HIDDEN TARGET DETECTION AND CLASSIFICATION USING MULTIPLE MODALITIES

by

Philip Saponaro

A dissertation submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Computer and Information Sciences

Winter 2016

HIDDEN TARGET DETECTION AND CLASSIFICATION USING MULTIPLE MODALITIES

by

Philip Saponaro

Approved: ____
Kathleen McCoy, Ph.D.
Chair of the Department of Computer and Information Sciences

Approved: ____
Babatunde Ogunnaike, Ph.D.
Dean of the College of Engineering

Approved: ____
Ann L. Ardis, Ph.D.
Interim Vice Provost for Graduate and Professional Education

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ____
Chandra Kambhamettu, Ph.D.
Professor in charge of dissertation

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ____
Jingyi Yu, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ____
Christopher Rasmussen, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ____
Kenneth Ranney
Member of dissertation committee

ACKNOWLEDGEMENTS

I wish to thank everyone who has helped me: my advisor, Dr. Chandra Kambhamettu, my colleagues at the Video/Image Modeling and Synthesis lab including Scott Sorensen, Abhishek Kolagunda, Stephen Rhein, Guoyu Lu, Xiaolong Wang, Wayne Treible, Rohith M.V., and Gowri Somanath; my colleague Dr. Robert Pavel from the Computer Architecture and Parallel Systems Laboratory. I also thank my mentors and colleagues at the Army Research Lab including Kyle Gallagher, Brian Phelan, Richard Pooler, Kristopher Young, Kenneth Ranney, Kelly Sherbondy, and Dr. Andy Sullivan. I thank my committee members not mentioned above, Dr. Jingyi Yu and Dr. Christopher Rasmussen. Finally, I would also like to thank my family for being supportive, Sherrie and Matthew Saponaro, and my fiance, Kelly Jennings.

TABLE OF CONTENTS

Appendix

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Hidden target detection and classification is an important task for many security and military applications. Long wave infrared (8-14 μm) cameras, otherwise known as thermal cameras, can be used towards hidden target detection and classification but are less studied in the Computer Vision literature due to their high cost and low resolution. Thermal imagery is able to reveal targets such as camouflaged or shallowly buried targets that would be hidden to optical band sensors. For this dissertation, I studied some of the problems in designing a computer vision system that uses the thermal modality along with other modalities to detect and classify hidden targets. Specifically, this dissertation seeks to address (1) calibration of multiple cameras both within the thermal modality and across modalities, (2) detection of hidden targets in the scene by identifying anomalous regions and known targets, and (3) classification of the hidden targets. I propose novel approaches towards solutions of these issues and argue for the efficacy of these approaches. Particularly, for calibration I used a ceramic backing and preprocessing technique for enhancing the contrast and its duration, and show that heating a printed calibration board is indeed viable for calibration in contrast to previous work. For detection, a dynamically updating Gaussian mixture model and sensor fusion was used to identify anomalous regions, while neural networks were used for fusing multimodal sensors and detecting known objects. Finally, for classification I developed novel thermal-based features such as water permeation and heating/cooling patterns to classify the materials. I developed the CHAracteristic Model of Permeation (CHAMP) for modeling both the rate and shape of water permeation, and use the heat equation for extracting physical material parameters for a heat feature. In each case, my results show that thermal is a useful modality for detection and classification of objects, and can be combined with other modalities to increase performance.

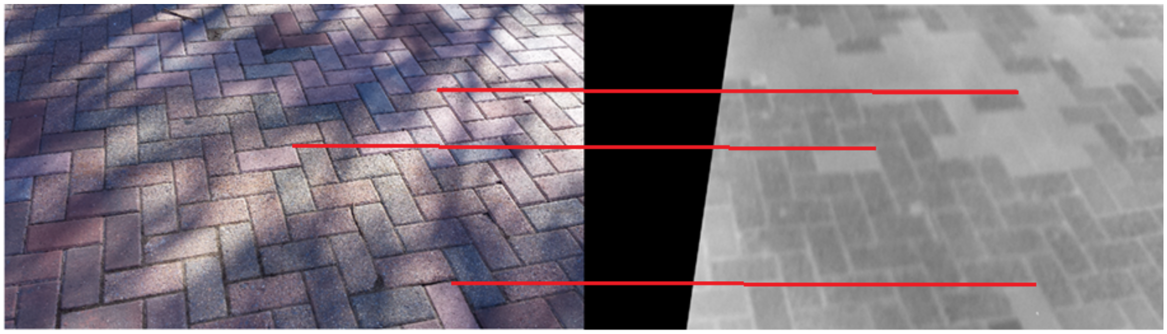
Chapter 1

INTRODUCTION

Humans use vision for gaining knowledge and interacting with the external world. Many computer vision algorithms try to mimic processes within humans to model the external world. Stereo vision [75] and structure-from-x [44] techniques are based on how humans observe geometry with their eyes. However, human vision is actually severely limited in the electromagnetic spectrum. Humans can only observe a small slice of the spectrum between 390-700nm [80]. Other wavelengths can contain important information about the external world, but cannot be detected without the aid of a sensor. Most current research focuses on the visible spectrum, but in this thesis I study the long-wave infrared spectrum of 8-14μm, otherwise known as thermal infrared. A history of thermal imaging is discussed in [7].

Thermal cameras can be used in a wide variety of applications. They were originally developed for military use, but have been extended to be used for building inspection [21], law enforcement [13, 42], pedestrian detection [54, 61, 82], medical imaging [16], astronomy [6], meteorology [37], and bank robbery detection [112] to name just a few. Thermal cameras have the advantage over color cameras for some applications because of their invariance to lighting changes, ability to work without any light, relative invariance to color changes, and the direct observation of thermal information.

Examples of phenomenon visible with a thermal camera and possible applications are shown in Figure 1.1, Figure 1.2, and Figure 1.3. In Figure 1.1, different material bricks and structure damage can be easily seen in the thermal imagery. In Figure 1.2, different material fruit can be seen after heating up, and objects inside an envelope are also visible. Figure 1.3 shows a dual stereo system that can be used to reconstruct both a reflecting surface and a reflective surface.



a)

b)

Figure 1.1: Color and thermal images of bricks. The thermal imagery shows that at least two different brick materials were used. Upon close inspection, the smooth brick appears brighter (hotter) than the rough bricks, and chipped corners of bricks appear brighter. **a)** Image correspondences. **b)** Aligned images via a homography with alpha blending.

Thermal imaging cameras can be divided into two groups: cooled and uncooled. Cooled detectors operate at around 100K and keep the sensor clear of their own thermal radiation with active cooling. They typically perform better with motion or quickly changing temperatures than uncooled and also have a greater thermal sensitivity. However, they are typically much more expensive and energy-intensive than uncooled detectors. The thermal cameras used in this dissertation are uncooled cameras. Uncooled cameras use a sensor operating at ambient temperature and measure changes in current or voltage when heated by infrared radiation. They are cheaper than cooled sensors but tend to have slightly lower image quality. The sensors are built using mainly pyroelectric and ferrorelectric materials, with the lenses made from Germanium (Ge), Chalcogenide glass, Zinc Selenide (ZnSe) and

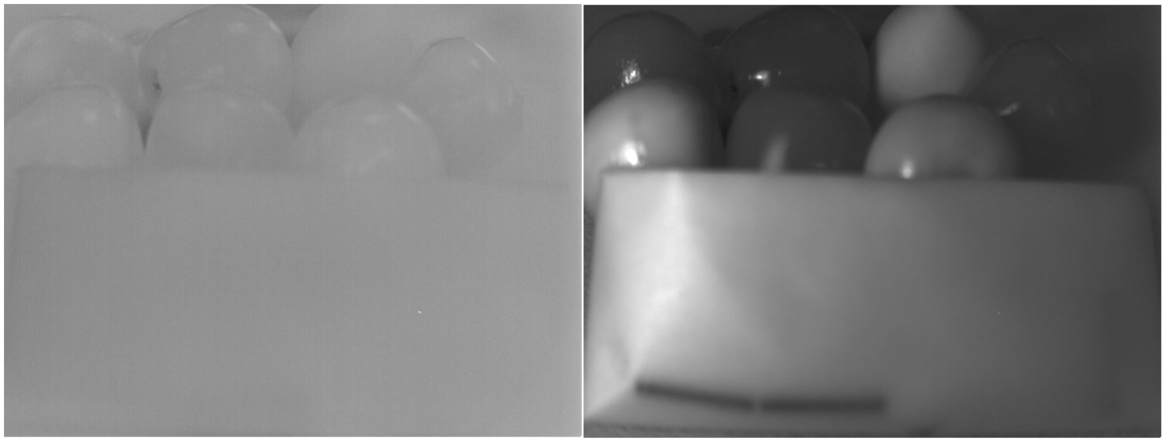


Figure 1.2: Thermal imagery of a scene containing plastic, styrofoam, real fruit, and an envelope containing batteries. Everything was left at room temperature for one day. **a)** Unheated **b)** Heated for 30 seconds with a heat lamp placed 0.6m away. The brightest 3 fruit are styrofoam, the darkest fruit is organic.

Zinc Sulfide (ZnS). [58] discusses the components of an uncooled thermal imaging system in more detail, while [53] discusses cooled sensors as well.

One of the reasons thermal imagery is less used in the computer vision literature is the high price and low resolution of current thermal cameras. Table 1.1 shows a comparison of prices and resolution. Typically, uncooled thermal cameras do not go above 640x480 resolution, and those cameras cost on the order of $20,000.00. However, recently FLIR has released an accessory that can connect to a smart phone which combines color (640x480) and thermal (80x60) information that can allow an artifical thermal resolution of 640x480 for under $350.00 [1]. With cheaper accessories, there is a greater possibility for mobile thermal applications.

1.1 Motivation

The application for the work in this thesis is hidden target detection, specifically improvised explosive device (IED) detection. An IED is a bomb made in an improvised manner from homemade or commercially sourced explosives. They were extensively used against US-led forces in the Middle East, and were the cause of over 40% of the casualties

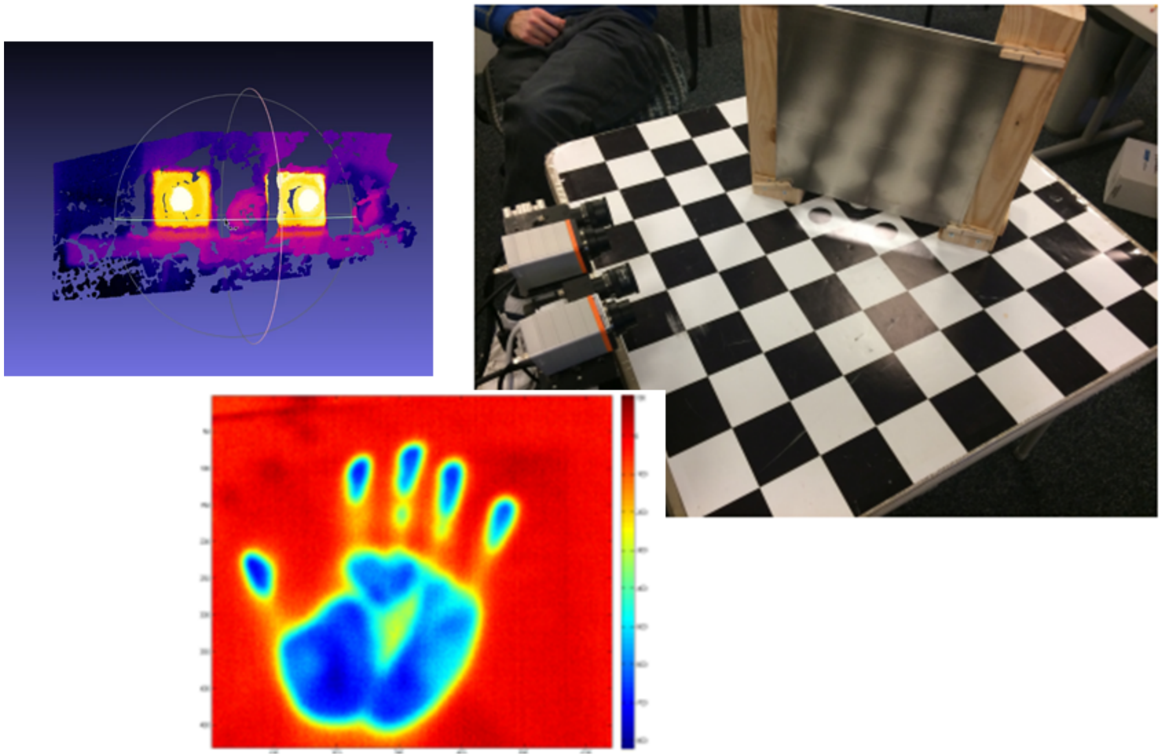


Figure 1.3: A four camera system with stereo thermal and stereo color cameras can be used to reconstruct both a reflecting surface and a reflective surface. Texture can be added in one modality to the reflecting surface that is invisible to the other modality. E.g. a thermal hand print can be added in thermal, or scribbles with marker can be added in color.

Brand	ID	Resolution	Price
FLIR	FLIR E4	80x60	$995.00
FLIR	FLIR E5	120x90	$1495.00
FLIR	FLIR E40BX - E40bx	160x120	$3995.00
FLUKE	FLK-TI200 60HZ	200x150	$6299.00
FLUKE	FLK-TI400 60HZ	320x240	$8495.00
Xenics	Gobi 640 GigE	640x480	$13115.00
FLIR	55903-5122-T620	640x480	$20950.00

Table 1.1: Comparison of uncooled thermal camera resolution and prices. Information in this table courtesty of [5] as of 12-29-1015

.

in the Afghanistan and Iraq wars from 2007-2013 [2]. IEDs were also heavily used in the Vietnam War, in Northern Ireland by the IRA, and more recently by Maoists in India [130].

The IED contains five main components: an activator, a fuse, a container, an explosive, and a power source. These various components can be detected more easily by certain modalities. For example, non-linear radar is adept to finding electronic components that can be used to detonate the IED, such as a cell phone [69, 32]. Thermal is adept at finding temperature gradients such as in the power supply or recent disturbances of earth (shown in Chapter 4 and Chapter 5). And ultra-wideband linear radar can see the explosive charge, even when it is buried up to 0.30 meters [83]. These technologies can be used to detect the target at a high standoff distance to keep any human personal safe.

The goal of this dissertation is to develop algorithms necessary for detection and classification of hidden targets using thermal imagery. This includes calibration of thermal cameras for their intrinsic parameters, calibration of thermal cameras for their extrinsic parameters between both color and thermal modalities, the fusion of modalities for detection of hidden targets, and the use of thermal imagery for material classification. The work in this dissertation is general enough to be applicable to other domains for detection and classification, and not necessarily IED detection, such as material classification.

1.2 Outline

Chapter 2 describes a technique for calibrating thermal cameras. The method allows for a printed calibration board to be used with standard toolboxes for calibration of stereo cameras in both color and thermal modalities. The performance is evaluated on both indoor and outdoor trials using both artificial and natural heating sources, respectively. Chapter 3 describes how to align multiple modalities in a dynamic scene. The alignment system was tested on data collected on the SIRE vehicle at both the Army Research Lab (ARL) and in Yuma, Arizona. Chapter 4 details the detection of anomalous pixels in the scene using Gaussian Mixture Models (GMMs). This system was tested at ARL in an indoor sand environment and an outdoor environment with targets buried underground and hidden in vegetation. Chapter 5 discusses detection of known targets using multiple modalities and neural networks. Chapter 6 discusses material classification using the thermal modality. I develop new features for thermal imagery based on water permeation and the heating/cooling cycle. Chapter 7 compares color-based stereo against thermal-based stereo for reconstruction of objects. Chapter 8 shows how to extend the thermal heating feature to curved objects in uncontrolled locations in the scene. Chapter 9 concludes the dissertation and discusses future work.

Chapter 2

THERMAL CALIBRATION

2.1 Introduction

Camera calibration is the process of determining the camera parameters which map 3D scene points onto a 2D image plane. For stereo or multi-camera systems, the translation and rotation between each camera is also calculated. These parameters are important for many computer vision algorithms such as 3D metric reconstruction, feature matching, and localization, which are useful in military, entertainment, medical, and industrial domains. Thermal or long wave infrared (LWIR) cameras can be complementary or even advantageous when compared to standard color cameras in certain applications. Thermal cameras can see a scene with no light, are invariant to lighting changes, and are robust to foggy conditions [10]. They are also useful for thermal analysis [129] and even for detecting hidden targets [8, 93].

With standard color cameras, the calibration process has effective methods of accurately determining the camera parameters. The most popular method applies the algorithm described in [133]. This requires reliably detecting corner points in a sequence of images, which is made easy by printing a chessboard calibration pattern with high contrast between the white and black squares. However, in thermal imagery the chessboard pattern is not visible due to a uniform temperature profile as shown in Figure 2.1. One abandoned approach uses a printed chessboard heated by a flood lamp which results in blurry, hard to detect corners that are only visible for a short period of time.

The purpose of this chapter is to discuss a method for increasing the contrast between the squares of the chessboard pattern while not requiring a custom calibration object to be built. This also will allow standard off-the-shelf calibration toolboxes to be more reliably

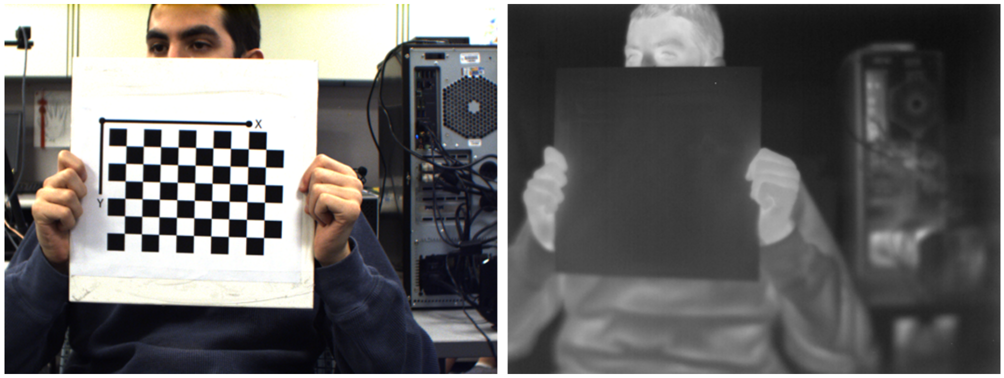


Figure 2.1: Comparison between color and thermal images of an unheated calibration board. The calibration board is not visible in thermal.

used. To achieve this I used a heat lamp to heat the calibration board. However, for this method to work a few problems needed to be overcome:

- Retaining heat long enough to record calibration images
- Correcting for non-uniform heating
- Enhancing/sharpening corners

I use a ceramic backing to retain heat longer. To correct non-uniform heating, I developed an iterative pre-processing technique combined with top-hat filtering. Finally, to enhance contrast between the squares, I use gamma correction.

The chapter is organized as follows. Section 2.2 gives related works and previous methods to calibrate thermal cameras. Section 2.3 details the setup and method for calibration. Section 2.4 explains the experiments and results, including how the calibration method performs in real calibration video sequences and how long it remains effective at successfully calibrating after cooling. I performed tests in both an indoor and outdoor setting.

2.2 Related Works

Standard calibration patterns have uniform temperature in thermal imagery. To get around this problem, a few techniques have been proposed which generally use heated, novel calibration boards made of varying materials. [118] creates a calibration board by cutting out a "mask" of squares to expose the background temperature. [46] mills a chessboard

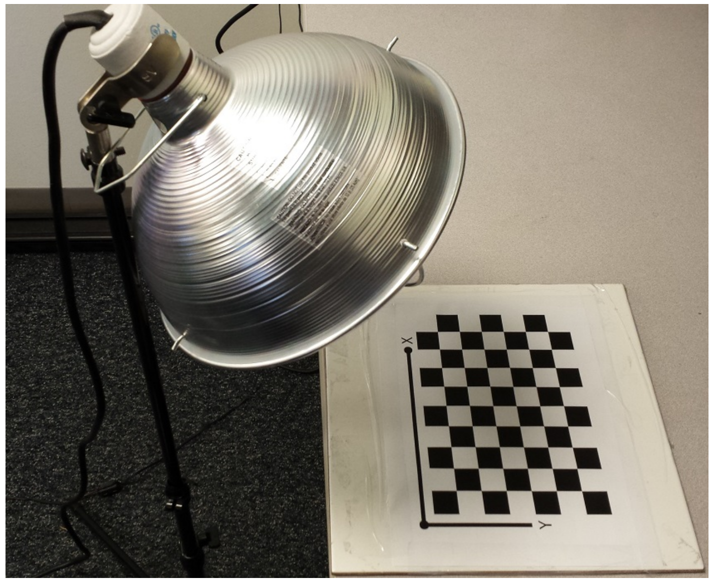


Figure 2.2: The setup includes a printed paper calibration pattern, a glazed finish ceramic tile backing to keep the pattern flat and retain heat, and a 250W heat lamp.

pattern into a printed circuit board with a high emissivity base material and low emissivity copper squares. [129] uses a wire net that is heated with a heat gun. [38] uses a set of resistors mounted in the center of each square. [60, 128, 28] use a grid of lightbulbs. [131] uses circular thermostatic heaters in a cross-shape pattern. The drawback to many of these methods is that the calibration board needs to be custom made and can be time consuming or expensive to make.

One method that does not require any changes to the standard chess calibration board simply heats it with a flood lamp, as in [20, 81]. However, this method was argued against by [118], which shows that the corners were not sharp enough to reliably detect. The contrast between the squares also decreased quickly – 30 seconds after heating the corners were hard to detect. I propose improvements to the heat-based method that makes it viable for calibration. The improvements allow calibration to be performed on a single sheet of printed paper instead of a custom made calibration object made of varying materials, and works with off-the-shelf calibration toolboxes.

2.3 Method

2.3.1 Physical Setup

The setup can be seen in Figure 2.2. A printed calibration board is taped to the glazed finish ceramic tile backing in order to keep the pattern flat. Ceramic is chosen because of it is inexpensive (often can be found as a free sample) and has a low thermal moment, which causes it to heat and cool slowly. I study this relationship between heating length, cooling time, and calibration quality in Section 2.4. The calibration pattern is then heated using a 250W heat lamp, which can reach temperatures up to 550°C, although at 2 feet away the calibration pattern reached 55°C. Next, pre-processing is applied to a video sequence of calibration images. Finally, corners are detected and calibrated using [133]'s method which is implemented in many off the shelf toolboxes.

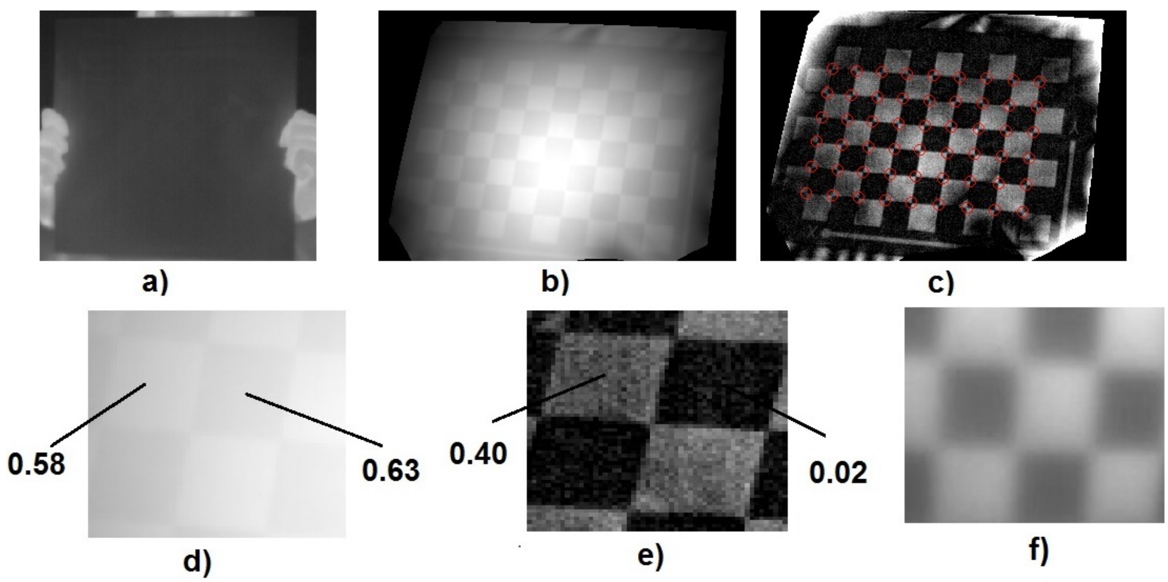


Figure 2.3: Comparison of heating effects; note that white pixels denote high temperature. **a)** Unheated calibration board. The intensity is mostly constant. **b)** Heated calibration board. The squares are much more visible, but uneven heating makes corner detection difficult. **c)** Output of the proposed method. Off-the-shelf toolboxes can now easily detect corners. **d)** Zoom in on b). A "white" square actually has a lower intensity than a "black" square. **e)** Zoom in on c). The light square is now much brighter than the dark square. **f)** The previous method's "best quality calibration image that could be produced using the heated chessboard method." [118]

2.3.2 Correction of Non-Uniform Heating

The processing can be summarized as: Mask out the calibration pattern $\rightarrow$ Iteratively fit a model to the image intensity and subtract the model from the image $\rightarrow$ Top hat filtering $\rightarrow$ Gamma correction. Each of these is described below.

Due to the size and position of the heat lamp, the center of the calibration object is heated more than the sides, as can be seen in Figure 2.4. Although off-the-shelf calibration toolboxes can find some corners, the corner detection is unreliable and inaccurate, as discussed in [118]. Moreover, with non-uniform lighting, standard contrast enhancement techniques fail. By correcting the non-uniform heating, there is an increase to the usefulness of standard contrast enhancement and reliability of obtaining more points.

To correct the non-uniform heating, the calibration pattern is masked out from the rest of the scene. Conveniently, the heated calibration pattern is much warmer than the rest of the scene. This assumption is used to automatically threshold out the calibration pattern using Otsu's method [78].

Next, a model is fit to the intensity data that remains after the masking. Let $I_{mask} = chess\_pattern\_data + parametric\_heat\_model + noise$. The goal is to model the non-uniform heat and noise, and subtract it out leaving only the chess pattern data. I chose to use a quadratic polynomial to model the intensity. That is

$$p\_h\_m(x,y) = p_{00} + p_{10} * x + p_{01} * y + p_{20} * x^2 + p_{11} * x * y + p_{02} * y^2, \quad (2.1)$$

where $p_{00}, p_{10}, p_{01}, p_{20}, p_{11}, p_{02}$ are the 6 parameters of the model. The fit is calculated using the Levenberg-Marquardt algorithm [67]. Next the model is subtracted from the image $I'(x,y) = I_{mask}(x,y) - p\_h\_m(x,y)$. To account for noise, this process of fitting and subtracting is repeated until the change is under a threshold, $I'_n - I'_{n-1} < \alpha$.

Finally, after this iterative fitting process, top hat filtering is performed. Top hat filtering is a morphological operation usually performed to remove non-uniform illumination and is defined as

$$I_{hat} = I - (I \circ S), \tag{2.2}$$

where $\circ$ is the morphological opening operator with structuring element S.

2.3.3 Contrast Enhancement

Standard contrast enhancement is performed via normalization and gamma correction. First, the intensity values are normalized to $[0, 1]$. Then the image is gamma corrected. This can be seen in Equation 2.3.

$$I_{gamma} = \left(\frac{I - min(I)}{max(I) - min(I)} \right)^{\gamma}. \tag{2.3}$$

The intensity values are then mapped again to $[0, 1]$.

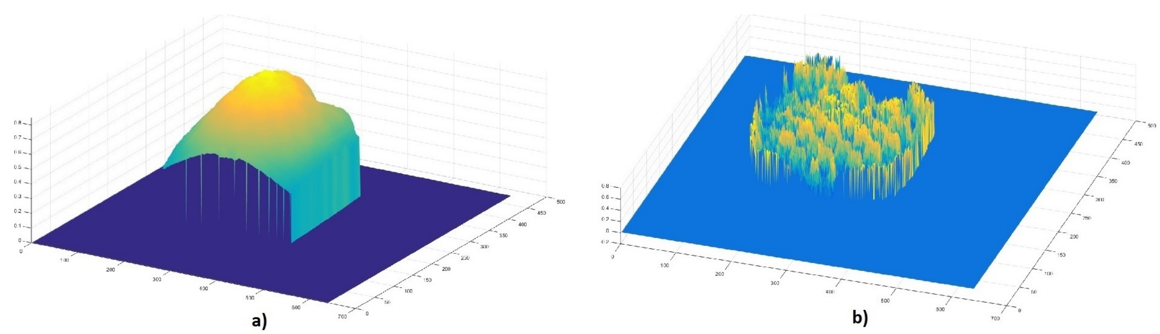


Figure 2.4: Heating correction. **a)** A surface plot of the intensity values after heating but before correction. The white/black squares lie along a surface and only slightly perturb the surface. **b)** Surface plot after correction. The ”white” squares rise up while the ”black” squares stay close to 0. Note that white and black are inverted from color imagery because the black squares absorb and emit more thermal energy.

