Multispectral Image Analysis for Object Recognition and Classification

Claude Viau

Thesis submitted to the Faculty of Graduate and Postdoctoral Studies in partial fulfillment of the requirements for the degree of

Master of Applied Science in Electrical and Computer Engineering

Ottawa-Carleton Institute for Electrical and Computer Engineering School of Electrical Engineering and Computer Science University of Ottawa

March 2016

© Claude Viau, Ottawa, Canada, 2016
Abstract

Computer and machine vision applications are used in numerous fields to analyze static and dynamic imagery in order to assist or automate some form of decision-making process. Advancements in sensor technologies now make it possible to capture and visualize imagery at various wavelengths (or bands) of the electromagnetic spectrum. Multispectral imaging has countless applications in various fields including (but not limited to) security, defense, space, medical, manufacturing and archeology. The development of advanced algorithms to process and extract salient information from the imagery is a critical component of the overall system performance.

The fundamental objectives of this research project were to investigate the benefits of combining imagery from the visual and thermal bands of the electromagnetic spectrum to improve the recognition rates and accuracy of commonly found objects in an office setting. The goal was not to find a new way to “fuse” the visual and thermal images together but rather establish a methodology to extract multispectral descriptors in order to improve a machine vision system’s ability to recognize specific classes of objects. A multispectral dataset (visual and thermal) was captured and features from the visual and thermal images were extracted and used to train support vector machine (SVM) classifiers. The SVM’s class prediction ability was evaluated separately on the visual, thermal and multispectral testing datasets. Commonly used performance metrics were applied to assess the sensitivity, specificity and accuracy of each classifier.

The research demonstrated that the highest recognition rate was achieved by an expert system (multiple classifiers) that combined the expertise of the visual-only classifier, the thermal-only classifier and the combined visual-thermal classifier.
Acknowledgment

I would like to offer my sincere gratitude to Dr. Pierre Payeur and Dr. Ana-Maria Cretu for all your support with this research project. You have provided invaluable insight, subject matter expertise and guidance along the way.

To my wife and daughters, many sacrifices were made on this journey and I am truly grateful for your patience and unconditional support.
Table of Contents

ABSTRACT .......................................................................................................................... II
ACKNOWLEDGMENT ......................................................................................................... III
LIST OF FIGURES ........................................................................................................... VI
LIST OF TABLES ............................................................................................................... VIII
LIST OF ACRONYMS ....................................................................................................... IX
CHAPTER 1 INTRODUCTION ............................................................................................. 1
  1.1. CONTEXT .................................................................................................................... 1
  1.2. RESEARCH OBJECTIVES ......................................................................................... 2
  1.3. THESIS ORGANIZATION ......................................................................................... 3
CHAPTER 2 LITERATURE REVIEW .................................................................................... 4
  2.1. SEGMENTATION ........................................................................................................ 4
  2.2. IMAGE FEATURES ..................................................................................................... 8
  2.3. CLASSIFIERS ........................................................................................................... 14
  2.4. PERFORMANCE METRICS ..................................................................................... 16
  2.5. SUMMARY OF LITERATURE REVIEW .................................................................... 18
CHAPTER 3 DATA COLLECTION AND DATASETS .............................................................. 20
  3.1. CAMERA SPECIFICATIONS AND IMAGE ANALYSIS SOFTWARE ....................... 20
  3.2. IMAGE PREPROCESSING ......................................................................................... 24
CHAPTER 4 METHODOLOGY ............................................................................................ 31
  4.1. OVERVIEW ............................................................................................................... 31
  4.2. SEGMENTATION ....................................................................................................... 31
    4.2.1. BASIC THRESHOLD ......................................................................................... 32
    4.2.2. K-MEANS ......................................................................................................... 32
    4.2.3. CONTOURS ....................................................................................................... 32
    4.2.4. WATERSHED WITH DISTANCE TRANSFORM ................................................ 32
    4.2.5. PERFORMANCE ASSESSMENT ....................................................................... 33
    4.2.6. WATERSHED WITH THERMAL MARKERS .................................................... 37
  4.3. IMAGE FEATURE SELECTION .................................................................................. 41
  4.4. CLASSIFIER SELECTION ......................................................................................... 47
CHAPTER 5 SOFTWARE IMPLEMENTATION AND DATA PROCESSING ......................... 48
5.1. OVERVIEW .................................................................................................................. 48
5.2. SEGMENTATION ........................................................................................................ 50
5.3. FEATURE EXTRACTION AND POST PROCESSING .................................................. 52
5.4. TRAINING AND TESTING THE CLASSIFIERS ....................................................... 56
CHAPTER 6 EXPERIMENTAL EVALUATION .................................................................. 57
6.1. EXPERIMENT DESIGN ............................................................................................. 57
6.2. CLASSIFICATION RESULTS ..................................................................................... 59
6.2.1. INDIVIDUAL FEATURES ...................................................................................... 59
6.2.2. VISUAL BAND ...................................................................................................... 64
6.2.3. THERMAL BAND ................................................................................................. 67
6.2.4. COMBINED VISUAL AND THERMAL BANDS .................................................. 71
6.2.5. EXPERT SYSTEM ................................................................................................. 76
6.2.6. FURTHER IMPROVEMENTS ................................................................................ 78
CHAPTER 7 CONCLUSIONS ......................................................................................... 80
7.1. SUMMARY OF FINDINGS ......................................................................................... 80
7.2. CONTRIBUTIONS ....................................................................................................... 81
7.3. FUTURE WORK .......................................................................................................... 82
REFERENCES ..................................................................................................................... 84
List of Figures

Figure 2-1: Generic machine vision process .......................................................................................... 4
Figure 3-1: Examples of the five classes of objects used for this project ................................................. 22
Figure 3-2: SmartView software user interface .................................................................................... 23
Figure 3-3: Various levels of blended visual-thermal images using SmartView. Full thermal (left), half- blended (center) and full visual (right). .......................................................................................... 23
Figure 3-4: Sample training dataset for Class 1 (Mobile Phone) ............................................................. 24
Figure 3-5: Sample training dataset for Class 2 (Coffee Mug) .............................................................. 25
Figure 3-6: Sample training dataset for Class 3 (Laptop Charger) ....................................................... 25
Figure 3-7: Sample training dataset for Class 4 (Desk Lamp) .............................................................. 25
Figure 3-8: Sample training dataset for Class 5 (Portable Heater) ....................................................... 26
Figure 3-9: Sample testing dataset for Class 1 (Mobile Phone) ............................................................ 27
Figure 3-10: Sample testing dataset for Class 2 (Coffee Mug) ............................................................ 27
Figure 3-11: Sample testing dataset for Class 3 (Laptop Charger) ...................................................... 27
Figure 3-12: Sample testing dataset for Class 4 (Desk Lamp) .............................................................. 28
Figure 3-13: Sample testing dataset for Class 5 (Portable Heater) ...................................................... 28
Figure 3-14: Original thermal image (left) and the matching saturated image (right) caused by a temperature range adjustment ....................................................................................... 29
Figure 3-15: Original thermal image (left) with reduced contrast and noise (right) caused by a temperature range adjustment ....................................................................................... 29
Figure 3-16: Visual band image (left) and its matching thermal image (right) mapped to a greyscale palette ...................................................................................................................................... 30
Figure 4-1: Principal workflow ............................................................................................................ 31
Figure 4-2: Example of the Distance Transform calculation. Original image (left), binary (center), distance transform (right) ............................................................................................... 33
Figure 4-3: Visual (top row) and thermal (bottom row) sample segmentation results (dark object on light background). ....................................................................................................... 34
Figure 4-4: Visual (top row) and thermal (bottom row) sample segmentation results (dark object on light multi-textured background) ....................................................................................... 35
Figure 4-5: Visual (top row) and thermal (bottom row) sample segmentation results (light object on light background). ....................................................................................................... 36
Figure 4-6: Visual (top row) and thermal (bottom row) sample segmentation results (light and dark objects on multi-textured background) .................................................................................... 36
Figure 4-7: Watershed with Thermal Markers flowchart ....................................................................... 38
Figure 4-8: *Watershed with Thermal Markers* segmentation algorithm applied to visual (top) and thermal (bottom) images (dark object on light background). Original image in the left column and segmented results in the right column. ................................................................. 39

Figure 4-9: *Watershed with Thermal Markers* segmentation algorithm applied to visual (top) and thermal (bottom) images (dark object on light background). Original image in the left column and segmented results in the right column. ................................................................. 40

Figure 4-10: *Watershed with Thermal Markers* segmentation algorithm applied to visual (top) and thermal (bottom) images (light object on light background). Original image in the left column and segmented results in the right column. ................................................................. 40

Figure 4-11: *Watershed with Thermal Markers* segmentation algorithm applied to visual (top) and thermal (bottom) images (light and dark objects on multi-textured background). Original image in the left column and segmented results in the right column. ................................................................. 41

Figure 5-1: Feature extraction application flowchart ............................................................................................................. 49
Figure 5-2: Classifier application flowchart ...................................................................................................................... 50
Figure 5-3: Original visual image (left), markers (middle) and *Watershed with Thermal Markers* segmentation results (right) ................................................................................................................................. 51
Figure 5-4: Improved segmentation using adjusted thresholds. Original visual image (left), thermal markers (middle) and *Watershed with Thermal Markers* segmentation results (right). ................................................................. 52
Figure 5-5: Feature extraction flowchart ............................................................................................................................. 53
Figure 5-6: Output data file (top) and segmented image (bottom) from the feature extraction application ................................................................. 54
Figure 5-7: Output data file containing multiple classes (0: Mug, 1: Heater, 2: Lamp). ................................................................. 55
Figure 6-1 : Sample class-conditional probability density function ......................................................................................................................... 60
Figure 6-2: Class-conditional probability density function graphs for Feature 1 and 2................................................................. 61
Figure 6-3: Class-conditional probability density function graphs for Feature 3 and 4................................................................. 61
Figure 6-4: Class-conditional probability density function graphs for Feature 5 and 6................................................................. 62
Figure 6-5: Class-conditional probability density function graphs for Feature 7 and 8................................................................. 62
Figure 6-6: Class-conditional probability density function graphs for Feature 9 and 10................................................................. 62
Figure 6-7: Class-conditional probability density function graphs for Feature 11 and 12................................................................. 63
Figure 6-8: Class-conditional probability density function graphs for Feature 13 ................................................................. 63
Figure 6-9: Visual band classifier TPR distribution ................................................................................................................................. 67
Figure 6-10: Thermal band classifier TPR distribution ................................................................................................................................. 70
Figure 6-11: Combined visual and thermal band classifier TPR distribution ................................................................................................. 75
List of Tables

Table 3-1: Fluke Ti10 camera specifications ................................................................. 21
Table 3-2: Dataset description ...................................................................................... 24
Table 4-1: Comparison between Cayouette et al.'s work and this research ...................... 42
Table 5-1: Software and hardware specifications........................................................... 48
Table 6-1: Classification TPR for single features ............................................................ 64
Table 6-2: Visual band feature classification results ....................................................... 65
Table 6-3: Most prominent features in the Top 30 visual-only feature combinations classification results .................................................................................................................... 66
Table 6-4: Visual band classifier performance metrics ..................................................... 67
Table 6-5: Thermal band feature classification results ..................................................... 68
Table 6-6: Most prominent features in the Top 30 thermal-only feature combinations classification results .................................................................................................................... 69
Table 6-7: Thermal band classifier performance metrics .................................................. 71
Table 6-8: Combined visual and thermal band feature classification results ...................... 72
Table 6-9: Most prominent features in the Top 50 visual-thermal feature combinations classification results .................................................................................................................... 73
Table 6-10: Combined visual-thermal classifier performance metrics .............................. 76
Table 6-11: Classifier F1 score comparison ...................................................................... 76
Table 6-12: Expert system ............................................................................................... 77
Table 6-13: Expert system performance metrics .............................................................. 77
List of Acronyms

AI     Artificial Intelligence
ATR    Automatic Target Recognition
BOF    Bag-of-Features
BRIEF  Binary Robust Independent Elementary Features
CAMSHIFT Continuously Adaptive Mean-Shift
CART   Classification and Regression Trees
CDMI   Concentric Discs Moment Invariants
CPU    Central Processing Unit
C-SVC  C-Support Vector Classifier
DOG    Difference of Gaussian
EM     Electromagnetic
FAST   Features from Accelerated Segment Test
FN     False Negative
FP     False Positive
FPA    Focal Plane Array
GPU    Graphic Processing Unit
GS     Global Silhouette
HOG    Histogram of Oriented Gradients
IFOV   Instantaneous Field of View
IR     Infrared
KNN    K-Nearest Neighbor
LDC    Linear Discriminant Classifier
MDC    Minimum Distance Classifier
MRI    Magnetic Resonance Imaging
NETD   Noise Equivalent Temperature Difference
NPV    Negative Predictive Value
ORB    Oriented FAST and Rotated BRIEF
PCA    Principal Component Analysis
PHOG   Pyramid Histogram Of Gradients
PHOW   Pyramid Histogram Of visual Words
PNN    Probabilistic Neural Network
PPV    Positive Predictive Value
QDC    Quadratic Discriminant Classifier
RBF    Radial Basis Function
SIFT   Scale Invariant Feature Transform
SLF    Sparse Localized Features
SURF   Speeded Up Robust Features
SVM    Support Vector Machines
TN     True Negative
TNR    True Negative Rate
TP     True Positive
TPR    True Positive Rate
UV     Ultraviolet
Chapter 1  Introduction

1.1.  Context

Computer and machine vision applications are used in numerous fields to analyze static and dynamic imagery in order to assist or automate some form of decision-making process. Some of the typical fields in which computer vision applications are used include artificial intelligence, medical, industrial, military, security and space. With the advances in computer hardware, vision systems are becoming more feasible and more commonly found in every day, real world applications. Advancements in computer processors alone are not the only reason for this recent surge in machine vision systems. A significant amount of interest from the research community in the last two decades has resulted in sophisticated and efficient processing algorithms in such systems.

A continuing challenge for computer and machine vision applications remains the recognition of objects of interest in real and complex scenes. Object recognition can be accomplished with a certain level of accuracy by using image or template matching whereby several images of the objects are stored in memory and compared to the presented scene. A correlation process is used to identify the object in the scene that appears most like the stored templates. The correlation process is typically performed in the spatial (pixel) domain but can also be performed in the frequency domain. One of the issues with template matching approaches is ensuring that the template or descriptor remains relatively accurate over time. As time unfolds, the object’s physical and visual properties can change and may no longer resemble the templates or the descriptors. Conversely, the templates or descriptors may have been obtained under specific conditions that do not match the current scene.

Alternatively, object recognition applications can be based on machine learning and artificial intelligence (AI) algorithms such as neural networks, decision trees, genetic algorithms, and support vector machines (SVM) to name a few. These AI algorithms are trained (typically in an offline process) to recognize features that distinguish the true object from its surroundings. These algorithms usually require a large dataset of imagery exposing the desired object under various viewing angles and conditions. If the object is previously known and a suitable training dataset is available, this type of algorithm can generate high recognition probabilities. However, as with the template matching algorithms, machine learning algorithms are likely to yield low success rate if the object is not previously known or if its appearance is different than in the training dataset.
Object recognition and classification research found in the open literature generally use image datasets from a specific band of the Electromagnetic (EM) spectrum [1] such as X-ray, ultraviolet (UV), visual (visible) or thermal (infrared). Multispectral image analysis is typically used in military and surveillance applications.

The following research investigates how features from visual and thermal imagery can be used jointly to improve the recognition rates of commonly found objects in an office setting. Naturally, the choice of objects was limited to those that radiate thermal energy. A multispectral dataset (visual and thermal) was captured and specific features were extracted to train several SVM classifiers. The SVM’s class prediction abilities were evaluated separately on the visual, thermal and the multispectral dataset. Commonly used performance metrics were applied to assess the sensitivity, specificity and accuracy of each classifier.

The intended application for this research is to support machine vision systems, such as mobile robots, trained to detect objects of interest in unknown environments for first responders and security forces. The research could find numerous other applications in medical image analysis, satellite imagery analysis, as well as in surveillance systems.

1.2. Research Objectives

The fundamental challenge of this research was to determine if the statistical classification and recognition rates of common objects could be improved by combining their visual (color) and thermal characteristics together. The challenge was not to find a new way to “fuse” the visual and thermal images together but rather to extract multispectral features in order improve a machine vision system’s ability to recognize specific classes of objects. As a result, the following objectives were established as part of this research effort:

[1] Identify segmentation algorithms that correctly extract the objects of interest from the background. The selected algorithm(s) needs to operate on both visual and thermal imagery.

[2] Acquire a meaningful collection of visual and thermal imagery that represent typical objects found in an office setting.

[3] Identify a series of image features (descriptors) that can be extracted from both visual and thermal imagery.
4 Identify optimum feature(s) from the visual-only, thermal-only and combined visual-thermal to maximize the classification results.

5 Demonstrate that the classification rates of a combined visual with thermal features are better than the individual classification rates of visual-only and thermal-only.

1.3. Thesis Organization

The remainder of this thesis is organized as follows. Chapter 2 presents a brief review of related works in the field of image segmentation, feature detection, classification and performance metrics. Chapter 3 discusses the data collection process, the datasets used for the experiments and the image preprocessing. Chapter 4 presents the methodology which discusses various segmentation algorithms implemented and compares the results to a new algorithm developed specifically for this research project. This chapter also discusses the choice of features and classifiers used as part of the experiments. Chapter 5 presents the implementation of the software specifically developed to support this research. Chapter 6 discusses the experimental evaluation procedures and experimental classification results obtained. Finally, Chapter 7 summarizes the experiments and suggests future work.
Chapter 2 Literature Review

2.1. Segmentation

The process of image segmentation consists of separating foreground objects in an image or scene from their background surroundings. This is often a critical first step in many computer and machine vision applications. Segmented images can subsequently be used to perform feature extraction, object detection and recognition, classification, motion estimation and tracking as illustrated in Figure 2-1.

