-r}^{r} A_k(x + p, y + q)}{(2r + 1)^2}$$

To fuse the coefficient maps, the max value is chosen:

$$C_{f,1:M}(x, y) = C_{k*,1:M}(x, y), \; k* = \arg\max_k (\bar{A}_k(x, y)) \tag{25}$$

And the detail layers are fused by:

$$F^d = \sum_{m=1}^{M} d_m * C_{f,m}$$

There are two possible fusion methods for the base layer, "choose-max" similar to Equation (25), or an averaging strategy. In regards of the base layer, the averaging strategy is applied as it's deemed more appropriate for multimodal fusion:

$$F^b(x, y) = \frac{1}{K} \sum_{k=1}^{K} I_k^b(x, y)$$

Lastly having obtained both the detail layer I_f^d and the base layer I_f^b, The fused image I_f is obtained by adding the two:

$$F = F^d + F^b$$

For an overview of the fusion process for method 3 see Figure 13.

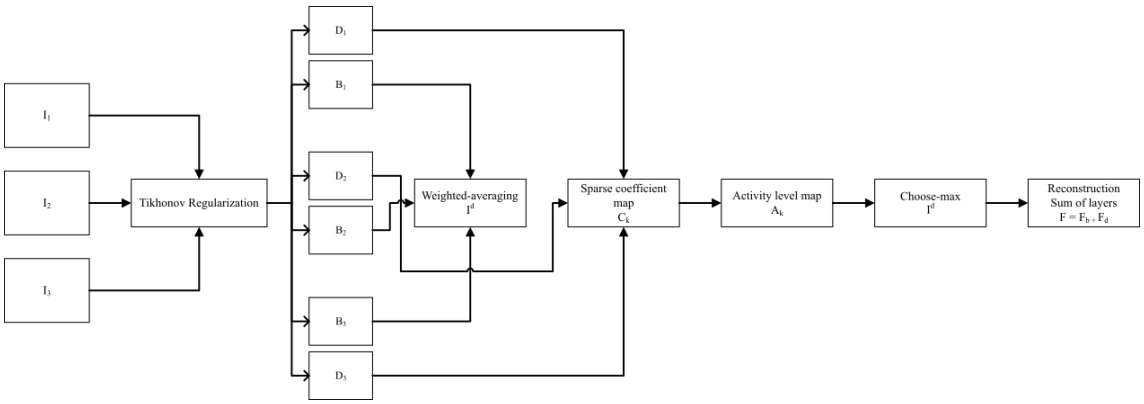


Figure 13: Illustration of Method 3 with three images as input. The input images are divided into base and detail layers by solving the Tikhonov Regularization problem. All detail layers are used to create sparse detail maps which are in turn used to create activity level maps. The sensor with highest values in the activity level map will then be selected at apropriate position, and fused with the averaged values from all base levels.

5.6.5 Method 4: Saliency Map with Weighted Least Squared Optimization (WLS)

Ma et al. [75] propose a conventional method for image fusion based on MSD and Visual Saliency Map (VSM). The method decomposes IR images and visual images into base layers and detailed layers using MSD. Decomposing is accomplished by utilizing a method proposed by Zhang et al. [76] named Rolling Guidance Filter (RGF). The motivation for RGF over other MSD methods separating detailed features from background structures is that RGF iteratively preserves both

edges and scale. For the number of input images $K = 3$ the images $I_k,\ k \in \{1, .., K\}$ is fed through a Gaussian filter G where x and y is the pixel coordinate system of the image and σ_s is the standard deviation, see Equation (26), removing small structures.

$$G(x) = \frac{1}{P_x} \sum_{y \in N(x)} exp\bigg(- \frac{||x - y||^2}{2\sigma_s^2} \bigg) I(y), \sigma_s = 0.05 \tag{26}$$

With normalization ($N(x)$ is a set surrounding pixels),

$$P_x = \sum_{y \in N(x)} exp\bigg(- \frac{||x - y||^2}{2\sigma_s^2} \bigg)$$

The next step is edge-recovery from the work of Zhang et al. [76]. The rolling guidance filter is given by Equation (27) where J^{t+1} (t-th iteration) is guided by the previous step J^t. J^t is the input image processed by Equation (26).

$$J^{t+1}(x) = \frac{1}{P_x} \sum_{y \in N(x)} exp\bigg(- \frac{||x - y||^2}{2\sigma_s^2} - \frac{||J^t(x) - J^t(y)||^2}{2\sigma_s^2} \bigg) \tag{27}$$

Where,

$$P_x = \sum_{y \in N(x)} exp\bigg(- \frac{||x - y||^2}{2\sigma_s^2} - \frac{||J^t(x) - J^t(y)||^2}{2\sigma_s^2} \bigg)$$

The image is processed with Equation (26) and Equation (27) for the number of decomposition levels l, in this thesis $l = 4$. When the base layers B_1, B_2 and B_3 given by J^{l+1} together with the detailed layers $d_1^1 ... d_1^l \in D_1$, $d_2^1 ... d_2^l \in D_2$ and $d_3^1 ... d_3^l \in D_3$ are obtained the base layers can be fused. The fusion of base layers are obtained by a weighted averaging strategy based on VSM by Ma et al. [75], this provides contrast information and overall structure of the fused image. The method operates at pixel-level and finds saliency features $V(p)$ based on a pixel p contrast to all other pixels in the image and is given by Equation (28). Assigning j for pixel intensity and M_j for the number of pixels with equal intensity, lastly L to the number of gray levels gives:

$$V(p) = \sum_{j=0}^{L-1} M_j |I_p - I_j| \tag{28}$$

The weight maps for each sensor are constructed as follows, with V_1, V_2 and V_3 representing the VSM of input source images:

$$\hat{B}_k = V_k * B_k,\ k \in \{1, 2, 3\} \tag{29}$$

Comparison between all sensors is achieved using max absolute strategy to calculate the coefficient matrix W_k^b and select the sensor with the highest pixel value at position x, y in the image. First coefficient matrices are created in Equation (30) for each of the sensors:

$$\begin{aligned} W_1^b &= \begin{cases} 1, & |\hat{B}_1 > \hat{B}_2| \ \& \ |\hat{B}_1 > \hat{B}_3| \\ 0, & otherwise \end{cases} \\ W_2^b &= \begin{cases} 1, & |\hat{B}_2 > \hat{B}_1| \ \& \ |\hat{B}_2 > \hat{B}_3| \\ 0, & otherwise \end{cases} \\ W_3^b &= \begin{cases} 1, & |\hat{B}_3 > \hat{B}_1| \ \& \ |\hat{B}_3 > \hat{B}_1| \\ 0, & otherwise \end{cases} \end{aligned} \tag{30}$$

And finally all base layers are fused together by applying max-absolute rule:

$$B_F = \sum_{n=1}^{K} W_n^b * \hat{B}_n \tag{31}$$

The detailed layers are combined with the WLS by Farbman et al. [77], fusing detailed data. Typically fusing an IR image with a visual image with max-absolute strategy, noise tends to appear in the result from the IR image and can appear grainy. With WLS optimization, the fused image can result in a more homogeneous image with less background noise. First, coefficients from the max-absolute strategy W^k is obtained, followed with a Gaussian smoothing of the coefficients W_d^k with $\sigma = 2$.

$$\begin{aligned} W_1^{d,1:l} &= \begin{cases} 1, & |d_1^{1:l} > d_2^{1:l}| \ \& \ |d_1^{1:l} > d_2^{1:l}| \\ 0, & otherwise \end{cases} \\ W_2^{d,1:l} &= \begin{cases} 1, & |d_2^{1:l} > d_1^{1:l}| \ \& \ |d_2^{1:l} > d_3^{1:l}| \\ 0, & otherwise \end{cases} \\ W_3^{d,1:l} &= \begin{cases} 1, & |d_3^{1:l} > d_1^{1:l}| \ \& \ |d_3^{1:l} > d_2^{1:l}| \\ 0, & otherwise \end{cases} \end{aligned} \tag{32}$$

Then applying max-absolute rule:

$$M^k = \sum_{n=1}^{K} \prod_{i}^{l} W_n^{d,i} * d_n^i \tag{33}$$

Equation (34) is a WLS cost function, the method minimizes the function to obtain D^j. The function minimizes Euclidean distance, between the detailed layer and the result obtained in Equation (33).

$$\sum_p \left((D_p^j - M_p^j)^2 + \lambda a_p^j \big(D_p^j - (d_2^j)_p\big)^2 \right) \tag{34}$$

Let ω_p represent a window centered around pixel p. If the window is to large blur can occur in the result, and if the window is too small noise can still be present. Given ω_p and ϵ as a small constant a_p^j is obtained by:

$$a_p^j = \left(\left| \sum_{q \in \omega_p} (d_1^j)_q \right| + \epsilon \right)^{-1} \tag{35}$$

If D^j, M^j and d_2^j is represented as vectors together with a diagonal matrix A^j as weights of a^j, Equation (34) can be expressed as:

$$(I + \lambda A^j) D^j = M^j + \lambda A^J d_2^j, \quad A^j = (A^j)^T \tag{36}$$

The last step is to reconstruct the image F as the fused result:

$$F = B_F + D^1 + ... + D^N$$

See Figure 14 for an overview of Method 4.

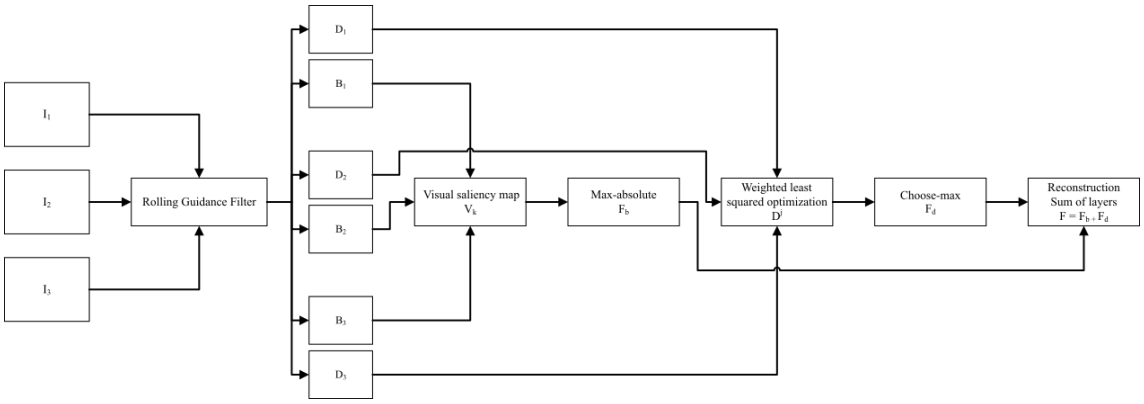


Figure 14: Illustration of Method 4 with three images as input. By applying a rolling guidance filter, source images are divided into base and detail levels. Visual saliency maps are created and max absolute rule is applied to select correct sensors input for the base layer at the corresponding position. For the detail layers choose max is implemented and the most prominent sensor is selected. Finally all detail and base layers are fused to create a fused output.