2.4 Experiments and Results

2.4.1 Indoor Experiments and Results

In all of the experiments I used two Xenics Gobi 640 GigE uncooled long wave infrared cameras, which each have a resolution of 640x480 and a 50mK sensitivity. I also used two Point Grey Flea 2G 5MP color cameras, which used a resolution of 1600x1200. All four cameras were synchronized with software triggers and placed on a baseline of 0.25m focused at 1m away. I performed multiple experiments: algorithm variations, calibration quality over time, calibration quality for multiple materials, and cross modality experiments.

2.4.2 Algorithm Variation Experiment

The first experiment was to compare different pre-processing variations for enhancing the thermal calibration results.

Algorithm	% Pairs	#PtsPerPair	RMS (px)
Proposed (Quad iter fit, top hat, gamma=.8)	**31.40**	**54**	0.48
Quad non-iter fit, top hat, gamma=.8	12.94	48	0.45
No fit, top hat, gamma=.8	1.18	**54**	N/A
Quad iter fit, no hat, gamma=.8	3.53	20	17.01
Quad iter fit, bot hat, gamma=.8	29.40	**54**	0.44
Quad iter fit, top and bot hat, gamma=.8	3.53	20	13.10
Cubic iter fit, top hat, gamma=.8	30.80	**54**	0.45
Quartic iter fit, top hat, gamma=.8	24.70	**54**	**0.41**
Quad iter fit, top hat, no gamma	16.46	**54**	0.47
Quad iter fit, top hat, gamma=.5	29.40	**54**	0.45
Quad iter fit, top hat, gamma=1.15	6.40	**54**	0.46
Color imagery, no processing	50.56	54	0.38

Table 2.1: Comparison of variations of our pre-processing method averaged over 3 trials of 61, 46, and 63 pairs of images. % pairs measures the number of pairs with detected corners divided by the total number of pairs. The actual number of corners is 54. Note the RMS error is similar for most methods, and any value under 1 is typically acceptable.

I placed the calibration board under a heat lamp for 1 hour before recording a calibration video where the board was rotated to different orientations. This experiment was repeated 3 times with trial 1 having 61 images per camera, trial 2 having 46 images, and trial 3 having 63 images. I averaged the results which are presented in Table 2.1. The quality of the reconstruction was measured by the percentage of image pairs with detected corners, the number of corners per image, and the root-mean-square (RMS) reprojection error. I tested a few variations of the method: shape fitting with and without iteration, higher order polynomial fitting, using bottom hat vs top hat, and different gamma values. The structuring element for the morphological operations was a disk of radius 15 pixels.

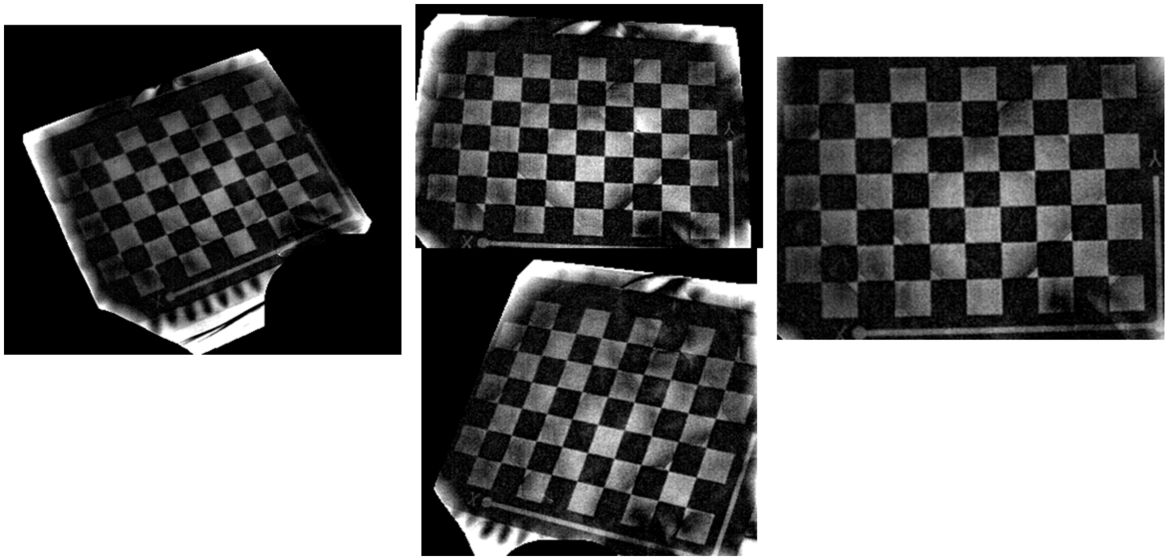


Figure 2.5: Example thermal calibration images after processing.

From Table 2.1 I noticed that when the number of detected pairs reached a certain threshold, the RMS error was similar between all methods. A more useful metric seemed to be the percentage of pairs of images with detected points. Typically, at least 6 different orientations are needed to give a robust, reliable calibration. In all of the trials, the proposed method gave more than double what is needed for a reliable calibration. Example thermal calibration images after processing are shown in Figure 2.5. A visual comparison of the quality of the proposed method versus the previous best heating method is shown in Figure 2.3.

I observed that the order of the fitted polynomial does affect the results, but only marginally so compared to other parts of the pipeline. Top hat filtering was necessary but not sufficient to acheive reliable calibration. The gamma value chosen for contrast enhancement was very important to the final results.

2.4.3 Calibration Quality Over Time Experiment

The second experiment I performed was to measure how well the calibration board retains heat to enhance contrast between the squares. To do this I performed four different trials where the calibration board was heated for 5, 10, 20, and 30 minutes and then was placed in a static, flat angle facing the camera. The camera recorded 14 minutes of video where the calibration board was cooling over time. The number of points detected over time are plotted in Figure 2.6. The number of detected points was measured in increments of 15 seconds. Note that curves were fit to the raw data for visualization purposes.

300 corner points is the cutoff for reliable calibration because it corresponds to about 6 image pairs where all corner points were detected. In the worse case of heating the calibration board for 5 minutes, the corner detection was reliable for over 5 minutes. This is significantly longer than the time calibration was reliable in other works (30 seconds in [118]). Note that the proposed method relies on the assumption that the calibration pattern is much hotter than the rest of the scene from being heated. Thus, if other objects in the scene are of similar temperature or if the pattern cools off too much our method will fail. However, in the data I recorded for this experiment, even if humans or computers were in the scene, the calibration board was significantly warmer than the rest of the scene for at least 5 minutes.

2.4.4 Material Type Experiment

For the third experiment I performed a similar process to the previous section (Calibration Quality Over Time), but using different material types as the backing to the calibration pattern. The printed calibration pattern was taped to wood, cardboard, and aluminum backings. The pattern was heated for 10, 20, and 30 minutes respectively, and then was placed at a static, flat angle facing the camera. The camera recorded 14 minutes of video where the calibration board was cooling over time. The number of points detected over time are plotted in Figure 2.7. The number of detected points was measured in increments of 15 seconds. Note that curves were fit to the raw data for visualization purposes.

From our observations, cardboard was only able to hold the heat on the order of seconds, not minutes which makes it inappropriate for the calibration task. The aluminum took the longest to heat up as the aluminum had more mass, thermal capicity, and is the most reflective. As a result, it can hold heat for a long time, but also requires a significantly longer time to heat up than other materials. Wood retained heat for about 3 minutes before it

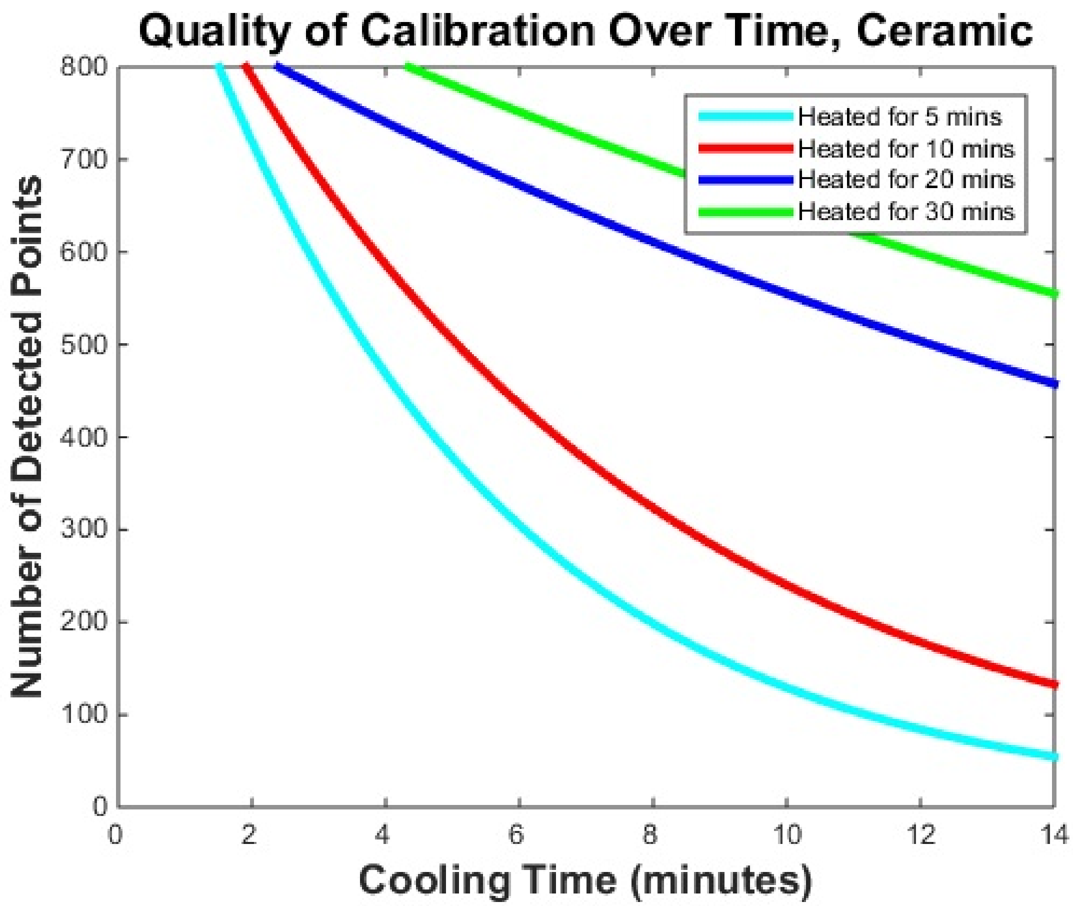


Figure 2.6: Detected points over time in increments of 15 seconds. The calibration pattern was in a static, flat pose facing the camera for 14 minutes. The calibration results would become unreliable under 300 detected points.

became unusable for calibration. The ceramic sample had the best tradeoff between heating up time and heat storage length.

2.4.5 Cross Modality Experiments

In this experiment, a color camera is calibrated with a thermal camera. The color images were resized to match the resolution of the thermal camera (640x480). The same dataset as the first experiment (Algorithm Variation Experiment) was used. There are three trials with trial 1 having 61 images per camera, trial 2 having 46 images, and trial 3 having 63 images. Table 2.2 shows the results.

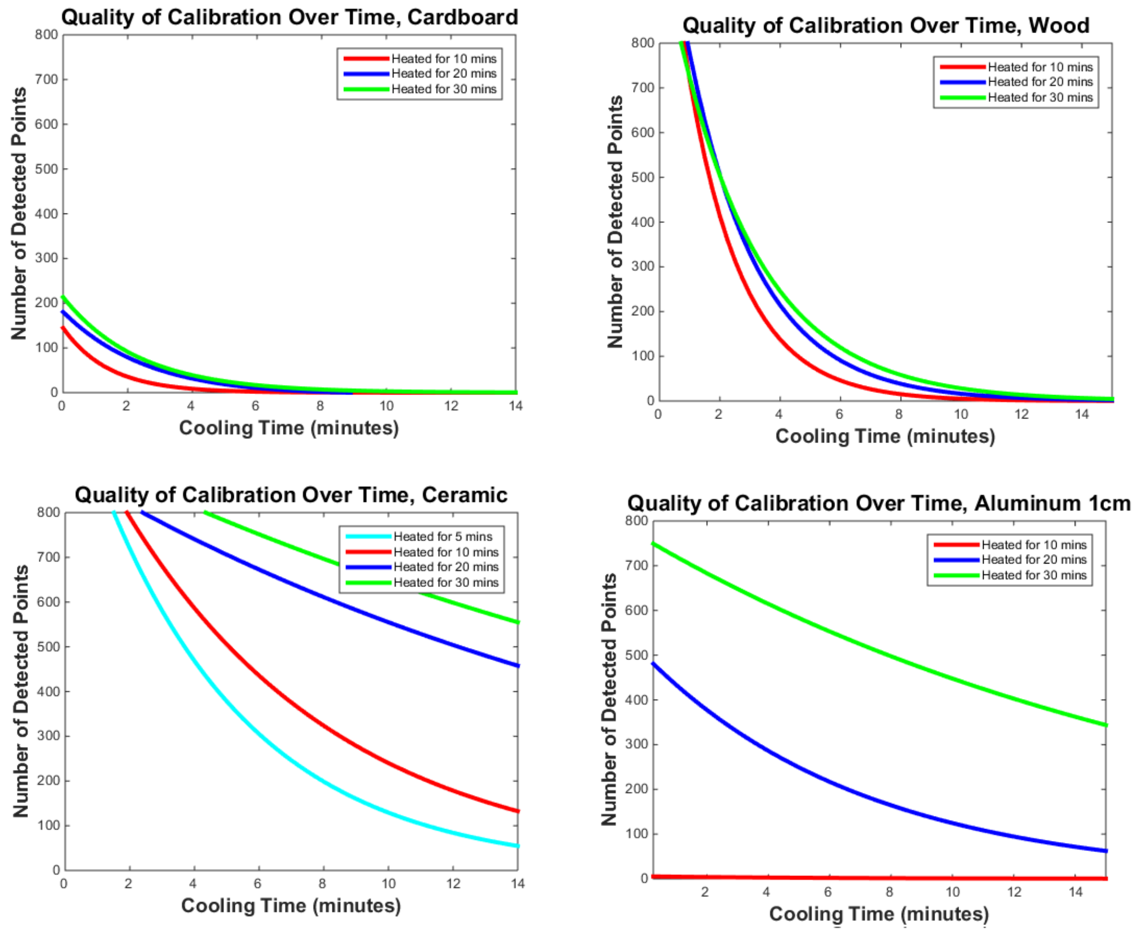


Figure 2.7: Detected points over time in increments of 15 seconds over various material types.

In general across all vairations, the percentage of pairs with detected points was higher with cross modality than with thermal-only. This is most likely due to the color calibration pattern being very sharp with low noise levels as compared to the processed thermal image, as shown in Figure 2.8. All of the other trends are similar to that in the algorithm variation experiment.

2.4.6 Outdoor Experiments and Results

We performed an experiment outside using the sun as the heat source. The weather outside was a mix of clouds and sun, windy, and a temperature of 65° F. The ceramic backed calibration board was placed facing the sun for 15 minutes before calibration. Then the

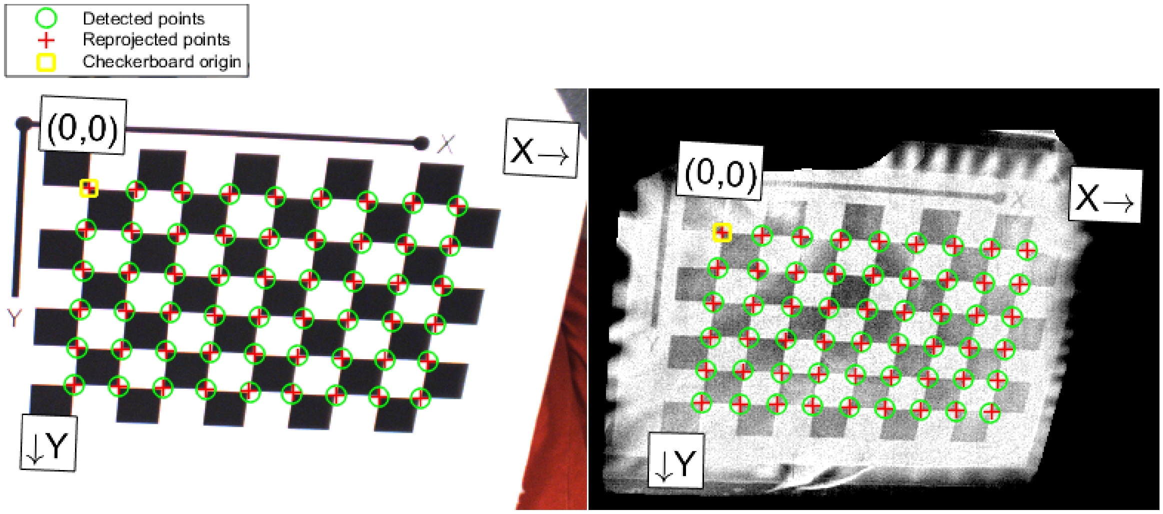


Figure 2.8: Example of cross modality calibration with detected and reprojected points.

Algorithm	% Pairs	#PtsPerPair	RMS (px)
Proposed (Quad iter fit, top hat, gamma=.8)	44.26	**54**	0.39
Quad non-iter fit, top hat, gamma=.8	24.60	54	0.40
No fit, top hat, gamma=.8	6.56	**54**	0.40
Quad iter fit, no hat, gamma=.8	0	0	N/A
Quad iter fit, bot hat, gamma=.8	31.14	**54**	0.41
Quad iter fit, top and bot hat, gamma=.8	19.67	54	0.42
Cubic iter fit, top hat, gamma=.8	40.98	**54**	0.39
Quartic iter fit, top hat, gamma=.8	**45.9**	**54**	**0.38**
Quad iter fit, top hat, no gamma	0	0	N/A
Quad iter fit, top hat, gamma=.5	29.40	**54**	0.45
Quad iter fit, top hat, gamma=1.15	9.84	**54**	0.48
Color imagery, no processing	50.56	54	0.38

Table 2.2: Comparison of variations of our pre-processing method on cross-modality image pairs of color and thermal. The results are averaged over 3 trials of 61, 46, and 63 pairs of images. % pairs measures the number of pairs with detected corners divided by the total number of pairs. The actual number of corners is 54. Note the RMS error is similar for most methods, and any value under 1 is typically acceptable.

board was recorded in different orientations and positions. This entire process was repeated 3 times with trial 1 having 56 images per camera, trial 2 having 64 images, and trial 3 having 62 images. The quality of the reconstruction was measured by the percentage of image pairs with detected corners, the number of corners per image, and the root-mean-square (RMS) reprojection error. These results are presented in Table 2.3. Qualitative results are shown in

Figure 2.9.

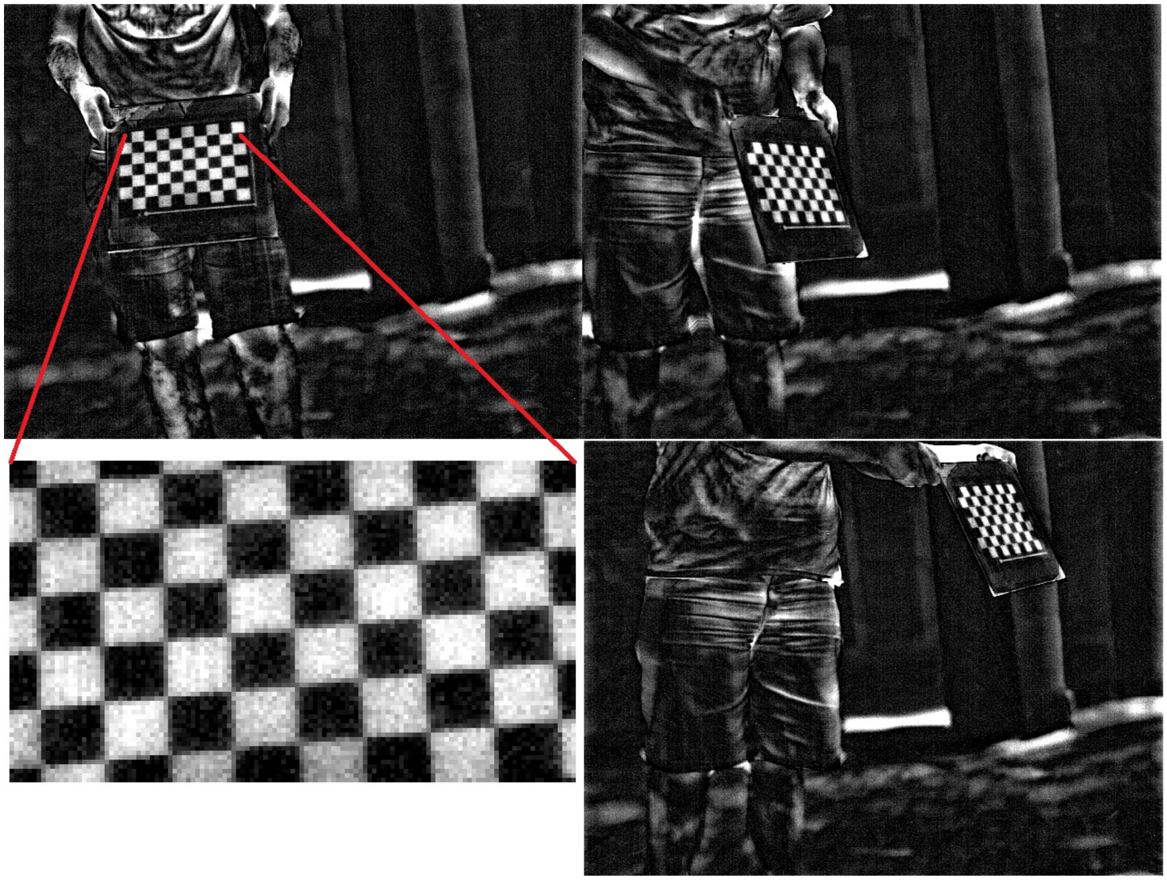


Figure 2.9: Qualitative results of outdoor calibration using the sun as the heat source and the proposed preprocessing algorithm.

The iterative shape fitting precedure does not affect the results for these outdoor experiments. This is because masking out the checkerboard using Otsu's method fails since the calibration board is not significantly hotter than the environment. The entire image is used to fit the polynomial, and since Otsu's method failed, that means most of the image falls within a single intensity distribution. The resulting model has minimal affect on the curvature of the intensity, and did not affect the calibration in any of our tests. The most important operation in the outdoor experiments was the tophat filtering.

Algorithm	% Pairs	#PtsPerPair	RMS (px)
Proposed (Quad iter fit, top hat, gamma=.8)	62.5	**54**	0.46
Quad non-iter fit, top hat, gamma=.8	62.5	**54**	0.46
No fit, top hat, gamma=.8	62.5	**54**	0.46
Quad iter fit, no hat, gamma=.8	0	0	N/A
Quad iter fit, bot hat, gamma=.8	33.9	**54**	0.49
Quad iter fit, top and bot hat, gamma=.8	**66.0**	**54**	0.49
Cubic iter fit, top hat, gamma=.8	62.5	**54**	0.46
Quartic iter fit, top hat, gamma=.8	62.5	**54**	0.46
Quad iter fit, top hat, no gamma	16.5	**54**	0.47
Quad iter fit, top hat, gamma=.5	0	0	N/A
Quad iter fit, top hat, gamma=1.15	0	0	N/A
Color imagery, no processing	50.56	54	0.38

Table 2.3: Comparison of variations of our pre-processing method performed outside with the sun as the heat source. Values are averaged over 3 trials of 56, 64, and 62 pairs of images. % pairs measures the number of pairs with detected corners divided by the total number of pairs. The actual number of corners is 54. Note the RMS error is similar for most methods, and any value under 1 is typically acceptable.

2.5 Conclusion

In this chapter, I described a physical setup and preprocessing technique to make calibration reliable for thermal cameras using off-the-shelf toolboxes. By taping a printed calibration board to a glazed finish ceramic tile backing, I was able to retain heat to reliably detect corner points for 10-20 minutes – much longer than other works reported [118]. The pre-processing technique involved masking out the calibration pattern using Otsu's method, iteratively fitting and then subtracting a quadratic polynomial surface from the intensity, applying top hat filtering, and performing gamma correction to the image. I experimented with different variations of these pre-processing steps to come to our final technique. In three different trials, we were able to successfully and reliably calibrate the thermal cameras using our method. I also experimented with calibration quality over time, cross-modal calibration, outdoor calibration using the sun, and calibration with different material backings. The results demonstate that cross-modality calibration is easier than thermal-only stereo calibration. Calibration outdoors using the sun can mostly uniformly heat the calibration board, which makes the iterative fitting unecessary. Finally, a ceramic backing was the most effective at balancing heating time with heat retention.

Chapter 3

ALIGNMENT OF MODALITIES

Detection of buried or hidden threats is an important military application. Buried or hidden improvised explosive devices (IEDs) can be detected by ground penetrating radar and thermal. The Synchronous Impulse Reconstruction (SIRE) [84] forward-looking radar can detect concealed in-road threats buried up to 0.3m in typical dirt road conditions. However, the Synthetic Aperture Radar (SAR) imagery is noisy and can contain many false positives due to brush, puddles, lightposts, or other objects. Multi-modal sensors, such as color and infrared cameras, can augment the SAR imagery to reduce false positives and help the user identify targets quickly, even in the dark at night. I have available 2 Point Grey Flea 2Gs and 2 Xenics Gobi 640 GigE long wave infrared cameras available, as well as SAR imagery from the SIRE system. All of this information introduces a problem with the amount of diverse information (SAR, color, and infrared imagery), which could overwhelm a user in a chaotic military setting. A solution to this problem is using Augmented Reality (AR) to combine all of the information streams into one image stream.

3.1 Background

Any two images of the same planar surface in space are related by a homography [44]. Typically, image correspondences need to be found to solve for the transformation that relates two images. With a single modality, this can be achieved with typical feature detection and matching across images. However, the problem is more difficult when aligning multiple modalities. [63] uses "a Bayesian framework for generating inter-subject large deformation transformations between multi-modal image sets of the brain". [102] models intensity correspondences across modalities using a joint density function that was tested o multi-modal brain scan images, satellite images, and the Yale Face database for variable illuminations.

[23] uses gradient orientations for multi-modal image registration. [114] uses tracking techniques to find correspondences across time between thermal and color imagery. However, many of these types of approaches rely on the assumption that these modalities have some commonality such as similar gradients or edges. With regards to synthetic aperture (SAR) imagery obtained from radar mounted on a vehicle, there is no such commonality.

Presented in this section is two methods for SAR, thermal, and color imagery alignment. The first method is designed to work on a dynamic scene with a moving sensor platform, but requires additional hardware communication. The second method assumes a static planar scene and sensor platform, and uses simple homography calculations to align the modalities.

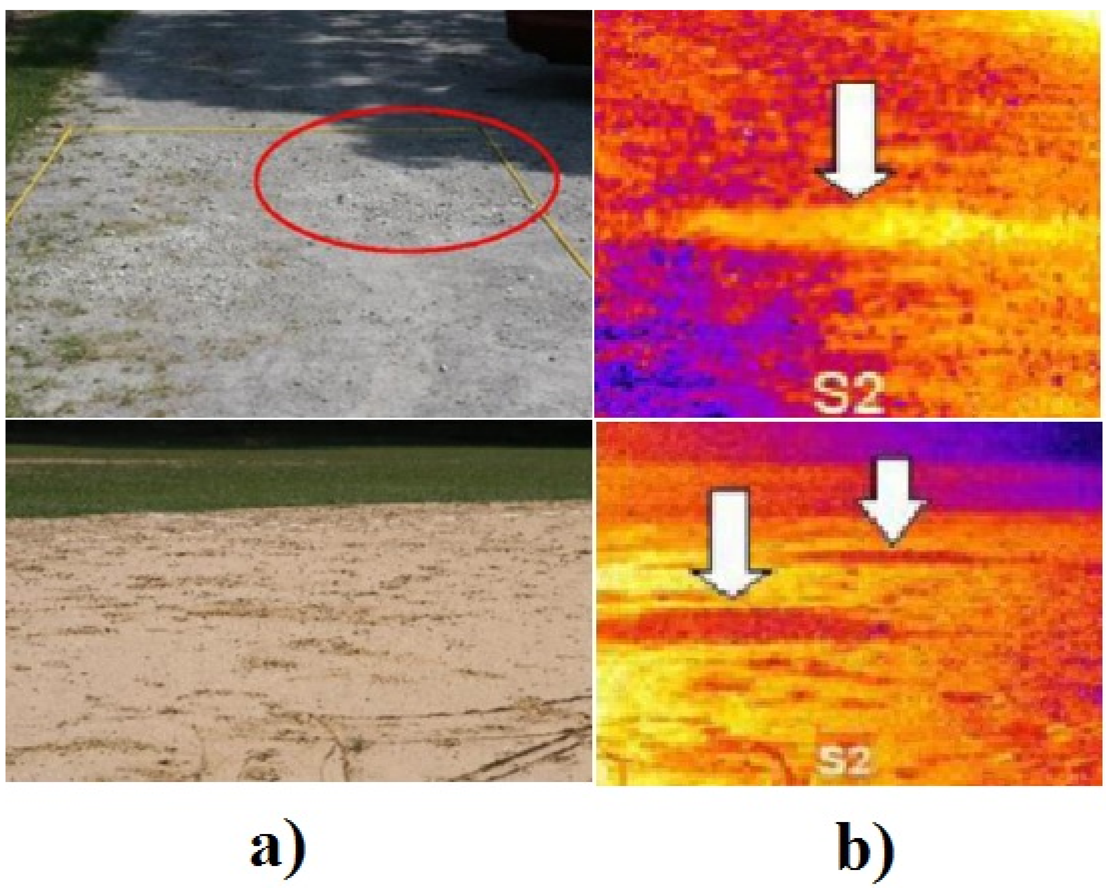


Figure 3.1: Two soil types with buried targets and corresponding LWIR imagery. a) gravel-clay-soil mixture, with a target buried 1 foot deep. b) Packed red-clay-silt soil with various buried targets under 1m.

