8-2014

Utilizing radiation for smart robotic applications using visible, thermal, and polarization images.

Ali H. Mahmoud
University of Louisville

UTILIZING RADIATION FOR SMART ROBOTIC APPLICATIONS
USING VISIBLE, THERMAL, AND POLARIZATION IMAGES

By

Ali H. Mahmoud
B.Sc. in Electrical Engineering, Alexandria University, 2005
M.Sc. in Electrical Engineering, Alexandria University, 2009

A Dissertation
Submitted to the Faculty of the
J. B. Speed School of Engineering of the University of Louisville
in Partial Fulfillment of the Requirements
for the Degree of

Doctor of Philosophy

Department of Electrical and Computer Engineering
University of Louisville
Louisville, Kentucky

August 2014

UTILIZING RADIATION FOR SMART ROBOTIC APPLICATIONS
USING VISIBLE, THERMAL, AND POLARIZATION IMAGES

By

Ali H. Mahmoud
B.Sc. in Electrical Engineering, Alexandria University, 2005
M.Sc. in Electrical Engineering, Alexandria University, 2009

A Dissertation Approved on

July 14, 2014

by the following Dissertation Committee:

_____
Dr. James Graham, Director

_____
Dr. Aly Farag, Co-Director

_____
Dr. Karla Welch

_____
Dr. Christopher Richards

_____
Dr. Prasanna Sahoo

DEDICATION

This dissertation is dedicated to my parents who gave me invaluable educational opportunities.

ACKNOWLEDGEMENTS

ABSTRACT

UTILIZING RADIATION FOR SMART ROBOTIC APPLICATIONS USING
VISIBLE, THERMAL, AND POLARIZATION IMAGES

Ali H. Mahmoud

July 14, 2014

The domain of this research is the use of computer vision methodologies in utilizing radiation for smart robotic applications for driving assistance. Radiation can be emitted by an object, reflected or transmitted. Understanding the nature and the properties of the radiation forming an image is essential in interpreting the information in that image which can then be used by a machine e.g. a smart vehicle to make a decision and perform an action. Throughout this work, different types of images are used to help a robotic vehicle make a decision and perform a certain action. This work presents three smart robotic applications; the first one deals with polarization images, the second one deals with thermal images and the third one deals with visible images. Each type of these images is formed by light (radiation) but in a way different from other types where the information embedded in an image depends on the way it was formed and how the light was generated.

For polarization imaging, a direct method utilizing shading and polarization for unambiguous shape recovery without the need for nonlinear optimization routines is

proposed. The proposed method utilizes simultaneously polarization and shading to find the surface normals, thus eliminating the reconstruction ambiguity. This can be useful to help a smart vehicle gain knowledge about the terrain surface geometry.

Regarding thermal imaging, an automatic method for constructing an annotated thermal imaging pedestrian dataset is proposed. This is done by transferring detections from registered visible images simultaneously captured at day-time where pedestrian detection is well developed in visible images. Histogram of Oriented Gradients (HOG) features are extracted from the constructed dataset and then fed to a discriminatively trained deformable part based classifier that can be used to detect pedestrians at night. The resulting classifier was tested for night driving assistance and succeeded in detecting pedestrians even in the situations where visible imaging pedestrian detectors failed because of low light or glare of oncoming traffic.

For visible images, a new feature based on HOG is proposed to be used for pedestrian detection. The proposed feature was augmented to two state of the art pedestrian detectors; the discriminatively trained Deformable Part based models (DPM) and the Integral Channel Features (ICF) using fast feature pyramids. The proposed approach is based on computing the image mixed partial derivatives to be used to redefine the gradients of some pixels and to reweigh the vote at all pixels with respect to the original HOG. The approach was tested on the PASCAL2007, INRIA and Caltech datasets and showed to have an outstanding performance.

TABLE OF CONTENTS

LIST OF FIGURES

xiv

CHAPTER 1

INTRODUCTION

Electromagnetic radiation [1] includes what is commonly referred to as 'light', in addition to longer and shorter wavelengths. Light plays an important role in the process of image formation which is one of the basic processes in computer vision [2] which is the science and technology of making machines that see. Light (radiation) can be emitted by an object, reflected or transmitted. Understanding the nature and the properties of the light forming an image is essential in interpreting the information in that image which can then be used by a machine e.g. a smart robot to take a decision and perform an action. This work will try to make use of different types of images to help a robot make a decision and perform a certain action. It will deal mainly with polarization images, thermal images and visible images. Each type of these images is formed by light but in a way different from other types where the information embedded in an image depends on the way it was formed and how light was generated.