5.6.6 Summary of Image Fusion Methods

Generally, there is a common thread amongst state-of-the-art fusion methods, including methods evaluated in this thesis. In practice, a fusion process begins with a separation of ”base layers” and ”detailed layers” with base layers meaning structures with a low frequency and detailed layers having a high frequency and rate of change compared to surrounding pixels. The process continues with a fusion of the base layers from the different image sources. This process is often a more straightforward fusion process than the fusion of detailed layers. The fused base layer provides the overall appearance and contrast of the image. The processing of detailed layers varies between methods, but the overall goal is to extract silent features from the extracted detail layer, representing them in layers as a subset of the detailed layer. Two of the methods evaluated in this thesis utilizes CNN for feature detection and extraction. The other two methods extract the features with conventional mathematical models and filters. The extracted features from the corresponding source image are fused to generate a detailed fused layer. Finally, the base layer and detailed layer are fused.

5.7 Image Quality Metrics

This section explains the methods used as quality metrics of the fused image. All methods are NR-IQA methods, as no comparison can be made to an optimally fused image. Two of the methods compare the output of the fusion method with the inputs from sensors. However, one method compares on a pixel-level, whereas the other compare edges, as edges are perceived as information. The third of the methods bases results of induced artifacts in the output image that have no way of appearing naturally.

5.7.1 Fast Feature Mutual Information

Fast-FMI is an NR-IQA method proposed by Haghighat and Razian [35] as an improvement of the method by Haghighat et al. [78], regarding complexity reduction. Fast-FMI is a metric on the similarity and dependency between two variables. It is measured with the Kullback Leibler divergence [79]. The Kullback Leibler divergence formula calculates the mutual information $I(X;Y)$ where X and Y is two random variables with a joint probability mass function $p(x,y)$ together with marginal probability mass functions $p(x)$ and $p(y)$, see Equation (37).

$$I(X;Y) = \sum_{x \in X} \sum_{y \in Y} p(x,y) log \frac{p(x,y)}{p(x)p(y)} \tag{37}$$

Features are extracted from the two source images and the fused image with metrics from Haghighat et al. [78]. With the features extracted $p(a)$, $p(b)$ and $p(f)$ is created with Marginal Probability Density Functions (MPDF). $p(a,f)$ and $p(b,f)$ is created with the Joint Probability Density Function (JPDF). JPDF is an estimate from $p(a)$ with $p(f)$ and $p(b)$ with $p(f)$. Furthermore, a JPDF estimation is quadratically related to MPDF regarding complexity and size of calculations. The issue of the complexity of the JPDF function is the motivation for Fast-FMI over FMI. Haghighat and Razian [35] reduces the overhead with a sliding window that calculates portions of the source and fused images. The calculated regional FMIs is normalized with Equation (38) where H is the entropy of the regional window. Lastly, the final Fast-FMI is the average of the regional FMIs.

$$\frac{H_i(A) + H_i(F)}{2} \tag{38}$$

5.7.2 Edge Preservation Quality Index

Xydeas and Petrović [36] propose a method for evaluating image quality at the pixel level. This method is based on a Sobel edge operator to find the approximate derivative of the input image in y- and x-direction. The Sobel kernels are defined in Equation (39).

$$G_x = \begin{vmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{vmatrix} * A \quad G_y = \begin{vmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{vmatrix} * A \tag{39}$$

The Sobel operator is applied on all pixels of the image $p(m,n)$ where $n \in N$ and $m \in M$ if the input image A has a size of NxM. The approximations combined gives the total magnitude of the gradient by applying Equation (40). The direction of the gradient can also be calculated with Equation (41).

$$G(n,m) = \sqrt{G_x^2(n,m) + G_y^2(n,m)} \tag{40}$$

$$\theta = atan\left(\frac{G_y(n,m)}{G_x(n,m)}\right) \tag{41}$$

The relative magnitude and direction of G^{AF} and Θ^{AF} with F being the fused image are given by:

$$G^{AF}(n,m) = \begin{cases} \frac{G_F(n,m)}{G_A(n,m)} & if \quad G_A(n,m) > G_F(n,m) \\ \frac{G_A(n,m)}{G_F(n,m)} & otherwise \end{cases} \tag{42}$$

The quality $\mathbf{Q^{AB/F}}$ index is based on the preservation of edge strength and orientation defined as:

$$Q^{AF}(n,m) = Q_G^{AF}(n,m) * Q_\Theta^{AF}(n,m) \tag{43}$$

where,

$$Q_G^{AF}(n,m) = \frac{\Gamma_G}{1 + e^{K_G(G^{AF}(n,m) - \sigma_G)}} \tag{44}$$

$$Q_\Theta^{AF}(n,m) = \frac{\Gamma_\Theta}{1 + e^{K_\Theta(\Theta^{AF}(n,m) - \sigma_\Theta)}} \tag{45}$$

Finally, the weighted and normalized metric is given by Equation (46) with $w(n,m) = [G(n,m)]^L$ where L is a constant.

$$Q^{AB/F} = \frac{\sum_{n=1}^{N}\sum_{m=1}^{M} Q^{AF}(n,m)w^A(n,m) + Q^{BF}(n,m)w^B(n,m)}{\sum_{i=1}^{N}\sum_{j=1}^{M}(w^A(i,j) + w^B(i,j))} \tag{46}$$

5.7.3 Natural Image Quality Evaluator

The last quality metric used in this thesis is NIQE from Mittal et al. [56]. This method is based on Natural Scene Statistics (NSS) [80]. NSS assumes that a distinction can be made between images captured from the real world and computer-generated images. In essence, processing images e.q. fusion can introduce errors that deviate from a naturalistic scene. The method divides the image in patches PxP and calculates NSS features from the patch coefficients. The patches are then fitted to a Multi-variate Gaussian (MVG) model and comparing the result to a natural MVG model. The method is not bound to any specific distortion variable compared to similar methods e.g., Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE). The spatial NSS model used in this method pre-processes the image with the removal of local mean and divisive normalization Equation (47) where w is circularly-symmetric Gaussian weighting function.

$$\hat{I}(i,j) = \frac{I(i,j) - \mu(i,j)}{\sigma(i,j) + 1} \tag{47}$$

$$\mu(i,j) = \sum_{k=-K}^{K}\sum_{l=-L}^{L} w_{k,l} I(i+k, j+l) \tag{48}$$

$$\sigma(i,j) = \sqrt{\sum_{k=-K}^{K}\sum_{l=-L}^{L} w_{k,l}[I(i+k,j+l) - \mu(i,j)]^2} \tag{49}$$

The final quality index NIQE is given with Equation (50).

$$D(v_1, v_2, \Sigma_1, \Sigma_2) = \sqrt{\left((v_1 - v_2)^T \left(\frac{\Sigma_1 + \Sigma_2}{2}\right)^{-1} (v_1 - v_2)\right)} \tag{50}$$

where v and Σ represents the mean and covariance matrix from the MVG model given in Equation (51).

$$f_x(x_1, ..., x_k) = \frac{1}{(2\pi)^{k/2}|\Sigma|^{1/2}} exp\Big(-\frac{1}{2}(x-v)^T\Sigma^{-1}(x-v)\Big) \tag{51}$$

5.7.4 Summary of Image Quality Metrics

The objective quality metrics in this thesis evaluates the task from different viewing points. Fast-FMI calculates mutual information on the feature level. The $\mathbf{Q^{AB/F}}$ index estimates image sharpness and edges on pixel level with Sobel operations. NIQE is not bound to a specific parameter and estimates induced error from image processing with the assumption that computer-generated errors diverse from natural images.

5.8 Object Detection

RQ4 requires the ability to detect objects present in the output from the evaluated image processing techniques. A comprehensive evaluation of object detection is out of the scope of this thesis. However, a provided object detection algorithm, based on Faster Residual-CNN architecture with a ResNet50-Feature pyramid network backbone, is used to evaluate the detection capacity. The algorithm is capable of detecting objects surrounding an airfield, including other aircraft, runways, and vehicles. The algorithm was initially trained on the same dataset used to fuse images in this thesis. A total of 8 images from each evaluated method is analyzed, with PyTorch Detectron 2 Application Program Interface (API) [81]. It's important to note that the network is not trained on the output from this thesis fused images.

5.9 Software Deployment

Software development in the avionics domain follows RTCA DO-178/C in compliance with ARP4754. A rigorous development process written by the standards and guidelines above is out of scope for this work. This thesis aims to evaluate different techniques in a "concept phase." Where in the first phase, simulations are conducted using MATLAB. In the second phase, adaption to the programming language C++ is conducted to evaluate the performance of the algorithms in a language commonly used in avionic systems. MATLAB coder [82] is applied to generate C++ code, and all previously implemented MATLAB code is optimized to ensure code generator compatibility. The generated code is only utilized for the fusion part of the program, and images are read using openCV. It should be noted, however, that the software is not deployed and executed on a CPU graded for avionic systems. However, deploying the software on specific hardware only requires minor changes. Further, the generated MATLAB code is not optimized for execution on embedded software systems with limited hardware capabilities. The software is also optimized to a minor extent using GPU acceleration with CUDA and OpenCL.

6 Ethical and Societal Considerations

The purpose of this thesis is to investigate technologies, to increase the safety of aircraft. The technology may be used and implemented in other areas without the knowledge of the authors, but it should be known that it is not the purpose of the technology.

As the test involved humans, all participating subjects were informed of the purpose of the thesis and asked for consent. Further, no data regarding identities is saved after the completion of the experiment, and no correlation between answers and identities is published in this paper. All participants were also informed that they could withdraw at any time during the test. It should be noted, however, as the correlation between answers and identities was removed when the test was completed that it was no longer possible to withdraw a single subject's answers.

7 Results

This section presents data and findings from work conducted in the thesis in a raw format. The data is obtained from the methods in section 5 and provides a basis for conclusions.

7.1 Systematic Review

Table 4 show a summary of the systematic review, where "HSR" is short for Human Subjective Ranking. Meaning that the methods in the literature are compared to human visual ranking.