In a 2008 publication [2], Zhang et al. stated that over 1000 references had been published on the subject of segmentation algorithms and, at the time, over 150 of those were specifically for visual images. Some of the more common segmentation algorithms are based on histogram thresholding, feature clustering, edge detection, region-based (region growing, region splitting/merging), fuzzy techniques and neural networks (supervised and unsupervised).

The subject of image segmentation has been thoroughly studied; however the selection and performance of the algorithms are very specific to the application for which it is being used. As an example, in a traffic sign recognition application, segmentation processes may use specific colors and shapes as the main discriminating factors to detect signs in the driver's field of view. The same algorithms would not necessarily be suitable for a Magnetic Resonance Imaging (MRI) processing or satellite-based remote sensing applications.

One type of algorithm commonly used is the K-means algorithm [3][4] which attempts to segment \( n \) data points into \( k \) clusters. This segmentation algorithm is an unsupervised learning technique but requires a general understanding of the dataset in order to determine the expected number of clusters. The centers of the \( k \) clusters are initialized randomly and eventually converge to final locations. The segmented results from a dataset can vary based on the number of clusters and their initial centers. The final location of the cluster centers is determined when the cluster error function is minimized. When the standard K-means
algorithm is applied to imagery, it typically requires that the color depth be converted to 8-bit greyscale imagery resulting in the potential loss of clustering information. Some of the principal drawbacks of the standard K-means are how to determine the correct number of clusters and the random initialization of the cluster centers. Many authors have focused their research on addressing these two issues and as a result several variations of adaptive K-means algorithms were proposed.

Bhatia [5] proposed two techniques based on minimum distances and thresholding to cluster a dataset without any prior information of the data required by the user. His approach consisted of creating, deleting and merging clusters until the Euclidean error function was minimized. Bhatia demonstrated his approach by segmenting color palettes into perceptually uniform colors and claimed it had the ability to effectively cluster multispectral data by minimizing the Euclidean distance \( E_k \) [5] between two \( n \)-dimensional data points (or a point and the center of a cluster) defined by \( E_1 = \{E_{i1},E_{i2},E_{i3},E_{i n}\} \) and \( E_2 = \{E_{21},E_{22},E_{23},E_{2 n}\} \):

\[
E_k = \sqrt{(E_{i1} - E_{k1})^2 + (E_{i2} - E_{k2})^2 + (E_{i3} - E_{k3})^2 + \ldots + (E_{i n} - E_{k n})^2}
\]  

(1)

This implementation makes use of the full color depth of an image (i.e. no need to convert to greyscale), but does not consider the clusters’ size, shape or individual pixel location within the image. This is a critical aspect of the segmentation required for this research and as such this approach is likely not suitable for this application.

Chen et al. [6] proposed an adaptive K-means algorithm that detected the number of clusters and their initial centers by analyzing the image’s histogram. Using a false-peak mean shift, their proposed algorithm detected the relevant peaks (number of clusters) in the histogram and their location (initial centers). A set of conditions were applied to determine the relevancy of the peaks based on their size and location with respect to other peaks. Similarly to Bhatia’s [5] approach, this technique does not require any prior knowledge of the imagery. However, the algorithm as presented requires an image to be converted into greyscale prior to processing which results in information loss.

Can et al. [7] proposed to use Scale Invariant Feature Transform (SIFT) features with the Bag-of-Features (BOF) technique for detection and tracking of sea-surface targets in infrared (IR) and visual band video streams. They used the K-means algorithm to generate clusters in the visual band. This manual process involved the input of an operator to select a \( k \) value based on the number of ships in the sensor’s field of
The detection and tracking was performed in the individual bands and did not combine the features or information from the different bands to improve the tracking results. The training and testing of the BOF was used to track the target from frame-to-frame as opposed to recognize the various classes of objects.

Sina [8] explored visual band satellite image segmentation using biologically-inspired concepts and algorithms. Biologically-inspired computer vision has grown into its own research area with a primary goal of reproducing the performance and capabilities of the human visual system. Humans can visualize continuously changing scenes while searching, identifying, recognizing and tracking multiple objects under various lighting, viewpoints, occlusions and backgrounds. Humans can observe a scene, understand it, navigate within it and most importantly learn from it. This is obviously just a glimpse of what the human visual system can do but is a very challenging task for even today's most sophisticated computer vision system. This research area encompasses experts from various fields such as cognitive science, neuroscience, psychophysical and physiological sciences as well as computer sciences and engineering. Bio-inspired computer vision applications are based on the same processes, stages and constraints of the human visual system. Such systems are characterized as "bio-inspired" if their design, implementation and results can be correlated or matched to biological research findings. A brief overview of the human visual system is provided in [9] and states that light captured by the human eye advances through various stages in the brain starting with a coarse detection of color, contrast [10] and orientation. This process is also known as early vision. As the information progresses in parallel to "higher stages" of the brain, visual attention or focus is established (with finer detail) on a specific area of the presented scene. The focus point is often driven by objects that stand out from the background or can be task-dependent. As described by Itti et al. [11], when presented with a visual input, it is believed that the human visual system performs parallel feature detection (shapes, color, orientation, size, etc.) and weighting operations to recombine the results into one or more "topographic" map(s). In computational visual systems, these maps are called saliency maps and are typically represented by greyscale images where the brightest areas indicate the locations in the image that significantly stand out from their surroundings. It is believed that visual attention is primarily conducted in a rapid bottom-up manner followed by a slower top-down manner. Bottom-up factors are elements in the scene that purely stand out from the background such as a red dot on a white sheet while top-down factors are built on prior knowledge of the searched objects (e.g. searching for a family member in a crowd). Saliency maps can be biased by task-dependent operations where prior knowledge of the object or the target and its background are used to segment and
ignore irrelevant objects. Bio-inspired algorithms could be used to support segmentation functions in this research as well as object recognition in the visual spectrum but because thermal vision is not an inherent human capability, it could be argued that it is not truly “bio-inspired”.

Another common segmentation algorithm is the Watershed [12] algorithm which was inspired by the field of topography whereby a geographical region is decomposed into peaks and valleys. The classic analogy is when water is dropped over an area and flows downwards to the lowest point which is called a catchment basin. As the water continues to flow, several localized basins (minima) eventually merge and create larger basins leaving only the highest points (maxima) or watershed lines unsubmerged. In image processing, the image topography is defined by the greyscale intensities of the pixel and this concept is used as a segmentation technique. Using classic mathematical morphology operations, the concepts of local minima and maxima, catchment basins and watershed lines can be extracted from digital images. The Watershed segmentation concept has since been exploited and robust algorithms [13][14] have been developed. The Watershed algorithm demonstrates a lot of potential for segmenting complete objects as it considers edges and gradient changes in the imagery unlike other thresholding algorithms that are only concerned with individual pixels.

Gupta and Mukherjee [15] proposed a segmentation algorithm based on Enhanced Fuzzy C-Means clustering for automatic detection system using thermal imagery. Fuzzy C-Means is closely related to K-means whereby a data point belongs to a cluster with a certain degree of certitude (fuzzy) instead of belonging to just one cluster (K-means). In their algorithm, the optimum number of clusters was estimated using the validity measures Global Silhouette (GS) Index and Separation Index (SI). The GS index was calculated for a large number of clusters (up to 20) and the number of clusters with highest index was chosen as the optimum number. The SI also provided a cluster quality measure. Although not specifically indicated, this type of implementation likely has a significant processing cost as compared to the standard and adaptive K-means. This type of algorithm could improve segmentation results but it is unclear how effective it could be when segmenting both visual and thermal imagery.

Hasanzadeh and Kasaei [16] proposed a multispectral segmentation method based on size-weighted fuzzy clustering and membership connectedness. This advanced fuzzy clustering technique took into consideration the local and global position of the image pixels as part of the segmentation process. This approach was developed primarily for thermal imagery and used the spectral and spatial content of the image to improve the clustering performance over the standard and adaptive K-means and Fuzzy C-means
This proposed algorithm has shown high potential for thermal imagery but it is unclear how it would perform on greyscale visual imagery.

2.2. Image Features

In computer vision applications, image features are properties of a scene or of a specific object within that scene that can be extracted to describe the entity. A feature can be something as simple as an object’s size or its intensity and can be as detailed as its texture. Features are typically regrouped into three categories; shape, color and texture. A series of features used to describe a scene or an object is referred to as a feature vector or descriptor. For example, a commonly used shape descriptor is the Hu Moments [17] which is composed of a seven dimensional vector. An example of a texture descriptor is the Legendre Moments [18] which is extracted from a local binary pattern and is invariant to translation, scaling and uniform contrast changes. Feature vectors or descriptors such as Hu and Legendre moments can be used by a feature matching algorithm in conjunction with a correlation process or by a machine learning system to identify objects in a scene that closely resemble an object of interest.

Many other object descriptors have been proposed such as SIFT, Speeded Up Robust Features (SURF), Features from Accelerated Segment Test (FAST), Binary Robust Independent Elementary Features (BRIEF), Oriented FAST and Rotated BRIEF (ORB). Others algorithms based on Mean-Shift, Continuously Adaptive Mean-Shift (CAMSHIFT), covariance, Principal Component Analysis (PCA), and various edge/corner detection methods such as Canny, Harris, Sobel and SUSAN have been used to detect and track objects of interest. In cases where feature extraction is used for tracking, these algorithms often work in conjunction with a variation of a Kalman or Particle filter to predict the size, position, velocity and acceleration of the targets.

Ebrahimi and Mayol-Cuevas [19] discussed the evolution of local feature detectors and their relative detection speed. SIFT is one of the most popular feature detector and descriptor with a feature length of 128. FAST is a corner detection method widely used to speed up and approximates SIFT. FAST uses box filters (computed using integral images) to approximate the Difference of Gaussian (DoG). SURF was initially inspired by SIFT and also uses box filters to approximate Hessian-Laplace detector. SURF is commonly described as being faster than SIFT and more robust. Machine learning based on WaldBoost is another technique for speeding up feature detectors. Ebrahimi and Mayol-Cuevas applied adaptive sampling to further speed up FAST and illustrate how it can be used to recognize objects in visual imagery.
but their work was focused on reducing the detection time of objects in long sequences of frames. Their work has relevancy to this research but cannot be compared directly.

Jang and Turk [20] developed a real time car recognition application based on the SURF feature descriptor. The feature descriptors were converted into a single value word using a vocabulary tree. The single value word was used to search an image database. Structural matching was then used to score the returned image search in a ranked list. They used three databases of toy cars and achieved between 65% and 92% accuracy over the various classes of cars to be recognized. Their work demonstrates the potential that SURF feature descriptors offer for this research but was conducted on synthetic visual imagery with controlled background and lighting.

Tsai et al. [21] proposed a novel feature descriptor called CDIKP which combined SIFT with compact feature descriptors (20-Dimension) in comparisons to the state-of-the-art (e.g. SIFT: 128-D, SURF: 64-D, and PCA_SIFT: 36-D) and showed many advantages. The algorithm was tested against several datasets acquired under varying conditions (rotation and distortion). The results achieved by Tsai et al. were comparable and sometimes better than the classic algorithms but using compact feature descriptors as proposed in this research.

Rublee et al. [22] proposed an alternative to SIFT and SURF called ORB. ORB is a binary descriptor based on BRIEF which is invariant to rotations (but not to scale) and resistant to noise. ORB claims to be two orders of magnitude faster than SIFT and SURF and was tested in real-time mobile phone applications using OpenCV 2.3. The focus of their work was not classification results but rather addressing deficiencies in existing feature detection algorithms in visual band imagery.

Pinto et al. [23] compared state-of-art visual features namely SIFT, Pyramid Histogram Of visual Words (PHOW), Pyramid Histogram Of Gradients (PHOG), Geometric Blur, and the bio-inspired Sparse Localized Features (SLF) against two baseline techniques (Pixels, V1-like) to determine the performance of features on image variation (position, scale, pose, illumination). They used a synthetic dataset made up of a series of 3D objects (cars, planes, boats, animals) rotated in various orientation and superimposed on various background types. For classification, they used the L2-regularized Support Vector Machines. They concluded that the bio-inspired SLF consistently performed better than the others in the majority of the tests conducted. They also noted that caution should be taken when making performance evaluation conclusions while using synthetic imagery. This was a significant motivation for using real visual and thermal imagery for this research.
Hartemink [24] used local covariance descriptors to compare bounding boxes (detected objects) in a visual maritime imagery dataset. The purpose of his work was to classify detected objects in individual frames as either a target class or background class. The features used to build the covariance matrix included (for each pixel in the bounding box) intensity, horizontal/vertical position, first derivative of horizontal/vertical position, second derivative of horizontal/vertical position, gradient magnitude and gradient orientation. The dataset included 800 images of maritime environments with some containing no objects (just background) and some containing up to 6 objects. This work was conducted on maritime visual imagery only but the author states that it is generic enough to be applied to thermal imagery as well.

Fehlman and Hinders [25][26] developed a classification system of outdoor non-radiating objects using thermal imagery for the purpose of a mobile robot navigation. The “extended” object classes were objects that extend laterally beyond the field of view of the camera and included brick walls, picket fences, hedges, and wooden walls. The “compact” object classes were objects that were completely laterally visible in the sensor’s field of view such as street poles and various types of trees. The authors showed in their work that no optimal feature vector existed but a most favorable subset could be used by each classifier to minimize the classification error. The most favorable subsets were extracted from a pool of 18 features computed from meteorological, micro and macro features. The meteorological features included ambient temperature and its associated rate of change. Micro features included radiance, background irradiance, emissivity and 3 others. Macro features included first and second order statistical features such as scene radiance, contrast, smoothness, third moment, uniformity, energy, entropy and 6 others. Textural features such as tree bark, brick walls and wood grain were available from the thermal images since the outdoor objects were non-heat generators and did not produce localized saturation in the sensor. The object’s emitted energy (signature) is a function of the amount of thermal radiation received from external sources during the previous diurnal cycle. In an earlier work, Hinders, Gao and Fehlman [27] used sonar and thermal imagery to classify cylindrical outdoor objects such as trees and smooth circular poles for the purpose of mobile robot navigation. The thermal imagery for this study was captured over a period of four months under various conditions. The thermal images were segmented with three center and three periphery segments. A Retinex algorithm was applied to the segmented images to enhance the details while a high-pass Gaussian filter attenuated the lower frequencies and sharpened the image. A median filter was applied as the final processing step to de-noise the image without reducing the sharpness of the image. Four features based on sparsity were extracted from the thermal images. Sparsity is commonly used in visual imagery to distinguish between manmade and
natural objects. Much can be learned from the work of Fehlman et al., specifically on which feature set to use with thermal imagery. However, their work was focused on recognition of objects found in an outdoor environment and as a result most of the thermal-physical features (temperature, emissivity, reflectivity, and surface reflection) cannot be extracted from indoor objects not subjected to diurnal cycle.

Cayouette, Labonté and Morin [28] investigated the possibility of incorporating a Probabilistic Neural Network (PNN) in an Automatic Target Recognition (ATR) system for an imaging IR seeker emulator. A seeker is the principal component of a missile system which performs the task of searching, acquiring and tracking a target of interest such as an aircraft. A seeker emulator is a hardware representation of a real system used to conduct hardware-in-the-loop or ground-based testing. In this case, the seeker emulator devised for their experiment consisted of a 256x256 focal plane array (FPA) IR camera operating in the mid-wave (3-5μm) band. The images acquired from the seeker emulator were processed and only the aircraft and decoy flares (used to protect aircraft from IR-guided missiles) remained for the discrimination by the neural network. They chose a variant of the PNN from several other types mainly for its ability to perform pattern classification. Unlike other types of neural networks, the PNN outputs a confidence level that the recognized patterns belong to a certain class of objects. They trained and tested the network with 758 aircraft images and 506 decoy flare images. They divided the images into four subsets: a training set and validation set for both the aircraft and the flare. They conducted several tests by shuffling the training and validation datasets and achieved between 95% and 99.43% success rate in correctly identifying the aircraft and the flares. The features they initially considered included:

**Intensity Features:**

- maximum intensity \( Z_{\text{max}} \)
- average intensity \( \bar{Z} \)
- intensity variance, or second moment \( \mu_{Z}^{2} = \frac{1}{n} \sum_{i=1}^{n} (Z_{i} - \bar{Z})^{2} \)
- third moment of the intensity distribution \( \mu_{Z}^{3} = \frac{1}{n} \sum_{i=1}^{n} (Z_{i} - \bar{Z})^{3} \)

**Shape Features:**

- area \( A \)
- coordinates \( (\bar{x}, \bar{y}) \) of the centroid
- perimeter $P$
- roundness $R = \frac{P^2}{4\pi A}$
- angle of the principal axis of minimum inertia $\theta$
- small principal moment of inertia $I_{\text{min}}$. That is the smallest possible moment of inertia about an axis that passes through the centroid of the blob.

$$I_{\text{min}} = \frac{1}{2}(I_{xx} + I_{yy}) - \frac{1}{2}\sqrt{(I_{xx} - I_{yy})^2 + 4I_{xy}^2}$$

where $I_{xx}$, $I_{yy}$ and $I_{xy}$ are components of the inertia tensor matrix.