A brief overview on optics which is essential to understand the process of image formation will be given in the next section of this chapter. This overview will cover geometrical optics and physical optics classically and modernly. Also some basics essential to understand the applications presented in this work will be discussed. Chapter 2 will give a review on the literature related to this work. Chapter 3 will discuss how light is radiated by an object and how this radiation is affected when incident on another

object. It will give the basics of thermal images, polarization images and visible images. These different types of images will be involved in some robotic applications in chapters 5, 6 and 7. The equipment used in this research including different robotic platforms and optical sensors will be discussed in chapter 4.

Polarization imaging will be used to develop a method for unambiguous shape recovery in chapter 5. Knowing whether the reconstructed surface is concave or convex is an important issue in many robotic applications e.g. an autonomous robotic vehicle can be able to discriminate between a ditch (negative obstacle) and a hill (positive obstacle). Proper discrimination leads to a proper decision and a proper action that helps the robot to complete its mission safely.

Chapter 6 proposes an automatic alternative for constructing an annotated thermal imaging pedestrian dataset by transferring detections from registered visible images simultaneously captured at day-time where pedestrian detection is well developed in visible images. Histogram of oriented gradients (HOG) is used for extracting features from images and discriminatively trained part based framework is used as the classifier. The proposed technique was implemented on the robotic vehicle ATRV2.

Chapter 7 proposes a new feature based on HOG to be used with visible images for the purpose of pedestrian detection. The proposed feature is used with two state of the art pedestrian detectors; the discriminatively trained part based framework for pedestrian detection and the integral channel features based framework for pedestrian detection. The proposed method is based on computing the image mixed partial derivatives to be used to redefine the gradients of some pixels and to reweigh the vote at all pixels with respect to the original HOG.

1.1. Optics

Optics is the branch of physics that deals with light which is important to understand the process of image formation. The word light will be used in this context to refer to any portion of the spectrum of the electromagnetic radiation (Figure 1.1) including the visible and invisible portions. Also the words light and radiation will be used interchangeably in this context. Optics can be divided into two categories; geometrical optics and physical optics (wave optics).

In geometrical optics [3], light is considered as rays that propagate in straight lines and are related by Snell's law at interfaces. Geometrical optics is useful in understanding optical instruments and image formation in a simple way provided that the dimensions of the instruments and the objects dealt with are much larger than the wavelength of the used light. This makes it possible to neglect the effect of diffraction that can only be explained through physical optics.

Radio TV	Microwave	Millimetre wave	Mid-infrared	Near-infrared	Visible	Near-ultraviolet	Far-ultraviolet	X-ray	γ-ray
NMR	ESR	Far-infrared							

mm	300	3	1 0.5						
µm				25	2.5	0.77			
nm						770	390	200	10 0.05
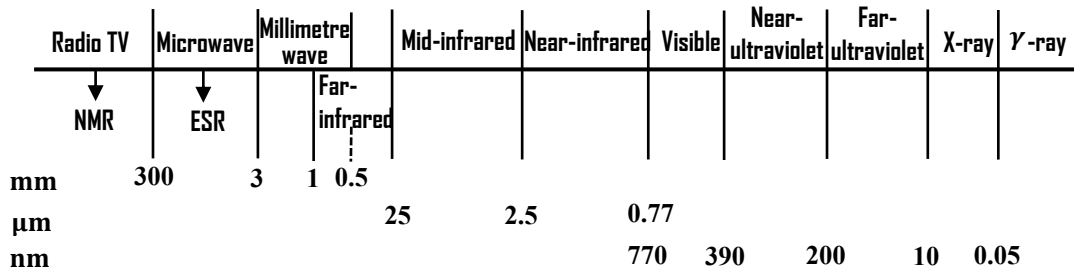
Figure 1.1. Electromagnetic radiation spectrum [1].