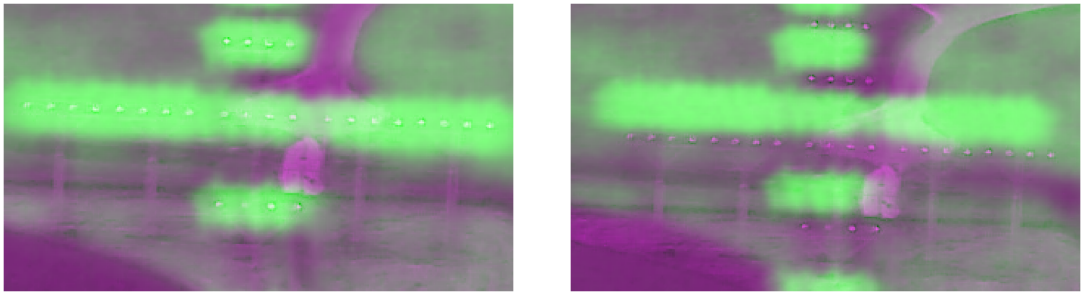
Author	Aims	HSR	Method	Conclusion
[53] Eskicioglu, Fisher. 1995	Evaluate quantitative image quality metrics.	Yes	Experiment by comparing algorithm with a set of images.	It is not possible to compare different techniques based on that one data point cannot describe several features.
[54] Pappas, Safranek. 1999	Migrate from MSR and PSNR to objective quality assessment based on human perception.	Yes	Examine objective criteria of existing models.	General models are more complex than specific methods. Most models are designed for a specific application.
[49] Wang, Bovik. 2002	Propose a new method for IQA with a mathematically approach that aims to be universal.	Yes	Testing the proposed method with "Lena" image.	The new method achieves better results comparing to MSR. However, a lot of research remains to be conducted but this is a great starting point.
[50] Wang et al. 2002	Propose a new philosophy of IQA that instead of looking at image errors, checks for structural errors.	Yes	Testing the philosophy with a simple evaluation of "Lena image.	Implementation with the new philosophy provides great results.
[55] Wang et al. 2003	Develop a multi-scale structural similarity method.	Yes	Comparing scales of different factors.	The multi-scale methods gains better result compared to single-scale with correct parameter settings.
[51] Wang et al. 2004	Finding a new approach to assess image quality based on structural similarity and not error visibility.	Yes	Construct a new quality index based on the philosophy by Wang et al. [50].	Some challenges remains, one that the SSIM index requires optimization for general purpose use. However, the performance is promising compared to traditional methods.

Table 4: Summary of systematic review in table format.

Author	**Aims**	**HSR**	**Method**	**Conclusion**
[48] Wang et al. 2006	Collect and summarize the current progress of IQA.	Yes	Evaluate FR-, RR- and NR-IQA together with challenges and possibilities.	A book of the current state-of-the-art practice with its conclusions.
[52] Hassen et al. 2010	Evaluate local phase coherence for image sharpness assessment.	Yes	Defining an image sharpness index.	This method decouples sharpness and blurriness.
[56] Mittal et al. 2012	Develop an NSS-based framework for distortions that are completely blind.	Yes	Fitting a collection of quality aware features in a multivariate Gaussian model.	The new method NIQE outperforms established full reference IQA.
[57] Fang et al. 2015	Develop a quality metric for contrast distorted images based on natural scene statistics.	Yes	Development of a NSS model with focus on contrast distortion.	NSS models requires future research but gained promising results for contrast distorted images.
[58] Gu et al. 2016	Evaluation of no-reference contrast distorted images metrics.	Yes	Designing a blind NIQMC with information maximization.	Promising results with the proposed method. Performing better than classical FR-IQA.
[59] Bosse et al. 2017	Utilizing CNNs for IQA.	Yes	Develop a network for FR and then adopting it to NR.	Experimental results outperforms traditional methods. However, there is room for improvements such as balancing ratio between parameters.

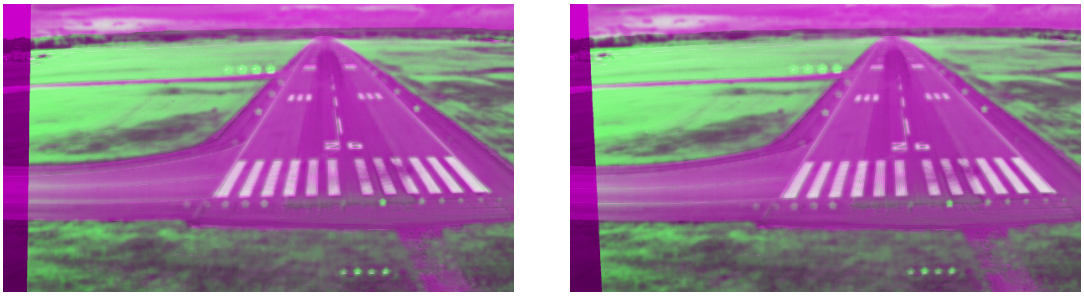
7.2 Image Registration

In this section results for image, registration are presented. The tests together with generated plots were run with MATLAB. The images in Figure 15 and 16 are presented with false-color overlay. One image has a magenta mask, and the other image has a green mask. When the intensity of the two images placed on top of each other is equal in an area, the color becomes gray/white. Otherwise, the image with higher intensity becomes the dominant color. False-color overlay is a good technique to distinguish features from their origin in different images. The false-color overlay also helps separate how one image is transformed when placed on top of the reference image, seen in Figure 15 and 16.



(a) Manual feature detection (b) SURF feature detection

Figure 15: False color overlay with runway lights comparing manual selection of features (15a) and SURF feature selection (15b).



(a) Affine (b) Projective

Figure 16: False color overlay shows comparison of Affine transformed image (16a) and Projective transormed image (16b).

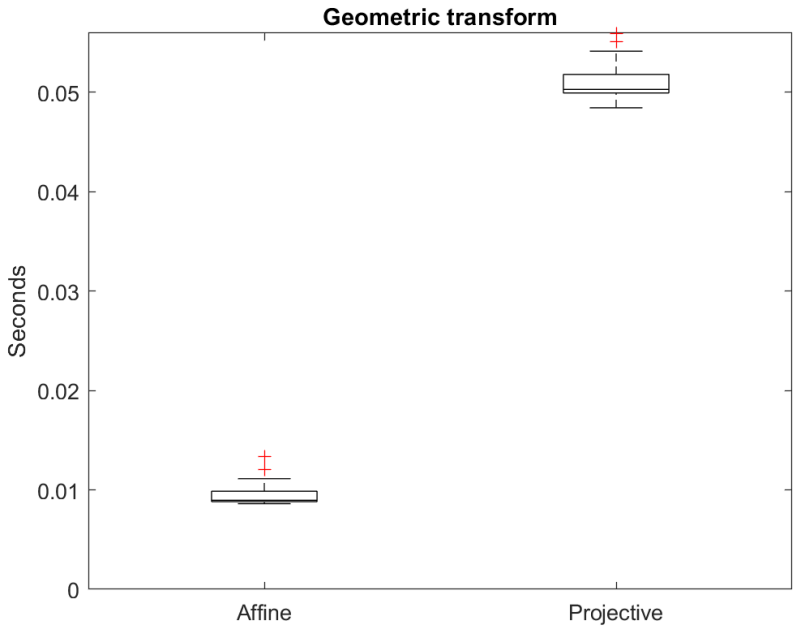


Figure 17: Boxplot of elapsed time during transform of images with a dataset of 23 images.

7.3 Image Fusion

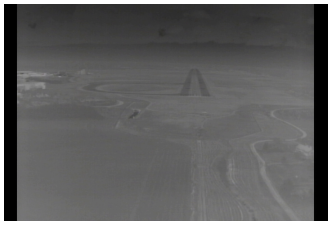
This section presents the major contribution of our work, image fusion. First, in Table 5 the relationship between IQA and image input size is displayed. Second, Figure 18 to 23 shows results of the fusion process with all four methods with respective input images. Its a result of the combination of three sensors with different spectral ranges sensing unique salient features. Furthermore, box-and-whisker plots are included in this section regarding the performance of the fused images with selected quality metrics together with time complexity plots (Figure 24 to 30). Lastly, a comparison between the generated C++ code and Matlab code is presented in Figure 31.

Table 5: This table shows the relation between input size of images and results from quality metrics. The size of images directly affects the quality metrics. A valid comparison between fusion methods requires the same input size of images.

Image Size Pixels	**Fast-FMI**	**Q^A/BF**	**NIQE**
94x54	0.85501	0.26116	18.8802
185x108	0.87332	0.26159	18.8784
278x162	0.89827	0.24856	6.44468
370x216	0.91069	0.22964	5.17057
462x270	0.91752	0.21760	5.62721
555x324	0.92176	0.20884	5.14801
739x432	0.92883	0.19509	3.60057
832x468	0.93174	0.19215	3.46398
924x540	0.93365	0.19380	3.15547
1016x594	0.93626	0.19456	3.03400
1109x648	0.93861	0.19746	2.79146
1201x702	0.93998	0.19855	2.85024
1293x756	0.94078	0.20100	2.83964
1386x810	0.94177	0.20308	2.67124
1478x864	0.94306	0.20467	2.75511
1570x918	0.94340	0.20650	2.97351
1663x972	0.94446	0.20845	2.84228
1755x1026	0.94532	0.21023	2.77491
1847x1079	0.94574	0.21340	3.02118

(a) Input visual image.
Size: 1920x1080 pixels.

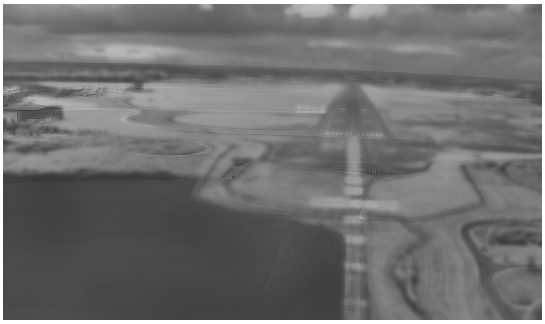
(b) Input SWIR image.
Size: 720x480 pixels.



(c) Input LWIR image.
Size: 720x480 pixels.

Figure 18: Input images for Figure 19

(a) Fused image with Method 1, section 5.6.2
Size: 1847x1079 pixels.



(b) Fused image with Method 2, section 5.6.3
Size: 1847x1079 pixels.

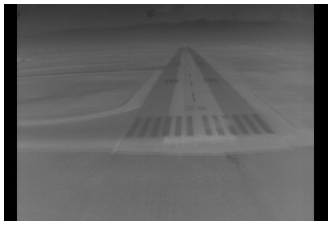
(c) Fused image with Method 3, section 5.6.4
Size: 1847x1079 pixels.

(d) Fused image with Method 4, section 5.6.5
Size: 1847x1079 pixels.

Figure 19: Fused images of Visual camera, SWIR and LWIR. The images displays the final approach of a landing sequence. The most salient features in this image is the ALS lights at the beginning of the runway including the 1000FT roll bar. At the left of the runway precision approach path indicators is also visible.

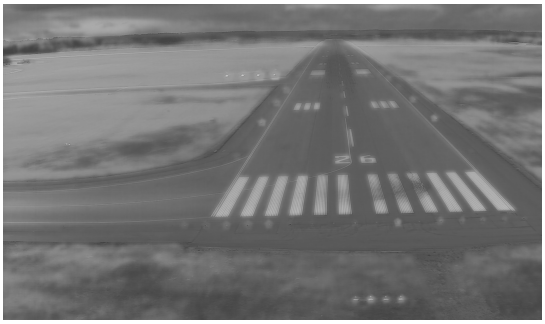
(a) Input visual image.
Size: 1920x1080 pixels.

(b) Input SWIR image.
Size: 720x480 pixels.