- large principal moment of inertia $I_{\text{max}}$. That is the largest possible moment of inertia about an axis that passes through the centroid of the blob.

$$I_{\text{max}} = \frac{1}{2}(I_{xx} + I_{yy}) + \frac{1}{2}\sqrt{(I_{xx} - I_{yy})^2 + 4I_{xy}^2}$$

- aspect ratio $AR = \sqrt{\frac{I_{\text{max}}}{I_{\text{min}}}}$
- maximum radial distance $D_{\text{max}}$
- minimum radial distance $D_{\text{min}}$
- average radial distance $\bar{D}$
- variance or second moment of the distance distribution $\mu_D^2 = \frac{1}{n} \sum_{i=1}^{n} (D_i - \bar{D})^2$

From this list of intensity and shape features, they observed the discriminability between the two classes for each feature and selected 13 features. The 13 features were normalized to make them invariant to rotation and translation. The 13 invariant features they chose were:

- normalized maximum intensity $\frac{Z_{\text{max}}}{Z}$
- normalized average intensity $\frac{Z}{A}$
- normalized variance of the intensity distribution \( \frac{\mu_Z^2}{A^2} \)
- normalized third moment of the intensity distribution \( \frac{\mu_Z^3}{A^3} \)
- normalized square root of the minimum moment of inertia \( \frac{\sqrt{I_{\text{min}}}}{A} \)
- normalized square root of the maximum moment of inertia \( \frac{\sqrt{I_{\text{max}}}}{A} \)
- normalized maximum radial distance \( \frac{D_{\text{max}}}{\sqrt{A}} \)
- normalized minimum radial distance \( \frac{D_{\text{min}}}{\sqrt{A}} \)
- normalized average radial distance \( \frac{\bar{D}}{\sqrt{A}} \)
- normalized second moment of the distance distribution \( \frac{\mu_D^2}{A} \)
- angle of the principal axis of minimum inertia \( \theta \)
- aspect ratio AR
- roundness R

Labonté and Morin [29] continued the work previously started with Cayouette [28] on the use of PNN to discriminate target aircraft and flares in static images. In the follow-up study, Labonté and Morin considered the time evolution of the image features from a series of frames. The target features identified in the previous study, were invariant under rotation and translation in the image but some were strongly dependent on the distance separating the object from the sensor. They defined temporal characteristics (independent of the separation distance) for the aircraft intensity and shape that differentiated it from the decoy flares. The purpose of their work was to determine if there were sufficient qualitative and quantitative differences in the temporal characteristics to differentiate the aircraft from the decoy flare. The imaging sensor acquired frames at a rate of 30 frames per second. At this rate, Labonté and Morin chose to use eight consecutive frames to evaluate and assess the dynamic characteristics. The complete dataset for this study consisted of 123 8-frame sequences for the aircraft and 50 8-frame sequences for the decoy flare. They used the same artificial neural network described in their previous work. In this
study, Labonté and Morin’s PNN achieved a 97.7% success rate in correctly identifying aircraft and decoy flares from the 173 sequences.

2.3. Classifiers

A classifier is an implementation of a classification scheme whereby an algorithm is used to learn the characteristics of a class or a pattern from a training dataset and subsequently attempts to recognize the pattern in a separate testing dataset. There are several types of classification schemes or machine learning algorithms such as decision trees, neural networks, SVM, probabilistic methods, Nearest Neighbor, Hidden Markov Model and Bayesian to name a few. A commonality between these machine learning algorithms is that they need to be trained prior to being capable of predicting class associations. The learning is typically done in a supervised or unsupervised way. Simply put, in supervised learning techniques, the class association or the class label is provided to the machine learning algorithm. In unsupervised learning approaches, the class association is not provided and the classification algorithm must look for similarities in the dataset for class prediction. Each method has its advantages over the other and ultimately the type of application and the available data greatly influence the method used for the training of the classification algorithm.

Mitri et al. [30] developed a color and scale independent object learning and detection system using a Sobel edge detector and threshold to remove noise from imagery. The application was used to detect soccer balls for the purpose of the Robot Soccer Cup. A Gentle AdaBoost learning technique was used in conjunction with a Classification and Regression Trees (CART) to identify the soccer balls in each frame. The authors claim that their learning and classification system was fast enough for real-time applications. This work was of interest to this research because it used indoor imagery and possible types of classifiers to evaluate.

Andreasson and Duckett [31] used a Minimum Distance Classifier (MDC) to recognize common office objects from an omni-directional camera (visual band) located on a mobile robot. The images were segmented by hand and low level features such as object corners (Sobel operator) were extracted and tracked. High level features such as velocity were also extracted and used as input vector in the pattern recognition classifier. The work is relevant to the current research because it uses indoor office objects such as chairs, tables, drawers, bottles and trash cans. The classification rates achieved varied between 63-100% depending on the object class.
Hartemink [24] used a classification system to assign detected objects in visual imagery as either a target class or background class. The classifiers considered in his work were Linear Discriminant Classifier (LDC), Quadratic Discriminant Classifier (QDC), Nearest Mean Classifier, K-Nearest Neighbour Classifier (KNN), Parzen Classifier and Fisher Classifier. Hartemink concluded that the linear discriminant classifier provided the best overall performance. Hartemink achieved a 51.6% recall with a 38.8% precision with his best classifier concluding that further research was required to increase the recognition rates while lowering the false alarm rate.

Kogut et al. [32] implemented and tested a real-time object classification system for mobile robots using a boosted Cascade of classifiers, which was first proposed by Viola and Jones, and trained with the Adaboost algorithm. They chose to use multiple strong classifiers for individual scales instead of a single strong classifier trained on multiple scales. They refer to a previous study that claims that this approach yields higher detection rates. Their training dataset included positive and negative examples to reduce the influence of the background clutter. Of particular interest in this work is that they used mobile robots to detect soda cans in indoor environments. The second part of their work was focused on detection of humans from a moving platform for navigation purposes by fusing laser and thermal sensors data. In this case, the thermal sensor was used to confirm or assist the laser scanner.

Duda [33] described some of the early work on SVM published in the early to mid-90s which were based on previous work done on margin classifiers (linear machines with margins). SVMs are categorized as linear discriminant classifiers and the general idea behind them is to map data patterns, which cannot be separated by a linear decision boundary, into a much higher dimension. The transformation to the higher dimensional space is achieved through a non-linear mathematical transformation (also known as a kernel) where the input patterns can then be separated by a linear decision boundary or hyperplane. The optimal hyperplane, in the new higher dimensional space, is determined by maximizing the margin (i.e. the separation distance) to the nearest training points. These training points used to compute the margin are known as the support vectors. As stated in [33], a larger margin between the support vectors and the optimal hyperplane typically results in better generalization ability by the classifier. Duda also notes that the support vectors are the hardest to classify but most useful in the design of the classifier.

There are numerous forms of kernels used to transform the feature space into a higher dimension and some of the typical ones include [33] Linear Radial Basis Function, Polynomial, Radial Basis Function (RBF) and Sigmoid. SVMs are, in their basic form, binary or two-category classifiers but can be extended to
handle multi-category classification problems. This is achieved by combining several binary classifiers [34] (i.e. Class A and not-A, Class B and not-B, Class C and not-C) where the output of the binary classifier carrying the largest weight is selected as the predicted class. The drawback to this approach is that there may be ambiguous regions which cannot be assigned to one class.

Fehlman and Hinders [25][26] conducted an extensive search for the most favorable feature set and totaled over 290,000 combinations reaching up to 18 dimensions. The favorable feature subsets differed for each of the three evaluated classifiers. The selection of classifiers is greatly dependent on the type of applications used, available features and nature of the classification problem. In their case, Fehlman and Hinders wanted to “achieve a minimum classification error while retaining the physical interpretation of the information in the signal data throughout the entire classification process”. They chose not to use the neural networks because “they tend to hide the physical interpretation”. They identified that pattern matching classifiers are “sensitive to intra-class variation” and therefore not suited either for their application which was greatly dependent on diurnal cycle of solar energy. They chose to use statistical classifiers, specifically nonparametric classifiers with probabilistic decision process. The three classifiers were Bayesian, K-Nearest-Neighbor (KNN) and Parzen. The authors observed that certain classifiers consistently misclassified objects class while others correctly classified objects. As a result they chose to form committees of experts for classifying specific classes. This was the baseline for their novel Adaptive Bayesian Classifier model. The use of a committee of expert for this thesis work was inspired by Fehlman and Hinders’ work.

2.4. Performance Metrics

The performance of classifiers is generally characterized by basic metrics that determine how well the classifier can correctly identify or reject samples belonging to a specific class. Several sources [24][35] define and provide examples and use these metrics as part of their analysis. The basic metrics listed below are typically used in binary classifier (i.e. two classes) but can also be extended to multi-class problems by comparing one class to the rest (A vs. not-A).

- True positive (TP): defines the number of correctly identified samples from the class of interest.
- True negative (TN): defines the number of correctly rejected samples from the class of interest.
- **False positive (FP):** defines the number of incorrectly identified samples from the class of interest. Also referred to as false alarm rate or Type I error.

- **False negative (FN):** defines the number of incorrectly rejected samples from the class of interest. Also referred to as Type II error.

From these four basic metrics, several other global performance metrics can be computed to assess the performance of a classifier. These metrics are not very useful individually and are typically quoted together when assessing the performance of a classifier.

- **True Positive Rate (TPR):** defines the classifier’s ability to correctly identify a specific class from a sample dataset. Also referred to as Sensitivity or Recall.

\[
TPR = \frac{TP}{TP + FN}
\]  

(2)

- **True Negative Rate (TNR):** defines the classifier’s ability to correctly reject a specific class from a sample dataset. Also referred to as Specificity.

\[
TNR = \frac{TN}{TN + FP}
\]  

(3)

- **Positive Predictive Value (PPV):** determines the chances that an identified sample truly belongs to the specific class. Also known as Precision.

\[
PPV = \frac{TP}{TP + FP}
\]  

(4)

- **Negative Predictive Value (NPV):** determines the chances that a rejected sample truly does not belong to the specific class.

\[
NPV = \frac{TN}{TN + FN}
\]  

(5)

- **Accuracy (Acc):** defines the number of correctly identified samples (true positives and true negatives) among the total number of samples.

\[
Acc = \frac{TP + TN}{TP + TN + FP + FN}
\]  

(6)
- **F1 score** ($F_1$): defines a weighted average of the precision and recall. F1 score is a measure of a test’s accuracy.

$$F_1 = 2 \cdot \frac{PPV \cdot TPR}{PPV + TPR}$$

(7)

The performance metrics listed above are typically used to assess the performance of classifiers or to compare different classifiers against a common dataset. Other metrics have been used such as the ones presented by Can et al. [7] which assessed the performance of tracking algorithm. Although tracking performance is beyond the scope of this research project, some of these metrics could potentially be used to compare and assess various segmentation algorithms. In their work, Can et al. proposed to use SIFT features with the BOF technique for detection and tracking of sea-surface targets in the thermal and visual band. They used four different evaluation criteria to compare the performance of their proposed algorithm to others. The four metrics described in their work are:

- **Metric 1 (M1):** M1 defined the Euclidean distance between the center of the ground truth of the target region and the center of the detected target region.
- **Metric 2 (M2):** M2 defined the city block distance between the center of the ground truth data and the center of the detected target area.
- **Metric 3 (M3):** M3 defined the ratio between the undetected target area and the total target area (false negative rate). This metric gives what percentage of the target is missed.
- **Metric 4 (M4):** M4 defined the true positive rate, which is calculated as the ratio between the correctly detected target area and the whole detected target area.

2.5. Summary of Literature Review

In terms of segmentation, numerous implementations of K-means, adaptive K-means and fuzzy variations have been proposed. The main issue with this type of algorithm is that they cluster pixels based on their pixel intensities and not on their spatial location within the image which makes it difficult to extract complete objects from their background. Conversely, the Watershed algorithm shows more potential to extract complete objects because it is based on gradients within the image. The majority of the segmentation algorithms reviewed were specifically designed for visual imagery and their performance on thermal imagery is unknown. The selected algorithm for this research had to effectively extract the
same object from both visual and thermal image pairs. A new variation of the Watershed was developed specifically for this research and demonstrated very good results.

The literature review found many different types of feature descriptors used for detection of various types of objects. Commonly used descriptors include SIFT, SURF, FAST, PCA with various types of edge detection such as Canny, Harris and Sobel. The focus of many publications has been to improve the performance of these classic algorithms. The issue with these descriptors is usually the large number of dimensions they use and the fact that their performance is unknown on thermal imagery. The features presented by Cayouette et al. [28] were compact (13-D) and demonstrated high recognition rates (>97%) on thermal imagery. For these reasons, the features used by Cayouette et al. were selected for this research instead of the classic descriptors.

Several classifiers namely Bayesian, KNN, Parzen, LDC, QDC and decision trees were used for various types of machine learning applications. It was decided to use a multi-class SVM for this research and demonstrate its ability to distinguish between multiple classes using features extracted from multispectral imagery. The literature review found several sources providing classification rates for visual imagery, but very few for thermal and even less for combined visual and thermal. The sources identified for combined visual and thermal were of particular interest to this research but none of the works were really comparable to this research. Many of the sources found used synthetic visual images which often yields better classification results in comparison to real imagery. In the case of thermal imagery, many of the sources found were based on outdoor environment [25][26][36] which significantly changes the thermal signature of the objects in comparison to an indoor setting and makes it difficult to compare the recognition rates. In the case of combined visual and thermal dataset, one publication [37] used the visual band imagery for daytime recognition while the thermal imagery was used for night time but the two were not combined in any way. In another source [38], the classification was performed individually on each band and features were not combined in any way. Data fusion is a complete research area on its own and one source [39] demonstrated how fusing visual and thermal imagery can improve recognition. Although image fusion showed interesting results, it was not the purpose of this research. Finally, the closest example [40] found of feature extraction from combined visual and thermal imagery used three color bands (RGB) and a near infrared band to segment an outdoor scene (background sky, trees, car and people). Due to a limited amount of directly comparable sources found in the open literature, the results presented in this research were not compared to any previous works.
Chapter 3  Data Collection and Datasets

A review of the open literature did not find a suitable dataset for the purposes of this research. The datasets found typically consisted of imagery from the visual or thermal bands of the EM spectrum but rarely from both. In the rare datasets that did have both visual and thermal imagery of the same scene, the viewpoint was very far away and made it very difficult to extract details from the potential objects. In another case, the imagery contained only one class of objects which was again not suitable for the purpose of this research. As a result, no suitable dataset was available and a custom set of matching visual and thermal imagery was captured.

In defining the objectives of this research, it was decided that the imagery collected would be in an indoor setting to have better control on the environment. Thermal imagery is greatly dependent on the ambient environment in which it is obtained. The thermal signature of an object changes considerably with time of day, time of year and atmospheric conditions (sunny day, rainy day or winter day). This requirement added additional challenges since only a limited class of objects actually radiate thermal energy. In outdoor settings, objects may reflect solar/lunar energy depending on their material and other surface properties which may produce distinct multispectral signatures. Other classes such as people and animals were considered, but in order to prevent privacy issues associated with photographing individuals, it was decided not to use a human class for this research.

This section discusses the process and equipment (hardware and software) used to acquire the dataset for the experimental evaluation described in Chapter 5. A total of 173 image pairs were acquired and pre-processed for this experiment. From this set, the image pairs were divided into a training set and testing set.

3.1. Camera Specifications and Image Analysis Software

The image dataset was acquired using a Fluke Ti10 Thermal Imager. The hand held camera encloses a visual and thermal detector which allows its user to acquire nearly co-located images of an object in both visual and thermal bands. The Ti10 specifications are tabulated in Table 3-1. The Ti10 allows the user to select one or more palette to display the apparent temperature of the scene in the camera’s field of view. For the purpose of this project, the thermal image was mapped to a greyscale palette to facilitate comparison to greyscale visual images.
### Table 3-1: Fluke Ti10 camera specifications

<table>
<thead>
<tr>
<th>Specification</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Field of view</td>
<td>23° x 17°</td>
</tr>
<tr>
<td>Spatial resolution (IFOV)</td>
<td>2.5 mRad</td>
</tr>
<tr>
<td>Minimum focus distance</td>
<td>Thermal lens: 15 cm (6 in)</td>
</tr>
<tr>
<td></td>
<td>Visible (visual) light lens: 46 cm (18 in)</td>
</tr>
<tr>
<td>Image frequency</td>
<td>9 Hz refresh rate</td>
</tr>
<tr>
<td>Detector type</td>
<td>160 X 120 FPA, uncooled microbolometer</td>
</tr>
<tr>
<td>Infrared lens type</td>
<td>20 mm F = 0.8 lens</td>
</tr>
<tr>
<td>Thermal sensitivity (NETD)</td>
<td>≤ 0.13 °C at 30 °C target temp. (130 mK)</td>
</tr>
<tr>
<td>Infrared spectral band</td>
<td>7.5 μm to 14 μm</td>
</tr>
<tr>
<td>Visual camera</td>
<td>640 x 480 resolution</td>
</tr>
</tbody>
</table>

The collected dataset consisted of common office items with a thermal signature. Examples of these items include a recently charged mobile phone, a coffee cup with a hot beverage, a laptop charger, a desk lamp and a portable heater. A sample of the five classes of objects is illustrated in Figure 3-1.
Fluke’s SmartView 3.7.23 was used to analyse the imagery captured by the Ti10 thermal imager. A screen shot of the SmartView user interface is illustrated in Figure 3-2. The image captured by the Ti10 camera was exported to the Fluke .ISO file format and using the SmartView software converted to visual and thermal (greyscale) bitmap images. The software permits blending of two bands (visual-thermal) into a single image as illustrated in Figure 3-3. SmartView also allows to set the color palette, the lower/upper scale limits, the object emissivity and background temperature to correctly estimate the object’s apparent temperatures.
Figure 3-2: SmartView software user interface

Figure 3-3: Various levels of blended visual-thermal images using SmartView. Full thermal (left), half-blended (center) and full visual (right).
3.2. Image Preprocessing

It was decided that the training and testing datasets would be captured using different environments in order to replicate real world scenarios where a machine vision system, such as a mobile robot, could be trained to detect an object in an unknown environment. The dataset consisted of a total of 173 image pairs captured using the Ti10 thermal imager and was divided into 44 training image pairs and 129 testing image pairs. All of the training images contained a single object class per image. In the case of the testing dataset, several images contained multiple object classes in the same image. This resulted in 44 instances of objects used in the training dataset and 165 instances of objects for the testing dataset. The breakdown of each class in the training and testing datasets is described in Table 3-2.