3.2 Dynamic Scene Alignment and Augmented Reality with SIRE radar

SIRE has been combined with a pan/tilt/zoom camera, a Global Positioning System (GPS), and an Inertial Measurement Unit (IMU), to test its capabilities of concealed target detection in a realistic environment. The thermal modality was not added to SIRE, but the same techniques will apply. As the vehicle moves forward, the system generates SAR imagery as well as displaying a live video stream. This information can be extremely useful to a soldier in the field; however, the information is displayed separately and can be confusing to interpret (see Figure 3.2).

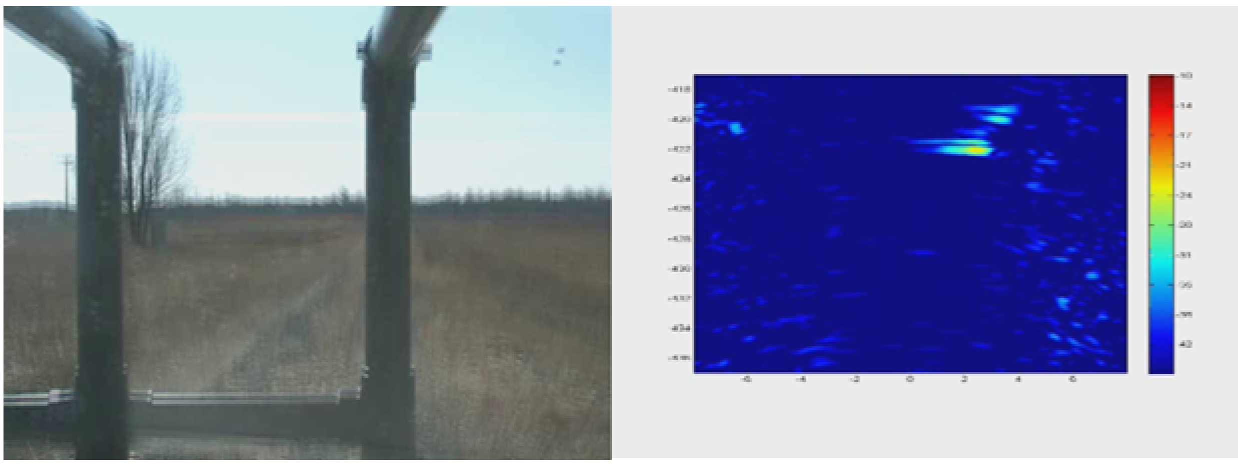


Figure 3.2: Video stream from camera and SAR imagery [77]. Mapping the location of targets in the SAR imagery to the ground in the video stream can be difficult in real time while under pressure.

Augmented Reality (AR) is a way to combine these data streams into one, easy to understand display; AR is a live view of the world that is augmented with additional information – in this case, SAR imagery. There are many modules that make up the total AR system. The system requires communications to the camera to send commands to pan, tilt, or zoom, and to receive a live video stream. It also requires the location and orientation of the camera in UTM coordinates, which is given by the GPS/IMU and a manually measured offset between the camera and GPS. Finally, it requires the SAR imagery and location in UTM coordinates of each pixel. An overview of the entire system can be seen in Figure 3.3.

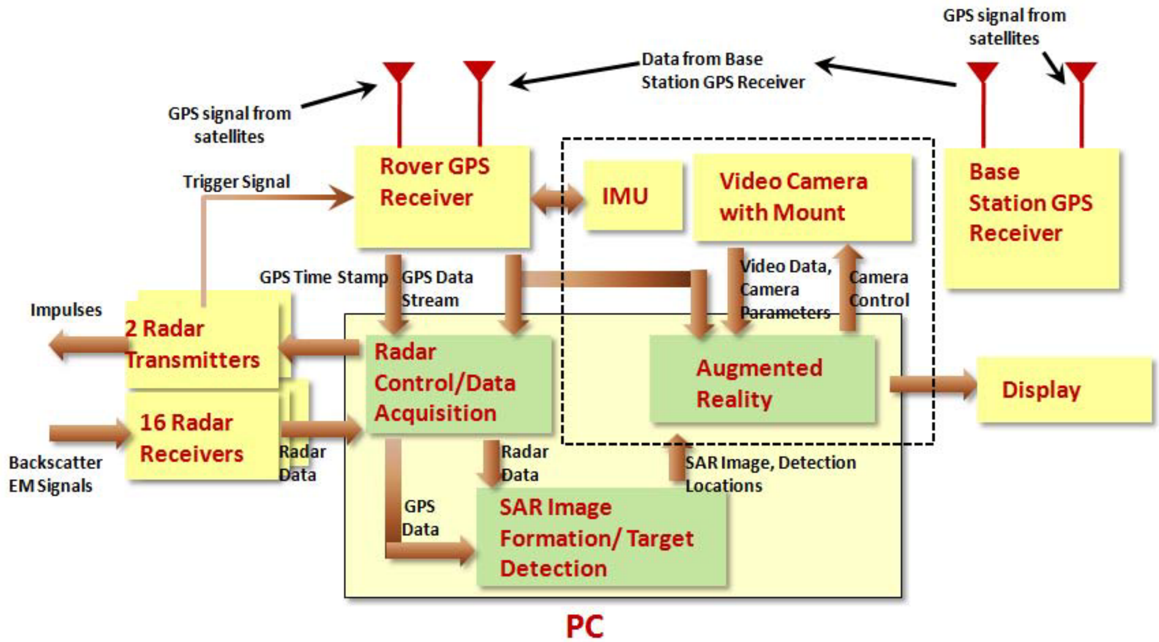


Figure 3.3: Block diagram for entire AR system [77]. The AR system requires the location and orientation of the camera relative to the truck from the GPS + IMU, requires the SAR imagery with the UTM coordinates of each pixel, and requires a live video stream from the Pan/Tilt/Zoom camera.

3.2.1 SIRE and SAR Imagery

The SIRE system has two modes of operation – forward looking mode and side looking mode. For the purposes of this project, only the forward looking mode is considered, which covers a down-range swath between 8m and 32m. More information about the SIRE hardware and image formation process can be found in work [84, 76], but for the purposes of my project the SAR image is output from a black box. It is important to note that all pixels in the resulting SAR image are referenced in the UTM coordinate system (easting and northing), and the entire image has an estimated altitude.

3.2.2 GPS and IMU

To know the location and orientation of the camera in the UTM coordinate system, the system was connected to a GPS and IMU combination from Novatel Incorporated. The GPS is comprised of two modules – a base station and rover. The base station is positioned at a known location in UTM coordinates and sends corrections via radio to the rover to achieve position accuracy within 2 cm. The rover, however, is attached to the vehicle near the camera, and the distance between the rover and camera is measured manually. The IMU is attached to the vehicle and is aligned with the axes of the cameras home position (i.e. facing the direction of travel of the vehicle). Reading data is performed at 20 Hz (the maximum rate) over a serial port.

3.2.3 Pan/Tilt/Zoom Camera

The pan/tilt/zoom camera by RVision Incorporated is mounted on the vehicle, facing forward with the direction of travel, and aligned with the axes of the IMU. Commands are sent and received over a serial port, while the video is sent over USB. The video is also sent at 20 Hz to coincide, while commands are sent asynchronously.

3.2.4 Augmented Reality Transformation Computation

Augmented Reality is achieved by mapping pixels in the SAR image to pixels in the video stream. This mapping can be computed by taking the UTM coordinate for each SAR image pixel and running it through a transformation. Using the pinhole camera model [31], the following equation can be derived:

$$\mathbf{y} = \mathbf{C}\mathbf{x} \tag{3.1}$$

where **x** is the 3x1 vector containing the UTM coordinates of the SAR image pixel, **y** is a 3x1 vector containing the corresponding image coordinates (with the third value being a scale factor), and **C** is the following 3x3 camera matrix for the pinhole camera model:

$$\mathbf{C} = \begin{pmatrix} \alpha_x & \gamma & u_0 \\ 0 & \alpha_y & v_0 \\ 0 & 0 & 1 \end{pmatrix} \tag{3.2}$$

where $\alpha_x = \lambda * f$ and $\alpha_y = f$. This formulation of the transformation from world to image coordinates assumes the optical center of the camera is at center of the world coordinate system, which is usually not the case. Therefore, to move the camera to the center of the coordinate system, two translation and rotation matrices need to be calculated. First, the GPS and IMU data can be used to transform into a coordinate system centered on the GPS rover antennae and aligned with the IMU axes. Next, using manual measurements and the pan/tilt values from the camera, the coordinate system can finally be transformed into camera coordinates.

The entire formulation for transforming SAR image pixels to video image pixels is given by the following equation:

$$\mathbf{y} = \mathbf{C}\mathbf{R_{cam}}[\mathbf{R_{gps}}(\mathbf{x} - \mathbf{T_{gps}}) - \mathbf{T_{cam}}] \tag{3.3}$$

where $\mathbf{R_{cam}}$ and $\mathbf{R_{gps}}$ are 3x3 rotation matrices for the pan/tilt and IMU rotations, respectively; $\mathbf{T_{cam}}$ and $\mathbf{T_{gps}}$ are 3x1 translation vectors from the GPS to optical center and from UTM $(0, 0, 0)$ to the GPS, respectively.

3.2.5 Combining Traditional Offline Camera Calibration with Dynamically Changing Parameters

Camera calibration is the process of finding the intrinsic parameters of the matrix, which, as described in equation 3.2, transforms points in 3D to the 2D image plane. Traditional calibration methods [115, 133] achieve high accuracy; however, these methods are offline and require the sensor parameters remain unchanged. Auto-calibration can handle dynamically changing intrinsic parameters, and is performed online with sequence of images instead of a calibration board [30, 73, 43], but these methods can be less robust and require computing resources while online. Since real time performance is desirable, I decided to calibrate the camera with the traditional offline method with a calibration board.

The software CalLab and CalDe from the Robotics and Mechatronics Center of the German Aerospace Center was used. The methods the software uses are described in other papers[108, 123, 109, 110]. Essentially, given a calibration checkerboard pattern on a board with known distances between the corners, I can create correspondences between estimated 2D corner points and 3D. Using a least squares minimization, the parameters of the camera are estimated. Since this parameter calculation is done by the software, only pictures of the calibration board must be taken and fed into the program. To calibrate, 15 images of the board were taken at varying rotations on varying axes, and used the software to get a reprojection error of less than 1 pixel. However, the intrinsic parameters change as a function of zoom so this method cannot be directly used, as it only gives the parameters at a specific zoom. To overcome this problem, the camera was calibrated at increments of 10% zoom. That is, the calibration process is performed 11 times, and the resulting intrinsic parameters are stored. When the augmented reality transformation is calculated, the intrinsic parameters are linearly interpolated between the two nearest zoom values.

3.3 Static Scene Alignment

With a static scene, the alignment problem is much easier. Only one transformation needs to be calculated and can be applied to subsequent images. Thus, the alignment can be calculated offline using manual correspondences. The modalities – color, thermal, linear, and nonlinear SAR imagery – were transformed onto the color imagery using Matlab's control point selection tool as shown in Figure 3.4. 15 correspondences were used in the alignment of each modality.

3.4 Experiments

3.4.1 Dynamic Alignment Experiments

The AR system was implemented in MATLAB [3]. This includes both the communications to the various hardware devices, and the Graphical User Interface (GUI). The (GUI) allows the user to pan/tilt/zoom the camera, change the transparency level of the displayed radar image, change the color map by changing the decibel (dB) range, and go into auto detect mode. In auto detect mode, the program automatically detects regions of high intensity values (which can be changed through the dB range bars), and draws a white X at the most

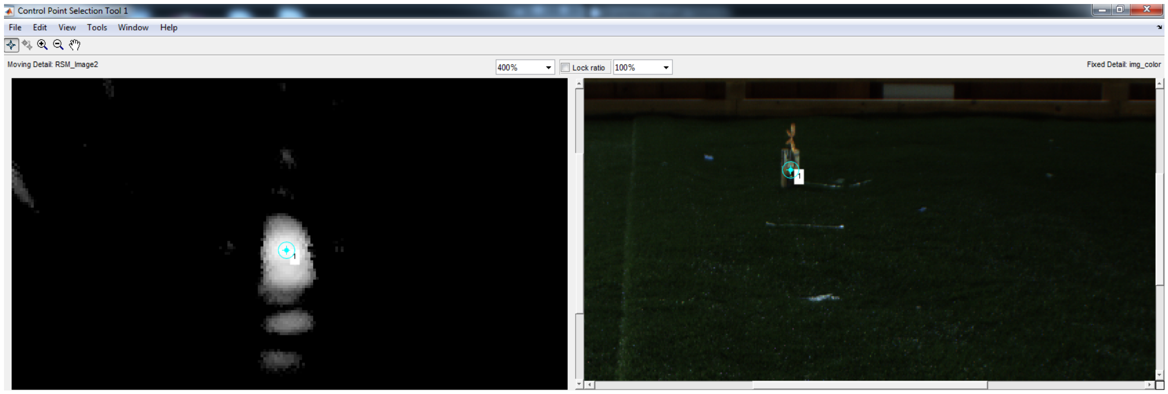


Figure 3.4: Matlab's control point selection tool with nonlinear SAR and color imagery. The peak intensity in the SAR imagery corresponds to the antennae of the device in the color imagery.

intense peak. Real data was collected which includes a SAR image with UTM coordinates for each pixel, GPS coordinates for the vehicle, IMU orientation of the vehicle, pan/tilt/zoom for the camera, the manual measurements from the GPS to the camera, and finally, a video recording. In this data, there are 5 targets that have high intensity values in the SAR image. To test the accuracy of the system, the ground truth location of these targets was measured. The AR system transforms the ground truth locations to image coordinates, and displays the results as gray triangles.

3.4.1.1 Qualitative Results

Below are figures which show the AR and GUI under various conditions. Figure 3.5 shows the results of transforming the 3D ground truth locations of the targets, and Figure 3.6 shows the same but with the transparency of the SAR image set to 0. Figure 3.7 shows the AR program still works under a change in pan, tilt, and zoom from the camera. Finally, Figure 3.8 shows the auto detect mode. Notice that by changing the dB range, the program can detect rocks that are on the side of the road.

3.4.1.2 Quantitative Results

6 video (720x480 resolution) and corresponding GPS/IMU data sets were recorded. Each test a different part of the Augmented Reality pipeline. In the first video, the vehicle is

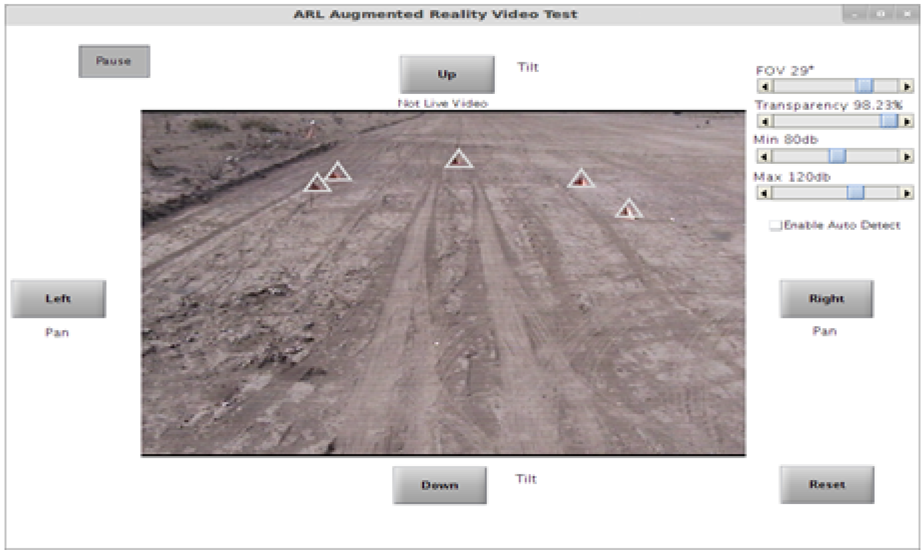


Figure 3.5: AR GUI with ground truth. The transparency of the SAR image is set high so that only the ground truth transformation is displayed. The gray triangles lie on top of the 5 targets in the image.

driven forward and the camera's pan, tilt, and zoom is not changed. This case tests the GPS with minimal changes in the IMU. In the second video, the vehicle was sharply turned in an S-shaped pattern; this tests the IMU and the GPS. The third through fifth videos change the camera's pan, tilt, and zoom, respectively, but the vehicle remains still. The final video tests everything together; the vehicle moves while the camera quickly pans, tilts, and zooms.

To quantify the errors involved with the system, the 3D ground truth of three targets was surveyed, and the center of each target in each 2D video frame was clicked manually. The errors are measured in two ways. First, the error was measured in 2D by calculating the Euclidean distance between the clicked 2D point and the transformed and projected 3D ground truth point. Secondly, the residue in 3D was measured, which is made more complicated since a 2D point relates to a line in 3D, and not a single point as shown in equation 3.4.

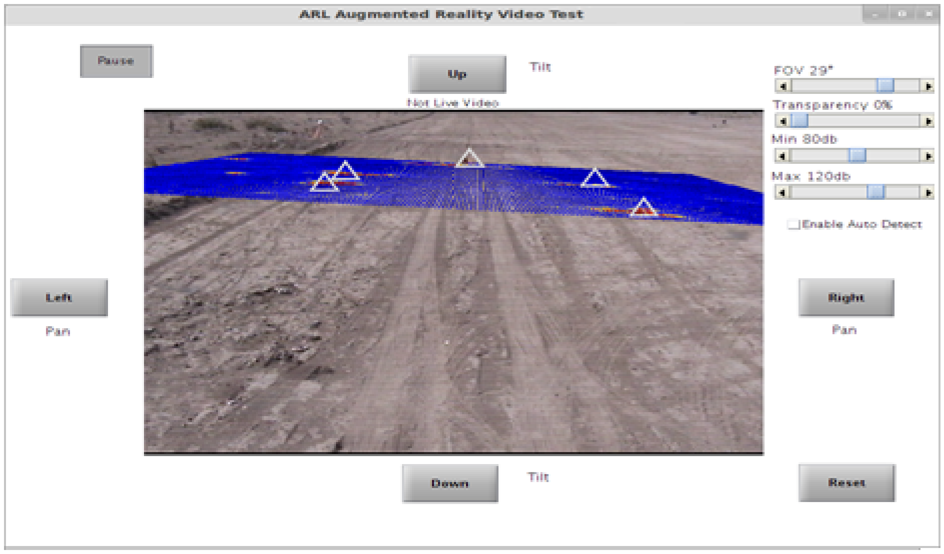


Figure 3.6: AR GUI with ground truth and SAR image. The transparency of the SAR image is set low so that the user can clearly see where the targets in red are.

$$\mathbf{T}^{-1}\mathbf{y} = \mathbf{x}$$
$$\mathbf{y} = \begin{bmatrix} u, v, s \end{bmatrix} \tag{3.4}$$

where $\mathbf{y}$ is the image coordinates with an unknown scale factor, s is the unknown scale factor, and $\mathbf{T}^{-1}$ is the inverse of the transformation described in equation 3.3. To remedy this problem, the Euclidean distance is minimized between the 3D ground truth point and the line. This distance is reported residue in Table 6.2.

Video	Num Frames	2D Avg Error	2D Std Deviation	3D Avg Residue	3D Std Deviation
IMU Straight	203	5.6598	2.9904	0.221	0.1216
IMU Turning	155	9.293	6.8921	0.2831	0.1508
Camera Pan	35	23.3271	14.2387	0.5014	0.2512
Camera Tilt	37	15.2676	6.459	0.3465	0.1311
Camera Zoom	62	41.8876	53.4842	0.2978	0.1336
Everything	155	48.0767	56.5175	0.5218	0.2242

Table 3.1: Results of 6 videos. 2D Error is measured in pixels, while the 3D residue is measured in meters.

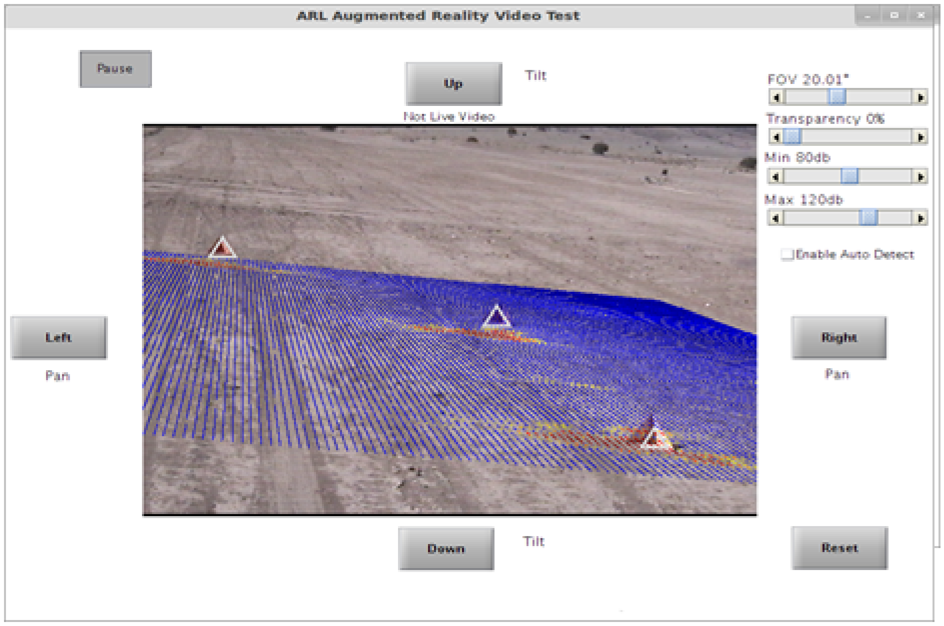


Figure 3.7: AR GUI with pan, tilt, and zoom. The camera is panned, tilted, and zoomed, and the ground truth gray triangles still lie on top of the targets

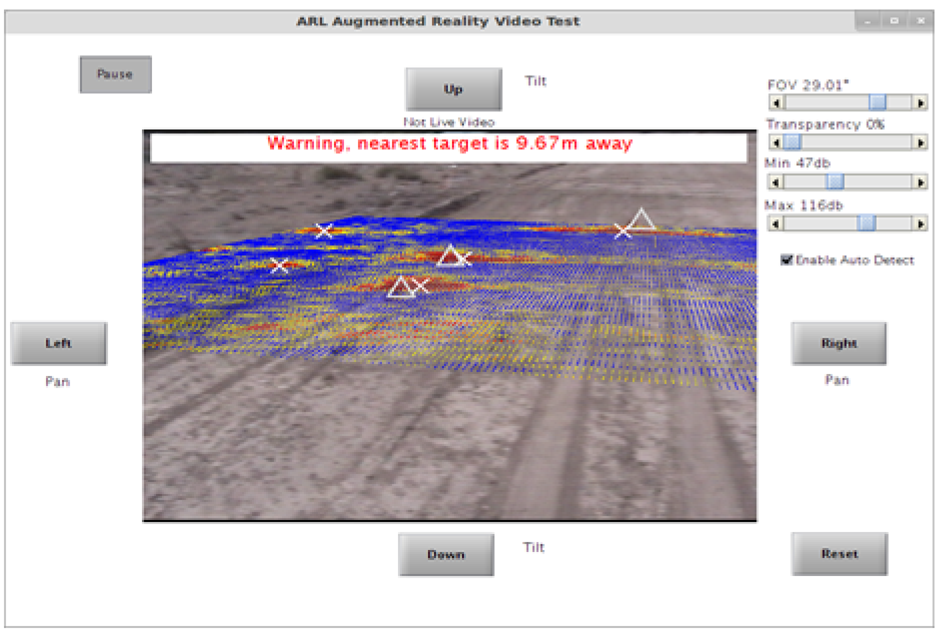


Figure 3.8: AR GUI with auto detect mode enabled. Auto detect mode detects the regions of highest intensity in the SAR image, and warns the user as the vehicle approaches them. Note the dB range was changed to include lower intensity objects, and thus a few rocks were detected on the side of the road.

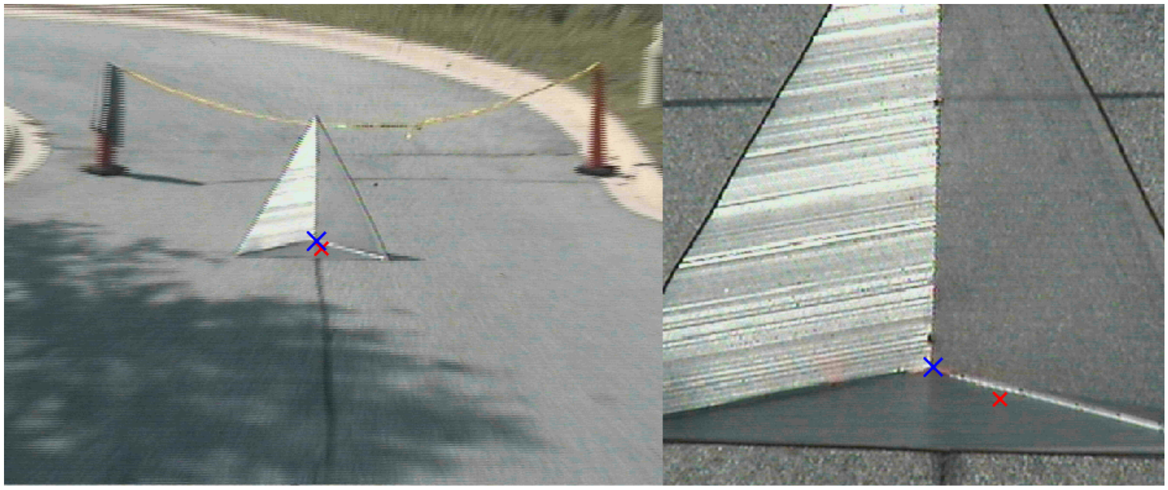


Figure 3.9: Comparison of 2D error while zoomed out and while zoomed in. Notice that a 10 pixel error becomes a 50 pixel error when zoomed in, even though the 3D residue remains the same.

From Table 6.2, there are a few observations. First, when the camera zooms, the 2D pixel error increases, while the 3D residue stays about the same. Suppose, for example, there was a 10 cm error involved with the transformation. When zoomed out, this error appears to be only a few pixels, but when zoomed in, the error is magnified, as 10 cm takes up more space in the image. This phenomenon can be observed in Figure 3.9. This is also the reason why the ”Everything” and ”Camera Zoom” videos have relatively high pixel errors – the camera zooms in at a high magnification, causing the pixel error to be more pronounced. Also, in the videos with camera movement, the error in pixels is not distributed evenly throughout the video. Rather, there are large errors when the camera suddenly moves, and relatively small errors otherwise. This is most likely caused by the camera reporting slightly old values for its pan,tilt, and zoom, and then, a few frames later, becoming up to date. These two sources of errors are tolerable for the application – single high error frames will hardly be noticed by a real time operation, and furthermore, relatively higher pixel error when zoomed in is not a problem since the 3D distances involved are small.