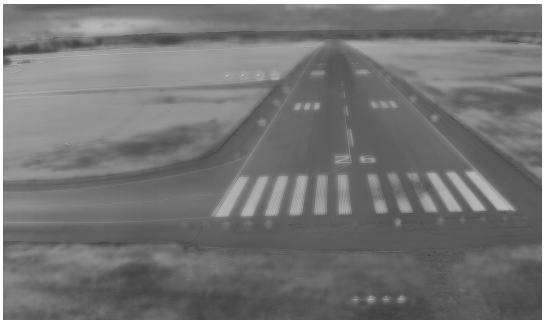


(c) Input LWIR image.
Size: 720x480 pixels.

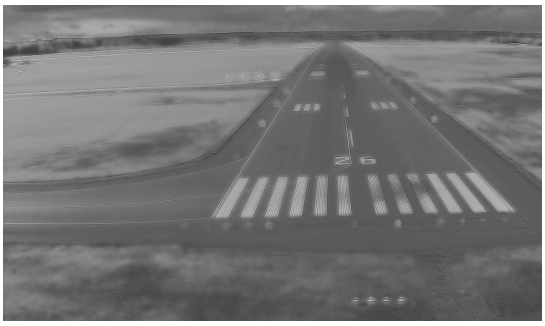
Figure 20: Input images for Figure 21



(a) Fused image with Method 1, section 5.6.2
Size: 1847x1079 pixels.



(b) Fused image with Method 2, section 5.6.3
Size: 1847x1079 pixels.



(c) Fused image with Method 3, section 5.6.4
Size: 1847x1079 pixels.

(d) Fused image with Method 4, section 5.6.5
Size: 1847x1079 pixels.

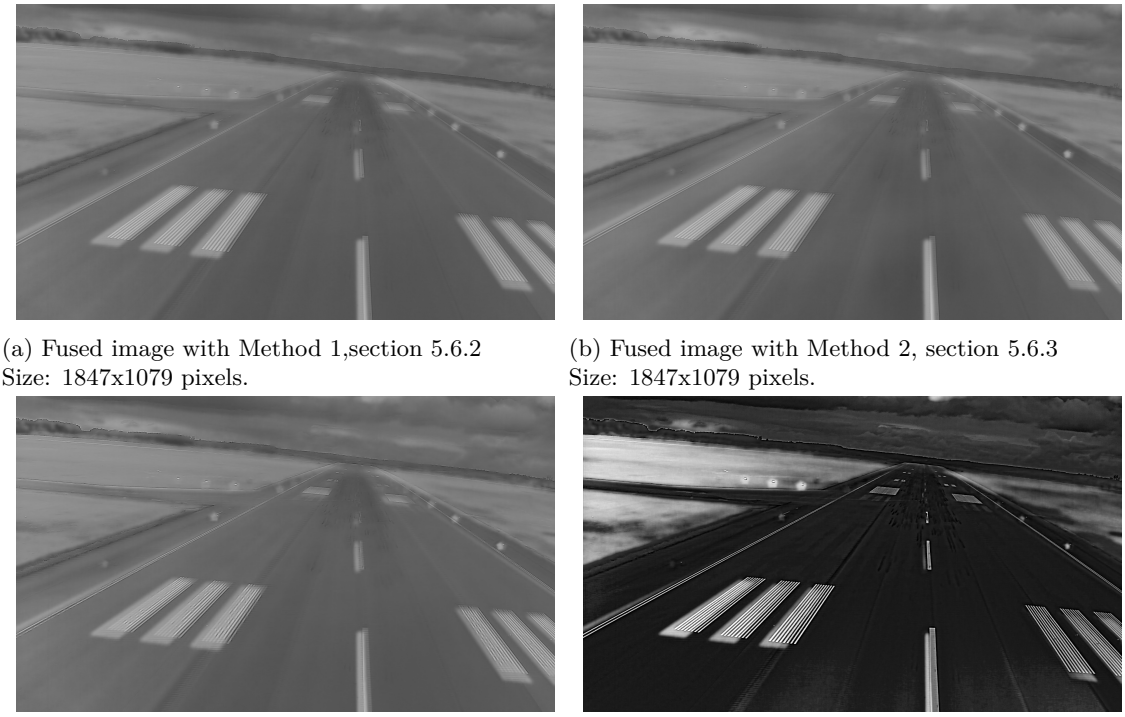
Figure 21: Fused images of Visual camera, SWIR and LWIR. This images show a close up on the runway, with the markings clearly visible. For example the precision approach path indicators can be identified at the left and at the bottom the runway threshold lights can be seen.

(a) Input visual image.
Size: 1920x1080 pixels.

(b) Input SWIR image.
Size: 720x480 pixels.

(c) Input LWIR image.
Size: 720x480 pixels.

Figure 22: Input images for Figure 23



(a) Fused image with Method 1,section 5.6.2
Size: 1847x1079 pixels.

(b) Fused image with Method 2, section 5.6.3
Size: 1847x1079 pixels.

(c) Fused image with Method 3, section 5.6.4
Size: 1847x1079 pixels.

(d) Fused image with Method 4, section 5.6.5
Size: 1847x1079 pixels.

Figure 23: Fused images of Visual camera, SWIR and LWIR. The images show an overview of the runway. With the precision approach path indicators and runway edge light's visible. Finally, the runway markings is also prominent in the picture.

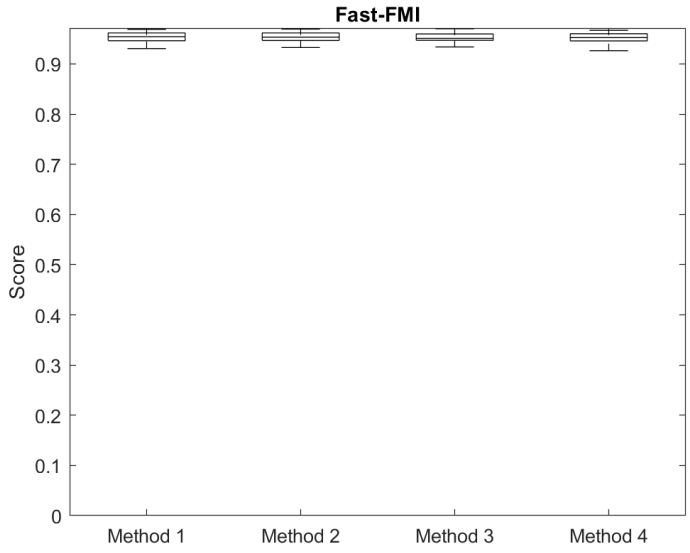


Figure 24: Boxplot of Fast-FMI. The y-axis indicates the percentage of features present in the fused image with respect to the source images. Hence, a score of 1 implies that all features from the source image are present in the fused result. The method used to find features are edge detection.

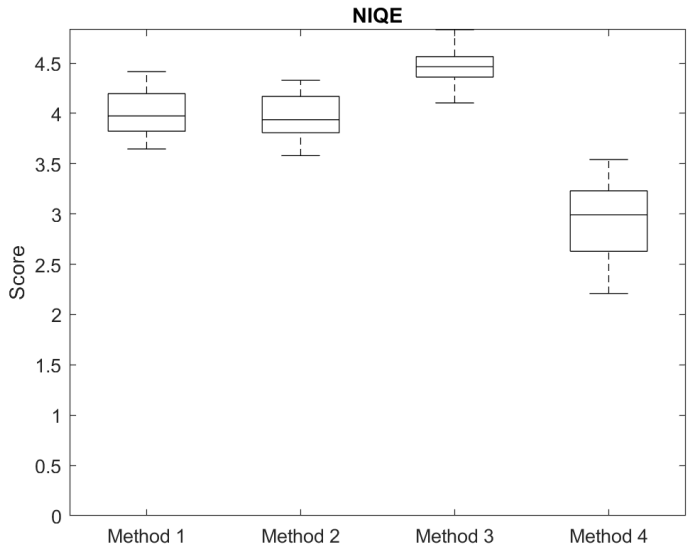


Figure 25: Boxplot of the NIQE evaluation. The y-axis is the NIQE score, a lower score indicates a better result. The score reflects the perceptual quality of the output image with respect to the NIQE model. In this thesis the model is based on NSS [80].

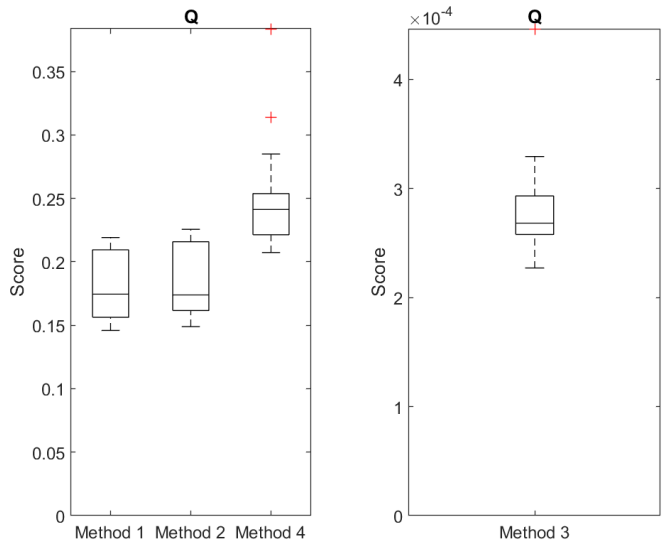


Figure 26: Boxplot of $Q^{AB/F}$. This index is similar to the Fast-FMI metric but operates on pixel-level, thus more sensitive to small dissimilarities. The y-axis comparable to Fast-FMI ranges from 0 to 1, with 1 meaning that all edges on pixel-level are persevered in the fused image from the source images.

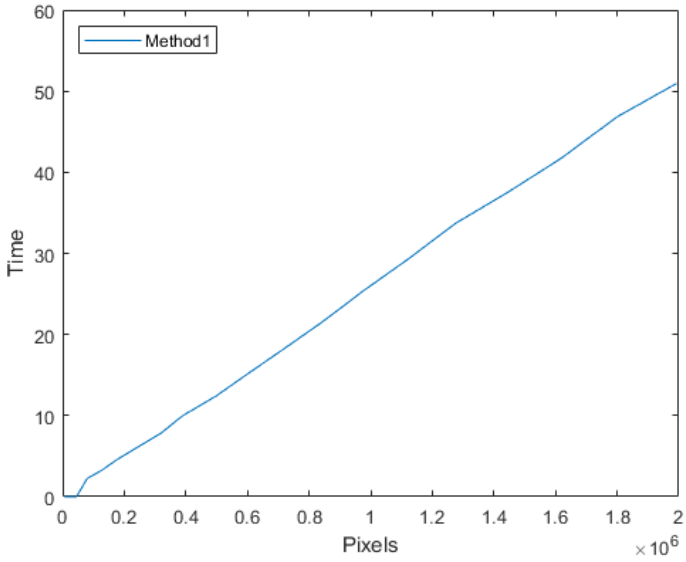


Figure 27: This plot illustrates the growth of time with larger image input for fusion Method 1. The x axis represents pixels With x Height of source/result image, and the y axis is the measured time in seconds of the fusion process. This line is approximately quasilinear in time, and the algorithm is estimated to $O(n \log n)$.