<table>
<thead>
<tr>
<th>Class</th>
<th>Name</th>
<th># of occurrence in the dataset</th>
<th>Object Temperature Range (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Training</td>
<td>Testing</td>
</tr>
<tr>
<td>1</td>
<td>Mobile Phone</td>
<td>7 (15.9 %)</td>
<td>23 (13.9 %)</td>
</tr>
<tr>
<td>2</td>
<td>Coffee Mug</td>
<td>9 (20.5 %)</td>
<td>44 (26.7 %)</td>
</tr>
<tr>
<td>3</td>
<td>Laptop Charger</td>
<td>10 (22.7 %)</td>
<td>28 (17.0 %)</td>
</tr>
<tr>
<td>4</td>
<td>Desk Lamp</td>
<td>9 (20.5 %)</td>
<td>40 (24.2 %)</td>
</tr>
<tr>
<td>5</td>
<td>Portable Heater</td>
<td>9 (20.5 %)</td>
<td>30 (18.2 %)</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>44</td>
<td>165</td>
</tr>
</tbody>
</table>

The training images were captured using the same background for all objects at various viewing angles and distances to the camera. Examples of the training dataset are illustrated in Figure 3-4 to Figure 3-8. Each sample contains a visual band image (top) and a thermal band image (bottom).
Figure 3-5: Sample training dataset for Class 2 (Coffee Mug)

Figure 3-6: Sample training dataset for Class 3 (Laptop Charger)

Figure 3-7: Sample training dataset for Class 4 (Desk Lamp)
The training dataset was not used for testing of the classifiers. Similarly, the testing dataset was not used to train the classifiers. To generate the testing dataset, the same five objects were positioned in different places within two different office spaces. The class object imagery was captured at different distances, under different lighting and in many cases with several objects in the same scene. In order to preserve the thermal signature of the coffee mug relatively constant in all of the testing dataset, the mug was refilled several times with boiling water. Similarly, the cell phone was placed back on a wireless charger for several minutes and the portable heater was restarted for several minutes as well.

In order to challenge the segmentation algorithms, the testing images were captured with the class objects positioned in typical office settings such as on bookshelves, on a work desk next to other objects of the same size and color as well as on a textured carpet. Several testing dataset examples are illustrated in Figure 3-9 to Figure 3-13. Each sample contains a visual band image (top) and a thermal band image (bottom).
Figure 3: Sample testing dataset for Class 1 (Mobile Phone)

Figure 3-10: Sample testing dataset for Class 2 (Coffee Mug)

Figure 3-11: Sample testing dataset for Class 3 (Laptop Charger)
Figure 3-12: Sample testing dataset for Class 4 (Desk Lamp)

Figure 3-13: Sample testing dataset for Class 5 (Portable Heater)

The SmartView software allows the user to correctly align (horizontally and vertically) the visual band image with the thermal band image to ensure both images are properly positioned prior to segmentation. As part of the preprocessing steps, some of the images in the training and testing datasets were realigned.

Prior to exporting each .ISO image to the thermal greyscale palette, the minimum and maximum temperature thresholds were adjusted to 40°C and 54°C respectively. This temperature range covered the majority of the thermal images and was selected to maintain as much of the details as possible without saturating too many pixels while allowing discriminability between intensity-based features. Adjusting the thresholds ensured that all image temperatures were compared using the same range. However, the selected temperature range still caused some saturation and loss of details in the images with very warm surfaces as illustrated in Figure 3-14.
In the case of cooler images, adjusting the minimum and maximum temperature threshold reduced the contrast but removed most of the noise in the image as illustrated Figure 3-15.

Figure 3-14: Original thermal image (left) and the matching saturated image (right) caused by a temperature range adjustment

Figure 3-15: Original thermal image (left) with reduced contrast and noise (right) caused by a temperature range adjustment
Once the minimum and maximum temperature thresholds of all the dataset had been adjusted, each .ISO image was exported to a visual image and thermal greyscale image as shown in Figure 3-16.

![Visual band image (left) and its matching thermal image (right) mapped to a greyscale palette](image)

Figure 3-16: Visual band image (left) and its matching thermal image (right) mapped to a greyscale palette

The process detailed in this section was applied to all of the image pairs to create complete and meaningful training and testing datasets. The following section describes the details of the experimental methodology used for this research.
Chapter 4  Methodology

4.1.  Overview

This section provides a general overview of the methodology used to develop the necessary software tools to achieve the objectives of the research. As a reminder, the main objective of this research was to determine if the statistical classification and recognition rates of common objects could be improved by combining their visual and thermal characteristics (features) together. The principal workflow used to meet the objective of the research is illustrated in Figure 4-1 and consists of first segmenting the image to extract the objects of interest, extracting the desired features from the segmented images, finally training the SVM classifiers and evaluating their class predictability against the testing dataset. Each of these main components of the research are discussed in Sections 4.2 to 4.4. The software implementation is discussed in detail in Chapter 5.

![Figure 4-1: Principal workflow](image)

4.2.  Segmentation

Segmentation of the foreground objects from their background is a critical first step for feature extraction and classification of an object. For the purpose of this research, several classic segmentation algorithms such as the Basic Threshold, K-means, Contours and Watershed with Distance Transformation Markers were implemented and tested using representative samples from the training and testing multispectral dataset. It was determined that these algorithms did not provide the segmentation capabilities required to meet the objectives of this research. As a result, a new segmentation algorithm called Watershed with Thermal Markers was developed as part of this research. The new algorithm is described in Section 4.2.6 and its performance is compared to the classic algorithms.
4.2.1. Basic Threshold

The classic Basic Threshold segmentation algorithm was implemented using the segmentation [41] function from the OpenCV library. In this algorithm, a user-defined threshold separates the pixels in an image (visual or thermal) into two groups based on the pixel intensity level. The pixels in the image with intensity levels above the user-defined threshold are assigned a color of white while those below are assigned a color of black. This classic algorithm works on color images but they must be converted to 8-bit greyscale prior to segmenting. The conversion to greyscale results in a loss of information as compared to other types of algorithms. Furthermore, the algorithm does not take into consideration the state of adjacent pixels as part of the segmentation process.

4.2.2. K-means

The K-means algorithm was implemented using the kmean [42] function from the OpenCV library. In this implementation, the image (visual or thermal) was first converted to greyscale and blurred to facilitate the clustering of pixels with similar intensities. The blurring function is implemented using the OpenCV blur function [43] which implements a normalized box filter. The algorithm separates the $n$ pixels in the image into $k$ clusters. For the purposes of this exercise, the $k$ value was set to 2 to separate the images into 2 clusters (background and foreground).

4.2.3. Contours

The Contours algorithm was implemented using the findcontours and drawcontours [44] functions from the OpenCV library. In this implementation, the image (visual or thermal) was first converted to greyscale and blurred using the OpenCV functions. The blurring function was implemented using the OpenCV blur function [43] which implements a normalized box filter. A Canny edge detection algorithm was then used to identify primary edges in the image prior to the findcontours algorithm that links these edges to highlight the outlines of various connected components in the scene.

4.2.4. Watershed with Distance Transform

The Watershed implementation for this research was based on an example [45] in the open literature whereby the basic watershed algorithm was enhanced with markers identifying clusters of pixels
belonging to the same object. There are several ways to create the markers but in this example the `distanceTransform` function from the OpenCV was used. The Distance Transform works on a binary image and converts each of the white pixels to greyscale value representing the smallest distance to the background (black pixels). An example of the Distance Transform operation is illustrated in Figure 4-2.

Figure 4-2: Example of the Distance Transform calculation. Original image (left), binary (center), distance transform (right).

The Watershed with Distance Transform implementation consisted of first converting the image to binary using a threshold of 40 (as suggested by the original author [45]) and then applying the `distanceTransform` function to the resulting binary image. A final threshold operation (using a threshold value 127 as the middle point on a scale of 0 to 255) was applied to the output of the `distanceTransform` function to create the markers for the watershed algorithm.

4.2.5. Performance Assessment

A version of the Basic Threshold, the K-means, the Contours and the Watershed with Distance Transform algorithms were implemented as part of this research to find a segmentation algorithm suitable for both visual and thermal images. Samples of the segmentation algorithm results are compared and illustrated in Figure 4-3 to Figure 4-6.

Figure 4-3 illustrates a training sample image (visual image on the top row and thermal image on the bottom row) from the Mobile Phone class. This sample was specifically selected because the visual image represented a dark object on a light background, which should not have been a real challenge for any segmentation algorithm. However, in the thermal image the radiance of the mobile phone was barely greater than its background. It can be observed that in this specific example, the K-means algorithm performed better over the other three algorithms in both the visual and thermal spectrum. The Basic Threshold, Contours and Watershed with Distance Transform correctly identified the outline of the object.
but were susceptible to the reflection of the light on the cell phone. In the thermal spectrum, the *Basic Threshold* provided a mediocre representation of the object while the *Contours* and *Watershed with Distance Transform* could not segment any part of the object.

![Segmentation Results](image)

Figure 4-3: Visual (top row) and thermal (bottom row) sample segmentation results (dark object on light background).

Figure 4-4 illustrates the performance of each segmentation algorithm against a training dataset sample from the Portable Heater class. This sample was specifically chosen to evaluate the segmentation capabilities of the algorithms on a dark object against a light multi-textured background. For the purpose of this research, the segmentation algorithm had to extract the complete outline of the object from its background in the visual and thermal spectrum. The challenges in this sample were the blinds and the other small dark object in the middle left-hand side of the scene. In the visual spectrum, the *K-means* provided the better segmentation of the object as it removed the majority of the blinds in the background and provided nearly a complete filled outline of the object. The other three algorithms all provided a good outline of the object but could not remove the background blinds from the segmentation. In the thermal spectrum, the *Basic Threshold* and the *Watershed with Distance Transform* provided a good representation of the object but the contour of the back of the heater was very grainy and not well defined which could make it difficult to extract dimensions and measurements from this segmented image. The *K-means* provided well defined outlines but only segmented the highly radiating elements of the object. Similarly, the *Contours* algorithm provided well defined outlines but struggled to capture the complete object in the thermal image.
Figure 4-4: Visual (top row) and thermal (bottom row) sample segmentation results (dark object on light multi-textured background).

Figure 4-5 illustrates a sample training dataset from the Desk Lamp class segmented by the various algorithms implemented. This sample was specifically selected because it illustrated a light colored object in front of a light multi-textured background. This is in contrast to the previous two examples presented in Figure 4-3 and Figure 4-4. It was expected that this type of image would be a greater challenge for the segmentation algorithms. In the visual spectrum, none of the algorithms were able to correctly segment the desk lamp from the background. The best approximation was probably the Basic Threshold but this result could not be used to easily extract features because of the large number of clusters in the image. The Contours algorithm provided a similar result again with a large number of Contours which would make it difficult to automatically identify the desk lamp. In the thermal image, the Basic Threshold and the Contours provided the best segmentation results in comparison to the other two algorithms. The shape of the lamp was clearly outlined and all the background clutter was removed. However, none of the algorithms provided adequate results in both the visual and thermal spectrum.
The last segmentation examples are illustrated in Figure 4-6 and were probably the most challenging for the algorithms. This testing dataset sample was specifically selected because it illustrated objects from all five classes in a very cluttered and multi-textured background. The visual image segmentation results showed that none of the algorithms tested were capable of extracting just objects of interest. A human observer could probably find the objects in the segmented images, but this would likely be difficult for an automated process. Conversely, the thermal image provided a very clear location of each of the objects of interest and all the algorithms were capable of identifying at least 3 of the five objects. The Basic Threshold and the Watershed with Distance Transform likely provided the better results for an automated feature extraction application. The Contour algorithm provided a general location of the objects but extracted many additional unnecessary outlines. Once again, none of the tested algorithms provided adequate segmentation capabilities in both the visual and thermal datasets.
In order to automatically segment objects of interest in both the visual and thermal spectrum, an alternate algorithm was developed. The segmentation results presented in Figure 4-3 to Figure 4-6 demonstrate that in the visual spectrum, an algorithm based on pixel intensity values only works if the object has similar colors and is presented against a contrasting background. Threshold-based algorithms used on visual imagery do not consider spatial content of the image nor the state of adjacent pixels and as a result were deemed unsuitable. Conversely, in thermal imagery the radiation emitted from a source directly affects its surrounding and consequently, a relationship exists between adjacent pixels with a similar greyscale intensity levels. In thermal imagery, a threshold-based algorithm can effectively segment related pixels from an object simply based on the image’s greyscale intensity.

The Watershed algorithm seemed to provide the most potential segmentation capability in the visual spectrum as this region growing algorithm takes into account the relationship between adjacent pixels. The concept of using markers to help the Watershed algorithm triggered the idea that perhaps the thermal image, which can be easily segmented, could be used as initial markers to enhance the segmentation in the visual image. This new segmentation algorithm was named Watershed with Thermal Markers.

4.2.6. Watershed with Thermal Markers

The Watershed with Thermal Markers algorithm proposed in this research uses the thermal image to produce markers that can be used by the watershed algorithm to segment either the visual or thermal images. The flowchart of the algorithm is presented in Figure 4-7.

The thermal markers are generated from the thermal image by separating the pixels into three greyscale intensity groups based on two user-defined thresholds. The pixels with a greyscale intensity above the upper thresholds are considered to be part of the object of interest in the thermal image and make up the first marker (color “white”). The pixels below the lower threshold are considered to be part of the background and make up the second marker (color “grey”). The rest of the pixels between the lower and upper thresholds are considered unassigned (color “black”) and could belong to the object(s) of interest or to the background. An example of the markers and unassigned pixels generated from the sample image from Figure 4-6 are illustrated in Figure 4-7.
Once the thermal markers are generated, the Watershed algorithm can be applied to either the visual or thermal image to complete the segmentation process. The algorithm is not automated and requires user interaction to optimize the segmentation results by adjusting the upper and/or lower thresholds. The software interface including the user-defined thresholds and additional examples of the Watershed with Thermal Markers are described in the software Implementation found in Section 5.2.

The segmentation examples presented earlier in Section 4.2.5 were reassessed against the Watershed with Thermal Markers and are illustrated in Figure 4-8 to Figure 4-11. It can be observed from these sample results that the new algorithm can segment the objects of interest in both the visual and thermal
datasets. Note that in the samples from Figure 4-8 to Figure 4-11, the segmented objects are “colored” white while the background is colored a shade of grey to easily identify the object from its background. In the actual implementation, the background is colored black and the objects of interest maintain their greyscale values in order to compute intensity features as shown in segmented image pair of Figure 4-7 (bottom image).

Figure 4-8: *Watershed with Thermal Markers* segmentation algorithm applied to visual (top) and thermal (bottom) images (dark object on light background). Original image in the left column and segmented results in the right column.
Figure 4-9: *Watershed with Thermal Markers* segmentation algorithm applied to visual (top) and thermal (bottom) images (dark object on light background). Original image in the left column and segmented results in the right column.

Figure 4-10: *Watershed with Thermal Markers* segmentation algorithm applied to visual (top) and thermal (bottom) images (light object on light background). Original image in the left column and segmented results in the right column.
Figure 4-11: *Watershed with Thermal Markers* segmentation algorithm applied to visual (top) and thermal (bottom) images (light and dark objects on multi-textured background). Original image in the left column and segmented results in the right column.

The next major component of the principal workflow previously illustrated in Figure 4-1 is the Feature Extraction. Section 4.3 discusses the feature selection for this research.

### 4.3. Image Feature Selection

As briefly described in Section 2.2, there are numerous types of features that can be extracted from visual and thermal imagery and are typically categorized by shape, color and texture. For the purpose of this research, it was necessary to identify a set of features that could be extracted from both visual and thermal imagery. Although beyond the scope of this research, the selected features should be easily and efficiently computed such that they could eventually be implemented in hardware for a real-time application. The final consideration was the dimensionality of the feature vector as it would be used to train classifiers.

It was decided to implement the features presented by Cayouette, Labonté and Morin [28] primarily because the features had already been evaluated on thermal imagery and demonstrated great discriminability in the context of the application presented. Cayouette et al. used the features on thermal images only and this research is an extension of their work. In this work, the same features are evaluated
on visual imagery as well as thermal imagery to demonstrate their discriminability capabilities. Furthermore, the classification performed in this work, which is described in Section 4.4, uses a different approach than Cayouette et al. The main differences between Cayouette et al.’s work and this research are summarized in Table 4-1.

<table>
<thead>
<tr>
<th></th>
<th>Cayouette et al.</th>
<th>Viau et al.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classes of objects</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Dataset imagery</td>
<td>Thermal only</td>
<td>Visual and thermal images</td>
</tr>
<tr>
<td>Samples</td>
<td>1264</td>
<td>209</td>
</tr>
<tr>
<td>Classifier</td>
<td>Probabilistic Neural Network</td>
<td>Support Vector Machine (SVM)</td>
</tr>
<tr>
<td>Descriptor</td>
<td>13 features</td>
<td>26 features</td>
</tr>
<tr>
<td>Classification experiment(s)</td>
<td>Thermal only</td>
<td>Visual-only, Thermal-only, Combined Visual-Thermal Expert System</td>
</tr>
</tbody>
</table>

The features as proposed by Cayouette et al. were:

**Intensity Features**

- normalized maximum intensity:

\[ F1 = \frac{Z_{\text{max}}}{Z} \]

\[ (8) \]

- normalized average intensity:

\[ F2 = \frac{Z}{A} \]

\[ (9) \]

- normalized variance of the intensity distribution:

\[ F3 = \frac{\mu_2}{A^2} \]

\[ (10) \]

- normalized third moment of the intensity distribution:

\[ F4 = \frac{\mu_3}{A^3} \]

\[ (11) \]
Shape Features

- normalized square root of the minimum moment of inertia:
  \[ F_5 = \frac{\sqrt{I_{\text{min}}}}{A} \]  (12)

- normalized square root of the maximum moment of inertia:
  \[ F_6 = \frac{\sqrt{I_{\text{max}}}}{A} \]  (13)

- normalized maximum radial distance:
  \[ F_7 = \frac{D_{\text{max}}}{\sqrt{A}} \]  (14)

- normalized minimum radial distance:
  \[ F_8 = \frac{D_{\text{min}}}{\sqrt{A}} \]  (15)

- normalized average radial distance:
  \[ F_9 = \frac{\bar{D}}{\sqrt{A}} \]  (16)

- normalized second moment of the distance distribution:
  \[ F_{10} = \frac{\mu_2}{A} \]  (17)

- angle of the principal axis of minimum inertia
  \[ F_{11} = \theta \]  (18)
• aspect ratio AR

\[ F_{12} = \sqrt{\frac{I_{\text{max}}}{I_{\text{min}}}} \]  

(19)

• roundness R

\[ F_{13} = \frac{P^2}{4\pi A} \]  

(20)

The list of features selected represent intensity (F1 to F4) and shape (F5 to F13) characteristics of the objects of interest. They do not however represent any texture characteristics and this was intentional because of the low imagery quality produced by the Ti10 camera. In the visual dataset, many of the images were blurred and contained very few textured details. In the thermal imagery, the low dynamic range of the sensor combined with highly radiating objects (such as the bulb of the desk lamp and the elements of the portable heater) resulted in localized pixel saturation in many images. For these reasons, only intensity (both in the visual and thermal images) and shape features were retained.