Regarding physical optics, there are two views to model the light; the classical view and the modern view. Classically, light can be considered as electromagnetic waves that are propagating continuously [4]. This was mathematically described by James Maxwell through his popular equations known as Maxwell's equations. The classical

3

electromagnetic view of light is used to understand many aspects about light polarization [5] and polarization imaging. For the modern view of physical optics (also known as Quantum Optics), light is considered to be formed of massless particles called photons (energy quanta). Energy is created and destroyed [4] in these photons non-continuously as opposed to the classical view of physical optics. Quantum optics was initiated by Max Planck [6] who discussed the blackbody radiation spectrum followed by Einstein who used what Planck reached to introduce the photon concept. Dealing with light as photons is useful in understanding the absorption and emission of atoms to light [4] which is important to understand thermal imaging. Also, quantum optics succeeded to explain some phenomena that classical physical optics failed to explain as the Ultra-violet catastrophe [3]. Schrodinger [4] stated that any particle is associated with a wave and from here dual nature (particle-wave nature) of photons arose. The wave nature of photons is clear as they prorogate in space, and the particle nature is clear during absorption and emission of photons by atoms [4].

1.1.1. Photons

Sometimes it is useful to deal with light as discrete quantities of energy (quanta). The quantum of light is known as photon and has the following properties:

- The relation between the photon wavelength λ and its frequency ϑ is given by:

$$\lambda\vartheta=c, \tag{1.1}$$

where c is the light speed ($c = 3 \times 10^8 \text{ m s}^{-1}$).

- A photon of frequency ϑ and wavelength λ has an energy E given as:

$$E = h\vartheta = h\frac{c}{\lambda}, \tag{1.2}$$

where h is the Planck's constant. $h = 6.625 \times 10^{-34}$ J s

- If p is the photon momentum, then:

$$p = \frac{h}{\lambda} = \frac{h\vartheta}{c}. \tag{1.3}$$

- For a system having many photons under the same conditions, the photon energy is statistically described using Bose-Einstein distribution.

1.1.1.1. Bose-Einstein distribution

Statistical mechanics is the branch of physics that uses the probability theory to describe the behavior of many identical particles under the same conditions on the average. To do this, the energy E of a particle is described by its probability density function $P(E)$. For photons, $P(E)$ is known as Bose-Einstein distribution and is given as [7]:

$$P(E) = \frac{1}{e^{\frac{E}{KT}}-1}, \tag{1.4}$$

where K is the Boltzmann constant ($K = 1.37 \times 10^{-23}$ J K^{-1}), and T is the temperature in Kelvins. From Equation (1.4), it is clear that $P(E)dE$ is the probability of having a photon with energy between E and $E + dE$. When dealing with radiation, an important quantity is the number of photons per unit volume per unit solid angle for energy between E and $E + dE$ which can be found as:

5

$$n_E(E)dE = P(E)G_E(E)dE = \frac{1}{e^{\frac{E}{KT}}-1}G_E(E)dE, \tag{1.5}$$

where $G_E(E)dE$ is the density of states per unit solid angle for energy between E and $E + dE$. Equation (1.5) can be expressed in terms of wavelength λ by substituting Equation (1.2) into Equation (1.5) which yields to:

$$n_\lambda(\lambda)d\lambda = \frac{1}{e^{\frac{hc}{KT\lambda}}-1}G_\lambda(\lambda)d\lambda, \tag{1.6}$$

where $n_\lambda(\lambda)d\lambda$ is the number of photons per unit volume per unit solid angle with wavelengths between λ and $\lambda + d\lambda$ and $G_\lambda(\lambda)d\lambda$ is the density of states per unit solid angle for wavelengths between λ and $\lambda + d\lambda$. The density of sates will be discussed below in some details.

1.1.1.2. Density of states

In 1927, Heisenberg [8] stated his principle for the uncertainty relation between the position and the momentum of a particle as it is impossible to exactly determine both the position and the momentum of a particle simultaneously. In other words, the more precisely the momentum is determined, the less precisely the position is determined and vice versa. Accordingly, for a system of identical particles, the number of states available to be occupied by a particle per unit volume in an elemental volume in momentum space i.e. between p and $p + dp$ is given by [9, 10, 7]:

$$g_p(p)dp = \frac{4\pi p^2}{h^3}dp, \tag{1.7}$$

where h is the Planck's constant and p is the particle momentum. In case of the particle is a photon, the photon momentum p is related to its wavelength λ by:

$$p = \frac{h}{\lambda}, \tag{1.8}$$

where c is the speed of light. From Equation (1.8), dp is related to $d\lambda$ by:

$$dp = -\frac{h}{\lambda^2} d\lambda. \tag{1.9}$$

The minus sign in Equation (1.9) will be ignored because it just accounts for direction when performing integration [11] and hence dp will be rewritten as:

$$dp = \frac{h}{\lambda^2} d\lambda. \tag{1.10}$$

Substituting Equation (1.8) and Equation (1.10) into Equation (1.7) gives the number of states per unit volume between λ and $\lambda + d\lambda$ as:

$$g_\lambda(\lambda)d\lambda = \frac{4\pi}{\lambda^4} d\lambda. \tag{1.11}$$

Equation (1.11) should be multiplied by 2 to account for the photon two independent planes of polarization [12, 7] leading to:

$$g_\lambda(\lambda)d\lambda = \frac{8\pi}{\lambda^4} d\lambda. \tag{1.12}$$

Divide Equation (1.12) by 4π to gives the number of states per unit volume per unit solid angle for wavelengths between λ and $\lambda + d\lambda$ as:

$$G_\lambda(\lambda)d\lambda = \frac{2}{\lambda^4} d\lambda. \tag{1.13}$$

The number of states per unit volume is sometimes referred to as the density of states and hence $G_\lambda(\lambda)d\lambda$ is then called density of states per unit solid angle between λ and $\lambda + d\lambda$.

1.1.2. Electromagnetic waves and polarization
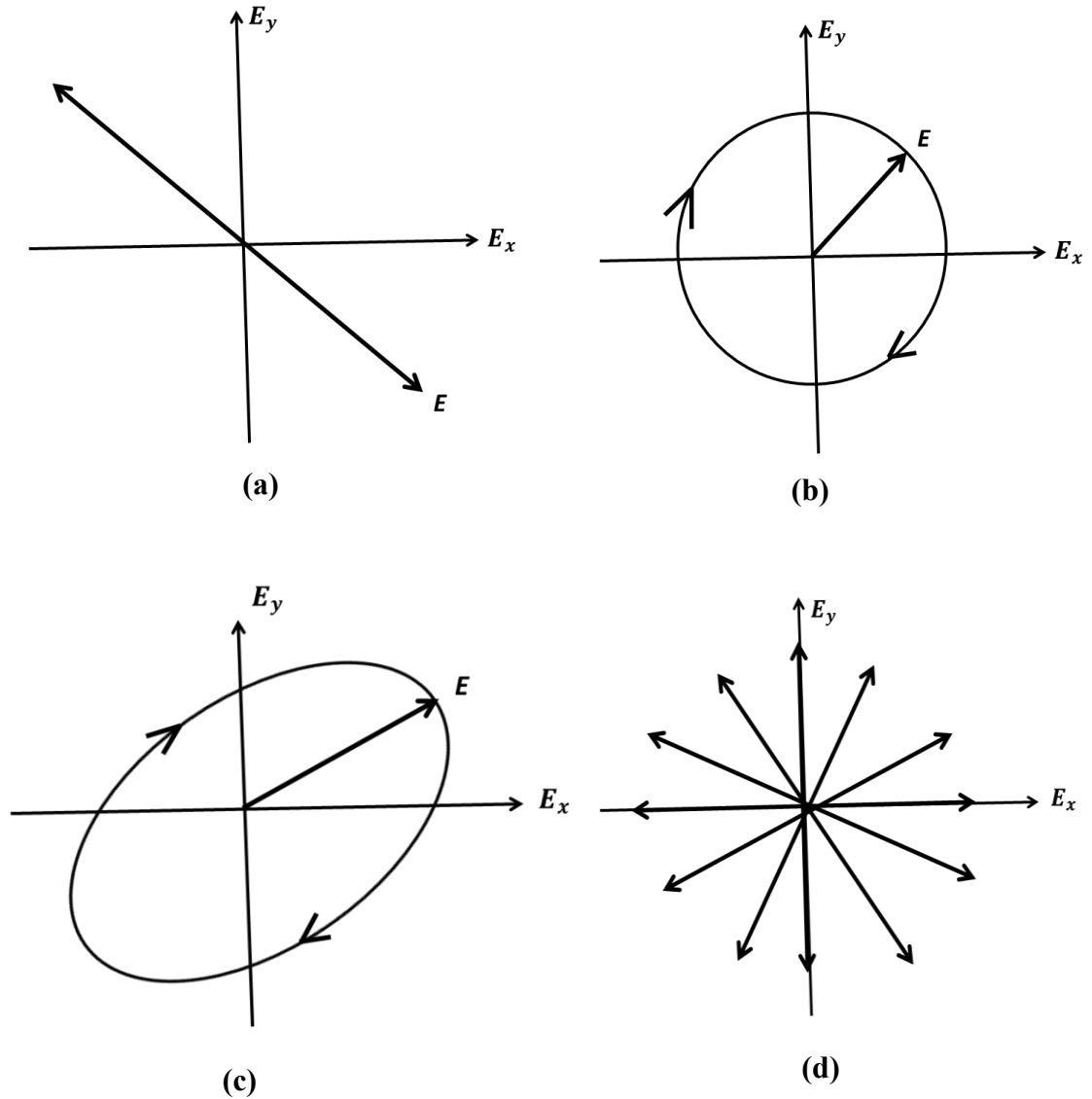
Figure 1.2. (a) Linear polarization. (b) Circular polarization. (c) Elliptical polarization. (d) Unpolarized [4].