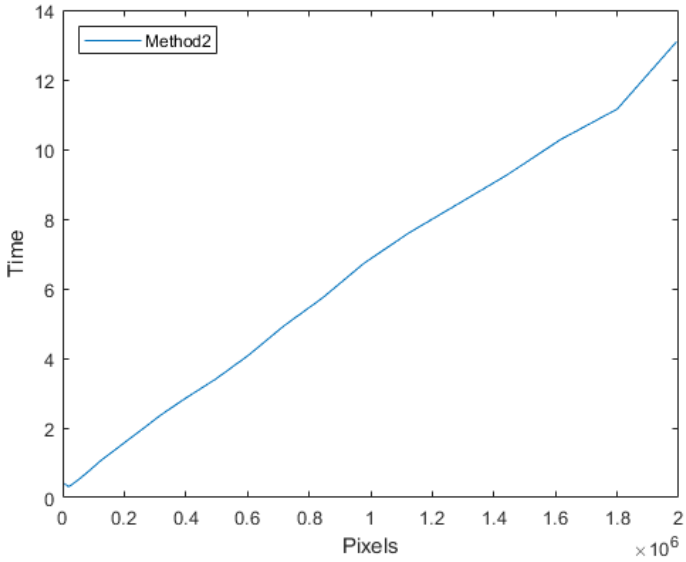


Figure 28: This plot illustrates the growth of time with larger image input for fusion Method 2. The x axis represents pixels With x Height of source/result image, and the y axis is the measured time in seconds of the fusion process. This line is approximately quasilinear in time, and the algorithm is estimated to $O(n \log n)$.

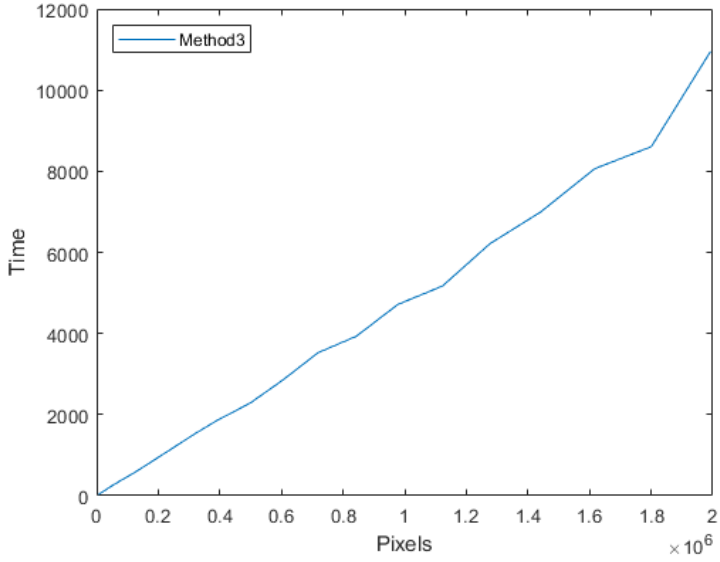


Figure 29: This plot illustrates the growth of time with larger image input for fusion Method 3. The x axis represents pixels With x Height of source/result image, and the y axis is the measured time in seconds of the fusion process. This line is approximately quasilinear in time, and the algorithm is estimated to $O(n \log n)$.

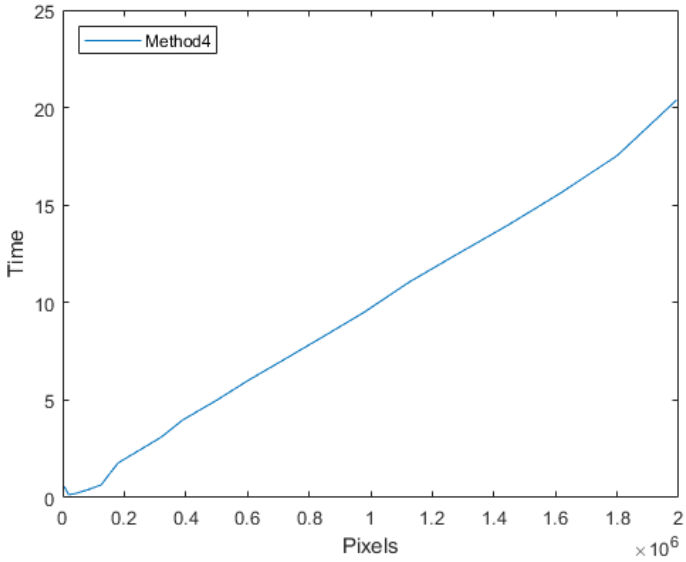
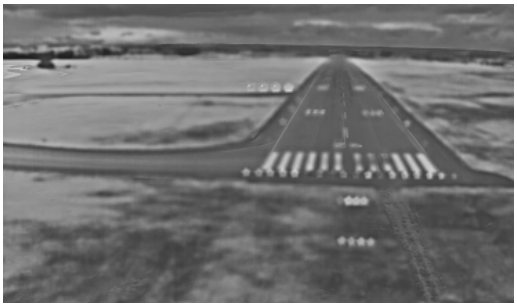
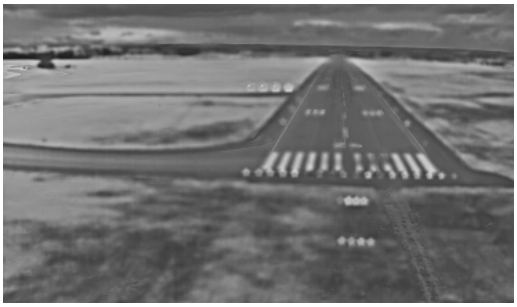


Figure 30: This plot illustrates the growth of time with larger image input for fusion Method 4. The x axis represents pixels With x Height of source/result image, and the y axis is the measured time in seconds of the fusion process. This line is approximately quasilinear in time, and the algorithm is estimated to $O(n \log n)$.



(a) Fused image with Method 3 in MATLAB



(b) Fused image with Method 3 in C++

(c) Fused image with Method 4 in MATLAB

(d) Fused image with Method 4 in C++

Figure 31: Comparison between fused images executed in MATLAB and in C++. Difference between Q for Method 3 is 0.026 and for Method 4 is 0.000. Regarding NIQE the difference for Method 3 is 0.030 and for Method 4 0.001. Finally for Fast-FMI the difference is also minor, Method 3 having a difference of 0.003 and Method4 0.000.

7.4 Subjective Ranking

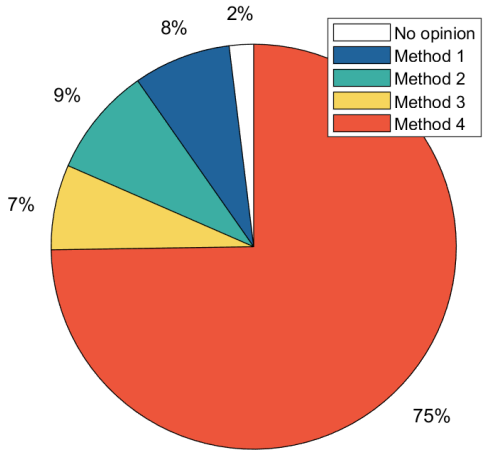


Figure 32: Pie Chart showing the results of the subjective ranking. A total of 104 data points was collected with an average decision time of 15 seconds per evaluated image. Information about the respondents can be seen in Table 3 in section 5.6.1.

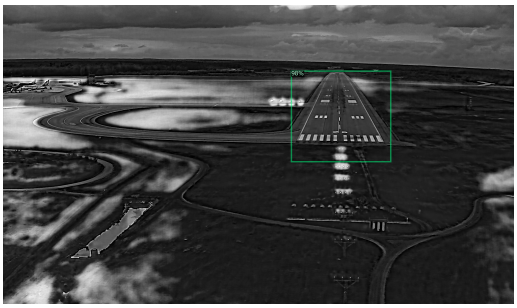
7.5 Object Detection

In this section, fused images applied with the object detection algorithm from section 5.8 is presented. Figure 33 displays the images with bounding boxes of each detected object in the image. In Table 6, a compilation of the results regarding object detection is shown.

(a) Object detection algorithm applied on fused image with Method 1, section 5.6.2

(b) Object detection algorithm applied on fused image with Method 2, section 5.6.3

(c) Object detection algorithm applied on fused image with Method 3, section 5.6.4



(d) Object detection algorithm applied on fused image with Method 4, section 5.6.5

Figure 33: Fused images with applied object detection algorithm.

Table 6: This table presents the output from each image with the evaluated fusion methods regarding number of objects detected and the mean certainty of each object detected.

Image	**Method 1**		**Method 2**		**Method 3**		**Method 4**	
	Detected Obj.	*Mean Certainty %*	*Detected Obj.*	*Mean Certainty %*	*Detected Obj.*	*Mean Certainty %*	*Detected Obj.*	*Mean Certainty %*
1	2	95	2	89	2	98.5	2	95.5
2	3	95	3	89.3	2	96.5	1	84
3	2	98.5	2	98	2	99.5	1	98
4	1	99	1	99	1	100	1	100
5	1	100	1	98	1	99	1	98
6	1	100	1	100	1	100	1	100
7	1	99	1	98	1	99	1	100
8	1	90	1	84	1	93	1	92
Total	12	96.9	12	94.4	11	98.2	9	95.9

8 Discussion

8.1 Systematic Review

Conducting a systematic review on IQA unfolds a clear pattern, particularly concerning assessment without reference. Literature tends to directly or as a reference compare algorithms with the human eye's ability to judge image quality. However, relying on human subjective assessment ranges from hard to impossible for embedded solutions, such as aviation. Therefore, there exists a demand for efficient and accurate quality assessment techniques. The methods implemented in this thesis originate from the systematic review and are possible to implement in an embedded environment. The methods evaluate images based on different parameters. The diversity of quality metrics provides a strong foundation for elaborating on the results. Another reflection essential to point out is that a majority of well-cited publications within the area of IQA are from a handful of authors pointing towards a narrow field within computer vision. The most recent publications provided a turning point in the literature when quality metrics developed implements neural networks instead of conventional algorithms.

8.2 Image Registration

Figure 15 shows that manual selection of feature matching outperforms the SURF method in this sensor setup. The finding is visible on salient features such as the ALS. One explanation of the poor performance of feature matching algorithms may be due to the differences in feature appearance in different modalities of sensors. The accuracy of point selection directly impacts the result of the transformation matrix, the choice of input images is vital as it enables a good selection process. For the transform matrix to be as correct as possible, input images need to contain details both in the near-field and far-field of the image. Another consideration regarding the evaluation of the transform matrix is the fact of good results in one sensor setup may not translate well to other setups. The comparison of applying affine transform and projective transform can be seen in Figure 16. Differences between the two transform methods are minimal, given the same control points from the feature matching. However, the affine method is more time-efficient, as seen in Figure 17.