The following describes how each of the features listed in Equations 8-20 were implemented in the Feature Extraction Application described in Section 5.3.

**Normalized maximum intensity (F1)**

To compute the normalized maximum intensity, a for-loop was used to parse through each pixel of each cluster identified in an image. A maximum intensity variable \(I_{\text{max}}\) and a total intensity variable were updated as each pixel was assessed. After all pixels from a cluster were evaluated, the average intensity \(\bar{Z}\) was calculated by dividing the total intensity by the pixel count. The normalized maximum intensity was determined by the maximum intensity of a cluster divided by the average intensity of the cluster.

**Normalized average intensity (F2)**

To compute the normalized average intensity of a cluster, a total intensity variable was updated as each pixel of the cluster was assessed. After each cluster was evaluated, the average intensity \(\bar{Z}\) was calculated by dividing the total intensity by the pixel count. The area, \(A\), of a cluster was extracted from
the first spatial moment (m00) of the Hu Moments [17] from the OpenCV library [46]. The normalized average intensity was the average intensity of a cluster divided by the average area of the cluster.

**Normalized variance of the intensity distribution (F3)**

The normalized variance of the intensity distribution or the second moment was calculated by computing the variance of the pixel intensity ($\mu^2_z$) for a cluster and dividing it by its area squared.

**Normalized third moment of the intensity distribution (F4)**

The normalized third moment of the intensity distribution was calculated by computing the third moment of the pixel intensity ($\mu^3_z$) for a cluster and dividing it by its area cubed.

**Normalized square root of the minimum moment of inertia (F5)**

The normalized square root of the minimum moment of inertia for each cluster was calculated by dividing the square root of the principal axis of minimum inertia ($\sqrt{I_{\text{min}}}$) by the cluster area (A).

**Normalized square root of the maximum moment of inertia (F6)**

The normalized square root of the maximum moment of inertia for each cluster was calculated by dividing the square root of the principal axis of maximum inertia ($\sqrt{I_{\text{max}}}$) by cluster area (A).

**Normalized maximum radial distance (F7)**

The normalized maximum radial distance was computed by first finding the centroid and perimeter coordinates of each pixel of a cluster. The radial distance between each perimeter pixel and the centroid were calculated. The maximum distance ($D_{\text{max}}$) was retained and divided by the cluster square root of the area (A).

**Normalized minimum radial distance (F8)**

The normalized minimum radial distance was computed by first finding the centroid and perimeter coordinates of each pixel of a cluster. The radial distance between each perimeter pixel and the centroid were calculated. The minimum distance ($D_{\text{min}}$) was retained and divided by the cluster square root of the area (A).
Normalized average radial distance (F9)

The normalized average radial distance was computed by first finding the centroid and perimeter coordinates of each pixel of a cluster. The radial distance between each perimeter pixel and the centroid were calculated. The average distance ($\overline{D}$) was retained and divided by the cluster square root of the area ($A$).

Normalized second moment of the distance distribution (F10)

The normalized second moment of the radial distance distribution was computed by first finding the centroid and perimeter coordinates of each pixel of a cluster. The radial distance between each perimeter pixel and the centroid were calculated and the variance distance ($\mu_{ij}^2$) was calculated. The variance was finally normalized using the area ($A$).

Angle of the principal axis of minimum inertia (F11)

The orientation of the cluster was computed from the central normalized moments ($\nu_{11}, \nu_{20}$ and $\nu_{02}$) of the Hu Moments [17] from the OpenCV library [46] for each cluster ($\mu[i]$) in the image. The angle theta was calculated using the following equation:

$$\theta = 0.5 \times \arctan\left(2 \times \frac{\mu[i].\nu_{11}}{\mu[i].\nu_{20}}\right).$$

Aspect ratio AR (F12)

The aspect ratio of the cluster was computed by dividing the square root of maximum moment of inertia, ($I_{\max}$) by the square root of the minimum moment of inertia ($I_{\min}$).

Roundness R (F13)

The roundness of a cluster was computed by dividing the perimeter length squared by $4\pi$ times the cluster area. The perimeter was extracted from the contour [44] function in OpenCV.

The number features and selection of feature combinations to maximize the performance of the classifiers is detailed in Section 6.2.
4.4. Classifier Selection

At this point, all but the last major component of the principal workflow illustrated Figure 4-1 have been described. The final component is the classifier selection and description. As presented in Section 2.3, there are several types of classifiers that could be used for this application and as the “no free lunch” theorem [47] suggests, no one classifier is above all in all cases. The choice of the Support Vector Machine (SVM) was based on the availability and well documented open-source library LIBSVM [48]. SVMs have been used extensively in the literature for various types of statistical pattern recognition applications and based on the author’s previous experience with the library, demonstrated a high potential for this specific application. SVMs are fast to train, can operate on linear and non-linear data in n-dimensional space, and are usually a top performer in statistical classification challenges. Specifically for these experiments, several multi-class SVMs were used with the polynomial kernel. The software implementation, the training and testing procedures of the classifiers are described in further details in Section 5.4.

The open literature [49] suggested that the classification results can be optimized by varying the cost and gamma parameters. As part of the Classifier application, a grid search was performed for each feature combination evaluated to optimize the classification results. The cost parameter was evaluated between $2e^{-4}$ and $2e^{5}$ while the gamma parameter was evaluated between $2e^{-5}$ and $2e^{4}$. The grid search initially consisted of 100 points and the SVM was trained and tested on each of the 100 grid points. From these 100 results, the highest classification rate was identified along with the associated cost and gamma parameters. An automated process refined the grid search until the best classification rate was achieved. Generally, the optimum cost and gamma parameters for a specific feature combination were identified within 300 to 600 iterations. Continuous optimization of the SVMs was performed until the best classification rate was achieved using the testing dataset. The stopping criterion for the optimization was the classification rate (i.e. change in TPR less than 0.1%). Once the optimum classification rate was identified, the specific feature combination along with the classification rate, the cost and gamma parameters were stored in an output text file for post classification analysis. The post-classification analysis consisted of determining the highest classification rates as well as the features that appeared the most and least often in the feature combinations. The classification analysis results are presented in Section 6.2.
Chapter 5  Software Implementation and Data Processing

5.1.  Overview

This section discusses the implementation of the application developed specifically for this research. Section 5.1 provides an overview of the software application, Section 5.2 presents the segmentation implementation while Section 5.3 presents the feature extraction details. Section 5.4 provides the classifier training and testing implementation details.

Table 5-1 summarizes the software and hardware specification on which the application was developed and tested.

<table>
<thead>
<tr>
<th>Feature Extraction Application</th>
<th>OpenCV 2.49, MS VS Express 2013</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classifiers</td>
<td>LIBSVM 3.20 (MATLAB interface)</td>
</tr>
<tr>
<td>Operating System</td>
<td>Windows 8.1, 64-bit</td>
</tr>
<tr>
<td>Processor</td>
<td>Intel® Core™ i7-4700MQ (2.4 GHz)</td>
</tr>
<tr>
<td>RAM</td>
<td>16 GB</td>
</tr>
</tbody>
</table>

The application was designed to extract features of interest from user-selected visual and thermal imagery, and from these features, train classifiers to automatically identify the objects from a separate dataset. The first module of the application (referred to as the “Feature Extraction Application”) was developed with OpenCV 2.49 [50] and allows the user to select a series of images, segment the objects of interest, extract the desired feature set from the objects of interest and store the features from each image in an external file for further processing and classification. The feature extraction process is illustrated in Figure 5-1.
The second module (referred to as the “Classifier”), developed in the MATLAB programming language imports the training dataset from the Feature Extraction Application and trains the SVM classifiers. The trained classifier loads the separate testing dataset and predicts the class association of all detected objects from the visual and thermal imagery. The classifier training and prediction process is illustrated in Figure 5-2.
5.2. Segmentation

The Feature Extraction Application uses a combination of command-line and graphical user interfaces and can be invoked in manual or automatic (batch run) mode. A typical function call to start the Feature Extraction Application from a command window is as follows:

```
> FeatExtApp path clrimg irimg bklev fglev band mode
```

The input parameters are as follows:

- `FeatExtApp` Name of the executable Feature Extraction Application
- `path` Directory name where the visual and thermal imagery are located
- `clrimg` Name of the visual image to use
- `irimg` Name of the thermal image to use
- `bklev` Initial background threshold level (value between 0 and 255)
The $fglev$ parameter is the initial foreground threshold level (value between 0 and 255).

The $band$ parameter is the Image ID to which the segmentation is to be applied (0: visual, 1: thermal).

The $mode$ parameter is the Mode to launch to application (0: manual, 1: automatic).

The $clrimg$ and $irmg$ parameters are the complementary visual and thermal image pair of the same scene. The application will not be able to segment the objects correctly if these two images are not a matching pair. In the Manual mode of operation ($mode = 0$), the $bklev$ and $fglev$ are the initial background and foreground intensity threshold levels to initiate the image segmentation process. These initial values can be any value set by the user. In this mode of operation, the user has the ability to adjust both thresholds using a graphical user interface to optimize the segmentation. The thresholds are used to create the background and foreground markers for the Watershed with Thermal Markers segmentation algorithm discussed in Section 4.2.6. Once the optimized threshold values have been determined using the manual mode, the thresholds can be reused with the Automatic mode ($mode = 1$) to bypass the user interaction and process all of the images rapidly in a batch run.

The application provides the flexibility to choose which of the image pairs to extract the features from. This functionality was implemented because in certain test scenarios it was necessary to repeat feature extraction on only the visual or only the thermal band. The $band$ parameter selects which of the two image pair (visual or thermal) to apply the segmentation algorithm on. An example of the application user interface is illustrated on the right side of Figure 5-3 and Figure 5-4.

Figure 5-3: Original visual image (left), markers (middle) and Watershed with Thermal Markers segmentation results (right).

Figure 5-3 illustrates a segmentation example where the background and foreground thresholds are not optimized. On the left side of the figure is the original image from the visual band while the middle image shows the markers generated from the thermal band image for the Watershed segmentation algorithm. In the middle figure, the white pixels identify the foreground object while the grey pixels identify the background objects. The black pixels have not yet been assigned as either foreground or background.
objects. When these specific markers are used in conjunction with the Watershed algorithm, the resulting segmentation from the visual image is illustrated in the far right of the figure.

The foreground objects are determined by identifying all the pixels with an intensity greater than or equal to the user-defined foreground threshold level ($fglev$). The background objects are determined by identifying all the pixels with an intensity less than or equal to the user-defined background threshold level ($bklev$). The segmentation by the Watershed with Thermal Markers can be improved by adjusting one or both thresholds and reducing the number of unassigned pixels (black) as illustrated in Figure 5-4.

![Figure 5-4: Improved segmentation using adjusted thresholds. Original visual image (left), thermal markers (middle) and Watershed with Thermal Markers segmentation results (right).](image)

5.3. Feature Extraction and Post Processing

After the image segmentation process is complete, the feature extraction process is initiated and autonomously creates a series of contours around each cluster. For each cluster in the visual and the thermal band images, the features described in Section 4.3 are extracted. Since the application works on image pairs (visual and thermal), a total of 26 features for each cluster are stored in a text file for post-processing. The feature extraction flowchart is displayed in Figure 5-5.
The first output of the Feature Extraction Application is a text file for each image processed containing a list of features for each of the clusters segmented from the image (one text file for the visual image and a separate text file for the thermal image). The second output is a segmented image identifying each of the clusters referenced in the output feature text file. An example of the output feature text file and of the segmented clusters are illustrated in Figure 5-6. The first column of the text file is the cluster identification number and the remaining columns are the feature values computed for each cluster. Each row in the output file represents the features of a separate cluster. The cluster identification number in the segmented image is positioned at the centroid of the segmented cluster.
A script file was used to automatically extract the features of all the image pairs (both visual and thermal) from the training and testing datasets. Once complete, a manual process was required to analyse each output data file such as the one presented in Figure 5-6 to locate the object of interest in the image. In the case of the training dataset, this process was essential to correctly establish the ground truths. This process removed all detected random shapes such as sensor noise and thermal reflections from the testing dataset. At the end of the process, only class objects remained in the dataset.

Using Figure 5-6 as an example, it can be observed that the object of interest is the Lamp and its features are identified by the cluster number 4. Row data number 4 was extracted from the output data file and stored as a separate Microsoft Excel worksheet. The other five clusters identified by the Feature Extraction Application were disregarded. These other clusters could have been used for a sixth object class named “other”. This was not done because the “other” clusters typically did not have a matching thermal image equivalent. In the case of thermal image, the “other” class may not have had a matching visual image equivalent. As per the objectives of this research, it was necessary that all objects had both a visual and thermal signature in order to be processed by the classifier. In the current software implementation, the identification of valid clusters was done manually. This process could possibly be
automated in the future by using the visual and thermal image pairs to verify the presence of the clusters in approximately the same location and with approximately the dimensions in both images.

As illustrated in Figure 5-7, many images were intentionally captured with several classes of objects in them. In these output data files, all of the objects of interest were extracted and added to the worksheet.

Figure 5-7: Output data file containing multiple classes (0: Mug, 1: Heater, 2: Lamp).

Once all of the output data files and segmented images were manually processed and the objects of interest were extracted and assigned a class identification number, the training and testing datasets were normalized such that all values were between 0 and 1. The Microsoft Excel worksheets were exported into separate text files, one for the training dataset and one for the testing dataset. The format of the training and testing data files consisted of 27 columns of data where each row represented the features
of a different object detected by the Feature Extraction Application. The first column represented the
class of the object, the next 13 columns were the visual features and the last 13 columns were the thermal
features.

5.4. Training and Testing the Classifiers

The Classifier application was developed in the MATLAB scripting language and was designed to
automatically evaluate thousands of feature combinations from the visual and thermal testing datasets.
The primary inputs to the Classifier application were the following:

- `runs`: number of feature combinations to evaluate
- `max_feats`: maximum number of features from the visual and thermal datasets
- `featlist`: feature combination list
- `trainfn`: training dataset filename
- `testfn`: testing dataset filename
- `outputfn`: output filename

At the core of the application was the LIBSVM library which allowed users to train and test various types
of SVM. For this research, a C-support vector classifier (C-SVC) multi-class SVM was used with a
polynomial kernel. The LIBSVM training function call in MATLAB was as follows:

```matlab
model = svmtrain(trainlabels, traindata, ['-c ' num2str(cost) ' -g ' num2str(gamma) ' -b 1' '-q']);
```

where
- `trainlabels` is a n x 1 vector of training labels
- `traindata` is a n x m array of m training instances of n features
- `cost` is the parameter C of C-SVC
- `gamma` is the gamma value of the polynomial kernel
- `model` is the output of the `svmtrain` function

The LIBSVM testing function call in MATLAB was as follows:

```matlab
[predicted_label, accuracy, prob_estimates] = svmpredict(testlabels, testdata, model, '-b 1');
```

where
- `testlabels` is a n x 1 vector of testing labels
- `testdata` is a n x m array of m testing instances of n features
- `model` is the output of the `svmtrain` function
- `predicted_label` is a vector of predicted labels
- `accuracy` is a 3-element vector with the prediction accuracy, mean-squared error, squared
correlation coefficient
- `prob_estimates` is the probability estimates
Chapter 6  Experimental Evaluation

6.1.  Experiment Design

For every training and testing imagery captured, 13 features were extracted from the visual image and 13 from the matching thermal image for a total of 26 features per object. Three separate experiments were conducted to determine if the proposed features were suitable for distinguishing between the five classes of objects selected.

The first experiment was to determine the best classification rates using only the visual features; the second experiment was to determine the best classification rates using only the thermal features. The third experiment was to find a feature vector or descriptor combination from both the visual and thermal image pairs which would hopefully produce better results than the visual or thermal alone.

For the first two experiments, the dimensionality of the feature vector could be up to 13 features. However, it was decided to limit the number of features to five in order to reduce the possible number of combinations. In the case of the third experiment, the dimensionality could be up to 26 features. It was decided to limit the feature vector to a maximum of five visual features and five thermal features for a maximum of ten features, once again to limit the possible number of feature combinations. Lottery mathematics [51] was used to determine the number of possible feature vector combinations for experiments 1 and 2.

Assuming that the feature vector had \( k = 5 \) features with \( n \) unique possible values between 1 and 13 then the number of possible combinations \( (c) \) is defined as follows:

\[
c(n, k) = \frac{n!}{k!(n-k)!}
\]  \hspace{1cm} (21)

However, since the feature vector can have up to \( k = 5 \) unique features with values between 1 and 13, Equation 21 has to be evaluated for all possible values of \( k \) as follows:

\[
c(n, k) = \sum_{k=1}^{5} \frac{n!}{k!(n-k)!}
\]  \hspace{1cm} (22)
Evaluating equation 22 for \( n = 13 \) from \( k = 1 \) to \( 5 \) results in 2,379 possible combinations. Therefore there are 2,379 possible feature combinations for the visual dataset and 2,379 possible feature combinations for the thermal dataset. As an example, possible visual feature combinations (cX) could be:

\[
\begin{align*}
    c_1 &= [F1, F3, F12] \\
    c_2 &= [F13] \\
    c_3 &= [F2, F3, F9, F11, F13] \\
    \ldots \\
    c_{2378} &= [F6, F8] \\
    c_{2379} &= [F4, F7, F10, F12, F13]
\end{align*}
\]

where \( F1 \) to \( F13 \) are the visual features defined in Section 4.3. The possible thermal feature combinations (tX) could be:

\[
\begin{align*}
    t_1 &= [F14, F20, F24, F26] \\
    t_2 &= [F16, F18] \\
    t_3 &= [F26] \\
    \ldots \\
    t_{2378} &= [F17, F23, F24, F25, F26] \\
    t_{2379} &= [F21, F25, F26]
\end{align*}
\]

where \( F14 \) to \( F26 \) are the thermal features similar to those defined in Section 4.3 (\( F1 \) to \( F13 \) respectively) but extracted from the thermal image.