The electric field of an electromagnetic wave propagating in the z direction $\boldsymbol{E}(z,t)$ can be represented by two orthogonal optical disturbances [4] $\boldsymbol{E}_x(z,t)$ and $\boldsymbol{E}_y(z,t)$ as:

$$\boldsymbol{E}(z,t) = \boldsymbol{E}_x(z,t) + \boldsymbol{E}_y(z,t), \tag{1.14}$$

where

$$\boldsymbol{E}_x(z,t) = \hat{x}E_{0x}\cos(kz - \omega t), \tag{1.15}$$

$$\boldsymbol{E}_y(z,t) = \hat{y}E_{0y}\cos(kz - \omega t + \delta), \tag{1.16}$$

where k is the propagation vector in the direction of motion, ω is the angular frequency and δ is the phase shift between $\boldsymbol{E}_x(z,t)$ and $\boldsymbol{E}_y(z,t)$. The polarization of the electromagnetic wave is a measure of the direction at which the electric field oscillates in space.

If $\delta = 0$, the orientation of the vector $\boldsymbol{E}(z,t)$ is constant over time and $\boldsymbol{E}(z,t)$ is considered linearly polarized as shown in Figure 1.2(a). If $\delta = 90^o$ and $E_{0x} = E_{0y}$, the orientation of the vector $\boldsymbol{E}(z,t)$ is plotting a circle over time and $\boldsymbol{E}(z,t)$ is considered circularly polarized as shown in Figure 1.2(b). A general case of circular and linear polarization is the elliptical polarization where the orientation of $\boldsymbol{E}(z,t)$ plots an ellipse over time as shown in Figure 1.2(c). This occurs when $\delta \neq 0$ and $E_{0x} \neq E_{0y}$. If δ varies randomly and rapidly [4] with time, this is the case of unpolarized electromagnetic wave e.g., natural light as shown in Figure 1.2(d).

This research is mostly interested in linear polarization of light. Unpolarized light can be converted into linearly polarized light by using an optical component called linear polarizer. It is worth mentioning that liquid crystal displays (LCD) that most people use every day are based on linear polarizers [13, 14, 15, 16].

1.2. Shape recovery using polarization imaging

Recently, polarization imaging has been used for shape recovery of a surface. The recovered shape can be expressed in many ways among which are [17] the depth, the surface normal, or the surface normal zenith angle and the surface normal azimuth angle. Throughout this work, the surface shape is described using the zenith and azimuth angles representation.