3.4.2 Static Alignment Experiment

For the static alignment experiments, objects that give off a response in multiple modalities were placed in a scene at different orientations and locations. The objects used in this experiment are shown in Figure 3.10. Each object was placed at 9 locations and 5 different orientations for the center location. A Harmonic Real Aperture Radar (RAR) has been developed at the Army Research Laboratory in Adelphi, MD. The radar utilizes 16 receive antennas spaced 3 inches for a total aperture length of 4 feet. A 16:1 switching network allows a single harmonic radar receiver to collect data from each of the 16 receive antennas. A back-projection algorithm [70] is used to form the radar image. The phase propagation for the image forming is outlined in [34, 33]. These details are outside the scope of this dissertation, and the imagery was treated as if it came from a "black box" sensor. This sensor was combined with a Gobi 640 thermal camera (640x480) and a Point Grey Flea 2G color camera (1600x1200) on a multi-sensor platform. The objects were placed approximately 12 meters from the sensor platform.

A GUI was created in Matlab for the purposes of testing the alignment. Different trials can be selected and the weight for alpha blending of each modality can be chosen by the user. A qualitative analysis is shown below in Figure 3.12.

3.5 Conclusion

In this section, multiple modalities including color, thermal, nonlinear radar, and linear radar were aligned together. A dynamic scene with the SIRE vehicle was tested qualitatively and quantitatively in a parking lot and a realistic setting in the desert. A pan/tilt/zoom video camera imagery was aligned with radar imagery from SIRE while the truck was moving and the camera was rotating. The algorithm used differential GPS/IMU data along with multi-zoom camera calibration to give an accurate alignment in real conditions. The methodology would allow for a different modality (e.g. thermal) to be used in place of the pan/tilt/zoom camera without any major changes as long as that sensor could be calibrated. Also tested was a static scene alignment algorithm using a homography transformation. This

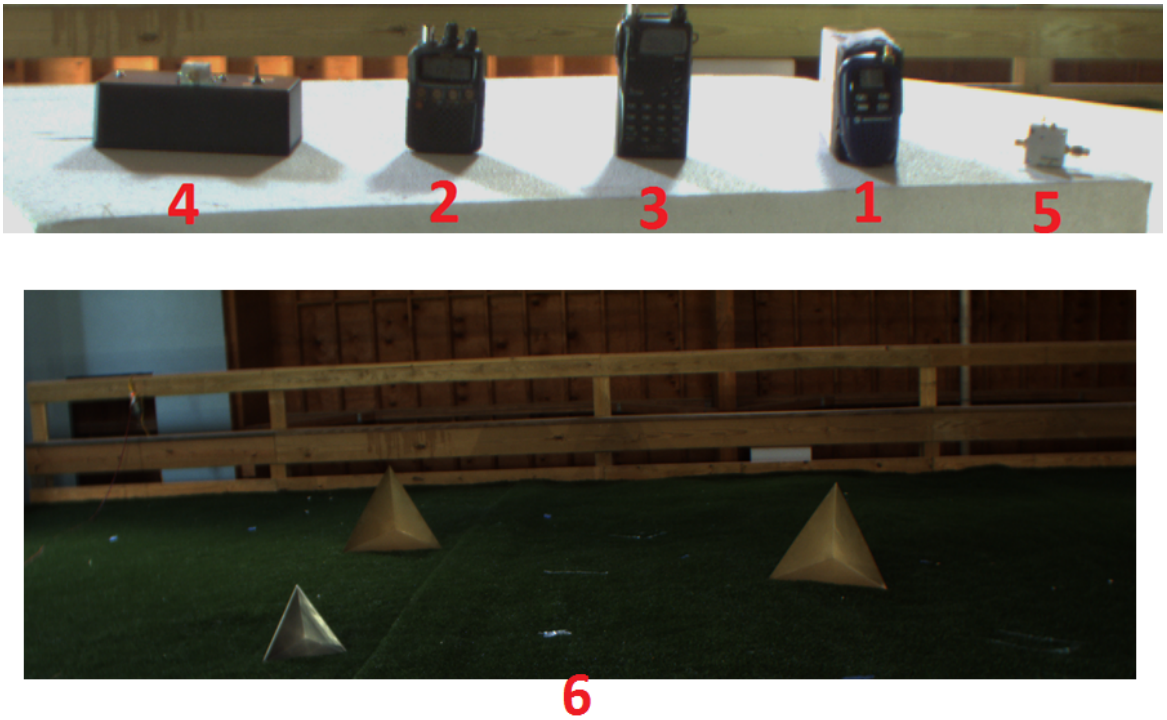


Figure 3.10: Objects used in static alignment. **1)** Small Motorola K7GFV300, Tx + Rx, 460 MHz Cut. **2)** VR120, Rx only, 460 MHz. **3)** Icom IC-T7H, Tx + Rx, 460 MHz. **4)** Mini-Circuits amplifier ZJL-4HG+, Rx only, Output terminated, Power is supplied via batteries. **5)** Mini-Circuits voltage control amplifier, ZFL-2000G+, No power is supplied, output is terminated, input antenna. **6)** Linear radar calibration trihedrals.

was tested on a sensor platform that included color, thermal, nonlinear, and linear radar imagery.

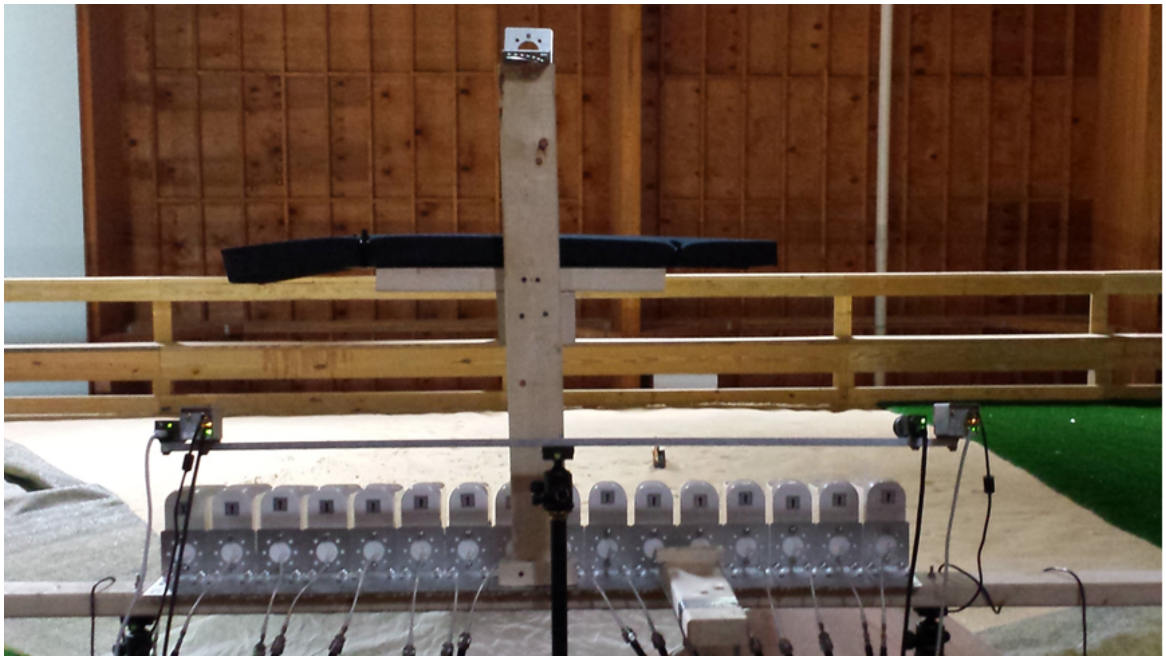


Figure 3.11: Multi-sensor platform with harmonic radar, linear radar, thermal, and color cameras.

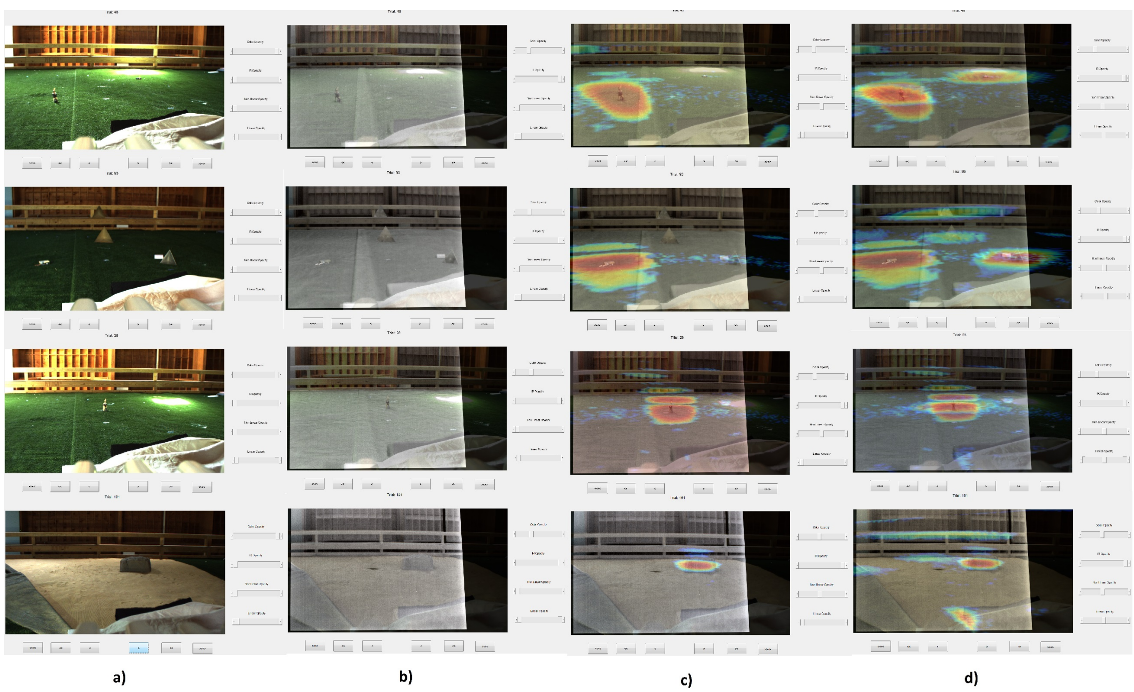


Figure 3.12: Representative qualitative results of static multi-modal alignment. **a)** Color only. **b)** Color and thermal. **c)** Color, thermal, and nonlinear radar. **d)** Color, thermal, nonlinear, linear radar. Note in the last row a buried target becomes visible in the thermal modality.

Chapter 4

MULTI-MODAL DETECTION OF ANOMALOUS OBJECTS

4.1 Introduction

The ability to detect buried or foliage-concealed explosive devices has become a high priority research objective in the last decade. A variety of sensors have been studied for detecting such targets [107, 124, 35, 126, 62, 14, 89, 52]. Vehicle based downward looking ground penetrating radar (GPR) systems are highly promising, but are slow moving with a low standoff distance. Uncooled long-wave infrared (LWIR) cameras, which sense wavelengths of 8-14m and can be thought of as passive temperature sensors, have been investigated for buried target detection [8, 101, 107]. These works found that shallowly buried ($< 10cm$) metal targets can be detected consistently. All objects above absolute zero emit infrared radiation at their surface, which increases with temperature. The emissivity is a property of a material which governs the effectiveness of emitting thermal radiation compared to a black body source at the same temperature. The infrared camera passively sees not only emitted thermal radiation from a source, but also reflected radiation, and transmitted radiation which travels through an object, as seen in Figure 4.1. Metallic objects possess different thermal properties than the surrounding soil. Soils have high emissivity values, while metallic objects are more reflective. The surrounding soil will heat or cool depending on the time of day, and give off a different thermal signature than the undisturbed background soil. Moreover, recently disturbed earth can be seen on an infrared camera, as the soil on the surface is a different temperature than the soil underneath during many parts of the day.

[41] tests the efficacy of using thermal imagery to detect buried IEDs and conclude that infrared improves detectability regardless of the tested soil type (gravel, clay, soil). They further say that "a consensus has emerged that two or more complementary technologies will

most likely be required to improve detectability while reducing false-alarm rates”. Their results show that the disturbed earth is visible in infrared imagery when the target is buried at 1 foot deep in the ground.

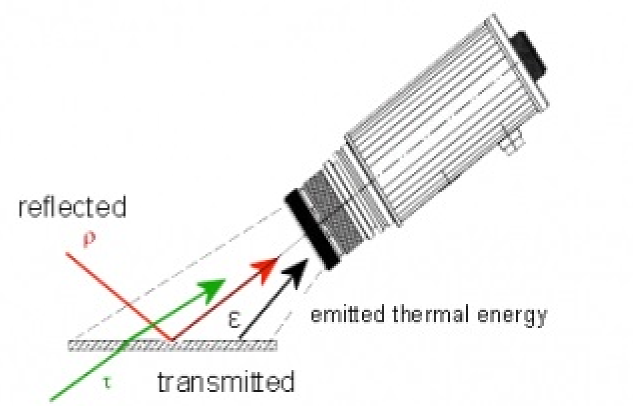


Figure 4.1: Three ways thermal radiation can travel to the infrared camera. Image courtesy of Keller M.S.R.

However, LWIR cameras are low resolution (640x480) and very expensive when compared to visible wavelength color cameras. For above ground targets or above ground clutter, the visible cameras can offer a higher resolution view of the scene and can classify objects as anomalous with a larger standoff distance. Visible cameras can also more easily detect metal targets inside or nearby bushes. Each modality is viewing different information about the scene that can be hopefully be used together to reduce false positives. This can be seen in Figure 4.2.

Gaussian mixture models (GMMs) can be used to adaptively estimate the background of scenes from a camera as in [105]. An infrared camera was mounted on a vehicle in [101] and a GMM was employed to detect buried targets that caused anomalous heat signatures in the scene. Their results show that a GMM model as a stand-alone detector works somewhat well, but has some limitations. The GMM finds all anomalous regions, not just those called by buried objects. The location, size, and shape of the results are not considered by the GMM, and high frequency changes in the scene will be reported (e.g. changes in the road type, channels caused by rain or erosion, brush, etc). They extend this work by including multiple detectors and classifiers in [8]. A pre-screener extracts possible target locations,

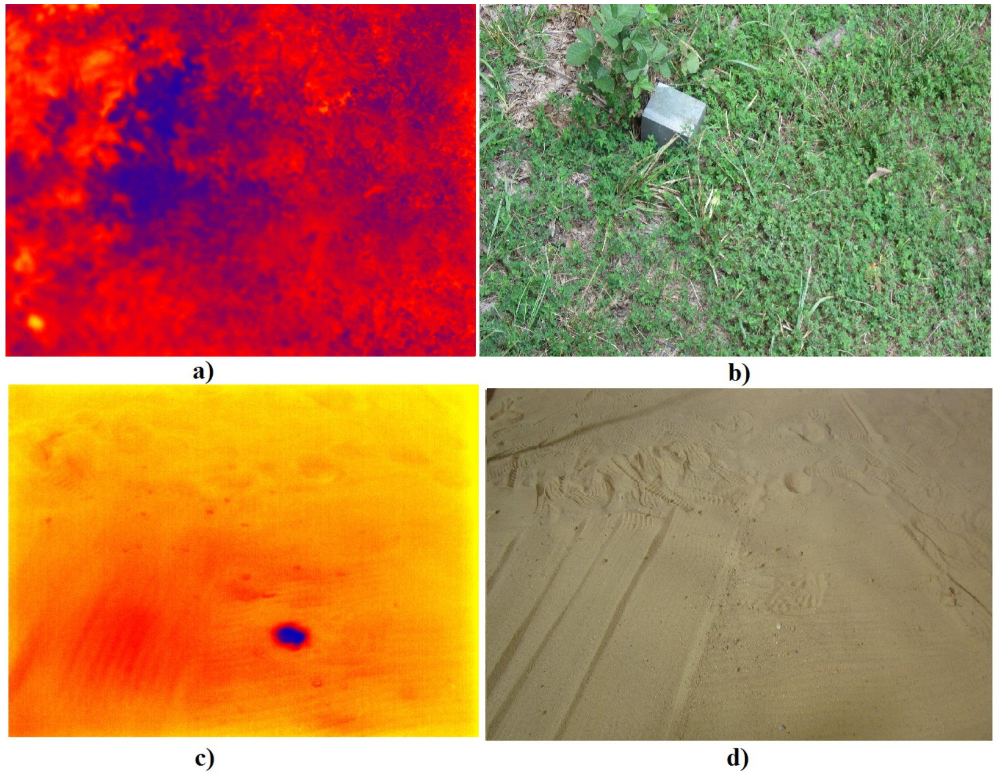


Figure 4.2: Each modality can see targets that are hidden in different ways. **a)**,**b)** are LWIR and visible wavelength imagery of the same outdoor scene, while **c)**,**d)** are LWIR and visible imagery of the same scene in a sandpit.

and the GMM is used to measure the amount of local change. The results in that work show increased detection rates ($\tilde{8}5\%$) on par with what a human manually observes in the imagery. However, the experiments show that infrared alone cannot detect metal objects nearby bushes, but can detect buried targets. Visible cameras cannot detect buried targets, but can detect metal objects nearby bushes. Therefore, multimodal fusion is a possible option.

Multimodal fusion has been studied in a variety of fields. A survey of fusion techniques along with considerations needed for fusion is given in [9]. There are many considerations needed for fusion:

- Level of Fusion. Fusion is possible in early stages of the pipeline (e.g. merging pixels or features), and is possible later in the pipeline (e.g. at the decision level, merging semantic objects). Fusing earlier cuts the amount of learning and model fitting needed, since it is only applied once, but decision level fusion is more flexible and allows for weighting modalities differently based on the current problem.

- How to Fuse. There are a few typical fusion methods. Rule based fusion includes statistical methods such as linear weighting (sum and product), MAX, MIN, AND, OR, and voting schemes. Rules based fusion works best for temporally aligned modalities. Other types of fusion methods include classification based fusion, such as support vector machines or Bayesian inferences.
- What to Fuse: Different modalities can contain complementary or contradictory information, and have different confidence values for a given context. Moreover, not all information (e.g. features) is relevant to a final assessment.

The purpose of this chapter is to implement the GMM detector in [101], which was used for infrared imagery, and fuse information from both LWIR and visible cameras on data collected from a realistic scene with buried or hidden targets. Fusion is performed at different stages: pixel, confidence map, and final decision level fusion. The result of each fusion type is compared against one another, and against the individual modalities in scenes with both hidden and ”out in the open” metal targets.

4.2 Methods

Often times the distribution of pixel values in a scene can be accurately modeled as Gaussian. The idea here is to create a set of Gaussian models of the background pixels in the first frame and then update the background model for successive frames. Multiple Gaussians can model scenes with multiple background colors (e.g. grass, dirt, sky) and can model multiple modalities, which typically do not have the same pixel intensities for the same background object. Once the background is modeled, foreground pixels (i.e. anomalous pixels) are identified and extracted. I also describe different methods for fusing the data between modalities: the initial images can be fused at a pixel level using a rule-based approach, the foreground confidence maps can be merged, and the final thresholded foreground pixels can be merged.

4.2.1 Modeling the Scene with GMMs

Given an image I_i for the ith frame of a sequence, histogram of the pixel intensity values is created. The goal is to obtain a set of k Gaussians $G_i = (\mu_{i1}, \sigma_{i1}, m_{i1})$, $(\mu_{i2}, \sigma_{i2}, m_{i2}), \ldots (\mu_{ik}, \sigma_{ik}, m_{ik})$ that best fits the histogram. Here μ_{ik} is the mean, σ_{ik} is the standard deviation, and m_{ik} is the mass, where mass is defined as the number of pixels within $M\sigma_{ik}$ of μ_{ik}.

For the first frame, G_1 is computed by optimizing the Gaussian mixture model likelihood using the iterative Expectation-Maximization (EM) algorithm as in [71]. This initial model is further improved by removing components with small mass and clamping the standard deviations. In the experiments I performed, the image size and expected intensity value range are constant, so Gaussian components with mass under 3000 are dropped, and standard deviation values are clamped to 1 and 20 for image intensities $\in [0, 255]$. These values were obtained through experimentation and will vary for different image sizes and intensity ranges.

The initial GMM G_1 is updated from frame to frame. A new GMM is calculated for frame 2 and merged with G_1 according to the rule described in Algorithm 1 below. The rule takes each new Gaussian component in the new frame and finds its closest match in the previous frame. The distance between two Gaussian components be defined as

$$d_{pq} = (\mu_p - \mu_q)(\sigma_p + \sigma_q)^{-1}(\mu_p - \mu_q). \tag{4.1}$$

This is similar to the Mahalanobis distance [66], which is a generalized metric for distance of a sample from a model. If the closest match is close enough ($d < 1$), then the two are merged based on a learning parameter ρ. A small value of ρ is more conservative and does not change the original GMM by much. If the closest match is far away ($d > 1$), then the component is added to the final merged model. A history of Gaussians are retained up to $n \geq k$ Gaussians. This process is continued for all other frames as described in Algorithm 1.

Algorithm 1 Update Gaussian Mixture Model

```
function UPDATEGMM(GMM1, GMM2)
    for each component g_{i+1,k} in GMM2 do
        Find min d from GMM1, call it g_{i,j}
        if d < 1 then
            μ_{i,j} = (1 − ρ)μ_{i,j} + ρμ_{i+1,k}
            σ_{i,j} = (1 − ρ)σ_{i,j} + ρσ_{i+1,k}
            Recalculate m_{i,j}
            if m_{i,j} < s then                ▷ s is component size threshold
                Remove g_{i,j}
            end if
        else
            Find empty slot in GMM1 or the slot with the smallest mass
            Add g_{i+1,k} to GMM1 in that slot
        end if
    end for
end function
```

4.2.2 Detection of Anomalous Regions

The next step is to take the GMM and determine if there are any anomalous regions – regions that do not match the background model of the scene. This is performed by computing a foreground confidence map. Let

$$BG_k(x, y) = 1 - min(|I(x, y) - \mu_k|, M\sigma_k)/M\sigma_k \tag{4.2}$$

be the confidence that a pixel is in the background given only the kth Gaussian. Let

$$BG(x, y) = max(BG_k(x, y)) \tag{4.3}$$

be the confidence that a pixel is in the background given the entire GMM. Finally, let

$$FG(x, y) = 1 - BG(x, y) \tag{4.4}$$

be the confidence that a pixel is anomalous given the entire GMM. Note that FG $\in [0, 1]$. Next, threshold FG based on a desired confidence level. Since anomalous pixel values do not necessarily lie next to each other in the image, I apply a spatial constraint. The spatial constraint removes anomalous pixels that are not part of a connected component of at least τ pixels. This gives the final anomalous pixel map.

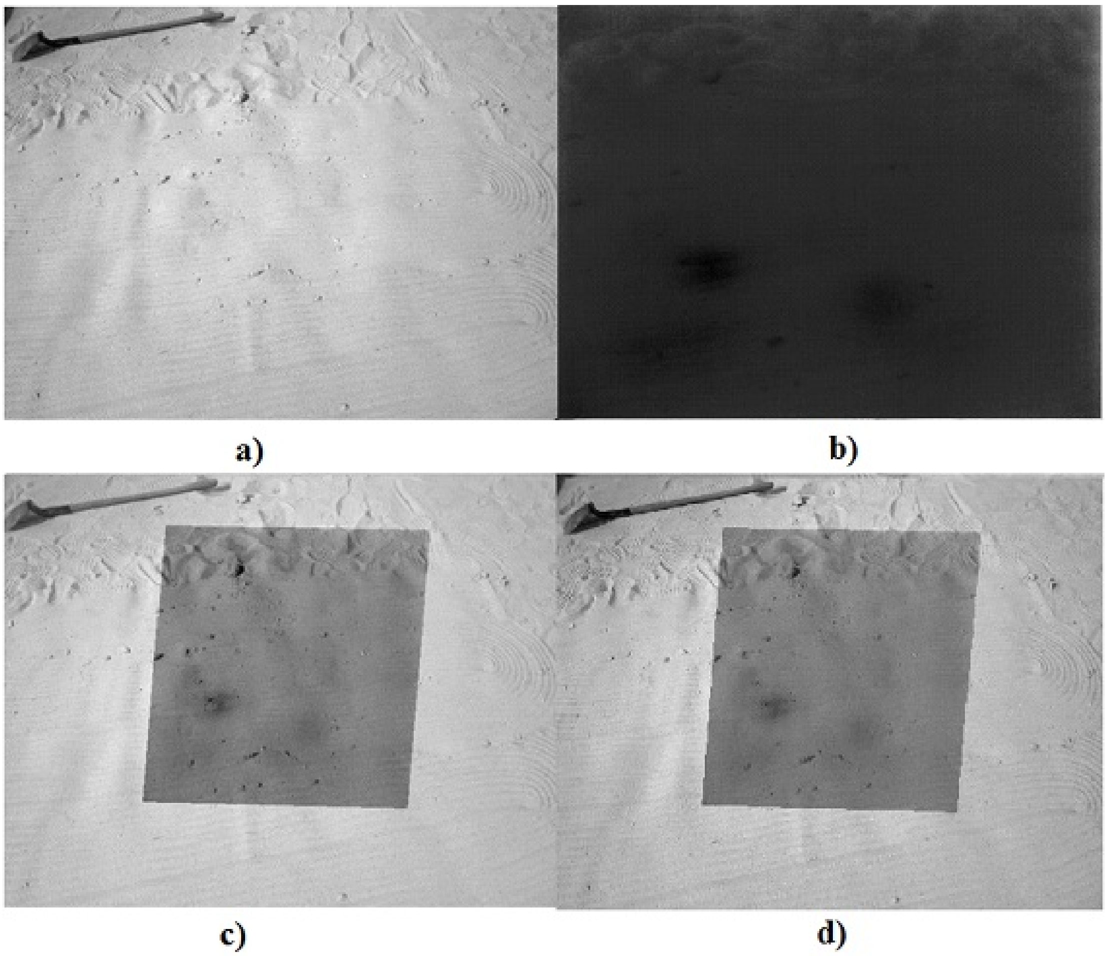


Figure 4.3: Comparison of Modalities and Pixel Fusion. **a)** Visible **b)** LWIR **c)** ADD **d)** MULT.

4.2.3 Fusion of Modalities

As described in [9], sensor fusion can occur at multiple levels and in multiple ways. The first way fusion is performed is at the pixel level. To achieve this, the IR and color imagery are aligned. This is accomplished solving for the best 3x3 projective transformation matrix (9 unknowns) with a set of pixel correspondences using the method described in [36]. Using an automatic method to detect and match pixels such as SIFT [64] did not lead to enough correct matches to reliably solve for the transformation matrix. Instead, 15 manually clicked correspondences are used which led to a correct alignment. Note this transformation need only be calculated once per camera setup. As long as the cameras remain rigid relative to each other, the transformation will be the same. Once the images are aligned, the images are combined with either a simple pixel-wise ADD or MULT operation. Figure 4.3 shows an example of individual modalities and pixel fusion. Both ADD and MULT are implemented and compared against one another. An average operation is not considered here because strong features in one modality can average with background color of the other modality. The effect is strong features become much weaker features after the fusion. The second way fusion is performed is at the confidence map level. The pipeline of fitting GMMs is applied to both the IR and color imagery separately and the confidence maps are merged. The same alignment via projective transformation matrix applies to the confidence maps as well. The maps are combined with a MAX operation. Other operations do not make as much sense. An average or MULT operation will again cause strong confidence to become much weaker since FG $\in [0, 1]$. The final way fusion is performed is at the final binary anomalous region map. The OR operation is considered at the final binary map level for fusion. Fusing at the confidence map and decision levels is similar. One difference between fusing at the confidence map level versus at the final binary image level is the spatial constraint. It is possible the spatial constraint is met only when the confidence maps are fused, and would not be met in the individual decision level maps. This logic would also apply to any further constraints added to the pipeline (e.g. shape).

4.3 Experiments and Results

To test the methods and compare the quality of different fusion methods, I collected two sets of imagery of real scenes. Both sets were collected using a Canon Powershot A1200 4000x2248 resolution visible wavelength camera and a Xenics Gobi 480 640x480 resolution LWIR camera viewing the same scene on a tripod. Synchronization was not needed since the scenes were static; single pairs of images were taken at discrete locations . The targets used were a solid brass sphere of 5cm radius and five aluminum blocks approximately 20x5x5cm.

	Sandpit	**Dirt + Bushes**
M	3	3
s	3000	3000
n	5	5
k	3	3
τ	100	100
Ambient Temp ($^\circ F$)	90	85
Targets above ground hidden in bushes	0	3
Targets above ground not hidden	3	1
Targets buried	3	2
Number of Image Pairs	10	33

Table 4.1: Parameters for two sets of imagery.

The targets were placed both above ground and buried at depths from 1-10cm. The first set of imagery was collected in a sandpit that contained no bushes or any other clutter. The second set of imagery was collected in an area with dirt, grass, and bushes. In both cases the objects were placed overnight. A summary of relevant parameters and conditions are in Table 4.1.