For the third experiment, the goal was to identify a feature vector that included both visual features and thermal features. In this experiment, the feature vector had to include a minimum of two features (one from the visual image and one from the thermal image) but was limited to 10 features (5 from the visual and 5 from the thermal). If there are 2,379 possible combinations in the visual dataset and the same number in the thermal dataset, the total possible combinations was determined by multiplying the possible combinations in the visual dataset (i.e. 2,379) by the possible combinations in the thermal dataset (i.e. 2,379) for a total possible combinations of 5,659,641.

To continue the previous example, if the visual features had the following possible combinations \{c1, c2, c3, \ldots c2379\} and the thermal features had the following possible combinations \{t1, t2, t3, \ldots, t2379\}, the combined visual and thermal feature combinations would be:

\[
\begin{align*}
    c_{1t1}, c_{1t2}, c_{1t3}, \ldots, c_{1t2379}, \\
    c_{2t1}, c_{2t2}, c_{2t3}, \ldots, c_{2t2379}, \\
    c_{3t1}, c_{3t2}, c_{3t3}, \ldots, c_{3t2379}, \\
    \ldots, \\
    c_{2379t1}, c_{2379t2}, c_{2379t3}, \ldots, c_{2379t2379}
\end{align*}
\]
For the first two experiments, all 2,379 visual band and 2,379 thermal band possible feature combinations were evaluated. In the case of the third experiment, all of the possible 5,659,641 feature combinations were evaluated. The feature combinations were evaluated by systematically combining the 2,379 visual features with the 2,379 thermal features, training, optimizing and testing the SVM classifier. The 5,659,641 combinations were divided into 40 subsets which were evaluated in parallel over a period of approximately 12 days. The classification results and analysis are provided in the Section 6.2.

The final experiment consisted of building an expert system using the visual-only, thermal-only and combined visual-thermal classifiers. It is expected that each classifier will be better at identifying certain classes of objects. By using the expertise of each classifier, it is expected that the overall classification results can be improved. The expert system results are discussed in Section 6.2.5.

6.2. Classification Results

Section 6.2 is divided into six subsections which analyse the results of the experiments. Section 6.2.1 looks at the classification results from the testing dataset using single features (no combination) to establish a baseline. Section 6.2.2 discusses the results of the visual classification results in isolation while Section 6.2.3 looks at thermal classification results in isolation. Section 6.2.4 presents the results of the combined visual and thermal classification results while Section 6.2.5 analyses the results of the expert systems. Finally Section 6.2.6 proposes future work to possibly further improve the classification rates.

6.2.1. Individual Features

This section establishes a classification baseline by analysing the F1 score of individual features from both visual and thermal bands. As a starting point in the selection of the optimum feature sets, the class-conditional probability density functions used in Bayesian probability theory [33] were generated to see if any feature or features stood out as a prominent class discriminant. The class-conditional probability density function graphs illustrated in Figure 6-2 to Figure 6-8 demonstrate the probability that a feature with a value of x belongs to the class \( \omega_i \). In each graph, the five classes are identified by separate colors and line style; the solid lines represent the features from the visual dataset while the dashed lines represent the features from the thermal dataset. The 13 graphs were generated by determining the mean
(μ) and standard deviation (σ) of each feature from the training dataset and plotting their probability density function \( p(x \mid \omega_i) \).

\[
p(x \mid \omega_i) = \frac{1}{\sigma \sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}
\]  

(23)

The variable \( x \) in Equation 22 was a vector with 200 values between the range of -1 and 1 (recall that all the features were normalized between 0 and 1 as described in Section 5.3). The range -1 to 1 was selected in order visualize the complete distribution function. A range of 0 to 1 could have been used but would have only displayed the distributions greater than 0. For example, the mean of Feature 4 (thermal features) from Figure 6-3 is almost 1 and the distribution values greater than 1 are not displayed. In this case, the range for variable \( x \) could have extended beyond 1.

The class-conditional probability density function graphs can be used to identify a feature that can provide discrimination between classes. For example, Figure 6-1 illustrates the class-conditional density function for Feature 11. It can be observed from this graph that Feature 11 visual band data alone could provide a discriminant between the mobile phone class (solid blue line) and the other classes because its distribution is clearly separated from the others. However, Feature 11 thermal band data does not provide any obvious discrimination capability between classes as all their distributions clearly overlap.

When observing the ensemble of the graphs (Figure 6-2 to Figure 6-8), some of the visual band features such as Feature 1 (Figure 6-2), Feature 7 (Figure 6-5) and Feature 11 (Figure 6-7) provide some discrimination between the Lamp, the Mobile Phone and the remaining classes. Feature 7 (Figure 6-5)
and Feature 13 (Figure 6-8) from the thermal band also provide potential discrimination capabilities for the Mobile Phone class. However, all the other features have a very similar distribution (no obvious discriminability) and it is difficult to identify any one feature that provides an overall discriminability between all classes in the training dataset.

Figure 6-2: Class-conditional probability density function graphs for Feature 1 and 2

Figure 6-3: Class-conditional probability density function graphs for Feature 3 and 4
Figure 6-4: Class-conditional probability density function graphs for Feature 5 and 6

Figure 6-5: Class-conditional probability density function graphs for Feature 7 and 8

Figure 6-6: Class-conditional probability density function graphs for Feature 9 and 10
To further demonstrate that no single feature provided a high discriminability rate, the individual features extracted from the training dataset were used to train the SVM classifiers. The classifiers’ prediction capabilities were evaluated on the testing dataset and the TPR (Equation 2) for the individual features are tabulated in Table 6-1. These results show, for example, that when only Feature 1 from the training dataset was used to train the visual classifier, it achieved a 37.6% correct TPR over the five classes of objects. Similarly, when only Feature 1 from the dataset was used to train the thermal classifier, it achieved a 32.1% TPR. A complete description of the features from Table 6-1 can be found in Section 4.3.
Table 6-1: Classification TPR for single features

| Visual | | | | | |
|--|---|---|---|---|
| Feat ID | TPR (%) | Feat ID | TPR (%) | Feat ID | TPR (%) |
| 1 | 37.6 | 14 | 32.1 | 15 | 28.5 |
| 2 | 33.3 | 16 | 27.3 | 17 | 27.3 |
| 3 | 33.9 | 18 | 27.3 | 20 | 49.7 |
| 4 | 27.9 | 21 | 38.2 | 22 | 37.0 |
| 5 | 38.8 | 23 | 42.4 | 24 | 44.2 |
| 6 | 30.9 | 25 | 23.6 | 26 | 46.1 |
| 7 | 55.8 | | | | |
| 8 | 42.4 | | | | |
| 9 | 46.1 | | | | |
| 10 | 47.3 | | | | |
| 11 | 37.6 | | | | |
| 12 | 26.7 | | | | |
| 13 | 35.8 | | | | |

One of the objectives of this research was to identify the optimum feature or features to maximize the classification results. In this section, individual features were evaluated and the highest TPR using a single feature in the visual band was 55.8% achieved using Feature 7 (Normalized Maximum Radial Distances). In the thermal dataset, the highest TPR was 49.7% also using Feature 7. It is clear from these results that using only one feature is not sufficient and that several visual features or several thermal features or a combination of visual and thermal features are required to achieve classification improvement. The following section presents the classification results achieved when combining several visual features to train the SVM classifier.

6.2.2. Visual Band

This section presents the classification results achieved using a combination of up to five visual band features. As discussed in Section 6.1, it was decided to limit the number of feature combinations in the visual band and thermal band to five features. Using between two and five features, a total of 2,379 possible feature combinations exists and all of them were evaluated for this research.

Table 6-2 illustrates the 30 best feature combinations with the highest TPR using visual band features only. Recall that the classification experiments were conducted on a testing dataset which contained images from all five classes and was not used to train the SVM classifiers. The first column of Table 6-2 represents the overall ranking; the second column represents the TPR while the next five columns represent the feature combination used to achieve the TPR result. As illustrated in Table 6-2, the highest
TPR achieved using up to five features in the visual band was 65.5%. This is a 9.7% improvement in comparison to the best of the individual features results (55.8%) reported in the previous section. As per the research objective #4, the highest TPR using a combination of visual band features was achieved with one intensity feature (F2) and four shape features (F5-F7-F10-F12). Finally, it can be noted that no combination with less than 4 features appeared in the top classification results. A complete description of the features from Table 6-2 can be found in Section 4.3.

Table 6-2: Visual band feature classification results

<table>
<thead>
<tr>
<th>Ranking</th>
<th>TPR (%)</th>
<th>Feat ID 1</th>
<th>Feat ID 2</th>
<th>Feat ID 3</th>
<th>Feat ID 4</th>
<th>Feat ID 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>65.5</td>
<td>2</td>
<td>5</td>
<td>7</td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td>2</td>
<td>64.2</td>
<td>2</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>10</td>
</tr>
<tr>
<td>3</td>
<td>63.6</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>4</td>
<td>63.0</td>
<td>7</td>
<td>8</td>
<td>10</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>62.4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>62.4</td>
<td>5</td>
<td>7</td>
<td>10</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>62.4</td>
<td>2</td>
<td>7</td>
<td>10</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>61.8</td>
<td>2</td>
<td>6</td>
<td>7</td>
<td>9</td>
<td>12</td>
</tr>
<tr>
<td>9</td>
<td>61.8</td>
<td>2</td>
<td>6</td>
<td>7</td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td>10</td>
<td>61.8</td>
<td>5</td>
<td>7</td>
<td>10</td>
<td>12</td>
<td>13</td>
</tr>
<tr>
<td>11</td>
<td>61.2</td>
<td>2</td>
<td>3</td>
<td>6</td>
<td>7</td>
<td>10</td>
</tr>
<tr>
<td>12</td>
<td>61.2</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td>13</td>
<td>61.2</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>10</td>
<td>13</td>
</tr>
<tr>
<td>14</td>
<td>61.2</td>
<td>2</td>
<td>6</td>
<td>7</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td>15</td>
<td>61.2</td>
<td>2</td>
<td>6</td>
<td>7</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>61.2</td>
<td>3</td>
<td>5</td>
<td>7</td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td>17</td>
<td>61.2</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td>18</td>
<td>61.2</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>61.2</td>
<td>5</td>
<td>7</td>
<td>9</td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td>20</td>
<td>61.2</td>
<td>2</td>
<td>7</td>
<td>9</td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td>21</td>
<td>61.2</td>
<td>2</td>
<td>7</td>
<td>9</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>61.2</td>
<td>6</td>
<td>7</td>
<td>10</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>60.6</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>9</td>
<td>12</td>
</tr>
<tr>
<td>24</td>
<td>60.6</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td>25</td>
<td>60.6</td>
<td>3</td>
<td>6</td>
<td>7</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>26</td>
<td>60.6</td>
<td>3</td>
<td>6</td>
<td>7</td>
<td>9</td>
<td>12</td>
</tr>
<tr>
<td>27</td>
<td>60.6</td>
<td>3</td>
<td>6</td>
<td>7</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td>28</td>
<td>60.6</td>
<td>2</td>
<td>3</td>
<td>7</td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td>29</td>
<td>60.6</td>
<td>1</td>
<td>3</td>
<td>7</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>60.6</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>12</td>
</tr>
</tbody>
</table>
It can be observed from Table 6-3 that the feature that appeared the most often in the top 30 is Feature 7 (Normalized Maximum Radial Distance) which was part of all 30 top combinations. The next best two features were Feature 10 (Normalized Second Moment of the Distance Distribution) with 23 appearances in the top 30 followed by Feature 6 (Normalized Square Root of the Maximum Moment of Inertia) with 19.

It is not surprising that the intensity features were not prominent in the visual band classifier as 4 of the 5 classes from the dataset had black colored surfaces. Similar surface colors would make it difficult for any type of classifier to distinguish between classes using only greyscale intensities or its derivatives.

<table>
<thead>
<tr>
<th>Feature ID</th>
<th># of Appearances in the Top 30 Combinations</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>30</td>
</tr>
<tr>
<td>10</td>
<td>23</td>
</tr>
<tr>
<td>6</td>
<td>19</td>
</tr>
<tr>
<td>12</td>
<td>18</td>
</tr>
<tr>
<td>5</td>
<td>12</td>
</tr>
<tr>
<td>2</td>
<td>11</td>
</tr>
<tr>
<td>9</td>
<td>11</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>13</td>
<td>3</td>
</tr>
<tr>
<td>11</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
</tr>
</tbody>
</table>

In comparing the TPR found in Table 6-1 and the results of Table 6-2, it is observed that in the individual feature classification results, Feature 7 (Normalized Maximum Radial Distance) achieved the highest TPR of 55.8% on its own and appeared in all of the top 30 feature combinations. This was somewhat anticipated as discussed in Section 6.2.1 where Feature 7 demonstrated some discriminability in its class conditional probability density function (Figure 6-5). In the individual feature TPR from Table 6-1, Feature 10 achieved the second highest rate and was also the second most prominent feature in the top 30 feature combinations. Conversely, Feature 12 (Aspect Ratio) had the lowest individual TPR (26.7%) but appeared in 10 of the 30 top feature combinations.

Figure 6-9 illustrates the visual band classifier TPR distribution of the 2,379 different feature combinations evaluated. The mean TPR of the visual band classifier is 47.2% with a variance of 48.1%.
Table 6-4 summarizes the performance metrics of the classifier by class, for the highest ranking visual band feature combination ([F2 F5 F7 F10 F12]) from Table 6-2. The description and equations for each performance metrics were defined in Section 2.4. Each row of the table illustrates the performance of the classifier for a specific class. Of particular interest, it is noted that the visual band classifier was better at identifying Class 3 (Laptop Charger) and Class 4 (Desk Lamp) samples (F1 = 80.5% and 73.3% respectively) but was not very good at identifying Class 1 (Mobile Phone) samples (F1 = 45.7%). Overall, the visual band classifier achieved an overall F1 of 65.5% with accuracy of 82.6% over the five classes.

<table>
<thead>
<tr>
<th>Class</th>
<th>Samples</th>
<th>TP</th>
<th>TN</th>
<th>FP</th>
<th>FN</th>
<th>TPR</th>
<th>TNR</th>
<th>PPV</th>
<th>NPV</th>
<th>Acc</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>23</td>
<td>8</td>
<td>100</td>
<td>4</td>
<td>15</td>
<td>34.8%</td>
<td>96.2%</td>
<td>66.7%</td>
<td>87.0%</td>
<td>85.0%</td>
<td>45.7%</td>
</tr>
<tr>
<td>2</td>
<td>44</td>
<td>29</td>
<td>79</td>
<td>18</td>
<td>15</td>
<td>65.9%</td>
<td>81.4%</td>
<td>61.7%</td>
<td>84.0%</td>
<td>76.6%</td>
<td>63.7%</td>
</tr>
<tr>
<td>3</td>
<td>28</td>
<td>22</td>
<td>86</td>
<td>10</td>
<td>6</td>
<td>78.6%</td>
<td>89.6%</td>
<td>68.8%</td>
<td>93.5%</td>
<td>87.1%</td>
<td>73.3%</td>
</tr>
<tr>
<td>4</td>
<td>40</td>
<td>31</td>
<td>77</td>
<td>6</td>
<td>9</td>
<td>77.5%</td>
<td>92.8%</td>
<td>83.8%</td>
<td>89.5%</td>
<td>87.8%</td>
<td>80.5%</td>
</tr>
<tr>
<td>5</td>
<td>30</td>
<td>18</td>
<td>90</td>
<td>19</td>
<td>12</td>
<td>60.0%</td>
<td>82.6%</td>
<td>48.6%</td>
<td>88.2%</td>
<td>77.7%</td>
<td>53.7%</td>
</tr>
<tr>
<td>Overall</td>
<td>165</td>
<td>108</td>
<td>432</td>
<td>57</td>
<td>57</td>
<td>65.5%</td>
<td>88.3%</td>
<td>65.5%</td>
<td>88.3%</td>
<td>82.6%</td>
<td>65.5%</td>
</tr>
</tbody>
</table>

6.2.3. Thermal Band

This section discusses the classification results and classifier performance when combining up to five thermal band features. As discussed in Section 6.1, it was decided to limit the number of feature
combinations in the visual band and thermal band to five features. A total of 2,379 possible feature combinations can be achieved using two to five features. Similar to the previous two experiments described in Sections 6.2.1 and 6.2.2, this experiment used the thermal band features from the training dataset to train the SVM classifier and assessed its prediction capabilities on the testing dataset.

Table 6-5 illustrates the 30 best feature combinations with the highest TPR using thermal band features only. Recall that the classification experiments were conducted on the testing dataset which contained images from all five classes and was not used to train the SVM classifiers. The first column represents the overall ranking, the second column represents the TPR while the next five columns represent the feature combination used to achieve the classification result. As illustrated in Table 6-5, the highest TPR achieved using up to five features in the thermal band was 69.1%. This is a 19.4% improvement in comparison to the best individual thermal band feature results (49.7%) reported in the Section 6.2.1. The TPR achieved using a combination of thermal band features only is better than the results achieved using a combination of visual band features only (65.5%) as reported Section 6.2.2. The highest TPR using a combination of thermal band features was achieved with one intensity feature (F14) and three shape features (F19-F20-F26). It is worth noting that Feature 20 and Feature 7 are the same (*Normalized Maximum Radial Distance*) and are prominent features in both the visual and thermal bands. Although the features are computed using the same mathematical formula in both bands, the extracted values are not necessarily the same since they may have different shapes in each bands. Finally, it can be noted that no combination with less than 3 features appeared in the top classification results. A complete description of the features from Table 6-5 can be found in Section 4.3.