A polarization imaging camera is theoretically composed from an intensity charge-coupled device (CCD) camera with a linear polarizer mounted in front of it. Theoretically, in order to obtain a polarization image, the CCD camera behind the polarizer should capture at least three intensity images that are spatially and temporally registered. Mathematical calculations are then done on these intensity images to compute the polarization image. The computed polarization image is composed of three components; Intensity, Degree of polarization and Angle of polarization. If using three CCD intensity images to form a polarization image, these images should be captured when the polarizer is making angles 0°, 45°, and 90° with respect to a given reference. If I_0, I_{45} and I_{90} are the intensities of the CCD images captured at angles 0°, 45°, and 90° respectively, the three components of the corresponding polarization image can be computed as follows [18]:

$$\varphi = \begin{cases} 0.5tan^{-1}\left(\frac{I_0+I_{90}-2I_{45}}{I_{90}-I_0}\right) & if\ I_{90} < I_0 < I_{45} \\ 0.5tan^{-1}\left(\frac{I_0+I_{90}-2I_{45}}{I_{90}-I_0}\right) + 180^o & if\ I_{90} < I_0\ and\ I_{45} < I_0 \\ 0.5tan^{-1}\left(\frac{I_0+I_{90}-2I_{45}}{I_{90}-I_0}\right) + 90^o & otherwise, \end{cases} \qquad (1.17)$$

$$\beta = \frac{I_{90}-I_0}{(I_{90}+I_0)cos2\varphi}, \qquad (1.18)$$

$$I = I_0 + I_{90}, \hspace{5cm} (1.19)$$

where φ, β and I are the angle of polarization, degree of polarization and intensity respectively.

When comparing the image resulting from a polarization camera (CCD camera coupled with a linear polarizer) with the image captured by a CCD intensity camera, it is clear that the polarization camera gives two more components which are φ and β. It was found that β has a one-to-one relation with the zenith angle of a surface normal and φ is related to the azimuth angle of a surface normal (not a one-to-one relation which gives ambiguity) [18]. And since φ and β are related to the surface normal zenith angle and the surface normal azimuth angle, they could be used in shape recovery but up to an ambiguity resulting from the non one-to one relation between the surface normal azimuth angle and φ. An approach is proposed throughout this work to remove this ambiguity as discussed in detail in chapter 5.

1.3. Pedestrian detection using visible and thermal images

Pedestrian detection is an important problem in the design of driving assistance systems that can reduce accidents and save lives. Both visible and thermal cameras have been used to capture images to be used with pedestrian detectors where thermal cameras play an important role during night when visible light is lack. A night vision pedestrian detection approach using thermal cameras is proposed in chapter 6.

The performance of any pedestrian detector is determined mainly by the detector learning algorithm and the used feature representation [19]. Throughout this work the feature representation problem was tackled. A new feature is proposed in chapter 7. This

feature is tested using the learning algorithms of two state of the art pedestrian detectors; the discriminatively trained Deformable Part based model (DPM) [20] and the Integral Channel Features (ICF) [19] using fast feature pyramids [21]. The proposed approach was tested on three publically available visible imaging pedestrian datasets. The performance of the approach was evaluated using Precision-Recall [20] curves which are described in this chapter.

1.3.1. Thermal Imaging versus Visible Imaging for Pedestrians

The concept of thermal imaging arose theoretically from studying the phenomenon of Blackbody radiation [22]. A blackbody at a temperature T above 0 K continuously absorbs and then diffusely reemits photons (light quanta) of all possible wavelengths λ (λ ranges from 0 to ∞). It was found that the number of photons emitted per unit wavelength at a certain wavelength is dependent on the temperature T. This is clear in Planck's law which gives the radiation spectral intensity of a blackbody at temperature [23]. And from Wien displacement law [24], the dominant emitted wavelength is in the visible light range for an extremely high temperature body (thousands of kelvins). For instance a blackbody at a temperature T=5800 K, emits radiation that peaks at wavelength λ =0.499μm. This is why the sun whose surface is at a temperature T=5800 K radiates dominantly green radiation. As the body temperature gets lower, the dominant wavelength moves towards the infrared regions which can be near-infrared (NIR), short-wavelength infrared (SWIR), mid-wavelength infrared (MWIR), and long-wavelength infrared (LWIR). For instance a blackbody at a temperature T=310 K, emits radiation that peaks at wavelength λ =9.345μm. This is why the radiation from a pedestrian whose temperature T=310 K peaks in the LWIR region.

For a real case (non-blackbody), beside the phenomenon of temperature based emission of photons, there is also reflection of photons incident from an external visible light source. Since the emitted photons carry information about the body temperature, succeeding in separating that portion of photons from the reflected portion gives the temperature of the body. Fortunately, at low temperature (few hundreds of kelvins), the dominant wavelength of the emitted photons lies in the infrared region, thus they are band separated from the reflected visible photons. Based on this, thermal infrared cameras can give the temperature of a body.