To compare the performance of various fusion techniques, a ROC curve was generated for each technique by varying the confidence threshold value for the foreground map. Ground truth was obtained by manually marking regions of the image. To identify buried targets, an extra image pair was taken with a marker on the spot the target was buried. The results for all the imagery in a dataset was averaged together for a final ROC curve, which is shown in Figure 4.4 and Figure 4.5. There are a few things to note in these results. Firstly, the confidence fused and decision fused methods gave the same curves. Secondly, in the sandpit dataset the ”IR only” curve is the best curve and even beats the fusion methods. This is most likely due to the fact that the IR camera can see both buried and above ground targets, while the visible camera adds nothing but false positives for buried targets. I observed the visible camera labeling shadows of clumps of dirt as anomalous. This trend is not observed in the dirt and bushes imagery. I observed the targets near or inside bushes were almost invisible

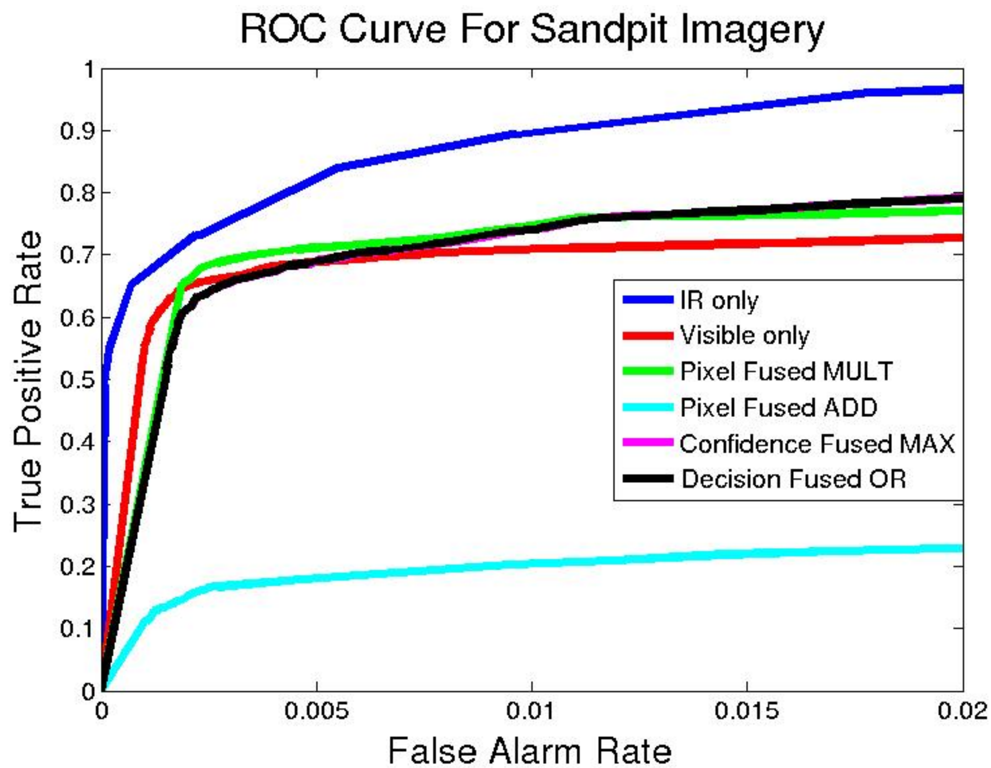


Figure 4.4: Comparison of fusion methods with sandpit imagery.

to the IR camera, while easily detectable in the visible wavelength camera. Since the metal targets were very reflective, they reflected the temperature of the nearby bushes. When ”out in the open”, the targets were easily detectable in the IR imagery. One other observation of note was that although the targets near bushes were invisible to the IR in single images, I noticed the reflections on the aluminum block change from frame to frame. This is a possible cue for using IR to detect metal targets inside bushes in the future.

Out of all the fusion methods, the pixel fusion using the ADD operation was the worst. As seen in Figure 4.3, darker features in the original image become lighter when using the ADD operation, but not the MULT operation. All other fusion methods gave similar results. If the pixel fusion gives comparable results to the decision level fusion methods, then pixel fusion would be preferable as less computing time is needed to get the same results. One technique not considered here is to apply each modality to different parts of the scene, as buried targets are more likely in the center of a road, while above ground targets are more

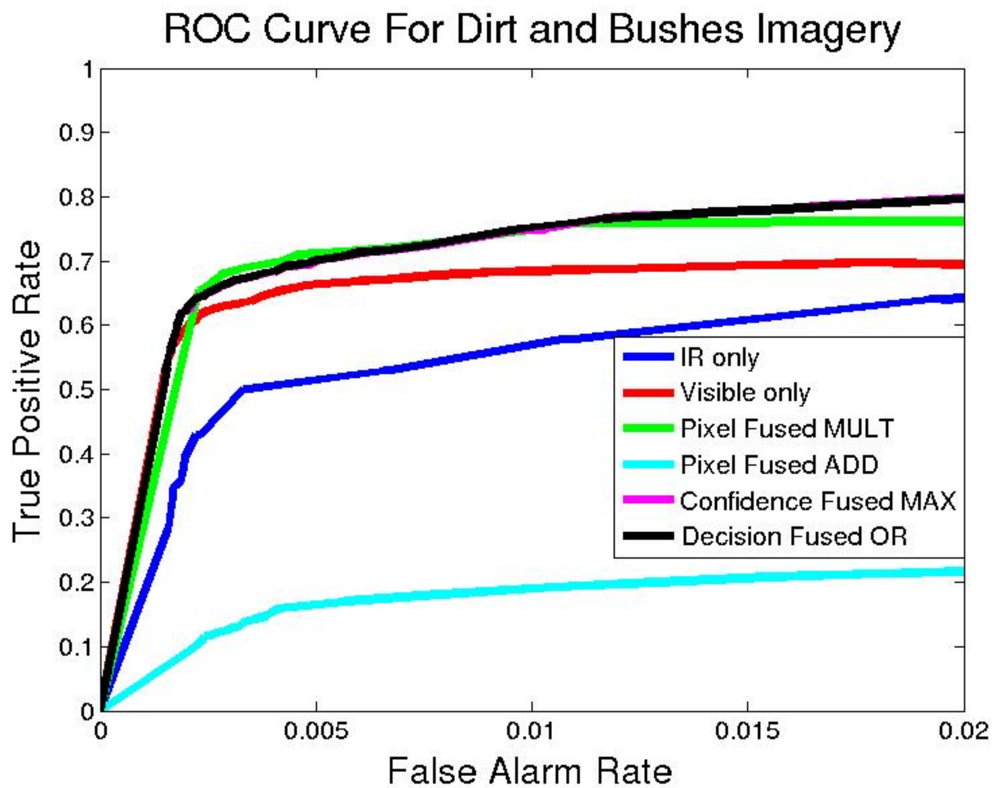


Figure 4.5: Comparison of fusion methods with dirt and bushes imagery.

likely in brush on the sides of the road.

4.4 Conclusion

Gaussian Mixture Models can be used to model the color intensities and can be dynamically updated to adapt to changing environments. The pixels that do not fit the model are anomalous. Both visible wavelength imagery and LWIR imagery can be fused in a multitude of ways. GMMs along with various fusion methods have been applied to scenes with metal targets both above ground and below ground. Results show that the IR was successful in detecting the buried target up to 10cm, and the GMM model was able to pull out the target in almost all cases. However, false positives do remain since the GMM alone only detects anomalous pixels, but not necessarily disturbed earth or dangerous targets. In the future, the GMM will be augmented with shape constraints and pre-screeners as in [8], along with multi-modal fusion techniques. One technique that was not considered here is to apply each modality to a different part of the scene. Buried targets are more likely in the road, while above ground targets are more likely in brush on the side of the road. This lends itself to applying IR to the center of the road, and applying the visible camera to the sides of the road.

Chapter 5

MULTI-MODAL DETECTION OF KNOWN OBJECTS

5.1 Introduction

The detection of electronic devices would help law enforcement agents locate devices whose emissions exceed those permitted by law, allow security personnel to detect unauthorized radio electronics in restricted areas, enable first-responders to pinpoint personal electronics during emergencies such as immediately after an avalanche or earthquake, or detect triggering devices on improvised explosive devices [68, 103]. Harmonic radar exploits harmonically generated returns from electronic targets to aid in their detection. The advantage of nonlinear radar over traditional radar is its high clutter rejection, as most naturally-occurring (clutter) materials do not exhibit a nonlinear electromagnetic response under illumination by radio-frequency (RF) energy [106]. The disadvantage of nonlinear radar is that the power-on-target required to generate a signal-to-noise ratio (SNR) comparable to linear radar is much higher than that of linear radar [56]. Nevertheless, nonlinear radar is particularly suited to the detection of man-made electronic devices, typically those containing semiconductors whose radar cross section is very low owing to their thin geometric profile. The goal of this chapter is to detect known targets using a combination of nonlinear radar and color.

5.2 Background and Related Works

The literature on object detection is vast, but a short overview is provided below. Clustering approaches such as Gaussian Mixture Models have been used to detect objects in a scene [104] by modeling the foreground/background. These methods can be adaptive to account for dynamic scenery, but heavily rely on color information and cannot distinguish similarly colored objects. Point feature matching is a class of methods that match correspondences between reference objects and a scene containing the object [27]. These methods work well on objects without non-repeating texture patterns and non-uniform color. Cascade object detector such as the Viola-Jones [119] detector which consists of stages that learn an ensamble of simple classifiers (weak learners) trained using boosting. These methods work when the target object does not rotate out of plane.

Neural networks are a class of machine learning architectures that provide high accuracy when learning non-linear functions, and have been shown to give high accuracy for object detection [12, 24, 90]. Neural networks are comprised of nodes ("neurons") that exchange information that is weighted and transformed by an activation function. Information is passed until it reaches the output node. Lately, deep neural networks with many layers between input and output have been shown to acheive state-of-the-art performance on large scale benchmarks [29].

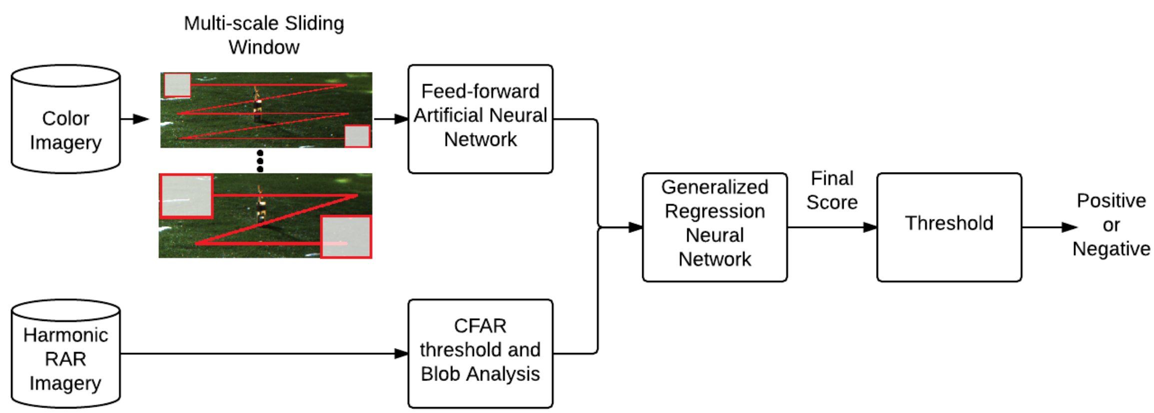


Figure 5.1: Pipeline for Harmonic Radar and Color Fusion

In this work, neural networks are used for both color-only object detection and for score-level fusion between the harmonic radar and color-based methods, as shown in Figure 5.1. The harmonic radar image formation process is outside the scope of this paper, but is discussed in [34, 33]. It can be thought of as a bird-eye view image of intensity values that represent the signal returned from the scene. For detection using only color images, a feed-forward artificial neural network is used with a multiscale sliding window. For score fusion, a generalized regression neural network [18] is used. The contributions in this chapter are the collection of a novel dataset containing synchronously obtained harmonic real aperture radar (RAR) and color imagery, and the comparison of fusion methods on this dataset.

5.3 Methods

5.3.1 Neural Networks for Object Detection

The architecture of the neural network can be seen in Figure 5.2. The input size is 7500 due to the window size of 50x50x3 applied to multiscale images, with image intensity used as the feature. There is a hidden layer with 10 perceptrons using a hyperbolic tangent sigmoid transfer function. The output layer has 2 perceptrons since detection is a binary problem. A softmax function is applied to the output layer to output normalized confidence scores $\in [0, 1]$. These two functions are defined below:

$$tansig(x) = \frac{2}{1 + 2^{-2*x}} - 1$$
$$softmax(\mathbf{x}_j) = \frac{e^j}{\sum_{k=1}^{K} e^k}, j = 1, ..., K. \tag{5.1}$$

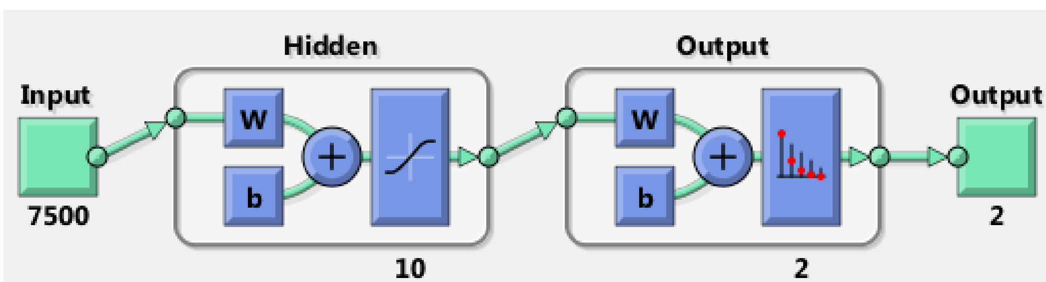


Figure 5.2: Neural network architecture. The input feature length is 7500, there are 10 nodes in the hidden layer, and 2 in the output layer.

Once the neural network is trained, detection is performed via a multi-scale sliding window approach. A confidence score for each patch is calculated and non-maximum suppression is performed across scales. To obtain a bounding box, a threshold is used on the confidence score. Possible threshold values are swept to generate a ROC curve.

5.3.2 Harmonic Radar Detection

Constant false alarm rate (CFAR) detection is a common method for radar modalities to detect objects in the presense of noise or clutter. The CFAR threshold determines the power above which any signal that is returned is considered to be a target. CFAR can be calculated adaptively for dynamic settings to maintain the constant probability of false alarm, but in these experiments the background scenery is static, and therefore a single CFAR threshold is used across all data. After the CFAR threshold is applied, blob analysis is performed to remove small blobs that are likely to be noise. Connected components under a size threshold are removed as positive detections. Possible values are swept for both the CFAR and the size threshold, and the chosen values maximize Youden's J index. The J index is defined as

$$\begin{aligned} J = Sensitivity + Specificity - 1 \\ Sensitivity = \frac{TP}{TP + FN} \\ Specificity = \frac{TN}{TN + FP}. \end{aligned} \tag{5.2}$$

The J index is a single statistic, $J \in [0, 1]$ that captures the performance of a detection or classification system, with 0 indicating the system performs no better than chance and 1 indicating the system performs perfectly (no false positives or false negatives). I chose the J index because multiple studies have shown it is a useful index for comparing ROC curves [74, 11]. The threshold that maximizes J is chosen, which can be thought of geometrically as the maximum value above the chance line on a ROC curve.

5.3.3 Fusion of Modalities

5.3.3.1 Alignment

In this data, the scene is static, and most targets lie along a plane. Therefore I calculate a homography between the modalities using manually obtained correspondences. To identify correspondences, the maximum intensity values in the radar imagery correspond to the antenna connectors on the electronic targets in the color imagery. See Figure 5.3 for an example. For a dynamic scene where radar imagery needs to be aligned with color imagery, see [97] who use hardware solutions combined with camera calibration techniques to calculate the transformation. For example alignment results, see Figure 3.12.

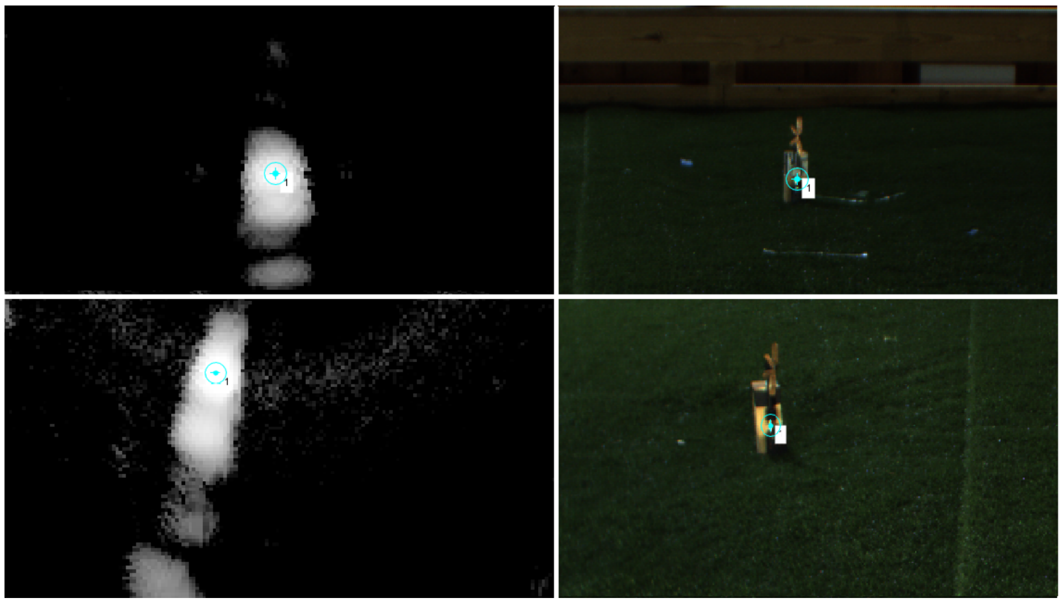


Figure 5.3: Example correspondences

5.3.3.2 Score Fusion via Generalized Regression Neural Networks

I perform fusion at the score-level because of scalability of adding more features or modalities, and because research has found score-level fusion to be "the most effective in delivering increased accuracy" [111]. A generalized regression neural network (GRNN) is trained, whose architecture is shown in Figure 5.4. Generalized regression neural networks can be used to approximate nonlinear functions by using radial basis functions as activation functions on the perceptrons [121]. In this case, the radial basis function is defined as

$$radbas(x) = e^{-x^2}.$$

The output is not a label, but rather a combined score between each modality. The scores from each modality are combined into a single vector per patch, and fed to the GRNN. For supervised labeling, positive patches are patches that overlap with the ground truth more than 50%.

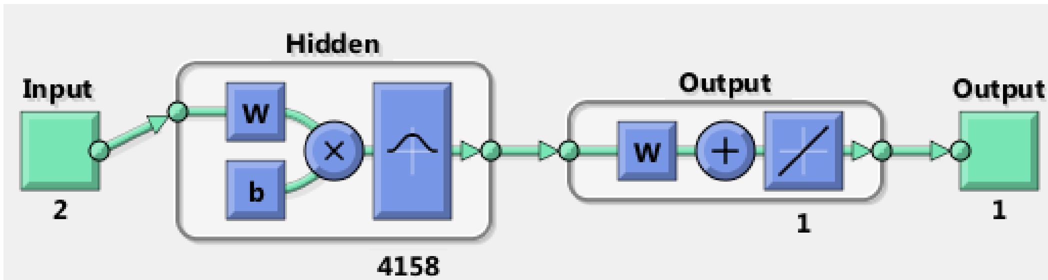


Figure 5.4: Score fusion neural network architecture. The input size is 2, with 4158 nodes in the hidden layer, and 1 in the output layer. The high amount of nodes in the hidden layer is due to the nature of general regression neural networks as seen in [49].

5.4 Experiments and Results

The setup contained two Point Grey Flea 2g cameras using a resolution of 1920x1080. A Harmonic Real Aperture Radar (RAR) has been developed at the Army Research Laboratory in Adelphi, MD. The radar utilizes 16 receive antennas spaced 3 inches for a total aperture length of 4 feet. A 16:1 switching network allows a single harmonic radar receiver to collect data from each of the 16 receive antennas. A back-projection algorithm [70] is used to form the radar image. The image formation process is outside the scope of this paper, but is outlined in [34, 33]. There were five different targets used (small radios or amplifiers): 1) Small Motorola K7GFV300, Tx + Rx, 460 MHz, 2) VR120, Rx only, 460 MHz, 3) Icom IC-T7H, Tx + Rx, 460 MHz, 4) Mini-Circuits amplifier ZJL-4HG+, Rx only with power is supplied via batteries, and 5) Mini-Circuits voltage control amplifier, ZFL-2000G+ with no power is supplied. The physical setup can be seen in Figure 5.6.

For one of the experiments I recorded each target at one of nine different locations, and at five different orientations at the center location. To obtain more data to train the neural network, I recorded 50 additional positions and locations per target with both cameras for an additional 50 captures $*5$ targets $*2$ cameras $= 500$ images. To train the color intensity-based neural network I split these into training and testing sets using 10 fold cross validation. To train the generalized regression neural network the only usable trials were when both radar and color imagery exists. However, the training is patch-based and not at the image level, so there are many examples to train from.

I ran various detector and fusion methods on the data. This includes each modality individually, weighted mean fusion, and neural network fusion. For weighted mean fusion, I used $\alpha * S_c + (1 - \alpha * S_r)$ where the optimal alpha was determined empirically. I created receiver operating characteristic (ROC) curves for each detector by varying the confidence score threshold, as seen in Figure 5.5. The corresponding Youden J-statistics are shown in Table 5.1. From these results, the nonlinear radar performs the worst. This is caused by artifacts such as ground bounces or reflections off of nearby walls. These cause real responses that are almost as strong as the actual target.

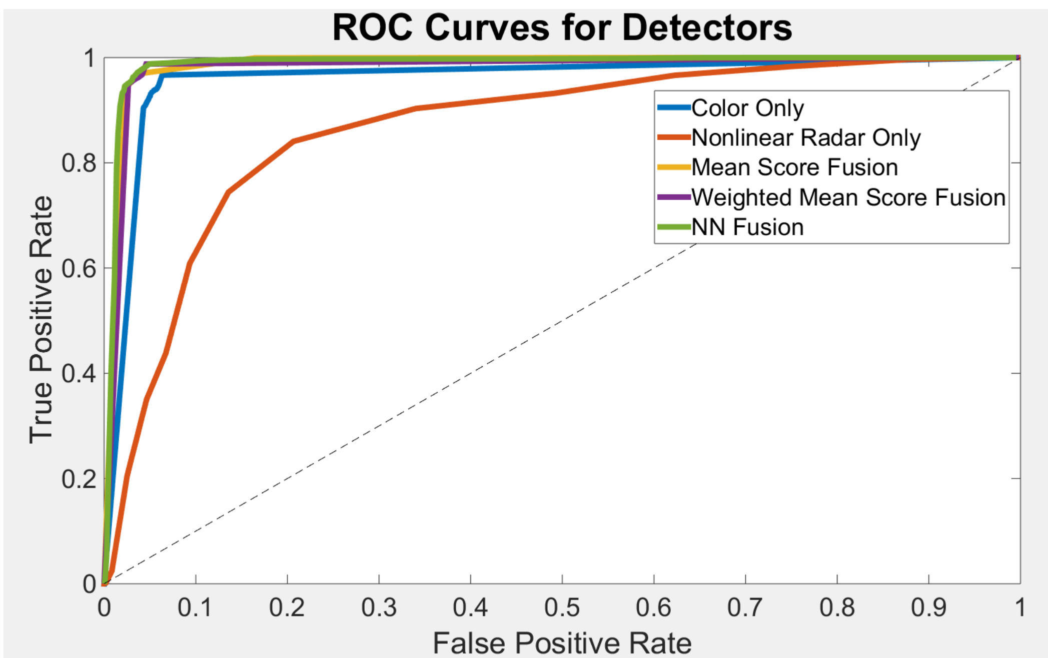


Figure 5.5: ROC curves for various detectors

Finally, as a proof of concept, I recorded 10 additional more realistic scenes to show how color based methods can fail while radar succeeds. Figure 5.7 shows representative samples.

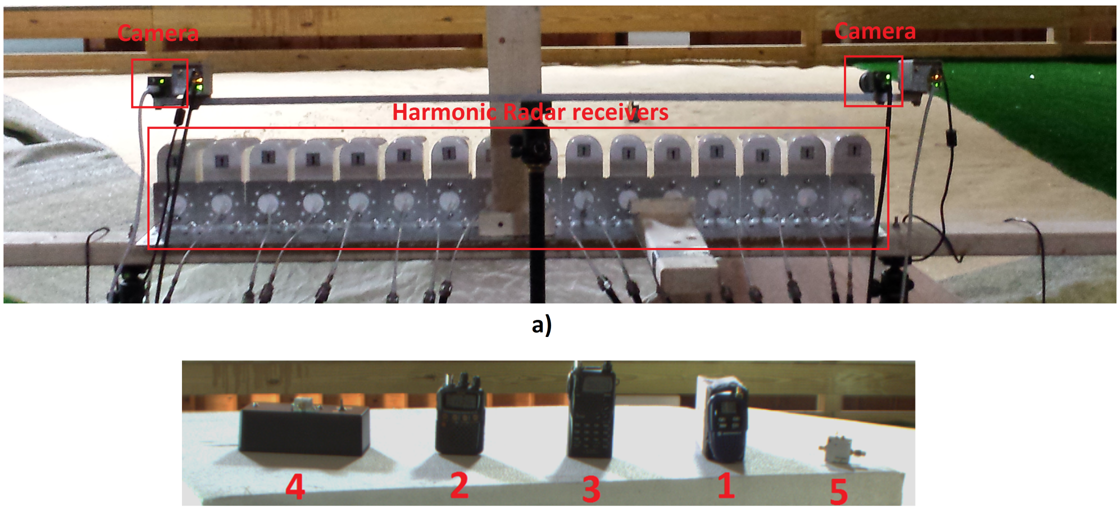
Method	J Statistic
Color only	0.9034
Nonlinear Radar	0.6342
Mean Fusion	0.9298
Weighted Mean Fusion	0.9383
GRNN Fusion	**0.9420**

Table 5.1: Youden J statistic of various detectors.

From these results, harmonic radar is needed to find electronic devices, but it alone produces too many false positives to be used alone in a general setting. After fusion with color, no matter the fusion method, there is an increase in performance. The weighted mean fusion and GRNN produced the best results for score fusion.

5.5 Conclusion

Nonlinear radar can be used to detect electronic devices even through obstructions. However, the radar imagery is noisy and contains artifacts due to ground bounces and coupling between the receiver and transmitter. Different classification schemes performed similarly in my tests. When combined with color imagery for detection of small electronic objects on the surface of the ground, my results show that fusion increases the performance with automatic detection.



a)

b)

Figure 5.6: **a)** Image of hardware setup. Note a pair of thermal cameras appear in the image, but were not used in this paper for fusion. **b)** Targets to detect (small radios and amplifiers). 1) Small Motorola K7GFV300, Tx + Rx, 460 MHz, 2) VR120, Rx only, 460 MHz, 3) Icom IC-T7H, Tx + Rx, 460 MHz, 4) Mini-Circuits amplifier ZJL-4HG+, Rx only with power is supplied via batteries, and 5) Mini-Circuits voltage control amplifier, ZFL-2000G+ with no power is supplied.

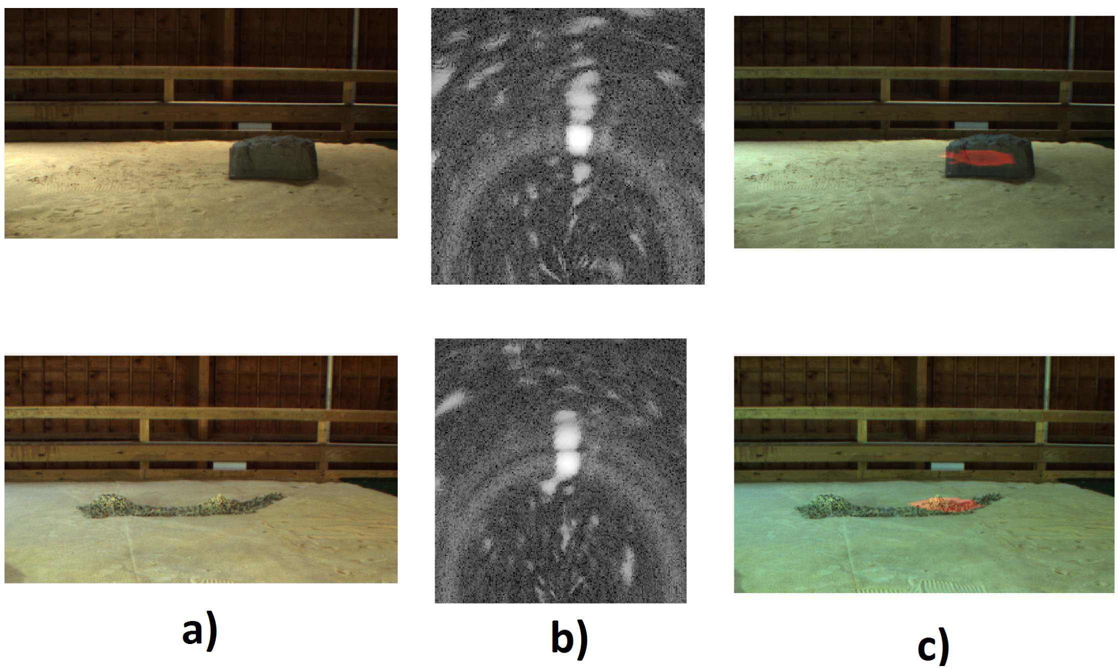


Figure 5.7: Realistic scenes to showcase the need for radar technology. The target is camoflauged or covered up completely, but the radar can still see it. **a)** The color image. **b)** The RAR image. **c)** The fused image after homography calculation and after thresholding+blob analysis.