8.3 Image Fusion

As mentioned by Luo and Kay [14], the use of multiple sensors may provide a more accurate output and, at the same time, the possibility of using sensors as a backup at sensor failure. The possibility of such benefits is a strong argument for implementing vision sensor fusion in the aviation domain. By implementing any of the tested methods, sensor failure would, by design, lead to other sensors in the system being dominant, and all redundant data still valid. It should further be noted that using vision sensors in different electromagnetic spectrums, can also lead to the fact of sensor-specific data being lost at sensor failure. Therefore if sensor-specific data is crucial to airplane operations, multiple sensors of the same type are needed to assure necessary data not being lost.

As previously stated in the background (section 3), the output from vision sensors is heavily dependent on the surrounding environmental conditions. In the quest to cope with this issue, sensors capturing different intervals of the electromagnetic spectrum are fused. A total of four methods is evaluated, and their quality assessed. The algorithms are based on [67, 71, 6, 75] with modifications to enable the fusion of three images instead of two images.

The fusion methods share similarities and variations amongst the implementations. First off, two methods implement a deep-learning-based feature detection approach. From a safety perspective in the aviation domain, introducing deep-learning to a system might complicate the process of development. The determinism of the output can prove harder to determine, due to the complexity of the neural network. Second, a common state-of-practice is to separate the images into detailed layers and base layers. For example, both Method 1 and Method 3 separate the layers with the Tikhonov's Regularization problem, while Method 4 separates the layers with multi-scale decomposition. The most significant differences in the process occur with the processing of detailed layers.

The methods have comparable results regarding the Fast-FMI index. See Figure 24. One explanation to this is that the output images from the method have a similar amount of features present concerning the source images. Preserving features from three source images is a difficult task, with the risk of the image appearing oversaturated when an overload of information is preserved.

On the contrary, regarding the NIQE index, Figure 25 shows that Method 4 deviates from the other methods with a lower score (in this case, a better result). Figure 19 to 23 have a frequent pattern being Method 4 stands out from the other methods. Method 4 possessing a higher contrast, with less blur present, indicating a better NIQE score with the assumption that natural images lack blurriness. With a high amount of distortions present in an image, the result might produce erroneous decision making, thus making the NIQE index relevant to this evaluation. Depending on the obstacle detection algorithm, and the nature of distortions, the risk of classifying distortions as obstacles exist.

The last quality metric $Q^{\mathbf{AB/F}}$ confirms a clear pattern with Method 4 producing a higher score, with higher contrast. A higher contrast can be related to the previously presented result of index NIQE. In comparison to the other methods, Method 4 appears to have a higher contrast ratio. Another abnormal finding is the result of Method 3. With this index, Method 3 is given a low score compared to the other methods. However, visually inspecting the result, Method 3 appears to have a similar contrast compared to Method 1 and 2. The results were repeated over several test runs, implying that the quality index provides a lower score for Method 3.

8.4 Subjective Ranking

All images were also graded by human subjects in subjective tests, with 13 participants. The participants all had previous knowledge in either computer science or aviation, the background of the participants strengthens the applicability of the method within its given fields, machine learning and aviation. A total of 104 datapoints was collected, and a majority of the data points showed that Method 4 is the preferred method. Method 4 is chosen as the top-ranked method in the subjective test, which may have two possible explanations. First, as the method contains more contrast compared to Method 1-3, resulting in runway lights being more prominent in the image, which participants with knowledge in aviation deem essential. And secondly, the human perception of image quality increases with the amount of visible contrast in an image. Method 1-3 showed very similar scores when compared to each other, which could be explained by the three methods of similar appearance in regards to clarity and contrast.

8.5 Object Detection

It's important to point out that the tests regarding object detection were not optimized for this thesis. The deep learning-driven approach for object detection is not trained on data produced from the output of this thesis. Lack of appropriate training might decrease the validity of the experiment, but accurate enough to produce result pointing towards a particular direction. The fusion methods produce similar results with minor deviation. Method 1 and Method 2 identify the highest quantity of objects while Method 3 having the highest certainty. Method 4 compared to the other methods performed at the lower bound. One rationale for this finding is that the object detection algorithm was trained on images similar to Method 1-3. However, all methods produce satisfying results, given the limitations of the tests.

8.6 Software Conversion

Two of the methods, Method 3 and Method 4, were also implemented in the compiled language C++. A visual comparison between MATLAB and the same methods in C++ can be seen in Figure 23, visually, the two images appear very similar. However, some minor differences for Method 3 can be seen when implementing $Q^{\mathbf{AB/F}}$, whereas Method 4 shows no difference when comparing the two outputs. When compared with NIQE, both methods show small differences. Method 3 has a difference of 0.03 and Method 4, differs by 0.001. Lastly, when comparing the methods using Fast-FMI, Method 3 again shows a small difference of 0.003 and Method 4, no difference at all. The small differences in output may be explained by the MATLAB coder's inability to directly

generate some functions leading to the need for alterations of the code. One other observation regarding the generated code, is the generator utilization of the stack for data storage, resulting in stack overflow if input images are big. Therefore the IQA comparison between MATLAB and C++ code, use smaller images as input for both implementations. However, when comparing the results of the three IQA methods, output values vary for all methods depending on input image size, as can be seen in Table 5. As the output of the IQA methods varies with regards to input size, it's important to note that the difference between C++ code and MATLAB may change depending on image size. While comparing images, it should further noted that comparing images of equal sizes is crucial to see what method yields better results.

8.7 Time Complexity

Analyzing the results of Figure 27 - 30 produces remarkable findings. The time complexity plot of all methods is similar to each other. However, the absolute time taken for a given set of images to be processed with the methods is very diverse. For example, implementing Method 3 compared to the faster method, Method 4, produces a significant time overhead. It's important to note that the time complexity (Ordo) is estimated with a measured time of execution and not calculated. On the contrary, the range of input size varies from a small input of approximately 5000 pixels to an extensive input of 2 million pixels. Another validity concern of this test is the platform that the software executed on. Executed on other platforms might provide other time samples. However, regardless of the platform and programming language, the algorithms follow the same characteristics and would produce the same Ordo.

9 Conclusions

This research aimed to show that a fusion method existed and was suitable to implement in an aircraft avionic system. All evaluated methods had to be able to receive input from three vision sensors, two in the IR spectrum and one in the visible spectrum to achieve visibility in low light and fog conditions. A pre-processing method was needed so that images from all three spectrums were registered correctly, as the three sensors differed in resolution and placement on the aircraft. By analyzing IQA results along with image grading from subjects who have a relevant background, it's possible to conclude which of the suggested methods have better application-specific quality. All methods were further evaluated by testing the time complexity of each of the methods. To show that it is possible to implement the suggested methods in a more extensive system, which includes object detection, the output from all methods was tested with a non-optimized object-detection algorithm.

9.1 RQ1

What image quality assessment techniques are required to determine the output quality of evaluated image processing methods? A systematic review was conducted to find adequate IQA methods. The review showed that several techniques existed, and some techniques were more appropriate for fusion-based image assessment. As no known optimally fused image existed for this setup, NR-IQA is chosen due to the findings of the review. Furthermore, the review proves that all of the investigated methods compare their results with human perception, and therefore it is concluded that a subjective test of perceived image quality is needed. The selected image quality methods $Q^{\mathbf{AB/F}}$, Fast-FMI, and NIQE is chosen based on findings in the review. $Q^{\mathbf{AB/F}}$ is chosen as its ability to show preserved edge information in the fused output indicates an output image with good clarity. NIQE gives a better score where a lack of artificial artifacts due to image processing, are present. As image fusion is an image processing technique, a conclusion based on the number of artificial artifacts helps in determining overall image quality. Fast-FMI is a method that compares the source images with the fused output and calculates the amount of preserved information. If the information in the fused output is lost, the risk of relevant data being neglected exists. To assure the different fusion methods ability to preserve information Fast-FMI is deemed appropriate. The review also showed that existing IQA methods usually compare their results with the results of human perception due to human ability to determine NR-IQA. A subjective test was conducted using subjects with prior knowledge in the field of avionics or computer science.

9.2 RQ2

What are the similarities and differences between state-of-the-art vision based fusion methods? One key finding is the similarities between the methods evaluated. It is concluded that the majority of image fusion processes separate the image into layers, more particularly detailed layers, and base layers. Furthermore, state-of-the-art methods usually represent the data in transforms. Both integral transform e.g., Laplace transform and Wavelet transform as well as discrete transform equation, Discrete Fourier transform. It is further concluded that a common practice for fusion is coefficient based, for example, weighted-average. On the other hand, methods implementing CNN has also emerged as a solution.

9.3 RQ3

What sensor fusion technique provides adequate results with respect to given quality metrics from RQ1 in an aircraft environment? Four state-of-the-art fusion methods were chosen to be evaluated, where two of the methods are CNN based and the other two implements more conventional methods. Method 4 showed promising results using NR-IQA, the method performed best in both NIQE and $Q^{\mathbf{AB/F}}$ and showed similar results to Methods 1,2 and 3 when evaluated using Fast-FMI. Method 4 was chosen in a majority of subjective tests as the technique with the best image quality, and it was chosen 75% of the time. The three other methods evaluated in the subjective tests showed similar results and were chosen 8%, 9%, and 7% of the times. As Method 4 showed

very promising results in subjective tests as well as non-subjective IQA, Method 4 is concluded to be the method with the best application-specific image quality. It should be noted, however, that when tested on the non-optimized object-detection algorithm, Method 3 showed the best results.

9.4 RQ4

What correlation exists between the detection capacity of objects and image quality? All four methods were tested using a non-optimized object detection algorithm, and Method 3 performed best regarding mean certainty, while Methods 1 and 2 had a higher amount of detected objects. As mentioned in section 8, the object detection algorithm being trained on images similar to the output of Methods 1,2, and 3 increases the score for those three methods. This seems to point towards the need of a properly trained network based on correct training data of the network or the fact of no correlation between the implemented image quality estimations and detection capacity. With the results being ambiguous, no conclusion about detection capacity and image quality is made in this thesis.

9.5 RQ5

What is the most correct way to assure that output from one sensor matches that of other sensors in a sensor setup of two IR sensors and one visual spectrum sensor? It is concluded that the ability to interpret the images for a correct matching between the sources is essential for a good result. The nature of the sensor setup in this thesis having differences in modality restricts the abilities of the SURF algorithm resulting in manual feature matching outperforming SURF. It is further concluded that a careful selection of images for feature matching is crucial and requires salient features both in the near-field and far-field of the image. A careful selection provides a more robust result and optimized transform matrices. Lastly, affine transform compared to projective transform did not change the outcome of the registration process, and therefore, nothing is concluded besides that affine transform is more time-efficient for this implementation.