Table 6-5: Thermal band feature classification results

<table>
<thead>
<tr>
<th>Ranking</th>
<th>TPR (%)</th>
<th>Feat ID 1</th>
<th>Feat ID 2</th>
<th>Feat ID 3</th>
<th>Feat ID 4</th>
<th>Feat ID 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>69.1</td>
<td>14</td>
<td>19</td>
<td>20</td>
<td>26</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>68.5</td>
<td>14</td>
<td>16</td>
<td>20</td>
<td>26</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>68.5</td>
<td>14</td>
<td>19</td>
<td>20</td>
<td>22</td>
<td>26</td>
</tr>
<tr>
<td>4</td>
<td>68.5</td>
<td>14</td>
<td>15</td>
<td>20</td>
<td>22</td>
<td>26</td>
</tr>
<tr>
<td>5</td>
<td>68.5</td>
<td>17</td>
<td>20</td>
<td>23</td>
<td>25</td>
<td>26</td>
</tr>
<tr>
<td>6</td>
<td>68.5</td>
<td>14</td>
<td>17</td>
<td>20</td>
<td>25</td>
<td>26</td>
</tr>
<tr>
<td>7</td>
<td>68.5</td>
<td>14</td>
<td>16</td>
<td>20</td>
<td>25</td>
<td>26</td>
</tr>
<tr>
<td>8</td>
<td>68.5</td>
<td>14</td>
<td>20</td>
<td>26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>67.9</td>
<td>14</td>
<td>20</td>
<td>25</td>
<td>26</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>67.9</td>
<td>14</td>
<td>20</td>
<td>22</td>
<td>25</td>
<td>26</td>
</tr>
<tr>
<td>11</td>
<td>67.3</td>
<td>14</td>
<td>15</td>
<td>20</td>
<td>26</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>67.3</td>
<td>14</td>
<td>16</td>
<td>19</td>
<td>20</td>
<td>26</td>
</tr>
</tbody>
</table>
It can be observed from Table 6-6 that the features that appear the most often in the top 30 are Feature 20 (*Normalized Maximum Radial Distance*) and 26 (*Roundness*) which both appear in all 30 of the top combinations. The next best feature is Feature 14 (*Normalized Maximum Intensity*), which appeared in 19 of the top 30, followed by Feature 22 (*Normalized Average Radial Distance*) which appeared 12 times.

In the individual results presented in Table 6-1, Feature 20 and 26 had the highest individual thermal TPR and as expected, both features were prominent in the thermal feature combinations. Features 16, 17 and 18 had some of the lowest individual TPR and as expected appeared 10, 7 and 0 times respectively in the top 30.

Table 6-6: Most prominent features in the Top 30 thermal-only feature combinations classification results

<table>
<thead>
<tr>
<th>Feature ID</th>
<th># of Appearances in the Top 30 Combinations</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>30</td>
</tr>
<tr>
<td>26</td>
<td>30</td>
</tr>
<tr>
<td>14</td>
<td>19</td>
</tr>
<tr>
<td>22</td>
<td>12</td>
</tr>
<tr>
<td>15</td>
<td>10</td>
</tr>
<tr>
<td>16</td>
<td>10</td>
</tr>
<tr>
<td>17</td>
<td>7</td>
</tr>
<tr>
<td>23</td>
<td>7</td>
</tr>
<tr>
<td>25</td>
<td>7</td>
</tr>
</tbody>
</table>
When comparing the performance of feature combinations from the visual (Table 6-2) and thermal (Table 6-5) band separately, it can be observed that the shape features specifically the *Normalized Maximum Radial Distance* (Feature 7 in visual and Feature 20 in thermal) provided the best discrimination in both bands. As anticipated, the intensity features specifically *Normalized Maximum Intensity* (Feature 14) and *Normalized Average Intensity* (Feature 15) are only useful discriminants in the thermal band and not in the visual band because each class has a slightly different thermal signature (i.e. apparent temperature mapped to a greyscale palette) which provides discriminability between the classes. In the visual band, 4 of the 5 classes had a black surface which makes it difficult to distinguish between classes solely based on greyscale intensities but thermal imagery can help in such circumstances, as demonstrated here.

Figure 6-10 illustrates the thermal band classifier TPR distribution of the 2,379 different feature combinations evaluated. The mean TPR of the thermal feature combinations is 47.6% with a variance of 76.5%. It can be observed that the mean TPR of the thermal band classifier nearly the same as the visual band classifiers (47.6% vs. 47.2%) but has a much wider distribution (76.5% vs. 48.1%).

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>19</td>
<td>6</td>
</tr>
<tr>
<td>21</td>
<td>1</td>
</tr>
<tr>
<td>18</td>
<td>0</td>
</tr>
<tr>
<td>24</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 6-7 summarizes the performance metrics of the classifier by class, for the highest ranking thermal features combination ([F14 F19 F20 F26]) from Table 6-5. The description and equations for each
performance metrics were defined in Section 2.4. Each row of the table illustrates the performance of the classifier for a specific class. Of particular interest, it is noted that the thermal classifier was good at identifying Class 4 (Desk Lamp) samples ($F_1 = 88.0\%$) but was very poor at identifying Class 1 (Mobile Phone) samples ($F_1 = 12.9\%$). The thermal band classifier also rarely misidentified samples as a Class 4 (Desk Lamp) (specificity = 97.6\%). Overall, the thermal band classifier achieved a $F_1$ score of 69.1\%, with an accuracy of 84.8\% over the five classes.

Table 6-7: Thermal band classifier performance metrics

<table>
<thead>
<tr>
<th>Class</th>
<th>Samples</th>
<th>TP</th>
<th>TN</th>
<th>FP</th>
<th>FN</th>
<th>TPR</th>
<th>TNR</th>
<th>PPV</th>
<th>NPV</th>
<th>Acc</th>
<th>$F_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>23</td>
<td>2</td>
<td>112</td>
<td>6</td>
<td>21</td>
<td>8.7%</td>
<td>94.9%</td>
<td>25.0%</td>
<td>84.2%</td>
<td>80.9%</td>
<td>12.9%</td>
</tr>
<tr>
<td>2</td>
<td>44</td>
<td>36</td>
<td>78</td>
<td>18</td>
<td>8</td>
<td>81.8%</td>
<td>81.3%</td>
<td>66.7%</td>
<td>90.7%</td>
<td>81.4%</td>
<td>73.5%</td>
</tr>
<tr>
<td>3</td>
<td>28</td>
<td>27</td>
<td>87</td>
<td>18</td>
<td>1</td>
<td>96.4%</td>
<td>82.9%</td>
<td>60.0%</td>
<td>98.9%</td>
<td>85.7%</td>
<td>74.0%</td>
</tr>
<tr>
<td>4</td>
<td>40</td>
<td>33</td>
<td>81</td>
<td>2</td>
<td>7</td>
<td>82.5%</td>
<td>97.6%</td>
<td>94.3%</td>
<td>92.0%</td>
<td>92.7%</td>
<td>88.0%</td>
</tr>
<tr>
<td>5</td>
<td>30</td>
<td>16</td>
<td>98</td>
<td>7</td>
<td>14</td>
<td>53.3%</td>
<td>93.3%</td>
<td>69.6%</td>
<td>87.5%</td>
<td>84.4%</td>
<td>60.4%</td>
</tr>
<tr>
<td>Overall</td>
<td>165</td>
<td>114</td>
<td>456</td>
<td>51</td>
<td>51</td>
<td>69.1%</td>
<td>89.9%</td>
<td>69.1%</td>
<td>89.9%</td>
<td>84.8%</td>
<td>69.1%</td>
</tr>
</tbody>
</table>

6.2.4. Combined Visual and Thermal Bands

In Section 6.2.1, the results of a classification experiment using individual visual and thermal band features were presented and the highest TPR achieved was 55.8\% with a single visual band feature and 49.7\% using a single thermal band feature. In Sections 6.2.2 and 6.2.3, multiple features (up to five) from the same band were combined to improve the recognition of the five classes of objects used for this research. The results demonstrated an improvement of 9.7\% (65.5\% vs. 55.8\%) when using a combination of visual features and an improvement of 19.4\% (69.1\% vs. 49.7\%) when using a combination of thermal features.

In this section, the results of the third experiment conducted to determine if one or more combination(s) of visual and thermal band features could further improve the TPR achieved in the previous sections are reported and discussed. In this experiment, the combination vectors contained between two (minimum of one from the visual band and one from the thermal band) and ten features (maximum of five from the visual and five from the thermal bands). As discussed in Section 6.1, a total of 5,569,641 feature combinations (composed of visual and thermal features) were evaluated to find the optimum combination to maximize the classification rates.

Table 6-8 illustrates the 50 best evaluated combinations (using visual and thermal feature descriptors) with the highest TPR. The first column represents the overall ranking, the second column represents the
TPR while the next ten columns represent the feature combination used to achieve the classification result. As illustrated Table 6-8, the highest TPR achieved using a combination of visual and thermal features was 71.5%. This is a 21.8% improvement (71.5% vs. 49.7%) over the individual thermal band TPR, a 15.7% improvement (71.5% vs. 55.8%) over the individual visual band TPR, a 6% improvement (71.5% vs. 65.5%) over the visual band combination and a 2.4% improvement (71.5% vs. 69.1%) over the thermal band combination results.

Table 6-8: Combined visual and thermal band feature classification results

<table>
<thead>
<tr>
<th>Ranking</th>
<th>TPR (%)</th>
<th>Feat ID 1</th>
<th>Feat ID 2</th>
<th>Feat ID 3</th>
<th>Feat ID 4</th>
<th>Feat ID 5</th>
<th>Feat ID 6</th>
<th>Feat ID 7</th>
<th>Feat ID 8</th>
<th>Feat ID 9</th>
<th>Feat ID 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>71.5</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td>14</td>
<td>19</td>
<td>20</td>
<td>22</td>
<td>26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>70.9</td>
<td>3</td>
<td>7</td>
<td>11</td>
<td>15</td>
<td>20</td>
<td>22</td>
<td>23</td>
<td>26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>70.9</td>
<td>3</td>
<td>7</td>
<td>14</td>
<td>15</td>
<td>20</td>
<td>22</td>
<td>26</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>70.3</td>
<td>6</td>
<td>7</td>
<td>10</td>
<td>11</td>
<td>13</td>
<td>17</td>
<td>20</td>
<td>21</td>
<td>25</td>
<td>26</td>
</tr>
<tr>
<td>5</td>
<td>70.3</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>6</td>
<td>14</td>
<td>20</td>
<td>22</td>
<td>25</td>
<td>26</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>70.3</td>
<td>3</td>
<td>7</td>
<td>11</td>
<td>15</td>
<td>17</td>
<td>20</td>
<td>25</td>
<td>26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>70.3</td>
<td>7</td>
<td>9</td>
<td>11</td>
<td>15</td>
<td>20</td>
<td>22</td>
<td>23</td>
<td>26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>70.3</td>
<td>2</td>
<td>3</td>
<td>7</td>
<td>14</td>
<td>15</td>
<td>20</td>
<td>22</td>
<td>26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>70.3</td>
<td>2</td>
<td>4</td>
<td>14</td>
<td>20</td>
<td>22</td>
<td>25</td>
<td>26</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>70.3</td>
<td>6</td>
<td>7</td>
<td>18</td>
<td>20</td>
<td>22</td>
<td>23</td>
<td>26</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>70.3</td>
<td>2</td>
<td>3</td>
<td>14</td>
<td>16</td>
<td>20</td>
<td>22</td>
<td>26</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>70.3</td>
<td>4</td>
<td>14</td>
<td>20</td>
<td>22</td>
<td>25</td>
<td>26</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>70.3</td>
<td>3</td>
<td>14</td>
<td>15</td>
<td>20</td>
<td>23</td>
<td>26</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>69.7</td>
<td>2</td>
<td>7</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>19</td>
<td>20</td>
<td>24</td>
<td>25</td>
<td>26</td>
</tr>
<tr>
<td>15</td>
<td>69.7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>17</td>
<td>20</td>
<td>24</td>
<td>25</td>
<td>26</td>
</tr>
<tr>
<td>16</td>
<td>69.7</td>
<td>7</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>13</td>
<td>17</td>
<td>20</td>
<td>21</td>
<td>25</td>
<td>26</td>
</tr>
<tr>
<td>17</td>
<td>69.7</td>
<td>2</td>
<td>3</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>17</td>
<td>20</td>
<td>21</td>
<td>25</td>
<td>26</td>
</tr>
<tr>
<td>18</td>
<td>69.7</td>
<td>2</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>16</td>
<td>17</td>
<td>19</td>
<td>20</td>
<td>26</td>
</tr>
<tr>
<td>19</td>
<td>69.7</td>
<td>3</td>
<td>8</td>
<td>9</td>
<td>11</td>
<td>16</td>
<td>17</td>
<td>20</td>
<td>25</td>
<td>26</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>69.7</td>
<td>2</td>
<td>7</td>
<td>9</td>
<td>11</td>
<td>16</td>
<td>20</td>
<td>22</td>
<td>23</td>
<td>26</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>69.7</td>
<td>2</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>17</td>
<td>19</td>
<td>20</td>
<td>26</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>69.7</td>
<td>3</td>
<td>7</td>
<td>9</td>
<td>15</td>
<td>20</td>
<td>24</td>
<td>25</td>
<td>26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>69.7</td>
<td>2</td>
<td>7</td>
<td>9</td>
<td>11</td>
<td>20</td>
<td>24</td>
<td>25</td>
<td>26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>69.7</td>
<td>2</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>16</td>
<td>20</td>
<td>24</td>
<td>26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>69.7</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>14</td>
<td>16</td>
<td>20</td>
<td>22</td>
<td>26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>69.7</td>
<td>3</td>
<td>7</td>
<td>11</td>
<td>15</td>
<td>17</td>
<td>19</td>
<td>20</td>
<td>26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>69.7</td>
<td>2</td>
<td>7</td>
<td>9</td>
<td>11</td>
<td>20</td>
<td>22</td>
<td>26</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>69.7</td>
<td>3</td>
<td>6</td>
<td>7</td>
<td>18</td>
<td>20</td>
<td>22</td>
<td>23</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>69.7</td>
<td>3</td>
<td>14</td>
<td>20</td>
<td>22</td>
<td>25</td>
<td>26</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>69.7</td>
<td>3</td>
<td>14</td>
<td>17</td>
<td>20</td>
<td>25</td>
<td>26</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>69.7</td>
<td>3</td>
<td>14</td>
<td>15</td>
<td>20</td>
<td>22</td>
<td>26</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
As illustrated in Table 6-9, the feature that appeared the most often in the top 50 feature combinations evaluated was *Roundness* (Feature 26) which appeared in 49 of the top 50 combinations. *Normalized Maximum Radial Distance* (Feature 20 in the thermal and Feature 7 in the visual band) appeared 47 and 33 times respectively in the top combinations. This is consistent with the results found in the previous two experiments where Feature 7 and 20 were the most common features in the visual-only and thermal-only combinations. It is also noted from Table 6-9 that the thermal band features (Features 14-26) were more often used in the top 50 combinations (240 vs. 153) as compared to the visual band features (Features 1-13). This was expected as the thermal band classifier achieved a slightly higher TPR (69.1% from Table 6-5) than the visual band classifier (65.5% from Table 6-2).

Table 6-9: Most prominent features in the Top 50 visual-thermal feature combinations classification results

<table>
<thead>
<tr>
<th>Feature ID</th>
<th># of Appearances in the Top 50 Combinations</th>
</tr>
</thead>
<tbody>
<tr>
<td>26</td>
<td>49</td>
</tr>
<tr>
<td>20</td>
<td>47</td>
</tr>
<tr>
<td>7</td>
<td>33</td>
</tr>
<tr>
<td>22</td>
<td>25</td>
</tr>
<tr>
<td>9</td>
<td>22</td>
</tr>
<tr>
<td>2</td>
<td>21</td>
</tr>
</tbody>
</table>
The highest (71.5%) combined visual and thermal band TPR from Table 6-8 is composed of three visual band features \([F2, F4, F6]\) and five thermal band features \([F14, F19, F20, F22, F26]\). The classification rates achieved by the visual-only combination \([F2, F4, F6]\) was 38.8% and by this thermal-only combination \([F14, F19, F20, F22, F26]\) was 68.5%. It is demonstrated by this experiment that adding the thermal data to the visual data improves the TPR by 32.7% but adding the visual data to the thermal data only improves the TPR by only 3%. Conversely, combining the feature vector that achieved the highest TPR in the visual band experiment \([F2, F5, F7, F10, F13]\) from Table 6-2) with the highest TPR feature vector from the thermal band experiment \([F14, F19, F20, F26]\) from Table 6-5), achieved a combined visual-thermal TPR of only 57.6% as compared to the maximum of 71.5% (from Table 6-8). Therefore it can be concluded that combining the best feature vectors from the visual band and thermal band experiment does not guarantee better TPR. An optimum feature combination of visual and thermal features needs to be evaluated as was done in the third experiment.

It can be demonstrated that the TPR improvement is due to the right combination of visual and thermal band features and not because the feature vectors contained more features. As a demonstration, if all 13 visual band features are combined (instead of 5 as used in the second experiment), the best TPR achieved is 58.2% as compared to 65.5% with only 5 features. Similarly, if all 13 thermal band features are
combined, the best TPR achieved is 55.8% as compared to 69.1%. Finally, if all 26 possible features (13 visual, 13 thermal) are combined, the best TPR achieved is 58.8% as compared to 71.5% achieved with only 8 features (3 visual and 5 thermal). Furthermore, the third highest TPR (70.3%) from Table 6-8 was achieved using only 6 of the possible 26 features which clearly demonstrates that more features does not necessarily result in better TPR. It is interesting to note that all of the combinations listed in Table 6-8 have two or less (many have none) feature pairs from both bands (e.g. Combination #4 [F6 F7 F10 F11 F13 F17 F20 F21 F25 F26] has 2 pairs of common features F7/F20 and F13/F26) which demonstrates that very different and useful classification information can be extracted from the individual bands.