Chapter 6

MATERIAL CLASSIFICATION USING THERMAL AND COLOR IMAGERY

6.1 Introduction

Automatically classifying materials impacts real world applications such as recycling [45], mineralogy [48], and robotics [51]. Computer vision methods typically use standard RGB camera imagery and rely on texture and lighting cues to distinguish the different materials. Appearance based classification is challenging due to the variety of colors and illuminations. Objects of different materials that have the same color intensity response can be almost impossible to distinguish. In this work the electromagnetic spectrum is considered between 8-14m with a long wave infrared (LWIR) camera, i.e. a thermal camera. I study thermal properties and how water permeates through different materials.

LWIR cameras detect infrared radiation, which is emitted by all objects above absolute zero according to Plancks black body radiation law [88]. The LWIR camera sees not only the emitted thermal radiation of a source object, but also reflected and transmitted thermal radiation. Emissivity is a property of a material which governs the effectiveness of emitting thermal radiation compared to a black body source at the same temperature. Metals have very low emissivity and high reflectance, while woods have high emissivity. Thus many metals appear mirrorlike in the LWIR imagery.

When water comes into contact with an object, the permeation behavior changes based on the material. For example, in wood the water follows along the grain of the wood and is jagged in appearance, while in metals the water stays above the surface and hardly moves if the surface is flat, and in paper materials, the water spreads radially.Permeation behavior is used as a cue for classification. A 3D model is constructed of the water permeation pattern for each class of material. Each frame of a video sequence of the water permeation is treated as a 2D slice of a 3D model. This model generalizes features of the permeation pattern such as rate of permeation and shape characteristics. Another cue used is the thermal heating and cooling cycle. I show that different materials heat and cool at different rates when placed under a heating lamp. The materials used for classification are shown in Figure 6.1.

The chapter is organized as follows. Section 6.2 gives previous works for material classification. Section 6.3 details the proposed method of feature extraction and model learning. Section 6.4 shows the data collected and discusses the results of the classifier. Section 6.5 concludes the paper and discusses future work.

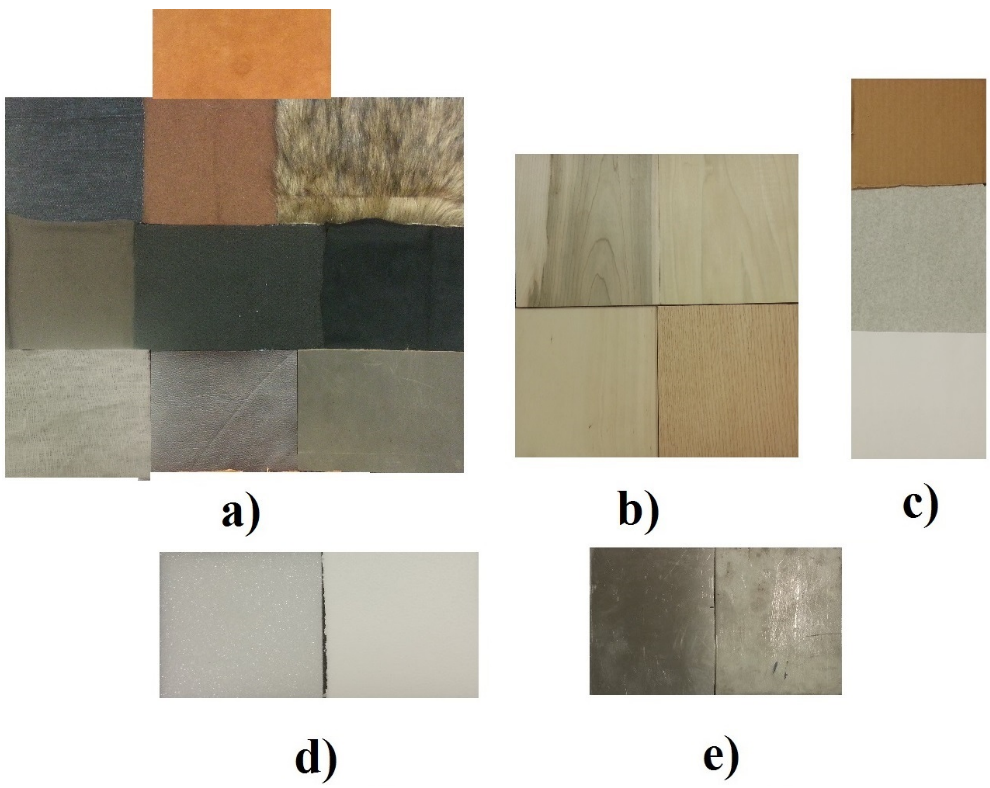


Figure 6.1: Materials used for classification. **a)** Cloth: suede, denim, wool, synthetic fur, cloth, felt, polyester, linen, synthetic leather, real leather. **b)** Wood: maple, poplar, birch, oak. **c)** Paper: corrugated cardboard, paper towel, printing paper. **d)** Plastic Foams: closed-cell expanded polystyrene, closed-cell extruded polystyrene. **e)** Metals: aluminum, steel.

6.2 Background and Related Works

Many previous works consider material classification with a standard RGB camera. Color and texture information were extracted in [26]. [39] uses a Bidirectional Reflectance Distribution Function (BRDF) as a feature for per-pixel classification. [?] uses a Bidirectional Texture Function (BTF). Both use coded light illumination in an LED dome. Light polarization was used in [19, 125]. BRDF slices were used in [120]. Visible spectral reflectance (400-720nm) was used in [50]. In most of these works, the lighting was very controlled to give cues for the classification. In this work, the lighting is less important compared to thermal properties for classification.

Thermal imagery has not been heavily studied for material classification. [91] uses near infrared (NIR) to get a more intrinsic image of the material sample. [85] uses mid wave infrared for paper and board identification for food packaging. [55] measures thermal conductivity of materials using tactile feedback; a robot touches the material with a probe. LWIR can easily detect water damage in buildings and can detect permeation in materials [127]. To the best of my knowledge, this is the first work to use long wave infrared cameras to obtain image-based thermal properties for material classification.

6.3 Methods

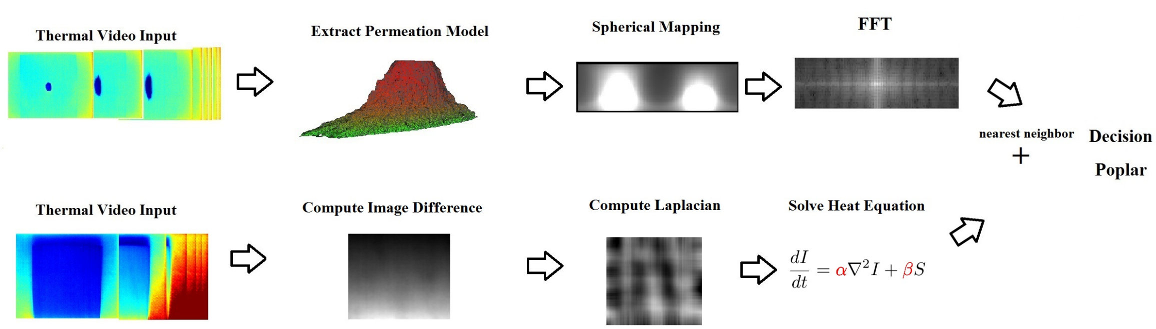


Figure 6.2: The proposed method consists of two types of features – water permeation and a heating/cooling cycle. For water permeation, 3D model is extracted called "CHAMP" which describes the water permeation rate and size. For heating and cooling, a variation of the heat equation [22] is solved for the constant parameters. Note that the permeation model is a mesh for display purposes, and the color refers to time (red = start). The FFT image is in a log-2 scale.

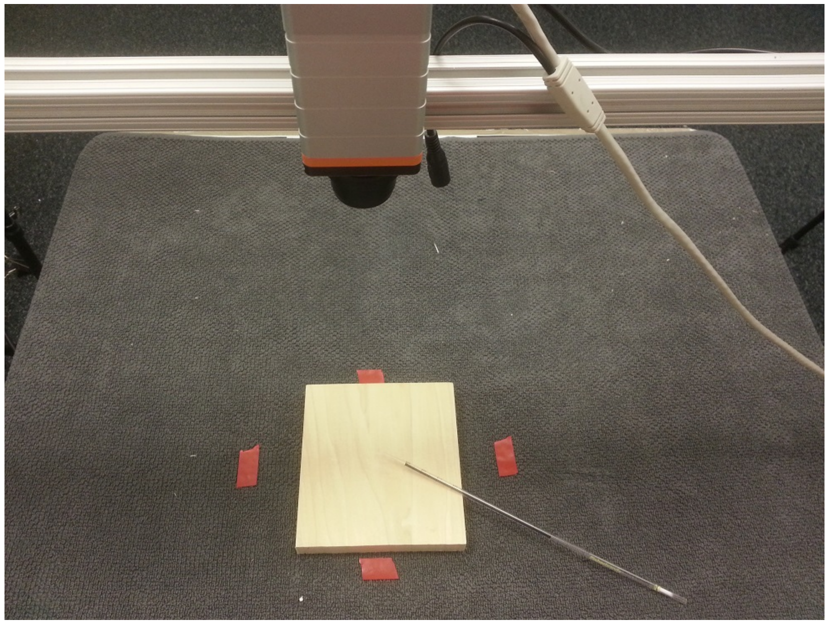


Figure 6.3: Camera setup for water permeation experiment. All materials were roughly centered in the image, and a pipette was used to control the amount of water.

An overview of the proposed methods can be seen in Figure 6.2. Water permeation behavior is extracted by computing a characteristic model of permeation as discussed in Section 6.3.1, and is further transformed into the FFT of a binned spherical map for comparison. I developed two heating/cooling features. The first is extracted by sampling small patches over time to obtain a temperature curve as discussed in Section 6.3.2.1. The second is extracted by solving the heat equation for unknown constant parameters in Section 6.3.2.2.

6.3.1 CHAMP - CHAracteristic Model of Permeation

There are a few interesting features that can be obtained for water permeation such as the rate of permeation of the water into the material and the shape characteristics of the permeation. These features are generalized by creating a 3D model which I call the CHAMP (CHAracteristic Model of Permeation). To compute the CHAMP, I first define a 2D indicator function f for a material as

$$f_i(X, Y) = I_i(x, y) < \tau_f, \tag{6.1}$$

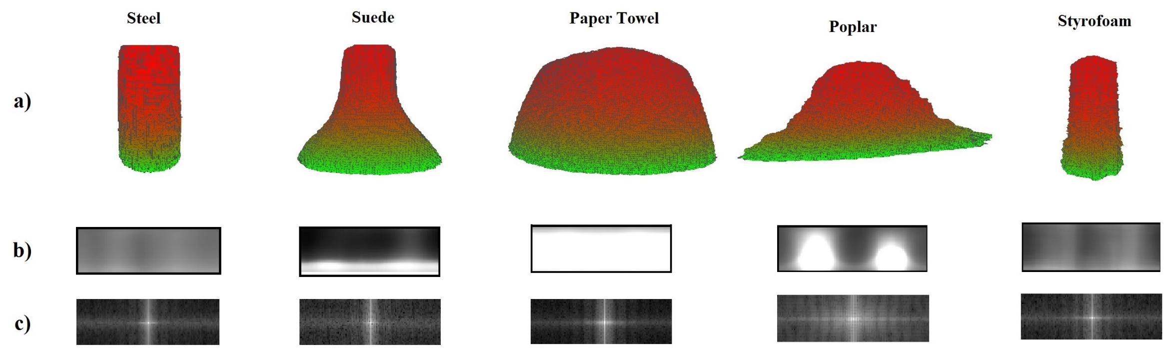


Figure 6.4: Comparison of select few materials from different coarse-grain classes. **a)** CHAMP. **b)** Binned spherical map. **c)** FFT of binned spherical map.

where τ_f is a material specific threshold value and i is the ith frame in the video sequence. The material threshold value can be calculated using Otsu's method [78], assuming a bimodal distribution of intensity values. The indicator function gives a value of 1 for points inside or on the CHAMP, and 0 for points outside the CHAMP. It is possible to use more complicated methods for calculating the indicator function, such as active contour models [17], but for most of my data the difference in temperature between the water and background is bimodal and significant enough that a simple threshold suffices. Moreover, the active contour model would lose some finer details depending on the snaxel resolution.

Once the indicator function is calculated, the boundary of the model can be quickly estimated using morphological operations [25, 72] as in

$$\Omega_i(X,Y) = f_i(X,Y) - f_i'(X,Y), \tag{6.2}$$

where $f_i'(X,Y)$ is obtained by eroding f using a small structuring element. To create the 3D model, simply concatenate the boundary $\Omega_i(X,Y)$ for each slice i along the Z dimension. The model can be "capped" by using the indicator function for $i = 1$ and $i = N$ for an N frame video sequence, instead of the boundary $\Omega_i(X,Y)$. That is

$$CHAMP(X,Y,i) = \begin{cases} f_i(X,Y), & \text{if } i = 1, N. \\ \Omega_i(X,Y), & \text{otherwise.} \end{cases}, \tag{6.3}$$

The CHAMP can be understood by visualizing each image in a video sequence as a 2D slice along the Z dimension. It implicitly contains the shape of the water permeation as well as the growth rate (i.e. curvature of the model).

6.3.1.1 Comparing CHAMPs with Binned Spherical Mapping

Once the models are created, a metric to compare the models is needed. A Fast Fourier Transform (FFT [117]) is computed of a modified binned spherical mapping of the CHAMP. This is a robust method since it is invariant to rotations and translations between CHAMPs and is fast to compute.

A spherical coordinate system is a coordinate system where points are represented by 3 parameters: the radial distance from the origin, the polar angle measured in the zenith direction, and the azimuth angle orthogonal to the zenith direction. Before performing this mapping, the CHAMP is centered around its centroid. Next, the Cartesian (X, Y, Z) coordinate system is mapped to (r, θ, ϕ) in the spherical coordinate system using simple trigonometric equations.

These spherical points are binned into a 2D histogram image. The intensity values of the histogram image are the r values multiplied by $cos(\phi_k)$, the rows are varying θ, and the columns are varying ϕ. This is performed by

$$\begin{aligned} SPH(x, y) &= cos(\phi_k) * avg(r_k) \\ \{k \mid \theta_k \pm \epsilon &= x * bin_x - \pi, \\ \phi_k \pm \epsilon &= y * bin_y - \frac{\pi}{4}\}, \end{aligned} \tag{6.4}$$

where bin_x and bin_y are the desired bin size. Each row corresponds to a slice of the model, and the values are the distances from the centroid of that slice to the edge of the model.

Next, the FFT is computed of these 2D histogram images and shift the zero-frequency component to the center. Since the only misalignment of the spherical maps will be in the horizontal direction, the phase can be ignored by taking only the amplitude of the FFT image. This allows the FFT images to be aligned even if the CHAMPs are misaligned due to rotations and translations. For display purposes in Figure 6.4, the base-2 log of the FFT image is shown. These FFT images are compared to each other using correlation, where a higher value corresponds to a better match. Figure 6.4 shows a comparison of the various models between a select few of the materials.

6.3.2 Material Heating and Cooling

In this section, I describe two features extracted to represent the heating and cooling of the materials. The first feature attempted was very simple, while the second one yielded higher accuracy and is more physically meaningful. These two features are described below.

6.3.2.1 Patched-Based Temperature Curves

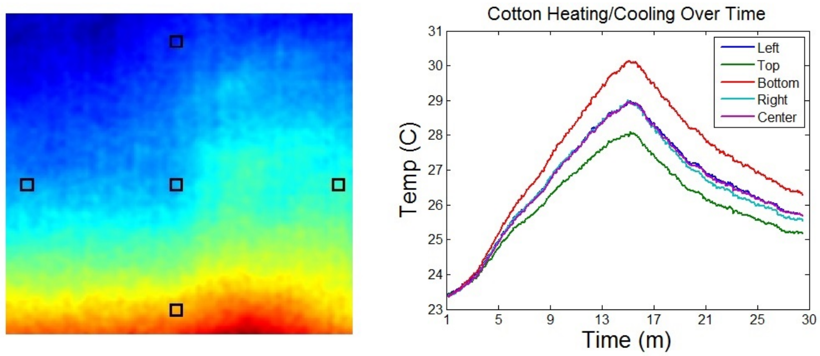
The first attempt of feature extraction of heating and cooling is quite simple. For each image in an infrared video stream, five patches are sampled as shown in Figure 6.5. For each patch, the mean temperature over the patch is plotted over time to give a temperature curve that should, ideally, be unique for each material. Patches were chosen over using all pixels to smooth over noise and to speed up the processing time.

To account for change in room temperature over the course of a day, the starting temperatures of each curve are aligned when comparing across materials, i.e. for materials x and y perform $T'_{xk}(t) = T_{xk}(t) + [T_{yk}(1) - T_{xk}(1)]$. Here $T_{xk}(t)$ is the temperature at time t for material x at patch k. Euclidean distance is used as a metric for comparison.

6.3.2.2 Solving the Heat Equation

The heat equation [22] is a parabolic partial differential equation that describes the distribution of heat over time. The standard heat equation is augmented to more closely describe the physical setup by adding a second term as in

$$\frac{dI}{dt} = \alpha \nabla^2 I + \beta S(t), \tag{6.5}$$



a) **b)**

Figure 6.5: Simple heating/cooling feature. A heat lamp was placed at the bottom of each material and was turned on at t=1m and turned off at t=16m. **a)** Location of sampled patches in infrared image for heating/cooling feature extraction. **b)** Corresponding graph of patch temperature over time.

where α, β are unknown constants, and S is a function which describes how heat is applied. In my setup, a heat lamp was the source of heat in the scene, and its temperature changed over time. To calculate S,the temperature of the heat lamp is sampled over time using an infrared thermometer, and fit a piecewise polynomial to the sample temperatures. Once S is known, α, β are calculated by setting up an overconstrained linear system and applying a Moore-Penrose pseudoinverse [15]. The system is set up as

$$\begin{pmatrix} \nabla^2 I & S(t) \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} \frac{dI}{dt} \end{pmatrix}. \tag{6.6}$$

The resulting parameters are the feature vector for comparison and Euclidean distance was used as a difference metric.

6.4 Experiments and Results

In my experiments I used a Xenics Gobi 640 GigE uncooled long wave infrared camera, which has a resolution of 640x480 and has a 50mC sensitivity at 30°C. The materials used were broken up into 5 coarse classes: cloth, wood, paper, plastic foams, and metal. Each coarse class was further broken up into a total of 21 subclasses as shown in Figure 6.1. For each type of material, 4 samples were imaged; this gives a total of 84 material samples. The

Method	Rank 1 Coarse	Rank 2 Coarse	Rank 1 Fine	Rank 2 Fine	Rank 3 Fine
Capped, FFT	79.8	92.8	59.5	70.2	82.1
Uncapped, FFT	77.4	97.7	57.1	73.8	86.9
Capped, SPH	83.3	91.7	63.1	81.0	83.3
Uncapped, SPH	82.1	89.3	61.9	76.2	85.7
Capped, FFTCOS	83.3	92.9	64.3	73.8	82.1
Uncapped, FFTCOS	84.5	92.9	64.3	73.8	83.2
Capped, SPHCOS	82.1	89.3	61.9	79.8	85.7
Uncapped, SPHCOS	80.1	89.3	60.7	76.2	85.6
Heat Equation 100x100, 50	78.6	85.7	42.9	57.1	69.1
Heat Equation 100x100, 25	66.7	76.6	35.7	50.0	61.5
Heat Equation 100x100, 75	69.1	88.1	28.6	40.5	42.9
Heat Equation 50x50, 15	64.3	76.2	33.3	45.2	57.1
Heat Equation 150x150, 85	54.8	83.3	35.7	40.5	45.5
Heating/Cooling Graphs 3x3	57.1	85.7	35.7	52.4	63.1
Heating/Cooling Graphs 9x9	59.5	88.1	35.7	52.4	63.1
Heating/Cooling Graphs 15x15	59.5	83.1	28.6	52.4	59.2
Heating/Cooling Graphs 1st Deriv 3x3	66.7	76.2	28.6	47.6	57.1
Heating/Cooling Graphs 1st Deriv 9x9	69.1	76.2	30.1	47.6	57.1
Heating/Cooling Graphs 1st Deriv 15x15	71.4	76.2	30.1	50.0	57.1
Heating/Cooling Graphs 1st Deriv 25x25	69.1	76.2	28.6	47.6	57.1
Combination	**95.3**	**100.0**	**71.4**	**85.7**	**92.9**

Table 6.1: Results using variations on the proposed features given as accuracy over entire dataset. Rank n means the correct class was in the top n choices. Coarse refers to wood vs metal vs cloth vs paper vs plastic foams, whereas fine refers to a specific class (e.g. poplar) against all other 20 classes.

Method	Rank 1 Coarse	Rank 2 Coarse	Rank 1 Fine	Rank 2 Fine	Rank 3 Fine
Ours + DCT	**98.8**	**100.0**	**81.0**	**91.7**	**94.0**
Ours	95.3	100.0	71.4	85.7	92.9
NIR [91]	92.9	96.4	76.2	88.1	92.9
HSL [87]	83.3	95.3	59.5	85.7	92.9
DCT [26]	92.9	95.2	61.9	80.9	83.3
Gabor [26]	82.1	90.5	61.9	79.8	81.0
Co-occurrence [26]	81.0	81.0	60.0	64.8	67.1

Table 6.2: Results of comparison to other works using the dataset. Rank n means the correct class was in the top n choices. Coarse refers to wood vs metal vs cloth vs paper vs plastic foams, whereas fine refers to a specific class (e.g. poplar) against all other 20 classes.

physical setup of the camera and materials is shown in Figure 6.3. The LWIR camera was 0.4m above the materials looking downwards. The boundaries of the image were marked so that materials can be roughly aligned to the center of each image.

For the water permeation experiment, a pipette was used to drop 0.4mL of water onto the center of the material. I recorded a ten minute long video at 1fps for each sample. For each video, the CHAMPs and FFT of binned spherical maps were extracted as described in

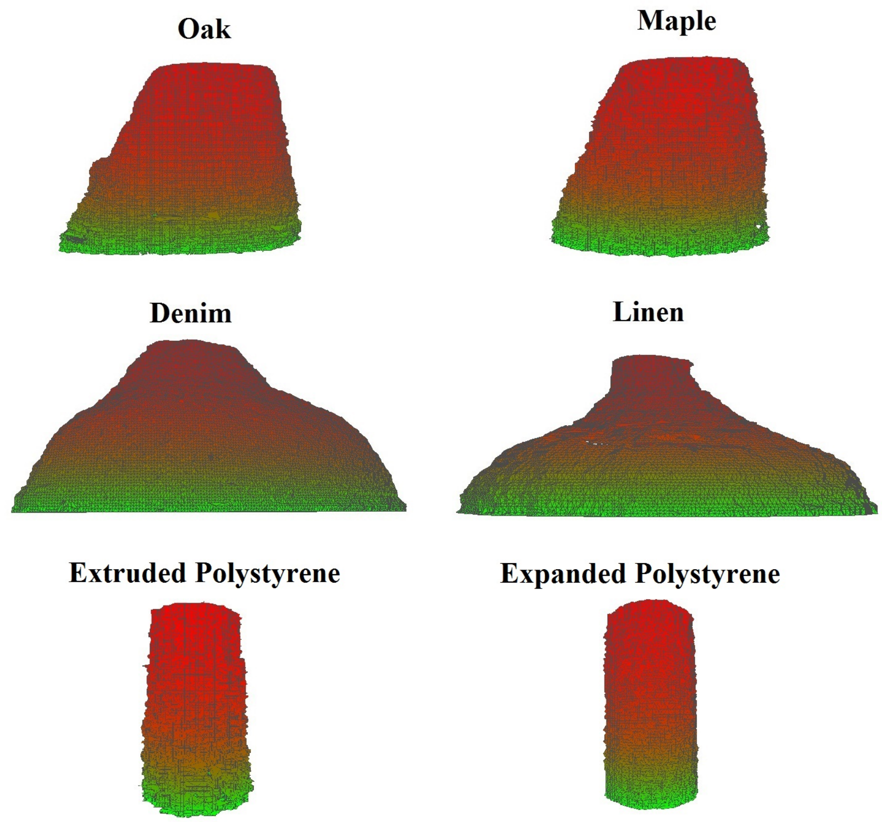


Figure 6.6: Misclassified CHAMPs of materials using only water permeation. When adding in heating/cooling information, these materials appear different.

Section 6.3.1. For each fine-grain class type, a mean model was created by averaging the FFT images for all samples. Leave-one-out cross validation was used in this procedure by leaving one sample out for testing, and three samples to create the mean model. The mean model is compared against all other material samples. I tested a few variations including: “capping” the CHAMP, leaving off the caps, using aligned spherical maps without FFT, using the FFT, and using aligned spherical maps without the $cos(\phi)$ term in Eq. 6.4. When uncapped, ϕ is restricted to $-\frac{\pi}{4}to\frac{\pi}{4}$ to avoid NaN results. These results are reported in Table 6.2.Accuracy is reported using rank1, rank 2, and rank 3 results, where rank n means if the algorithm's choice was in the top n choices it is marked as correct. This is a useful metric to see how much each feature can narrow down the possible choices.

For the heating and cooling experiment, 30 minute videos were recorded of the heating and cooling of each material sample. Each sample was placed in the center of the camera's view and heated with a heat lamp placed in front and above the material. The heat lamp was turned on for 15 minutes, and then switched off at the 15 minute mark. This process was automated using a programmable Arduino microcontroller to ensure precise timing. I tested a few different sized Laplacian of Gaussian (LoG) filter sizes and sigma values for the feature described in Section 6.3.2.2. I also tested a few different patch sizes for the feature described in Section 6.3.2.1. The results are included in Table 6.2.

To combine the features, the top $n = 5$ choices are taken from the water permeation features. Then, a simple linear combination is used of the normalized distances of the best water permeation variation and best heating/cooling patch size as in $\alpha_w *$UncappedFFT $+(1 - \alpha_w)*$(HeatEquation [100 100],50). The weighting parameter α_w and n chosen were the ones with the highest average accuracy scores across all categories.

I compare to other material classification works that use color, texture, and near infrared. To the best of my knowledge there is no other work on image-based thermal properties for material classification. [26] gives a comparative study of approaches for classification of color texture images. I implemented the three features they recommended across all three color channels – Discrete Cosine Transform (DCT), Gabor Filters, and Co-occurance. [91] uses NIR for classification due to its relative independence from color imagery. The materials were recorded with the Sony EVI-D70, which is a color video camera with a NIR mode and a resolution of 640x480.

The infrared results were combined with color results to further improve accuracy. The way this is performed is similar as described above – a linear combination of the normalized distances of the best infrared features and best color features as in $1 - \beta_w *$Combination $+(\beta_w)*$DCT. In the presented experiments, $n = 5$, $\alpha_w = 0.14$, and $\beta_w = 0.25$. The smaller β_w is, the less color and texture information is relied on for classification.

6.4.1 Discussion

When water permeation is taken as a feature alone, the best rank 1 accuracy over the 5 coarse classes is 84.5% when using the FFT of the spherical map multiplied by $cos(\phi)$. This is also the best version of the feature when looking at the best rank 1 accuracy over all 21 classes at 64.3%. Water permeation may be better utilized when combined with other features. This is because it is able to narrow down the possible classes very effectively. Using an uncapped CHAMP with the FFT of the spherical map 97.7% rank 2 accuracy is achieved for coarse classes; this means the correct material class is almost always in the top 2 choices. Similarly for fine classes, this variation maximizes the accuracy where the correct class is in the top 3 choices about 87% of the time. That is why UncappedFFT was chosen for the combination – it was able to narrow down the possible choices the best. Similarly, using the heat equation with a LoG filter of size 100x100 and $\sigma = 50$ led to the highest results across the categories. When combined together, the results are improved up to 16%, which implies these features are complimentary to each other.

When comparing to other works, the proposed method gives the best result across all categories, although in some cases the gain is only slight. However, when color is added using the linear combination described above, the results are significantly improved up to 10%. Moreover, weighting parameter β_w can be controlled to give more invariance to color and texture information, depending on the dataset being used.