9.6 Future Work

The work of this thesis shows promising results with regards to realize computer vision with image fusion in aviation. However, further research is necessary. For example, addressing the validity concerns is a great starting point. The sensors and chosen fusion method require extensive tests in different weather conditions. More research is also required in the field of identifying objects with the output from the fusion method to provide an accurate estimation between image quality and detection capacity. More thorough image registration and alignment of sensors is also a factor that can improve future results. The software needs to be evaluated in an embedded avionics system environment with limited computational resources. Furthermore, additional safety-related activities need to be performed to achieve compliance with industry standards and guidelines. Regarding the complexity of the methods, they all share complexity within the range of inputs that are tested. When comparing the methods, actual run-time of the methods differ vastly.

References

[1] O. Vygolov, "Enhanced and synthetic vision systems development based on integrated modular avionics for civil aviation," 10 2013.

[2] S. de Aviacion Civil, *KLM, B-747, PH-BUF and Pan Am B-747 N736 collition at Tenerife Airport Spain on 27 March 1977*. Subsecretaria de Aviacion Civil, Spain, 1978.

[3] "Statistical summary of commercial jet airplane accidents," 2019. [Online]. Available: http://www.boeing.com/resources/boeingdotcom/company/about_bca/pdf/statsum.pdf

[4] M. R. Endsley, "Toward a theory of situation awareness in dynamic systems," in *Situation awareness*. Routledge, 2017, pp. 9–42.

[5] J. Ma, Y. Ma, and C. Li, "Infrared and visible image fusion methods and applications: A survey," *Information Fusion*, vol. 45, pp. 153–178, 2019.

[6] Y. Liu, X. Chen, R. K. Ward, and Z. J. Wang, "Image fusion with convolutional sparse representation," *IEEE signal processing letters*, vol. 23, no. 12, pp. 1882–1886, 2016.

[7] SAAB. (2020) Company page. Date Accessed: 2020-05-14. [Online]. Available: https://saabgroup.com/

[8] "Decision of the Govering Board adopting the updated Clean Sky 2 Development Plan," Clean Sky 2, Tech. Rep. CS2DP, Oct. 2019. [Online]. Available: https://www.cleansky.eu/sites/default/files/inline-files/CS-GB-2019-11-21%20Decision%20CS2DP%20adoption_0.pdf

[9] O. A. Yakimenko, "Pilots requirements to universal airborne intelligent pilot decisions support system," in *Proceedings of the IEEE 1995 National Aerospace and Electronics Conference. NAECON 1995*, vol. 1. IEEE, 1995, pp. 446–452.

[10] D. L. McCallie, "Exploring potential ads-b vulnerabilites in the faa's nextgen air transportation system," AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OH DEPT OF ELECTRICAL AND ..., Tech. Rep., 2011.

[11] FAA, *Airworthiness Approval of Traffic Alert and Collision Avoidance Systems (TCAS II)*, 2014.

[12] J. Lee, "Modeling terrain awareness and warning systems for airspace and procedure design," in *2012 IEEE/AIAA 31st Digital Avionics Systems Conference (DASC)*. IEEE, 2012, pp. 2B2–1.

[13] B. Fleming, "New automotive electronics technologies [automotive electronics]," *IEEE Vehicular Technology Magazine*, vol. 7, no. 4, pp. 4–12, 2012.

[14] R. Luo and M. Kay, "Data fusion and sensor integration: State-of-the-art 1990s," *Data fusion in robotics and machine intelligence*, pp. 7–135, 1992.

[15] D. Drajic and N. Cvejic, "Adaptive fusion of multimodal surveillance image sequences in visual sensor networks," *IEEE Transactions on Consumer Electronics*, vol. 53, no. 4, pp. 1456–1462, 2007.

[16] T. Kim, S. Kim, E. Lee, and M. Park, "Comparative analysis of radar-ir sensor fusion methods for object detection," in *2017 17th International Conference on Control, Automation and Systems (ICCAS)*. IEEE, 2017, pp. 1576–1580.

[17] S. J. Krotosky and M. M. Trivedi, "Person surveillance using visual and infrared imagery," *IEEE transactions on circuits and systems for video technology*, vol. 18, no. 8, pp. 1096–1105, 2008.

[18] *Digital avionics handbook*, 3rd ed. Boca Raton, Florida: CRC Press, 2015.

[19] B. Keelan, *Handbook of image quality: characterization and prediction.* CRC Press, 2002.

[20] ICAO, "Meteorological service for international air navigation," *Annex 3*, vol. Sixteenth Edition, 2007.

[21] K. Beier and H. Gemperlein, "Simulation of infrared detection range at fog condition for enhanced vision system in civil aviation," *Aerospace Science and Technology*, vol. 8, pp. 63–71, 01 2004.

[22] A. Dogra, B. Goyal, and S. Agrawal, "From multi-scale decomposition to non-multi-scale decomposition methods: a comprehensive survey of image fusion techniques and its applications," *IEEE Access*, vol. 5, pp. 16 040–16 067, 2017.

[23] Q. Zhang, Y. Liu, R. S. Blum, J. Han, and D. Tao, "Sparse representation based multi-sensor image fusion for multi-focus and multi-modality images: A review," *Information Fusion*, vol. 40, pp. 57–75, 2018.

[24] B. Yang and S. Li, "Multifocus image fusion and restoration with sparse representation," *IEEE Transactions on Instrumentation and Measurement*, vol. 59, no. 4, pp. 884–892, 2009.

[25] P. Chai, X. Luo, and Z. Zhang, "Image fusion using quaternion wavelet transform and multiple features," *IEEE access*, vol. 5, pp. 6724–6734, 2017.

[26] S. Li, X. Kang, and J. Hu, "Image fusion with guided filtering," *IEEE Transactions on Image processing*, vol. 22, no. 7, pp. 2864–2875, 2013.

[27] Y. Liu, X. Chen, Z. Wang, Z. J. Wang, R. K. Ward, and X. Wang, "Deep learning for pixel-level image fusion: Recent advances and future prospects," *Information Fusion*, vol. 42, pp. 158–173, 2018.

[28] L. Yujiri, M. Shoucri, and P. Moffa, "Passive millimeter wave imaging," *IEEE microwave magazine*, vol. 4, no. 3, pp. 39–50, 2003.

[29] X. Song, L. Li, and J. Yang, "Image fusion algorithm for visible and pmmw images based on em and ncut," in *2013 International Conference on Computational Problem-Solving (ICCP)*. IEEE, 2013, pp. 319–323.

[30] Y. Xia, H. Leung, and E. Bossé, "Neural data fusion algorithms based on a linearly constrained least square method," *IEEE Transactions on neural networks*, vol. 13, no. 2, pp. 320–329, 2002.

[31] B. Zitova and J. Flusser, "Image registration methods: a survey," *Image and vision computing*, vol. 21, no. 11, pp. 977–1000, 2003.

[32] B. Putz, M. Bartyś, A. Antoniewicz, J. Klimaszewski, M. Kondej, and M. Wielgus, "Real-time image fusion monitoring system: Problems and solutions," in *Image Processing and Communications Challenges 4*. Springer, 2013, pp. 143–152.

[33] A. Antoniewicz, "Fpga implementation of decomposition methods for real-time image fusion," *Advances in Intelligent Systems and Computing*, vol. 184, pp. 163–170, 01 2013.

[34] G. Qu, D. Zhang, and P. Yan, "Information measure for performance of image fusion," *Electronics letters*, vol. 38, no. 7, pp. 313–315, 2002.

[35] M. Haghighat and M. A. Razian, "Fast-fmi: non-reference image fusion metric," in *2014 IEEE 8th International Conference on Application of Information and Communication Technologies (AICT)*. IEEE, 2014, pp. 1–3.

[36] C. Xydeas, , and V. Petrovic, "Objective image fusion performance measure," *Electronics letters*, vol. 36, no. 4, pp. 308–309, 2000.

[37] D. Liu, B. Wen, X. Liu, Z. Wang, and T. S. Huang, "When image denoising meets high-level vision tasks: A deep learning approach," *arXiv preprint arXiv:1706.04284*, 2017.

[38] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising," *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3142–3155, 2017.

[39] H. Li, "Deep learning for image denoising," *International Journal of Signal Processing, Image Processing and Pattern Recognition*, vol. 7, no. 3, pp. 171–180, 2014.

[40] C. Tian, Y. Xu, L. Fei, and K. Yan, "Deep learning for image denoising: a survey," in *International Conference on Genetic and Evolutionary Computing*. Springer, 2018, pp. 563–572.

[41] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 2. IEEE, 2005, pp. 60–65.

[42] X. Lan, S. Roth, D. Huttenlocher, and M. J. Black, "Efficient belief propagation with learned higher-order markov random fields," in *European conference on computer vision*. Springer, 2006, pp. 269–282.

[43] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," *IEEE Transactions on image processing*, vol. 16, no. 8, pp. 2080–2095, 2007.

[44] D. Park and H. Ko, "Fog-degraded image restoration using characteristics of rgb channel in single monocular image," in *2012 IEEE International Conference on Consumer Electronics (ICCE)*. IEEE, 2012, pp. 139–140.

[45] B. Li and Z. Zhan, "Research on motion blurred image restoration," in *2012 5th International Congress on Image and Signal Processing*. IEEE, 2012, pp. 1307–1311.

[46] F. Middleton. (2019) The four types of validity. Date Accessed: 2020-05-14. [Online]. Available: https://www.scribbr.com/methodology/types-of-validity/

[47] J. Webster and R. T. Watson, "Analyzing the past to prepare for the future: Writing a literature review," *MIS quarterly*, pp. xiii–xxiii, 2002.

[48] Z. Wang and A. C. Bovik, "Modern image quality assessment," *Synthesis Lectures on Image, Video, and Multimedia Processing*, vol. 2, no. 1, pp. 1–156, 2006.

[49] ——, "A universal image quality index," *IEEE signal processing letters*, vol. 9, no. 3, pp. 81–84, 2002.

[50] Z. Wang, A. C. Bovik, and L. Lu, "Why is image quality assessment so difficult?" in *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 4. IEEE, 2002, pp. IV–3313.

[51] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

[52] R. Hassen, Z. Wang, and M. Salama, "No-reference image sharpness assessment based on local phase coherence measurement," in *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2010, pp. 2434–2437.

[53] A. M. Eskicioglu and P. S. Fisher, "Image quality measures and their performance," *IEEE Transactions on communications*, vol. 43, no. 12, pp. 2959–2965, 1995.

[54] T. N. Pappas, R. J. Safranek, and J. Chen, "Perceptual criteria for image quality evaluation," *Handbook of image and video processing*, vol. 110, 2000.

[55] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*, vol. 2. Ieee, 2003, pp. 1398–1402.

[56] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a "completely blind" image quality analyzer," *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 209–212, 2012.

[57] Y. Fang, K. Ma, Z. Wang, W. Lin, Z. Fang, and G. Zhai, "No-reference quality assessment of contrast-distorted images based on natural scene statistics," *IEEE Signal Processing Letters*, vol. 22, no. 7, pp. 838–842, 2014.