Figure 6-11 illustrates the combined visual-thermal classifier TPR distribution of all the feature combinations evaluated. The mean TPR of the combined visual-thermal classifier is 51.6% with a variance of 38.1%. As discussed in Section 6.2.3, the mean TPR of the thermal band classifier was 47.6% (with a variance of 76.5%) while the visual band had a mean of 47.2% (with a variance of 48.1%). Although these are not significant rates, it does show an upward trend and an improvement in the average performance and a decrease in the variance from the combined visual and thermal classifiers.

![Figure 6-11: Combined visual and thermal band classifier TPR distribution](image)

Table 6-10 summarizes the performance metrics of the combined visual and thermal bands classifier by class, for the highest ranking feature combination ([F2 F4 F6 F14 F19 F20 F22 F26]) from Table 6-8. Of particular interest is the improvement in the F1 score of the combined classifier for the Class 2 which went from 63.7% and 73.5% for the visual and thermal band classifiers respectively to 76.0% for the combined.
The combined classifier also improved the Class 5 accuracy to 88.1% by reducing the number of False Positive (FP) from 19 down to 4. The combined classifier was not as good as the visual-only classifier at identifying Class 1 samples (F_1 score of 45.7% down to 13.3%). The overall average accuracy of the combined classifier improved from 82.6% (visual-only) to 86.3% which is an improvement of 3.7%.

Table 6-10: Combined visual-thermal classifier performance metrics

<table>
<thead>
<tr>
<th>Class</th>
<th>Samples</th>
<th>TP</th>
<th>TN</th>
<th>FP</th>
<th>FN</th>
<th>TPR</th>
<th>TNR</th>
<th>PPV</th>
<th>NPV</th>
<th>Acc</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>23</td>
<td>2</td>
<td>116</td>
<td>5</td>
<td>21</td>
<td>8.7%</td>
<td>95.9%</td>
<td>28.6%</td>
<td>84.7%</td>
<td>81.9%</td>
<td>13.3%</td>
</tr>
<tr>
<td>2</td>
<td>44</td>
<td>38</td>
<td>80</td>
<td>18</td>
<td>6</td>
<td>86.4%</td>
<td>81.6%</td>
<td>67.9%</td>
<td>93.0%</td>
<td>83.1%</td>
<td>76.0%</td>
</tr>
<tr>
<td>3</td>
<td>28</td>
<td>25</td>
<td>93</td>
<td>17</td>
<td>3</td>
<td>89.3%</td>
<td>84.6%</td>
<td>59.5%</td>
<td>96.9%</td>
<td>85.5%</td>
<td>71.4%</td>
</tr>
<tr>
<td>4</td>
<td>40</td>
<td>35</td>
<td>83</td>
<td>3</td>
<td>5</td>
<td>87.5%</td>
<td>96.5%</td>
<td>92.1%</td>
<td>94.3%</td>
<td>93.7%</td>
<td>89.7%</td>
</tr>
<tr>
<td>5</td>
<td>30</td>
<td>18</td>
<td>100</td>
<td>4</td>
<td>12</td>
<td>60.0%</td>
<td>96.2%</td>
<td>81.8%</td>
<td>89.3%</td>
<td>88.1%</td>
<td>69.2%</td>
</tr>
<tr>
<td>Overall</td>
<td>165</td>
<td>118</td>
<td>472</td>
<td>47</td>
<td>47</td>
<td>71.5%</td>
<td>90.9%</td>
<td>71.5%</td>
<td>90.9%</td>
<td>86.3%</td>
<td>71.5%</td>
</tr>
</tbody>
</table>

6.2.5. Expert System

By observing the performance metrics of the three classifiers discussed up to this point (see Table 6-4, Table 6-7 and Table 6-10), it can be observed that the combined visual-thermal provided a F_1 score improvement of 6% over the visual-only classifier and 2.4% improvement over the thermal-only classifier. It is clear that each classifier has its strength and can identify specific classes of objects better than others. Table 6-11 compares the F_1 score of each classifier against each class. The highlighted cells indicate the classifier that performed the best in each class. For example, the combined visual-thermal classifier achieved the highest F_1 score for Class 2 while the thermal classifier achieved the highest for Class 3.

Table 6-11: Classifier F_1 score comparison

<table>
<thead>
<tr>
<th>Class</th>
<th>Samples</th>
<th>Visual</th>
<th>Thermal</th>
<th>Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>23</td>
<td>45.7%</td>
<td>12.9%</td>
<td>13.3%</td>
</tr>
<tr>
<td>2</td>
<td>44</td>
<td>63.7%</td>
<td>73.5%</td>
<td>76.0%</td>
</tr>
<tr>
<td>3</td>
<td>28</td>
<td>73.3%</td>
<td>74.0%</td>
<td>71.4%</td>
</tr>
<tr>
<td>4</td>
<td>40</td>
<td>80.5%</td>
<td>88.0%</td>
<td>89.7%</td>
</tr>
<tr>
<td>5</td>
<td>30</td>
<td>53.7%</td>
<td>60.4%</td>
<td>69.2%</td>
</tr>
</tbody>
</table>

Based on the F_1 score achieved from each classifier, an expert system was assembled using the best classifiers. The expert system operated by presenting the testing dataset to all three classifiers (visual, thermal and combined) and using their expertise to identify their own specific classes. Once all three individual classifiers completed their prediction, a script compiled each experts’ predictions into a single
expert system prediction output. For example, if the combined classifier predicted that a sample belonged to Class 2, then its prediction was used as the output of the expert system. Conversely, if the combined classifier predicted a sample belonged to Class 3, its output was overwritten by the output of the thermal classifier as it is the expert on Class 3 recognition. With this expert system, each testing sample was evaluated by each of the three classifiers and the final prediction was made by the class specific expert. Table 6-12 illustrates the expert classifier for each of the five classes investigated in this research.

Table 6-12: Expert system

<table>
<thead>
<tr>
<th>Class</th>
<th>Name</th>
<th>Classifier</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Mobile Phone</td>
<td>Visual band</td>
</tr>
<tr>
<td>2</td>
<td>Coffee Mug</td>
<td>Combined visual-thermal bands</td>
</tr>
<tr>
<td>3</td>
<td>Laptop Charger</td>
<td>Thermal band</td>
</tr>
<tr>
<td>4</td>
<td>Desk Lamp</td>
<td>Combined visual-thermal bands</td>
</tr>
<tr>
<td>5</td>
<td>Portable Heater</td>
<td>Combined visual-thermal bands</td>
</tr>
</tbody>
</table>

The performance of the expert system is compiled in Table 6-13. It is observed from this table that the expert system classifier achieved a TPR of 76.4% with a specificity of 93.0% and an overall accuracy of 89.1%. The expert system demonstrated a small TPR improvement of 4.9% over the combined classifier (76.4% vs. 71.5%) which was previously the best classifier from this experiment. The complete performance metrics of the expert system are tabulated in Table 6-13 and can be compared to the metrics achieved by the combined classifier from Table 6-10, the thermal classifier from Table 6-7 and the visual band classifier metrics from Table 6-4.

Table 6-13: Expert system performance metrics

<table>
<thead>
<tr>
<th>Class</th>
<th>Samples</th>
<th>TP</th>
<th>TN</th>
<th>FP</th>
<th>FN</th>
<th>TPR</th>
<th>TNR</th>
<th>PPV</th>
<th>NPV</th>
<th>Acc</th>
<th>F1</th>
<th>Expert</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>23</td>
<td>8</td>
<td>118</td>
<td>5</td>
<td>15</td>
<td>34.8%</td>
<td>95.9%</td>
<td>61.5%</td>
<td>88.7%</td>
<td>86.3%</td>
<td>44.4%</td>
<td>Visual</td>
</tr>
<tr>
<td>2</td>
<td>44</td>
<td>38</td>
<td>88</td>
<td>13</td>
<td>6</td>
<td>86.4%</td>
<td>87.1%</td>
<td>74.5%</td>
<td>93.6%</td>
<td>86.9%</td>
<td>80.0%</td>
<td>Combined</td>
</tr>
<tr>
<td>3</td>
<td>28</td>
<td>27</td>
<td>99</td>
<td>12</td>
<td>1</td>
<td>96.4%</td>
<td>89.2%</td>
<td>69.2%</td>
<td>99.0%</td>
<td>90.6%</td>
<td>80.6%</td>
<td>Thermal</td>
</tr>
<tr>
<td>4</td>
<td>40</td>
<td>35</td>
<td>91</td>
<td>3</td>
<td>5</td>
<td>87.5%</td>
<td>96.8%</td>
<td>92.1%</td>
<td>94.8%</td>
<td>94.0%</td>
<td>89.7%</td>
<td>Combined</td>
</tr>
<tr>
<td>5</td>
<td>30</td>
<td>18</td>
<td>108</td>
<td>6</td>
<td>12</td>
<td>60.0%</td>
<td>94.7%</td>
<td>75.0%</td>
<td>90.0%</td>
<td>87.5%</td>
<td>66.7%</td>
<td>Combined</td>
</tr>
<tr>
<td>Overall</td>
<td>165</td>
<td>126</td>
<td>504</td>
<td>39</td>
<td>39</td>
<td>76.4%</td>
<td>92.8%</td>
<td>76.4%</td>
<td>92.8%</td>
<td>89.0%</td>
<td>76.4%</td>
<td></td>
</tr>
</tbody>
</table>

If additional classes were to be added in the future, the classification process used in this study (determine classification rates for visual-only, thermal-only and combined visual-thermal) would have to be repeated before re-assembling a new expert system. The addition of new classes of objects would impact the performance for each of the classifiers.
6.2.6. Further Improvements

The results achieved by the Expert System (76.4%) described in Section 6.2.5 were an improvement over the classification rates achieved by the single feature classifiers (55.8%) presented in Section 6.2.1, the visual band classifier (65.5%) from Section 6.2.2 and the thermal band classifier (69.1%) from Section 6.2.3 but only a small improvement over the combined visual-thermal classifier discussed in Section 6.2.4. A fundamental question remains which is how to further improve these results?

The features selected for this work were not typical or commonly used features such as SIFT, FAST, SURF, Hu or Hough Moments discussed in Section 2.2, but rather a unique set of features that were previously used in another classification study using thermal imagery only [28]. In Cayouette et al.’s study, classification rates between 95% and 99.43% were achieved using only two classes of objects (aircraft and flare) that clearly differentiated from each other. A direct comparison of the accuracy achieved with this study is difficult to make because in this study, 5 different classes were used, visual and thermal imagery were used and much fewer training and testing samples were available.

It is possible that other feature descriptors (different features such as SURF, SIFT or FAST) could achieve better classification rates than the descriptors selected for this work. Furthermore, it was decided to limit the dimension of the descriptor length to 5 features (up to 13 available) for the visual and thermal-only classifiers and to 10 features (up to 26 available) for the combined visual-thermal classifier. It is possible that a feature descriptor with a dimension greater than 10 could provide better classification rates. However as demonstrated in Section 6.2.4 and as is known, additional features do not necessarily result in better classification (the curse of dimensionality).

In the early stages of the experimental design, it was decided that two separate datasets would be acquired; one to train the classifiers and one to test the classifiers. The training dataset was generated in a different environment setting than the testing dataset in order to replicate real world scenarios. As described in Chapter 3, the training imagery was captured using a common background while the testing dataset was captured with different backgrounds intended to create different levels of difficulty for the segmentation algorithm. It is possible that if the training dataset had been captured in the same setting as the testing dataset, that the classifier might have achieved better classification rates (more similarities between the training and testing samples). In addition, the size of the training dataset was relatively small in comparison to the testing dataset. The results presented in this work were achieved using between 7
and 10 training samples per class. It is possible that better classification rates could be achieved if more training samples were used.

In many classification studies, a single large dataset is divided into a training and testing subset. In these studies, the datasets are randomly divided and re-tested with new subsets each time to obtain an average classification rate using cross-validation techniques such as n-fold, hold-out and leave-one-out. Various techniques of meta-learning such as Arcing, Bagging and Boosting can be used to further enhance the classification results by choosing an optimum subset to train the classifiers. It is possible that by combining cross-validation and boosting techniques, better classification rates could have been achieved.

The segmentation of the objects of interest was performed using a custom algorithm developed specifically for this research. Although the semi-automatic algorithm performed better than the other tested segmentation techniques, it was not perfect by any means. Segmentation is a critical first step in extracting features from an image and if the segmentation is not done accurately, the extracted features will also be less accurate. It is possible that segmenting the objects of interest by hand or applying a different automated segmentation algorithm may result in more accurate features which would likely lead to better classification results.

Finally, the image quality of the Ti10 camera was poor at best. In many image pairs, the visual image was blurry and had low level of details. In the thermal samples, the small dynamic range of the sensor resulted in limited thermal details of the objects. A better quality imager could very likely improve the classification results achieved in this study and could potentially allow for additional features based on object texture in both the visual and thermal bands to be used as part of the descriptor.
Chapter 7  Conclusions

7.1.  Summary of Findings

The fundamental objectives of this research were to investigate the potential offered by combining features from visual and thermal imagery to improve the recognition rates and accuracy of commonly found objects in an office setting. The initial expectation was that the best classifier for this specific application would be a combined visual-thermal classifier using features extracted from both spectral bands. The research demonstrated that the best results were instead achieved by an expert system (multiple classifiers) that combined the expertise of the visual-only classifier, the thermal-only classifier and the combined visual-thermal classifier. As demonstrated in Section 6.2.5, the expert system designed for this research improved the overall $F_1$ score of the system to 76.4% with an accuracy of 89.0%. This is an improvement of 10.9% and 7.3% respectively over the visual-only and thermal-only classifiers.

When observing the results of the individual classes, the classifiers were able to recognize certain classes of objects with a higher certainty. For example, the Laptop Charger (Class 3) was recognized 96% of the time while the Desk Lamp (Class 4) was recognized over 88% of the time. However, in the case of the Mobile Phone (Class 1) and the Portable Heater (Class 5), the classifiers only achieved a TPR of 35% and 60% respectively.

In terms of feature descriptor, the feature combinations that achieved the highest classification rates were:

**Visual band only:** [F2 F5 F7 F10 F12] (1 intensity feature, 4 shape features)

**Thermal band only:** [F14 F19 F20 F26] (1 intensity feature, 3 shape features)

**Combined visual-thermal:** [F2 F4 F6 F14 F19 F20 F22 F26] (3 visual band, 5 thermal band features)

It was determined that the *Normalized Maximum Radial Distance* feature (Feature 7 in the visual band and Feature 20 in the thermal band) was the most prominent feature and the feature that provided the most discriminability between the classes. The *Normalized Maximum Radial Distance* feature was part of the best feature combination in the visual-only, in the thermal-only and combined visual-thermal classifiers. Also worth noting that the *Roundness* feature (Feature 26) extracted from the thermal band appeared in 49 of the top 50 combined visual-thermal combinations.
In summary, it can be concluded that each of the research’s initial objectives have been met:

[1] A segmentation algorithm was developed specifically for this research and used to extract objects of interest in both the visual and thermal bands of the EM spectrum.

[2] A dataset composed of five classes of objects was acquired to conduct this experimental study. The dataset was divided into 88 training occurrences (44 visual and 44 thermal), 330 testing occurrences (165 visual and 165 thermal).

[3] A feature descriptor composed of visual and thermal features was identified which yielded better classification rates than the visual-only or thermal-only feature descriptor.

[4] It was determined that best feature for optimized classification in the visual-only, thermal-only and combined visual-thermal bands was the shape feature *Normalized Maximum Radial Distance*.

[5] It was demonstrated that using an expert system classifier composed of visual-only, thermal-only and combined visual-thermal expert classifiers, can achieve better classification rates than any of its individual classifiers.

7.2. Contributions

The following contributions were made to the field of object recognition and classification as a result of this research:

- A new segmentation algorithm, *Watershed with Thermal Markers*, extending the classic *Watershed* approach whereby a thermal image is used to enhance the segmentation in the visual image.

- A demonstration of how a thermal image can be used to identify objects of interest in a visual band image.

- An experimental study demonstrating the use of combining features from the visual and thermal bands to improve object recognition and classification rates.

- A new feature descriptor based on visual and thermal characteristics of an object.
- A new approach for object recognition which combines support vector machine (SVM) classifiers with multispectral imagery.

- An Expert System that combines classifiers on visual and thermal images to improve classification performance.

7.3. Future Work

In Section 4.2.6, a new segmentation algorithm based on the classic Watershed algorithm was proposed. In the current work, the algorithm required the input of an operator to find the best upper and lower thresholds in order to optimize the performance of the segmentation. An extension of this work could be to investigate the feasibility and subsequently the robustness of an automatic threshold detection algorithm to support the Watershed with Thermal Markers algorithm.

In Section 5.3, a process was described by which the user had to review the output of the feature extraction function to eliminate all the non-class objects. This process is currently a manual process and is very time consuming. An extension of this research could be to develop an automated process to analyse and extract just the objects of interest following the feature detection process.

The presented research used a dataset that contained a single object from each of the five classes and the proposed algorithms recognized those specific objects in their environment. An extension of this work could be to increase the number of object per class to generalize the recognition system and attempt to identify classes of objects as opposed to specific objects. The challenge with this suggested future work will be to find relevant objects with thermal signatures observable with the available sensors. Additional classes such as humans and household animals could be considered in the future but research ethics and personal privacy policies may present a challenge in acquiring and using this type of data.

The 13 image features selected for this study were not classic features descriptors such as SIFT, FAST or Hu moments typically used for object recognition applications. An extension of this work could be to compare the 13 features used in this study against other classic object descriptors such as SURF or Legendre moments for example. Similarly, the classifier used for this study was the SVM. Future work could also be to compare the performance of the features used in this study with other types of classifiers to see if better classification rates can be achieved.
Another suggested extension of this work would be to implement the detection and recognition algorithms presented in this body of work on a real-time embedded system with the ability to capture visual and thermal imagery simultaneously in order to detect, recognize and classify various types of objects.
References