One drawback of the proposed method is the amount of time it takes to record the water permeation and heating/cooling videos, which were 10 minute and 30 minute respectively. The length of videos recorded for this project were conservative, and it may be possible to decrease the video length. Also, the heating/cooling can be sped up by placing the heating element closer to the material.

6.5 Conclusions and Future Work

In this chapter, I described features that can be extracted from thermal imagery which give results that not only outperform other color and texture features, but also are complimentary to them, and can be combined to increase performance. I collected a dataset of 21 different classes with 84 total samples, and recorded thermal video of water permeation and heating/cooling, as well as color and NIR photographs. I presented CHAMPs, which model the water permeation, and a method to extract the heat equation constants. Combining these features together results in higher accuracy than using either one individually.

Chapter 7

THERMAL STEREO

7.1 Introduction

The thermal features discussed in Chapter 6 assumed a planar material sample placed in a standardized location and orientation to the heat source. However, in more realistic scenarios, the material sample will have a non-planar, possibly curved surface, and can be in an arbitrary location relative to the heat source. In this case, the thermal features will fail. To take into account the location and curvature of objects, stereo reconstruction can be applied. In this chapter, I discuss a comparison between using stereo color cameras and stereo thermal cameras.

7.2 Background and Related Works

[57] uses stereo infrared cameras and color cameras for pedestrian detection. Standard block matching with sum of square differences was used for stereo matching. However, an analysis of the stereo matching phase is not given. [79] estimates dense disparity using block matching with a 7x7 correlation window on 480x512 images of various types of surfaces (grass, concrete, dirt). They observe that if the signal to noise ratio is over 30/1, 90% disparity accuracy is achieved. However in the middle of the night, the temperature tends to equalize and compress the dynamic range, thus decreasing the signal to noise ratio. They conclude that uncooled infrared cameras have too low of a signal to noise ratio for dense disparity results. [40] deals with the problem of noisy infrared imagery by first transforming the image with phase congruency. The transformed image is less noisy and has a more distinct and smoother background. They use standard block matching with sum of absolute differences and show improved results over block matching without the phase congruency transform.

More work has been done in cross-modality matching or multi-modal fusion for depth estimation. [116] Compute depth using a fusion of multiple sensors, including color, thermal, and time of flight sensors. Thermal is used to increase information in textureless regions when fused with color imagery. [113] Compares various features for cross modality (color and thermal) matching for indoor scenes humans and found local self-simlarity to outperform Normalized Cross-Correlation (NCC), and was slightly more robust for human reconstruction than Histograms of Oriented Gradients (HOG) and Mutual Information (MI).

For evaluating thermal-only stereo, these works did not consider global and semi-global matching techniques [132]. Only local methods have been used, which perform poorly in textureless region [86].

7.3 Experiments

7.3.1 Color Vs. Thermal Stereo Comparison

In this section, color-based stereo reconstructions are compared to thermal-based stereo reconstructions on various material types. The reconstructed objects were spherical balls which were chosen since they are easy to model, easy to measure their physical diameter, and have heavy uniform curvature. The spheres are made of seven different materials: aluminum, stainless steel, polystyrene foam, vinyl, glass, maple, butyl rubber, as shown in Figure 7.2. The same spheres are shown in thermal after heating in Figure 7.3. The advertised diameter was 127mm, but calipers were used to find the exact diameter as shown in Table 7.1 .

Material	Diameter
Wood	128.7
Vinyl	122.2
Butyl Rubber	138.7
Aluminum	127.4
Glass	127.4
Polystyrene Foam	121.6
Steel	126.7

Table 7.1: Measured ground truth sphere diameters.

For sensors, I used a stereo pair of both thermal and color cameras. The thermal cameras used were Gobi 640 GigE long wave infrared (8 μm to 14 μm) uncooled microbolometer cameras, which have a resolution of 640x480 and a thermal sensitivity of 50mK. The color cameras used were Point Gray Flea2 FL2G-50S5C Firewire cameras with a resolution of up to 2448x2048, but for synchronization purposes I used 1280x960. All four cameras were synchronized with a software trigger.

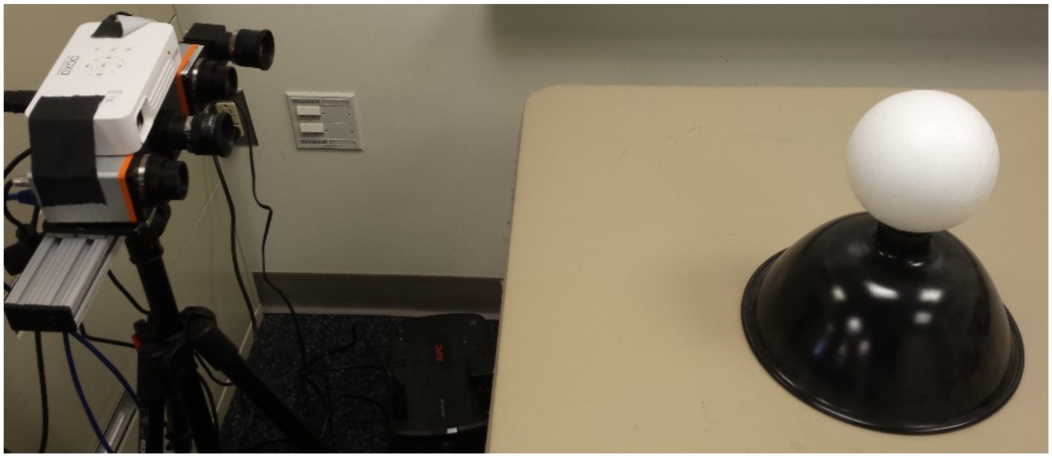
In addition to the sensors, I used a pico projector and a heat gun to add texture to the spheres in each modality. The projector was an AAXA ST200 1280x720, 150 lumens LED projector, and projected a highly textured, randomized pattern onto the sphere. The pattern was projected with the lights in the room turned off. The heat gun was a Genesis GHG1500A Dual-Temperature Heat Gun (1500W/750W). The heat gun was applied to the surface for 5 seconds at a high temperature setting from 0.6m away. The setup is shown in Figure 7.1.

For each material type, the sphere was imaged nine times – three standard lighting, three using the projector, three using the heat gun. A stereo reconstruction was performed in each modality. For calibration of the thermal cameras, I used the method described in [95], in which the calibration board is attached to a ceremic backing and heated with a heat lamp. The resulting calibration is used to perform calibrated rectification. Semi-global block matching [47] was performed on the rectified images since it is the highest performing algorithm that is readily available in many languages and toolboxes.

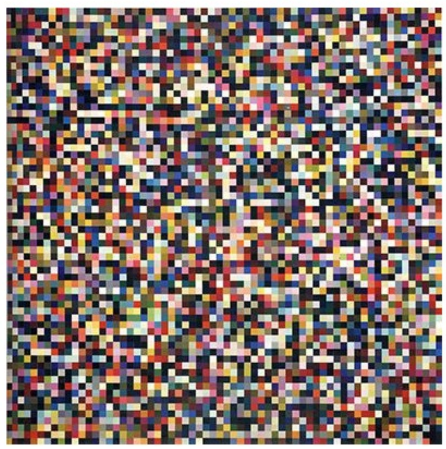
Two spheres were fit to the resulting point clouds. The first sphere is fit using the known ground truth radius to measure point-wise error. The second sphere is fit without knowledge of the ground truth radius in order to compare the model radius to the ground truth. The following equation is minimized to obtain the best fit sphere:

$$\min_{\mathbf{c},r} f(\mathbf{x}) = [(\sum \mathbf{x} - \mathbf{c}) - r]^2, \tag{7.1}$$

where **c** is the 3D center of the sphere and r is the radius. This equation is minimized using the simplex search method of Lagarias et al. [59]. Table 7.2 shows the results of varied materials with and without added texture in each modality. The results are averaged over all



a)



b)

Figure 7.1: **a)** Setup for sphere stereo experiments. **b)** The projected pattern.

relevant samples, and both pixel root mean square error and radius error are shown.

7.3.1.1 Discussion

If the Pixel RMS is considered (the best sensor per material is highlighted in red in Table 7.2), the visible with a projected texuture performs better on about half of the materials (wood, styrofoam, steel, aluminum), while the thermal performs better on the other materials (vinyl, glass, rubber). The materials that thermal performed better were materials where the projected pattern did not appear as distinct due to the dark colors in the material. On the

Figure 7.2: Materials used are 1 polystyrene foam, 2 aluminum, 3 vinyl, 4 glass, 5 maple, 6 butyl rubber, 7 stainless steel.

other hand, thermal performs poorly on metallic materials which are mostly reflective in the long infrared range.

When using the radius estimation percent error, the results are the same for all materials except aluminum and styrofoam; however, in aluminum the lowest radius estimation percent error (0.93) has the highest standard deviation (3.24mm), which is unreliable. Thus the radius error estimation supports the per pixel error across all materials. I also observed in all of the highlighted cases that most of the error was caused by a relatively few outlier 3D points which were incorrect by 20mm+.

The density of the reconstruction is defined as the number of pixels with estimated disparity over the total number of possible disparity estimations. In general, high density is correlated with higher accuracy, but this is algorithm dependent. In these experiments, higher

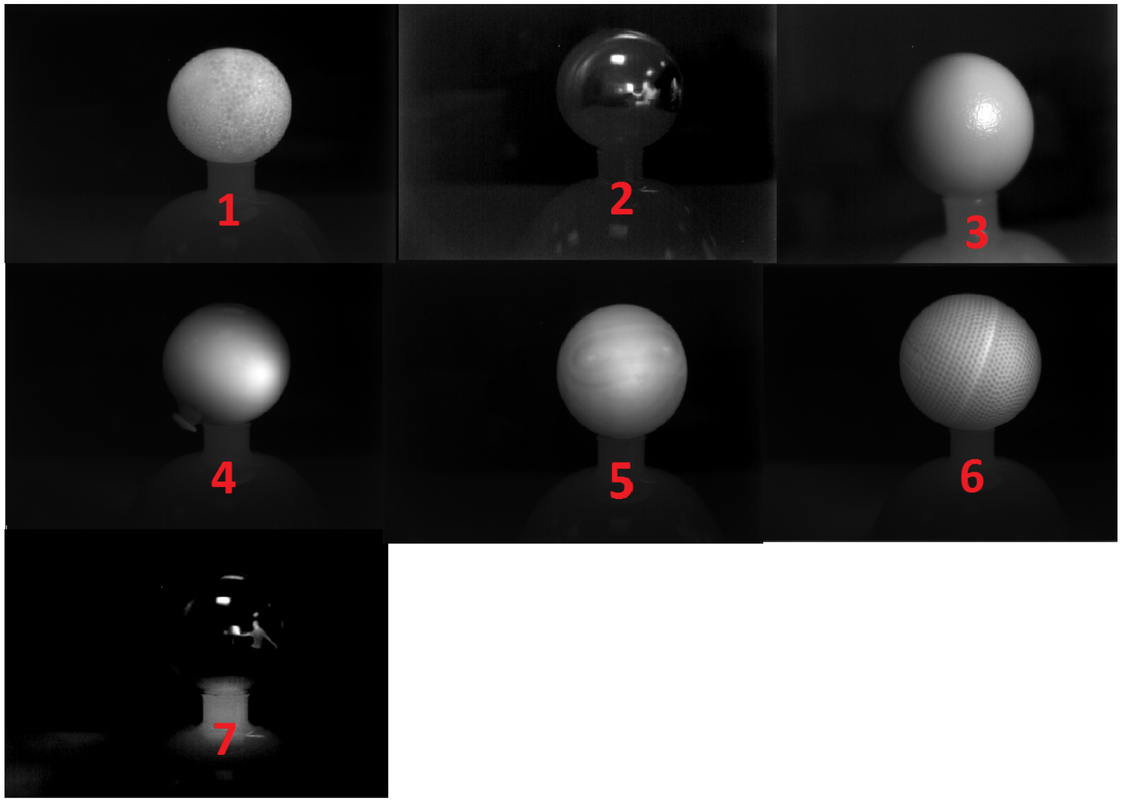


Figure 7.3: Materials in thermal after heating for one minute. 1 polystyrene foam, 2 aluminum, 3 vinyl, 4 glass, 5 maple, 6 butyl rubber, 7 stainless steel.

density is correlated with higher accuracy. In almost all of the cases, the highest performing sensor w.r.t pixel-wise acheived the highest density in all except 2 materials; however, within these two materials (vinyl, rubber), the density is very high and close to the other value.

7.4 Conclusion

In this chapter, I compared stereo thermal and stereo color reconstructions. I reconstructed spheres with a known radius and compared the reconstructed points against ideal sphere points, and I compared the reconstructed radius against the measured radius. I also used projected light patterns and a heat gun to add texture to each modality respectively. I found that thermal is competitive with and even outperforms color based reconstructions in some materials (vinyl, glass, rubber) when using the pixel RMS error as a metric.

Sensor	Material	Projector?	Heater?	GT r	Pred r	% err	Pred r σ	RMS (mm)	RMS σ	Density
Visible	wood	N	N	64.35	67.6657	5.1526	3.6133	7.0051	9.9601	0.80435
Visible	wood	Y	N	64.35	64.4238	0.11468	0.26117	0.78274	2.5245	0.95235
Thermal	wood	N	N	64.35	69.2361	7.593	11.7897	12.9168	9.1761	0.85487
Thermal	wood	N	Y	64.35	63.5329	1.2697	0.73276	2.8354	3.5446	0.91183
Visible	vinyl	N	N	61.1	56.1651	8.0767	0.68092	18.8616	13.5417	0.87342
Visible	vinyl	Y	N	61.1	64.1672	5.02	0.37074	1.3837	2.2512	0.96159
Thermal	vinyl	N	N	61.1	69.1941	13.2473	15.9771	11.8484	11.8501	0.91368
Thermal	vinyl	N	Y	61.1	60.8838	0.3538	0.025835	0.80622	1.1564	0.93312
Visible	glass	N	N	63.7	65.8552	3.3833	5.0303	23.6173	17.2831	0.81172
Visible	glass	Y	N	63.7	72.1337	13.2397	4.6557	17.5617	14.7453	0.65006
Thermal	glass	N	N	63.7	67.756	6.3674	6.6014	16.5256	11.0016	0.85364
Thermal	glass	N	Y	63.7	61.9532	2.7422	2.0967	6.8163	7.2075	0.92477
Visible	rubber	N	N	69.35	64.0468	7.647	1.8262	13.144	11.341	0.75911
Visible	rubber	Y	N	69.35	65.1223	6.0962	1.1989	1.9272	3.7583	0.9479
Thermal	rubber	N	N	69.35	67.7898	2.2498	2.5959	10.0504	9.7062	0.89729
Thermal	rubber	N	Y	69.35	68.7436	0.87438	0.31867	1.7727	2.5442	0.92587
Visible	styro	N	N	60.8	58.5957	3.6256	1.1518	10.5935	9.2631	0.84136
Visible	styro	Y	N	60.8	63.8694	5.0484	0.46937	1.071	2.0229	0.94307
Thermal	styro	N	N	60.8	67.1997	10.5258	14.6867	11.8861	10.0602	0.85152
Thermal	styro	N	Y	60.8	60.0238	1.2767	0.034588	1.8583	2.5724	0.91828
Visible	steel	N	N	63.35	57.6273	9.0335	0.57655	12.9085	11.4058	0.88286
Visible	steel	Y	N	63.35	64.5509	1.8956	1.0508	2.4085	5.7781	0.94145
Thermal	steel	N	N	63.35	61.2164	3.368	0.18211	10.8778	8.1621	0.91767
Thermal	steel	N	Y	63.35	60.7334	4.1304	0.74124	6.6754	7.0092	0.898
Visible	alum	N	N	63.7	58.8309	7.6438	2.3186	13.3053	11.6614	0.84449
Visible	alum	Y	N	63.7	65.0809	2.1679	0.97299	2.6691	6.802	0.92921
Thermal	alum	N	N	63.7	63.1059	0.9327	3.2403	12.4797	10.7564	0.84168
Thermal	alum	N	Y	63.7	59.505	6.5855	1.3503	11.5691	10.1813	0.86647

Table 7.2: Results of sphere reconstruction experiments in each modality using semi-global block matching [47]. Error statistics are given with both pixel-wise RMS to a sphere model with ground truth radius and with predicted radius error. Density is the number of pixels with an estimated disparity over the total number of possible disparity estimations. Error is averaged over 3 samples for each row in the table. The radius and RMS values are measured in mm. The red bordered rows correspond to the sensor that had the lowest Pixel RMS (and in all cases the Pixel RMS σ).

Chapter 8

MATERIAL CLASSIFICATION ON CONVEX OBJECTS IN ARBITRARY POSITIONS

In this chapter I describe a method for handling material classification on convex objects in arbitrary positions using a four camera dual-stereo system. By calibrating the cameras and heat source together, the material properties can estimated by modeling the scene using a modified version of the heat equation in Chapter 6. This equation is derived from the radiosity equation used in computer graphics, and is described below. However, implementation and testing of this system is still in progress, and the experiment outlined below is still pending results.

8.1 Background on Radiosity

Radiosity is 3D computer graphics technique for rendering scenes. It is a global illumination method in that energy comes from not only the light source, but other surfaces. In fact, light sources are treated no differently from other surfaces, and we can model their shape. The method is view independent as energy transfer is calculated between all surfaces. Radiosity calculation is a modeling process and involves modeling the geometry of the scene. Conveniently, the radiosity method has its basis in thermal radiation heat transfer. I will summarize the radiosity method below, but for an in depth review of radiosity, see [122, 4].

The heat emitting from a surface i is the radiosity and is made up of the heat reaching it from other surfaces and the heat it directly generates if it is a heat source. The fraction of heat that reaches surface i from surface j is form factor and depends on the geometry of the scene. Let A_i be the area of surface i, B_i be the radiosity, E_i be the generated energy per area per time for heat sources, let R_i be the fraction of reflected incident heat, and let F_ij be the form factor, the fraction of heat which gets to i from j. Then the radiosity (heat emitting from a surface area) equation is given in the following equation as

$$B_iA_i = E_iA_i + R_i \int F_ijB_jdA_j \tag{8.1}$$

To calculate the form factors, the hemisphere method is employed. A unit-hemisphere is placed around the surface area A_i , and the heat that passes into the hemisphere is the heat arriving at the surface. For another surface A_j, it is projected onto i's hemisphere and then downwards onto the surface of i. This will give a high heat transfer if the projected area is near the center, and a low if near the edge based on the cosine of the orientation. See Figure 8.1. The two angles are needed to encapsulate both the orientation of j and is position in the hemisphere of i.

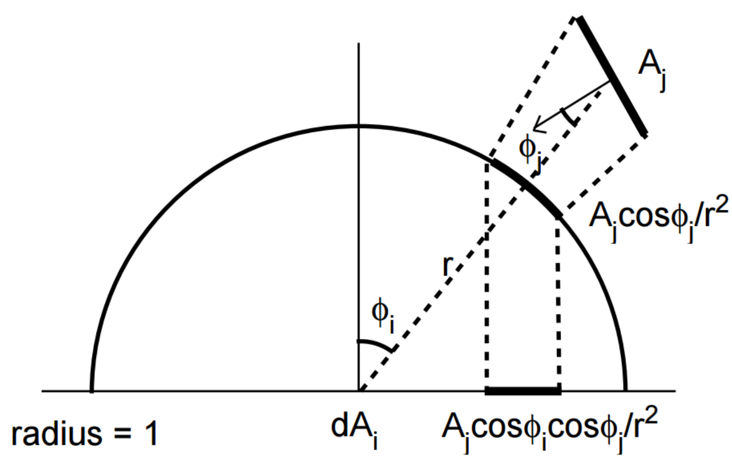


Figure 8.1: Hemisphere method for calculating the form factor in the radiosity equation. r is the distance between A_i and A_j. Image courtesy of [4]

The form factor is then approximated by integrating over the hemisphere as in

$$F_ij \approx \int_{A_j} \frac{cos\phi_i cos\phi_j}{\pi r^2} dA_j. \tag{8.2}$$

To approximate the integrals, it is assumed that the surfaces i and j are small patches that are energetically uniform.

8.2 The Radiosity Technique for Thermal Features

The radiosity technique in computer graphics assumes that R_i is known (along with the distance and angle between scene structures), but in the case of material classification case this is an unknown material property. Moreover, the distance and angle between the surface patches are unknown. To calculate the surface patch geometry, stereo reconstruction is performed to reconstruct the shape of the objects in the scene. Using the techniques for calibration earlier in this dissertation (Chapter 2), as well as thermal stereo reconstruction, everything in the scene can be accounted for except the heat source. The heat source is at an unknown location and orientation.

To calculate the heat source's location and orientation, a technique developed in [99] is used. The idea is to use a multi-modal stereo system consisting of stereo color and stereo thermal cameras to reconstruct the reflecting surface and the reflected scene. In one modality, texture is added to the mirror surface that does not appear at all in the other modality (e.g. temperature from hand transferred to mirror, which is invisible to the color modality). Then, a ray trace reconstruction technique is used to calculate 3D points in the coordinate system of the left camera. With this it is possible to reconstruct the scene next to the cameras, behind the cameras, or even the cameras themselves.

The radiosity equation as stated above for computer graphics encapsulates radiative heat transfer between objects. However, it does not account of conduction across a single object. For that, I take into account the heat equation [22]. The heat equation is a parabolic partial differential equation that describes the distribution of heat over time. The heat equation is

$$\frac{dI}{dt} = \alpha\nabla^2 I, \tag{8.3}$$

where α is an unknown constant material property known as thermal diffusivity. To approximate the contribution of each method of heat transfer (conduction and radiation), the two equations are additively combined to give

$$\frac{dI}{dt} = \alpha\nabla^2 I + \int_i E_i A_i + R_i \int_{A_j} F_i j B_j dA_j, \tag{8.4}$$

In the experiments outlined in this chapter, the objects imaged are convex, so there is no radiative exchange between surface patches. Moveover, there is only a single heat source, which can be appomixated with a single patch. This allows us to reduce the equation to

$$\frac{dI}{dt} = \alpha\nabla^2 I + R_i(F_{ih}B_h), \tag{8.5}$$

where h is the heat source. By renaming the terms, we can get a similar equation to Chapter 6. This new equation is the same as 6.5 except it has the form factor, which includes distance and angle components. The equation is

$$\frac{dI}{dt} = \alpha\nabla^2 I + \beta(F_{ih}S(t)), \tag{8.6}$$

where $S(t)$ is a function that outputs the heat given off by the light source at time t and F_{ih} is the form factor of the heat source with a patch i as described in equation 8.2. α and β are material properties for conduction and radiation, respectively. In my setup, a heat lamp was the source of heat in the scene, and its temperature changed over time. To calculate $S(t)$,the temperature of the heat lamp is sampled over time using an infrared thermometer, and fit a piecewise polynomial to the sample temperatures.

Once S is known, α, β are calculated by setting up an overconstrained linear system and applying a Moore-Penrose pseudoinverse [15]. The system is set up as

$$\begin{pmatrix} \nabla^2 I & F_{ih}S(t) \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} \frac{dI}{dt} \end{pmatrix}. \tag{8.7}$$

8.3 Experiments

In this section, the radiosity technique described above was used in a controlled experiment to classify material types. The setup is very similar to the previous chapter for thermal stereo analysis. The reconstructed objects were spherical balls as in the previous chapter, because of the easy to obtain ground truth shape. The spheres are made of seven different materials: aluminum, stainless steel, polystyrene foam, vinyl, glass, maple, butyl rubber, as shown in Figure 7.2.

A stereo pair of both thermal and color cameras were used to reconstruct the scene, including the location and orientation of the heat lamp using [99] . The thermal cameras used were Gobi 640 GigE long wave infrared (8 μm to 14 μm) uncooled microbolometer cameras, which have a resolution of 640x480 and a thermal sensitivity of 50mK. The color cameras used were Point Gray Flea2 FL2G-50S5C Firewire cameras with a resolution of up to 2448x2048, but for synchronization purposes I used 1280x960. All four cameras were synchronized with a software trigger. A 250W heat lamp was used to heat the materials for 10 minutes. A 20 minute video consisting of 10 minutes of heating and 10 minutes of cooling was recorded.

The results of classification using the above technique and the experiment described are still pending and will be released in future work.

8.4 Conclusion

The radiosity technique is a potential solution to handling material classification in curved surfaces. Thermal energy transfer is calculated between all surfaces and iterated over. The equations can be used to model radiative heat transfer and when added to the heat equation for conductive heat transfer, a similar equation is formed to the one described in Chapter 6, but with the addition of a form factor. This form factor handles the angle and distance between the surface and the heat source and allows the classification technique from Chapter 6 to be used on curved objects at arbitrary locations. Experiments have been performed but the results are still pending.

Chapter 9

CONCLUSION AND FUTURE WORK

Thermal cameras are a complementary technology to color cameras that can increase performance and robustness. They allow us to view information that is otherwise invisible from color cameras, which opens up the opportunity for new features and fusion between modalities. However, in general thermal cameras are less studied in computer vision applications than color-based approaches due to their high costs and low resolution. In this dissertation, I explored calibration, alignment, detection, and classification of hidden targets with multiple modalities and a focus on the thermal modality.

For thermal calibration, I used a printed calibration board on a ceramic backing with a simple pre-processing method to expose the checkerboard pattern. I successfully harnessed the sun as a heat source outside, while using a heat lamp inside. I found this method even works when different modalities are used for the left and right cameras.

To align modalities when there are no visible correspondences, I used a differential GPS/IMU along with camera calibration to construct a transformation matrix to project pixels in one image to the other. Since the camera I used could pan/tilt/zoom, the intrinsics and extrinsics change between frames. To solve this problem, I calibrated at multiple zoom levels and interpolated the corresponding intrinsics.

For detection with multiple modalities of anomalous objects, I used Gaussian Mixture Models (GMMs). GMMs can model typical pixel intensities in the scene. When pixel intensities lie outside the model, they are flagged as anomalous. The GMMs are updated between frames to include both new and old information. This allows the GMMs to model new intensities if they are in the scene enough. I tested various fusion schemes to combine the color and thermal imagery and found that fusion at the decision level and pixel level gave similar results, and that fusion was useful if the scene had targets that could only be seen by one modality.

For detection of known objects, I used a neural network to fuse the scores between color and nonlinear radar. Testing was performed on 5 small electronic devices, and I found that nonlinear radar had artifacts that lowered the performance of automatic detection, but it was valuable because it can see the targets even through obstructions such as camouflage or plastic hollow rocks. The fusion results outperformed the individual results no matter the fusion technique used.

For material classification, two novel thermal features were developed – water permeation and heating/cooling. For water permeation, a pipette was used to put a few drops of water onto a material and the thermal camera could easily detect the water's spread. A feature I call the CHAMP (CHAracteristic Model of Permeation) encapsulates information about the speed and shape of the water's spread. For heating/cooling, I solved a variation heat equation for thermal diffusivity and absorptivity. The results show that the water permeation performs better, but fusing the two scores performs even better. Color based techniques perform better than the individual thermal features, but not than the fused thermal features. Fusing color and the two thermal features performs the best.

Since the material classification in Chapter 6 relies on planar material samples in a standarized location and orientation, it will fail when a non-planar sample is used. To remedy this, stereo reconstruction can be used to calculate the location and orientation of the sample. How to adjust the heat equation to use the angle and distance information is discussed in Chapter 8. Stereo cameras would be used to reconstruct the object and obtain the surface normals and distance from the heat source. In Chapter 7, I compared thermal stereo against color stereo and found that each performs better on certain material types when using projection and active heating. Moreover, thermal stereo mostly fails in an indoor setting at room temperature where all objects have similar temperatures. Only under active heating does enough texture appear for stereo matching.

Future work is implementing the thermal feature for curved objects and testing in both indoor and outdoor conditions. After that, the four-camera duel stereo system can be put on a moving vehicle with radar. This sensor system would be tested in a more realistic environment with real targets. Another improvement to the heating feature can be using convection from e.g. a heat gun to quickly heat an object. This reduces the time needed to calculate features, but also reduces the modeling potential. Rather, machine learning techniques, especially deep learning, can be used to automatically pick the best feature from a sequences of thermal images instead of the current method of hand picking features.

BIBLIOGRAPHY

[1] Flir one launch faq. http://www.flir.com/flirone/press/FLIRONE_Launch_FAQ.pdf. Accessed: 12-29-2015.

[2] icasualities. icasualities.org/OEF/Index.aspx. Accessed: 6-29-2015.

[3] Matlab and statistics toolbox release 2014b. The MathWorks, Inc., Natick, Massachusetts, United States.

[4] Radiosity rendering. http://www.cs.bath.ac.uk/~pjw/NOTES/75-ACG/ch5-radiosity.pdf. Accessed: 12-30-2015.