[58] K. Gu, W. Lin, G. Zhai, X. Yang, W. Zhang, and C. W. Chen, "No-reference quality metric of contrast-distorted images based on information maximization," *IEEE transactions on cybernetics*, vol. 47, no. 12, pp. 4559–4565, 2016.

[59] S. Bosse, D. Maniry, K.-R. Müller, T. Wiegand, and W. Samek, "Deep neural networks for no-reference and full-reference image quality assessment," *IEEE Transactions on Image Processing*, vol. 27, no. 1, pp. 206–219, 2017.

[60] Marshall. (2019) Cv342-csb/cs hd-sdi compact tube camera. Date Accessed: 2020-05-14. [Online]. Available: http://www.marshall-usa.com/discontinued/cameras/CV342-CSB.php

[61] Xenics. (2019) Rufus-640-analog. Date Accessed: 2020-05-14. [Online]. Available: http://www.g4.com.tw/userfiles/files/Datasheet/xenics_swir_rufus-640-analog.pdf

[62] ——. (2019) Raven-640-analog. Date Accessed: 2020-05-14. [Online]. Available: https://www.g4.com.tw/userfiles/files/Datasheet/xenics_lwir_raven-640-analog_17%CE%BCm.pdf

[63] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," in *European conference on computer vision*. Springer, 2006, pp. 404–417.

[64] G. K. Matsopoulos, N. A. Mouravliansky, K. K. Delibasis, and K. S. Nikita, "Automatic retinal image registration scheme using global optimization techniques," *IEEE Transactions on Information Technology in Biomedicine*, vol. 3, no. 1, pp. 47–60, 1999.

[65] *9.7.0.1296695 (R2019b) Update 4*, The Mathworks, Inc., Natick, Massachusetts, 2019.

[66] V. Petrović, "Subjective tests for image fusion evaluation and objective metric validation," *Information Fusion*, vol. 8, no. 2, pp. 208–216, 2007.

[67] H. Li, X.-J. Wu, and J. Kittler, "Infrared and visible image fusion using a deep learning framework," in *2018 24th International Conference on Pattern Recognition (ICPR)*. IEEE, 2018, pp. 2705–2710.

[68] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[69] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A Large-Scale Hierarchical Image Database," in *CVPR09*, 2009.

[70] H. Li and X.-J. Wu, "Multi-focus image fusion using dictionary learning and low-rank representation," in *International Conference on Image and Graphics*. Springer, 2017, pp. 675–686.

[71] H. Li, X.-J. Wu, and T. S. Durrani, "Infrared and visible image fusion with resnet and zero-phase component analysis," *Infrared Physics & Technology*, vol. 102, p. 103039, 2019.

[72] K. He, 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, pp. 770–778.

[73] "Nearest Neighbor, Bilinear, and Bicubic Interpolation Methods," The Mathworks, Inc., Natick, Massachusetts, 2020. [Online]. Available: https://se.mathworks.com/help/vision/ug/interpolation-methods.html

[74] B. Wohlberg, "Efficient algorithms for convolutional sparse representations," *IEEE Transactions on Image Processing*, vol. 25, no. 1, pp. 301–315, 2015.

[75] J. Ma, Z. Zhou, B. Wang, and H. Zong, "Infrared and visible image fusion based on visual saliency map and weighted least square optimization," *Infrared Physics & Technology*, vol. 82, pp. 8–17, 2017.

[76] Q. Zhang, X. Shen, L. Xu, and J. Jia, "Rolling guidance filter," in *European conference on computer vision*. Springer, 2014, pp. 815–830.

[77] Z. Farbman, R. Fattal, D. Lischinski, and R. Szeliski, "Edge-preserving decompositions for multi-scale tone and detail manipulation," *ACM Transactions on Graphics (TOG)*, vol. 27, no. 3, pp. 1–10, 2008.

[78] M. B. A. Haghighat, A. Aghagolzadeh, and H. Seyedarabi, "A non-reference image fusion metric based on mutual information of image features," *Computers & Electrical Engineering*, vol. 37, no. 5, pp. 744–756, 2011.

[79] T. M. Cover and J. A. Thomas, *Elements of information theory*. John Wiley & Sons, 2012.

[80] D. L. Ruderman, "The statistics of natural images," *Network: computation in neural systems*, vol. 5, no. 4, pp. 517–548, 1994.

[81] Y. Wu, A. Kirillov, F. Massa, W.-Y. Lo, and R. Girshick, "Detectron2," https://github.com/facebookresearch/detectron2, 2019.

[82] *MATLAB Coder Generate C and C++ code from MATLAB code*, The Mathworks, Inc., Natick, Massachusetts, 2019.

A MATLAB Graphical User Interface

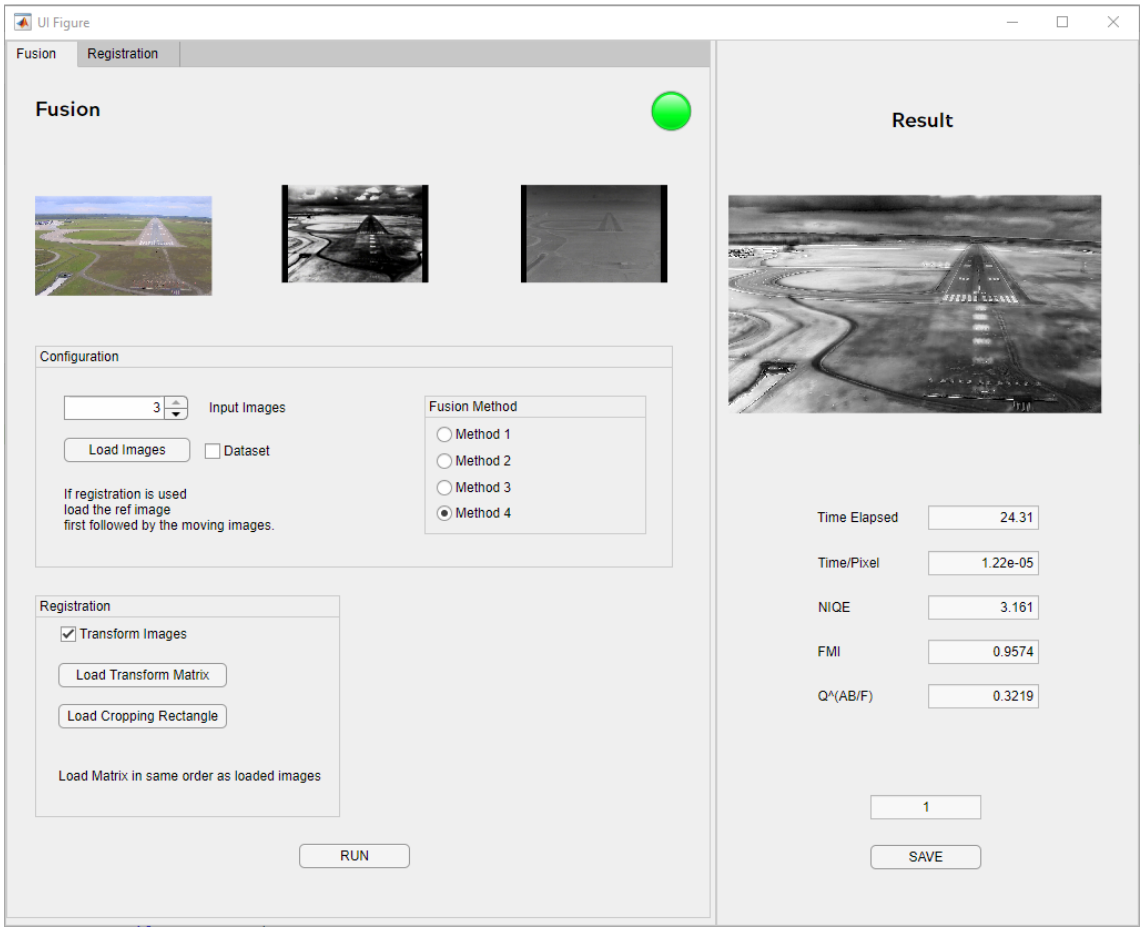


Figure 34: Develoepd software for a structured evaluation approach.

B Subjective Ranking Enviroment

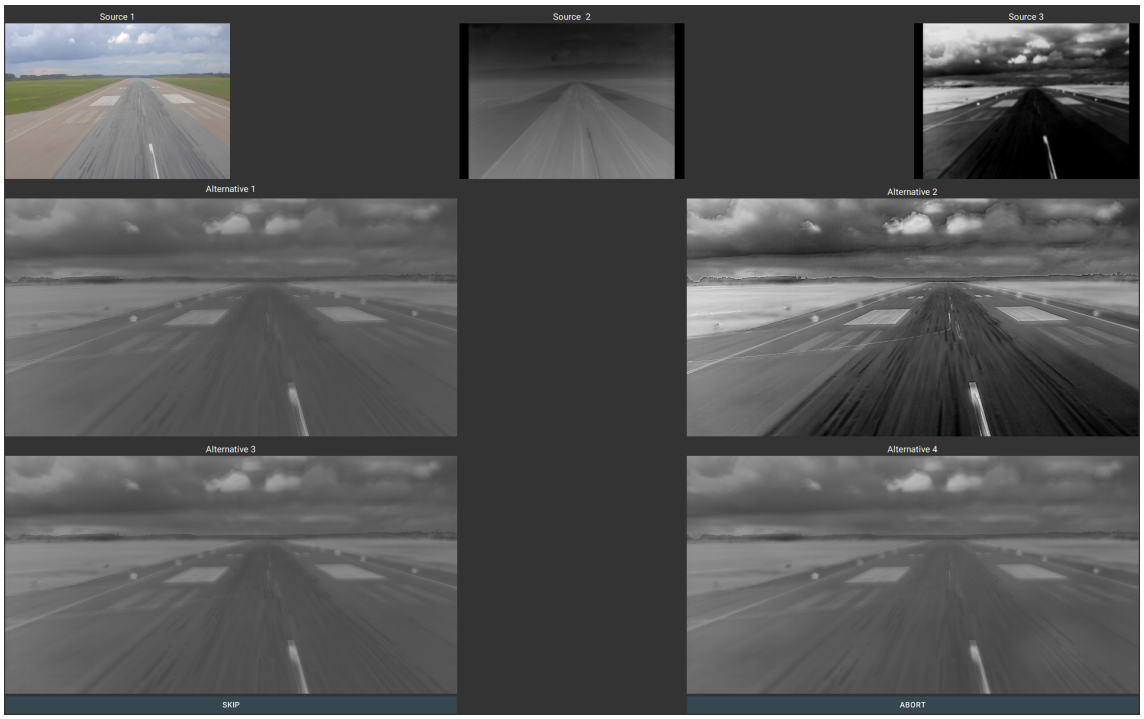


Figure 35: Developed software for the subjective ranking tests. Here the subjects selects the preferred image and a new set of samples is loaded.