[5] Thermal imaging cameras: Specifications comparison. http://www.all-spec.com/compare/TST-28/. Accessed: 12-29-2015.

[6] Infrared waves. National Aeronautics and Space Administration, Science Mission Directorate., 2010.

[7] K. Ammer. *The Thermal Image in Medicine and Biology, Chapter: The History of Thermal Imaging*. Proceedings of the Meeting of the Royal Photographic Society, Imaging Science & Technology Group, 11th-12th October 1994 in Bath, and the 6th European Congress of Thermology, 1995.

[8] D.T. Anderson, K.E. Stone, J.M. Keller, and C.J. Spain. Combination of anomaly algorithms and image features for explosive hazard detection in forward looking infrared imagery. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, 5(1):313–323, Feb 2012.

[9] P. Atrey, A. Hossain, A. El Saddik, and M. Kankanhalli. Multimodal fusion for multimedia analysis: a survey. *Multimedia Systems*, 16(6):345–379, 2010.

[10] K. Beier and H. Gemperlein. Simulation of infrared detection range at fog conditions for enhanced vision systems in civil aviation. *Aerospace Science and Technology*, 8(1):63 – 71, 2004.

[11] D. Bhning, W. Bhning, and H. Holling. Revisiting youden's index as a useful measure of the misclassification error in meta-analysis of diagnostic studies. *Statistical Methods in Medical Research*, 17(6):543–554, 2008.

[12] M. Bianchini, M. Maggini, L. Sarti, and F. Scarselli. Recursive neural networks for object detection. In *Neural Networks, 2004. Proceedings. 2004 IEEE International Joint Conference on*, volume 3, pages 1911–1915 vol.3, July 2004.

[13] T.L. Bisbee and D.A. Pritchard. Today's thermal imaging systems: background and applications for civilian law enforcement and military force protection. In *Security Technology, 1997. Proceedings. The Institute of Electrical and Electronics Engineers 31st Annual 1997 International Carnahan Conference on*, pages 202–208, Oct 1997.

[14] A P. Bowman, E.M. Winter, AD. Stocker, and P.G. Lucey. Hyperspectral infrared techniques for buried landmine detection. In *Detection of Abandoned Land Mines, 1998. Second International Conference on the (Conf. Publ. No. 458)*, pages 129–133, Oct 1998.

[15] S. Campbell and C. Meyer. *Generalized Inverses of Linear Transformations: The MoorePenrose or Generalized Inverse*, chapter 2, pages 8–27.

[16] H.F. Carlak, N.G. Gencer, and C. Besikci. Medical thermal imaging of electrically stimulated woman breast: A simulation study. In *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*, pages 4905–4908, Aug 2011.

[17] V. Caselles, R. Kimmel, and G. Sapiro. Geodesic active contours. In *Computer Vision, 1995. Proceedings., Fifth International Conference on*, pages 694–699, Jun 1995.

[18] H. Celikoglu. Application of radial basis function and generalized regression neural networks in non-linear utility function specification for travel mode choice modelling. *Mathematical and Computer Modelling*, 44(78):640 – 658, 2006.

[19] H. Chen and L.B. Wolff. Polarization phase-based method for material classification and object recognition in computer vision. In *Computer Vision and Pattern Recognition, 1996. Proceedings CVPR '96, 1996 IEEE Computer Society Conference on*, pages 128–135, Jun 1996.

[20] S.Y. Cheng, S. Park, and M.M. Trivedi. Multiperspective thermal ir and video arrays for 3d body tracking and driver activity analysis. In *Computer Vision and Pattern Recognition - Workshops, 2005. CVPR Workshops. IEEE Computer Society Conference on*, June 2005.

[21] S. Choi and E. Kim. Design and implementation of vision-based structural safety inspection system using small unmanned aircraft. In *Advanced Communication Technology (ICACT), 2015 17th International Conference on*, pages 562–567, July 2015.

[22] J. Crank and P. Nicolson. A practical method for numerical evaluation of solutions of partial differential equations of the heat-conduction type. *Advances in Computational Mathematics*, 6(1):207–226, 1996.

[23] D. De Nigris, D.L. Collins, and T. Arbel. Multi-modal image registration based on gradient orientations of minimal uncertainty. *Medical Imaging, IEEE Transactions on*, 31(12):2343–2354, Dec 2012.

[24] B. Do and S. Huang. Dynamic background modeling based on radial basis function neural networks for moving object detection. In *Multimedia and Expo (ICME), 2011 IEEE International Conference on*, pages 1–4, July 2011.

[25] E.R. Dougherty. *An introduction to morphological image processing*. Tutorial texts in optical engineering. SPIE Optical Engineering Press, 1992.

[26] A. Drimbarean and P.F. Whelan. Experiments in colour texture analysis. *Pattern Recognition Letters*, 22(10):1161 – 1167, 2001.

[27] A. Drobchenko, J. Ilonen, J. Kamarainen, A. Sadovnikov, H. Klviinen, and M. Hamouz. Object class detection using local image features and point pattern matching constellation search. In BjarneKjr Ersbll and KimSteenstrup Pedersen, editors, *Image Analysis*, volume 4522 of *Lecture Notes in Computer Science*, pages 273–282. Springer Berlin Heidelberg, 2007.

[28] A. Ellmauthaler, E.A.B. da Silva, C.L. Pagliari, J.N. Gois, and S.R. Neves. A novel iterative calibration approach for thermal infrared cameras. In *Image Processing (ICIP), 2013 20th IEEE International Conference on*, pages 2182–2186, Sept 2013.

[29] D. Erhan, C. Szegedy, A. Toshev, and D. Anguelov. Scalable object detection using deep neural networks. In *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pages 2155–2162, June 2014.

[30] O. Faugeras, Q. Luong, and S. Maybank. Camera self-calibration: Theory and experiments. In G. Sandini, editor, *Computer Vision ECCV'92*, volume 588 of *Lecture Notes in Computer Science*, pages 321–334. Springer Berlin Heidelberg, 1992.

[31] D. Forsyth and J. Ponce. *Computer Vision: A Modern Approach*. Prentice Hall Professional Technical Reference, 2002.

[32] K. Gallagher, G. Mazzaro, K. Ranney, L. Nguyen, A. Martone, K. Sherbondy, and R.. Narayanan. Nonlinear synthetic aperture radar imaging using a harmonic radar, 2015.

[33] K.. Gallagher, G. Mazzaro, K Ranney, L Nguyen, A Martone, K Sherbondy, and R. Narayanan. Nonlinear synthetic aperture radar imaging using a harmonic radar, 2015.

[34] K.A. Gallagher, R.M. Narayanan, G.J. Mazzaro, K.I. Ranney, A.F. Martone, and K.D. Sherbondy. Moving target indication with non-linear radar. In *Radar Conference (RadarCon), 2015 IEEE*, pages 1428–1433, May 2015.

[35] R. Gasser and T.H. Thomas. Prodding to detect mines: a technique with a future. In *Detection of Abandoned Land Mines, 1998. Second International Conference on the (Conf. Publ. No. 458)*, pages 168–172, Oct 1998.

[36] A. Goshtasby. Image registration by local approximation methods. *Image and Vision Computing*, 6(4):255 – 261, 1988.

[37] B. Goswami and G. Bhandari. Temperature induced mean based cloud motion prediction model from satellite infrared images. In *India Conference (INDICON), 2012 Annual IEEE*, pages 719–723, Dec 2012.

[38] M. Gschwandtner, R. Kwitt, A. Uhl, and W. Pree. Infrared camera calibration for dense depth map construction. In *Intelligent Vehicles Symposium (IV), 2011 IEEE*, pages 857–862, June 2011.

[39] J. Gu and C. Liu. Discriminative illumination: Per-pixel classification of raw materials based on optimal projections of spectral brdf. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 797–804, June 2012.

[40] L. Guo, G. Zhang, and J. Wu. Infrared image area correlation matching method based on phase congruency. In *Artificial Intelligence and Computational Intelligence (AICI), 2010 International Conference on*, volume 1, pages 488–491, Oct 2010.

[41] K. Gurton and M. Felton. Remote detection of buried land-mines and ieds using lwir polarimetric imaging. *Opt. Express*, 20(20):22344–22359, Sep 2012.

[42] S. Hackwood and P.A. Potter. Signal and image processing for crime control and crime prevention. In *Image Processing, 1999. ICIP 99. Proceedings. 1999 International Conference on*, volume 3, pages 513–517 vol.3, 1999.

[43] R. Hartley. Self-calibration of stationary cameras. *Int. J. Comput. Vision*, 22(1):5–23, February 1997.

[44] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge books online. Cambridge University Press, 2004.

[45] K. A. Hendrik, R. Massey, B. Bras Whitham, and M. D. Russell. Technologies for the identification, separation, and recycling of automative plastics. In *International Journal of Environmentally Conscious Design and Manufacturing*, page 3750, 1997.

[46] V. Hilsenstein. Surface reconstruction of water waves using thermographic stereo imaging, 2005.

[47] H. Hirschmuller. Accurate and efficient stereo processing by semi-global matching and mutual information. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 2, pages 807–814 vol. 2, June 2005.

[48] R. Hogan and T. Roush. Mineral emittance spectra: Clustering and classification using self-organizing maps. In *Aerospace conference, 2009 IEEE*, pages 1–7, March 2009.

[49] H. Husain, M. Khalid, and R. Yusof. Automatic clustering of generalized regression neural network by similarity index based fuzzy c-means clustering. In *TENCON 2004. 2004 IEEE Region 10 Conference*, volume B, pages 302–305 Vol. 2, Nov 2004.

[50] A. Ibrahim, S. Tominaga, and T. Horiuchi. Spectral imaging method for material classification and inspection of printed circuit boards. *Optical Engineering*, 49(5):057201–057201–10, 2010.

[51] N. Jamali and C. Sammut. Material classification by tactile sensing using surface textures. In *Robotics and Automation (ICRA), 2010 IEEE International Conference on*, pages 2336–2341, May 2010.

[52] A Jeremic and Arye Nehorai. Landmine detection and localization using chemical sensor array processing. *Signal Processing, IEEE Transactions on*, 48(5):1295–1305, May 2000.

[53] C. Jha. *Thermal Sensors: Principles and Applications for Semiconductor Industries*. Springer Publishing Company, Incorporated, 2015.

[54] V. John, S. Mita, Zheng Liu, and Bin Qi. Pedestrian detection in thermal images using adaptive fuzzy c-means clustering and convolutional neural networks. In *Machine Vision Applications (MVA), 2015 14th IAPR International Conference on*, pages 246–249, May 2015.

[55] E. Kerr, T.M. McGinnity, and S. Coleman. Material classification based on thermal properties x2014; a robot and human evaluation. In *Robotics and Biomimetics (ROBIO), 2013 IEEE International Conference on*, pages 1048–1053, Dec 2013.

[56] J. A. Kosinski, W. D. Palmer, and M. B. Steer. Unified understanding of RF remote probing. *IEEE Sensors Journal*, 11(12):3055–3063, Dec. 2011.

[57] S.J. Krotosky and M.M. Trivedi. A comparison of color and infrared stereo approaches to pedestrian detection. In *Intelligent Vehicles Symposium, 2007 IEEE*, pages 81–86, June 2007.

[58] P.W. Kruse. *Uncooled Thermal Imaging: Arrays, Systems, and Applications*. SPIE tutorial texts. Society of Photo Optical, 2001.

[59] J. Lagarias, J. Reeds, M. Wright, and P. Wright. Convergence properties of the nelder–mead simplex method in low dimensions. *SIAM J. on Optimization*, 9(1):112–147, May 1998.

[60] S. Lagela, H. Gonzlez-Jorge, J. Armesto, and P. Arias. Calibration and verification of thermographic cameras for geometric measurements. *Infrared Physics and Technology*, 54(2):92 – 99, 2011.

[61] W. Li, D. Zheng, T. Zhao, and M. Yang. An effective approach to pedestrian detection in thermal imagery. In *Natural Computation (ICNC), 2012 Eighth International Conference on*, pages 325–329, May 2012.

[62] G. J. Lockwood, S.L. Shope, J. C. Wehlburg, M. M. Selph, J. M. Jojola, B. N. Turman, and J. A Jacobs. Field tests of x-ray backscatter mine detection. In *Detection of Abandoned Land Mines, 1998. Second International Conference on the (Conf. Publ. No. 458)*, pages 160–163, Oct 1998.

[63] P. Lorenzen. *Multi-modal Image Registration and Atlas Formation*. PhD thesis, Chapel Hill, NC, USA, 2006. AAI3207334.

[64] D.G. Lowe. Object recognition from local scale-invariant features. In *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, volume 2, pages 1150 –1157 vol.2, 1999.

[65] G. Lu, Y. Yan, Li Ren, P. Saponaro, N. Sebe, and C. Kambhamettu. Where am i in the dark: Exploring active transfer learning on the use of indoor localization based on thermal imaging. *Neurocomputing*, pages –, 2015.

[66] R. De Maesschalck, D. Jouan-Rimbaud, and D.L. Massart. The mahalanobis distance. *Chemometrics and Intelligent Laboratory Systems*, 50(1):1 – 18, 2000.

[67] D. Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *SIAM Journal on Applied Mathematics*, 11(2):431–441, 1963.

[68] A. Martone and E. Delp. Characterization of RF devices using two-tone probe signals. In *Proceedings of the 14th IEEE/SP Workshop on Statistical Signal Processing, (SSP'07)*, pages 161–165. Madison, WI,, Aug. 2007.

[69] A. Martone, K. Ranney, G. Hedden, A.and Mazzaro, and D. McNamara. Cognitive processing for nonlinear radar, 2013.

[70] J. McCorkle and L. Nguyen. Focusing of dispersive targets using synthetic aperture radar. Technical Report ARL-TR-305, Army Research Laboratory, August 1996.

[71] G. Mclachlan and D. Peel. *Finite Mixture Models*. Wiley Series in Probability and Statistics. Wiley-Interscience, 1 edition, October 2000.

[72] M. Mirmehdi and M. Petrou. Segmentation of color textures. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(2):142–159, Feb 2000.

[73] T. Moons, L. Van Gool, M. Proesmans, and E. Pauwels. Affine reconstruction from perspective image pairs with a relative object-camera translation in between. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 18(1):77 –83, jan 1996.

[74] C. Nakas, Todd A. Alonzo, and C. Yiannoutsos. Accuracy and cut-off point selection in three-class classification problems using a generalization of the youden index. *Stat Med*, 29(28):2946–2955, Dec 2010. 20809485[pmid].

[75] L. Nalpantidis, G. Sirakoulis, and A. Gasteratos. Review of Stereo Vision Algorithms: From Software to Hardware. *International Journal of Optomechatronics*, 2(4):435–462, 2008.

[76] L. Nguyen. signal and image processing algorithms for the u.s army research laboratory ultra-wideband (uwb) synchronous impulse reconstruction (sire) radar. 2009.

[77] L. Nguyen, F. Koenig, and K. Sherbondy. Augmented reality using ultra-wideband radar imagery. pages 802104–802104–6, 2011.

[78] N. Otsu. A threshold selection method from gray-level histograms. *Systems, Man and Cybernetics, IEEE Transactions on*, 9(1):62–66, Jan 1979.

[79] K. Owens and L. Matthies. Passive night vision sensor comparison for unmanned ground vehicle stereo vision navigation. In *Robotics and Automation, 2000. Proceedings. ICRA '00. IEEE International Conference on*, volume 1, pages 122–131 vol.1, 2000.

[80] R. Porter and M.J. Nye. *The Cambridge History of Science: Volume 5, The Modern Physical and Mathematical Sciences*. Cambridge histories online. Cambridge University Press, 2003.

[81] Surya Prakash, Pei Yean Lee, Terry Caelli, and Tim Raupach. Robust thermal camera calibration and 3d mapping of object surface temperatures, 2006.

[82] B. Qi, V. John, Z. Liu, and S. Mita. Pedestrian detection from thermal images with a scattered difference of directional gradients feature descriptor. In *Intelligent Transportation Systems (ITSC), 2014 IEEE 17th International Conference on*, pages 2168–2173, Oct 2014.

[83] K. Ranney, A. Martone, L. Nguyen, B. Stanton, M. Ressler, D. Wong, F. Koenig, C. Tran, G. Kirose, G. Smith, K. Kappra, and J. Sichina. Recent mti experiments using arl's synchronous impulse reconstruction (sire) radar, 2008.

[84] M. Ressler, L. Nguyen, F. Koenig, D. Wong, and G. Smith. The army research laboratory (arl) synchronous impulse reconstruction (sire) forward-looking radar. pages 656105–656105–12, 2007.

[85] J.-R. Riba, T. Canals, and R. Cantero. Recovered paperboard samples identification by means of mid-infrared sensors. *Sensors Journal, IEEE*, 13(7):2763–2770, July 2013.

[86] M.V. Rohith, G. Somanath, C. Kambhamettu, and C. Geiger. Towards estimation of dense disparities from stereo images containing large textureless regions. In *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, pages 1–5, Dec 2008.

[87] Zhang Ronghua, Chen Hongwu, Zhang Xiaoting, Pan Ruru, and Liu Jihong. Unsupervised color classification for yarn-dyed fabric based on fcm algorithm. In *Artificial Intelligence and Computational Intelligence (AICI), 2010 International Conference on*, volume 1, pages 497–501, Oct 2010.

[88] G.B. Rybicki and A.P. Lightman. *Radiative Processes in Astrophysics*. Physics textbook. Wiley, 2008.

[89] J.M. Sabatier and Ning Xiang. Acoustic-to-seismic coupling and detection of landmines. In *Geoscience and Remote Sensing Symposium, 2000. Proceedings. IGARSS 2000. IEEE 2000 International*, volume 4, pages 1646–1648 vol.4, 2000.

[90] R. Sabzevari, A. Shahri, A.R. Fasih, S. Masoumzadeh, and M.R. Ghahroudi. Object detection and localization system based on neural networks for robo-pong. In *Mechatronics and Its Applications, 2008. ISMA 2008. 5th International Symposium on*, pages 1–6, May 2008.

[91] N. Salamati, C. Fredembach, and S. Ssstrunk. Material classification using color and nir images.

[92] P. Saponaro and C. Kambhamettu. Towards auto-calibration of smart phones using orientation sensors. In *2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 20–26, June 2013.

[93] P. Saponaro, K. Sherbondy, and C. Kambhamettu. Concealed target detection with fusion of visible and infrared. In *Advances in Visual Computing*, volume 8888 of *Lecture Notes in Computer Science*, pages 568–577. Springer International Publishing, 2014.

[94] P. Saponaro, S. Sorensen, A. Kolagunda, and C. Kambhamettu. Material classification with thermal imagery. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.

[95] P. Saponaro, S. Sorensen, S. Rhein, and C. Kambhamettu. Improving calibration of thermal stereo cameras using heated calibration board. In *2015 IEEE International Conference on Image Processing (ICIP)*, 2015.

[96] P. Saponaro, S. Sorensen, S. Rhein, A.R. Mahoney, and C. Kambhamettu. Reconstruction of textureless regions using structure from motion and image-based interpolation. In *2014 IEEE International Conference on Image Processing (ICIP)*, pages 1847–1851, Oct 2014.

[97] Philip Saponaro, Chandra Kambhamettu, Kenneth Ranney, and Anders Sullivan. Concealed target detection using augmented reality with sire radar, Proc. SPIE 8714, Radar Sensor Technology XVII, 87140S (31 May 2013).

[98] S. Sorensen, Kolagunda A., P. Saponaro, and C. Kambhamettu. Refractive stereo ray tracing for reconstruction underwater structures. In *2015 IEEE International Conference on Image Processing (ICIP)*, 2015.

[99] S. Sorensen, P. Saponaro, S. Rhein, and C. Kambhamettu. Multimodal stereo vision for reconstruction in the presence of reflection. In *British Machine Vision Conference (BMVC)*, 2015.

[100] S. Sorensen, P. Saponaro, S. Rhein, and C. Kambhamettu. Multimodal stereo vision for reconstruction in the presence of reflection. In Mark W. Jones Xianghua Xie and Gary K. L. Tam, editors, *Proceedings of the British Machine Vision Conference (BMVC)*, pages 112.1–112.12. BMVA Press, September 2015.

[101] Christopher J. Spain, Derek T. Anderson, James M. Keller, Mihail Popescu, and Kevin E. Stone. Gaussian mixture models for measuring local change down-track in lwir imagery for explosive hazard detection, 2011.

[102] Diga Spiclin, Botjan Likar, and Franjo Pernus. Groupwise registration of multimodal images by an efficient joint entropy minimization scheme. *Trans. Img. Proc.*, 21(5):2546–2558, May 2012.

[103] C. Stagner, A. Conrad, C. Osterwise, D.G. Beetner, and S. Grant. A practical superheterodyne-receiver detector using stimulated emissions. *IEEE Transactions on Instrumentation and Measurement*, 60(4):1461–1468, Apr. 2011.

[104] C. Stauffer and E. Grimson. Learning patterns of activity using real-time tracking. *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, 22:747–757, 2000.

[105] Chris Stauffer and W. E L Grimson. Adaptive background mixture models for real-time tracking. In *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.*, volume 2, pages –252 Vol. 2, 1999.

[106] D. W. Steele, F. S. Rotondo, and J. L. Houck. Radar system for manmade device detection and discrimination from clutter. US Patent 7,830,299,, Nov. 9 2010.

[107] K. Stone, J. M. Keller, M. Popescu, and C. J. Spain. Buried explosive hazard detection using forward-looking long-wave infrared imagery, 2011.

[108] K.H. Strobl and G. Hirzinger. Optimal hand-eye calibration. In *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on*, pages 4647 –4653, oct. 2006.

[109] K.H. Strobl and G. Hirzinger. More accurate camera and hand-eye calibrations with unknown grid pattern dimensions. In *Robotics and Automation, 2008. ICRA 2008. IEEE International Conference on*, pages 1398 –1405, may 2008.

[110] K.H. Strobl and G. Hirzinger. More accurate pinhole camera calibration with imperfect planar target. In *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*, pages 1068 –1075, nov. 2011.

[111] M.J. Sudhamani, M.K. Venkatesha, and K.R. Radhika. Revisiting feature level and score level fusion techniques in multimodal biometrics system. In *Multimedia Computing and Systems (ICMCS), 2012 International Conference on*, pages 881–885, May 2012.

[112] M. Szczodrak and G. Szwoch. An approach to the detection of bank robbery acts employing thermal image analysis. In *Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA), 2013*, pages 297–301, Sept 2013.

[113] A. Torabi, M. Najafianrazavi, and G.-A. Bilodeau. A comparative evaluation of multimodal dense stereo correspondence measures. In *Robotic and Sensors Environments (ROSE), 2011 IEEE International Symposium on*, pages 143–148, Sept 2011.

[114] Atousa Torabi, Guillaume Massé, and Guillaume-Alexandre Bilodeau. An iterative integrated framework for thermal-visible image registration, sensor fusion, and people tracking for video surveillance applications. *Comput. Vis. Image Underst.*, 116(2):210–221, February 2012.

[115] R. Tsai. A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses. *Robotics and Automation, IEEE Journal of*, 3(4):323 –344, august 1987.

[116] J. van Baar, P. Beardsley, M. Pollefeys, and M. Gross. Sensor fusion for depth estimation, including tof and thermal sensors. In *3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT), 2012 Second International Conference on*, pages 472–478, Oct 2012.

[117] Charles Van Loan. *Computational Frameworks for the Fast Fourier Transform*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1992.

[118] S. Vidas, R. Lakemond, S. Denman, C. Fookes, S. Sridharan, and T. Wark. A mask-based approach for the geometric calibration of thermal-infrared cameras. *Instrumentation and Measurement, IEEE Transactions on*, 61(6):1625–1635, June 2012.

[119] P. Viola and M. Jones. Robust real-time object detection. In *International Journal of Computer Vision*, 2001.

[120] O. Wang, P. Gunawardane, S. Scher, and J. Davis. Material classification using brdf slices. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 2805–2811, June 2009.

[121] Philip D. Wasserman. *Advanced Methods in Neural Computing*. John Wiley & Sons, Inc., New York, NY, USA, 1st edition, 1993.

[122] A.H. Watt. *3D Computer Graphics*. Number v. 1 in 3D Computer Graphics. Addison-Wesley, 2000.

[123] J. Weng, P. Cohen, and M. Herniou. Camera calibration with distortion models and accuracy evaluation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 14(10):965–980, October 1992.

[124] AJ. Wilkinson and M.R. Inggs. Radiometry for landmine detection. In *Communications and Signal Processing, 1998. COMSIG '98. Proceedings of the 1998 South African Symposium on*, pages 477–482, Sep 1998.

[125] L.B. Wolff. Polarization-based material classification from specular reflection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 12(11):1059–1071, Nov 1990.

[126] I J. Won and D. Keiswetter. Electromagnetic induction spectroscopy [for landmine amp; uxo detection]. In *Geoscience and Remote Sensing Symposium Proceedings, 1998. IGARSS '98. 1998 IEEE International*, volume 1, pages 517–519 vol.1, Jul 1998.

[127] Lixin Wu, Shanjun Liu, and Yuhua Wu. Infrared imaging detection of water permeation on field large-scale rock relics. *Geoscience and Remote Sensing, IEEE Transactions on*, 49(1):581–590, Jan 2011.

[128] Rongqian Yang, Wei Yang, Yazhu Chen, and Xiaoming Wu. Geometric calibration of ir camera using trinocular vision. *Lightwave Technology, Journal of*, 29(24):3797–3803, Dec 2011.

[129] Yiu-Ming, Harry Ng, and R. Du. Acquisition of 3d surface temperature distribution of a car body. In *Information Acquisition, 2005 IEEE International Conference on*, pages 5 pp.–, June 2005.

[130] N. Youngblood. *The Development of Mine Warfare: A Most Murderous and Barbarous Conduct*. Praeger Security International. Praeger Security International, 2006.

[131] Zhang Yu, Shen Lincheng, Zhou Dianle, Zhang Daibing, and Yan Chengping. Camera calibration of thermal-infrared stereo vision system. In *Intelligent Systems Design and Engineering Applications, 2013 Fourth International Conference on*, pages 197–201, Nov 2013.

[132] X. Zhang and Z. Liu. A survey on stereo vision matching algorithms. In *Intelligent Control and Automation (WCICA), 2014 11th World Congress on*, pages 2026–2031, June 2014.

[133] Z. Zhang. A flexible new technique for camera calibration. *IEEE Trans. Pattern Anal. Mach. Intell*, pages 1330–1334, 2000.

Appendix

PUBLICATIONS

Below is a list of my publications with the thesis related publications highlighted with a bold font.

1. **S. Sorensen, P. Saponaro, S. Rhein, and C. Kambhamettu. Multimodal stereo vision for reconstruction in the presence of reflection. In Mark W. Jones Xianghua Xie and Gary K. L. Tam, editors, *Proceedings of the British Machine Vision Conference (BMVC)*, pages 112.1–112.12. BMVA Press, September 2015**

2. G. Lu, Y. Yan, Li Ren, P. Saponaro, N. Sebe, and C. Kambhamettu. Where am i in the dark: Exploring active transfer learning on the use of indoor localization based on thermal imaging. *Neurocomputing*, pages –, 2015

3. **P. Saponaro, S. Sorensen, A. Kolagunda, and C. Kambhamettu. Material classification with thermal imagery. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015, ©2015 IEEE. Reprinted, with permission.**

4. S. Sorensen, Kolagunda A., P. Saponaro, and C. Kambhamettu. Refractive stereo ray tracing for reconstruction underwater structures. In *2015 IEEE International Conference on Image Processing (ICIP)*, 2015

5. **P. Saponaro, S. Sorensen, S. Rhein, and C. Kambhamettu. Improving calibration of thermal stereo cameras using heated calibration board. In *2015 IEEE International Conference on Image Processing (ICIP)*, 2015, ©2015 IEEE. Reprinted, with permission.**

6. **P. Saponaro, K. Sherbondy, and C. Kambhamettu. Concealed target detection with fusion of visible and infrared. In *Advances in Visual Computing*, volume 8888 of *Lecture Notes in Computer Science*, pages 568–577. Springer International Publishing, 2014 ©2014 Springer.**

7. P. Saponaro, S. Sorensen, S. Rhein, A.R. Mahoney, and C. Kambhamettu. Reconstruction of textureless regions using structure from motion and image-based interpolation. In *2014 IEEE International Conference on Image Processing (ICIP)*, pages 1847–1851, Oct 2014

8. **Philip Saponaro, Chandra Kambhamettu, Kenneth Ranney, and Anders Sullivan. Concealed target detection using augmented reality with sire radar, Proc. SPIE 8714, Radar Sensor Technology XVII, 87140S (31 May 2013) ©2013 Society of Photo Optical Instrumentation Engineers.**

9. P. Saponaro and C. Kambhamettu. Towards auto-calibration of smart phones using orientation sensors. In *2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 20–26, June 2013