Thus, a human subject can be considered as a source of two band separated optical signals, one in the visible region of the electromagnetic spectrum and the other in the infrared region. The visible signal can be captured by a visible camera while the infrared signal can be captured by a LWIR thermal camera.

1.3.2. Discriminatively trained deformable part based models (DPM)

DPM described by Felzenszwalb et al. [20] is an object detection system that uses multiscale deformable part models. DPM is based on pictorial structures [25] which consider an object as a collection of parts arranged in a deformable configuration. This deformable configuration specifies the connection between parts. The DPM object detection system has two major stages; the training stage in which a model is built and the detection stage in which the built model is used to detect objects. To build a DPM model, Felzenszwalb et al. [20] used only boxes bounding the objects (e.g. human) in the set of the training images. This means that the parts locations of an object are hidden which they treated as latent variables. The obtained model contains information mainly

about a root filter and part filters. The root filter defines the entire object while the root filters have higher resolution and deal with the object parts.

A filter (either root filter or part filter) is a rectangular array of weight vectors that are used to compute the response or the score of the filter given the feature pyramids of an image. The feature pyramid of an image is an arrangement of a finite number of rectangular arrays called feature maps; each computed at a different image scale. For a certain scale, [20] used a variation of the Histogram of Oriented Gradients (HOG) features [26] to compute the feature map at that scale.

For a certain subwindow of the feature pyramid, the response of a filter is defined as [20] the dot product of the filter and that subwindow. Let H defines a filter of size $w \times h$, F defines a feature pyramid. For a subwindow of size $w \times h$ located at the l-th level of F, if (x, y) denotes the position of the top-left corner of the subwindow, the feature map at that subwindow will be denoted by $F(x, y, l)$. if $Q(x, y, l)$ is a vector obtained by concatenating the features vectors corresponding to $F(x, y, l)$, then the response $R(x, y, l)$ of the filter H at that subwindow is given as [20]:

$$R(x, y, l) = H \cdot Q(x, y, l), \tag{1.20}$$

where $(\cdot)$ is the dot product operator. Placing a root filter at a certain location in the feature pyramid defines a detection window taking into account that the corresponding part filters are put down to the root filter by an octave i.e. at double resolution [20]. The score corresponding to that detection window is found based on the previous equation.

If the object of interest e.g. pedestrian is to be specified using N parts, the corresponding DPM model should contain $(N + 2)$ terms which are $H_0, P_1, P_2, \dots, P_N$ and

b. H_0 is the root filter, P_1 through P_N are the part models and b is a bias term [20]. For the i-th part, the corresponding part model P_i consists of three terms which are the i-th root filter H_i, the anchor position v_i of the i-th part with respect to the root filter and d_i which is the deformation cost of all possible placements of the i-th part relative to its anchor position. Using the information embedded in the model, the *overall score* corresponding to a certain detection is window is given as [20]:

$$overall\ score =$$

$$H_0 \cdot Q(x_0, y_0, l_0) + \sum_{i=1}^{N}(H_i \cdot Q(x_i, y_i, l_i) - d_i \cdot Q_d(dx_i, dy_i)) + b, \qquad (1.21)$$

where (dx_i, dy_i) represents the displacement of the i-th part with respect to its anchor position and is given as [20]:

$$(dx_i, dy_i) = (x_i, y_i) - (2(x_0, y_0) + v_i). \qquad (1.22)$$

$Q_d(dx_i, dy_i)$ are the deformation features given as:

$$Q_d(dx_i, dy_i) = (dx_i, dy_i, dx_i^2, dy_i^2). \qquad (1.23)$$

In order to allow for spatial uncertainty [20], the computed responses of the part filter are transformed which modifies the *overall score* to:

$$overall\ score = H_0 \cdot Q(x_0, y_0, l_0) + \sum_{i=1}^{N}(\max_{(dx_i, dy_i)} (H_i \cdot Q(x_i + dx_i, y_i +$$

$$dy_i, l_i)) - d_i \cdot Q_d(dx_i, dy_i)) + b \qquad (1.24)$$

The subwindows whose *overall score* are above a certain threshold are selected as true positive detections. Figure 1.3 shows a block diagram of a DPM detector at one scale given the input image and the model.
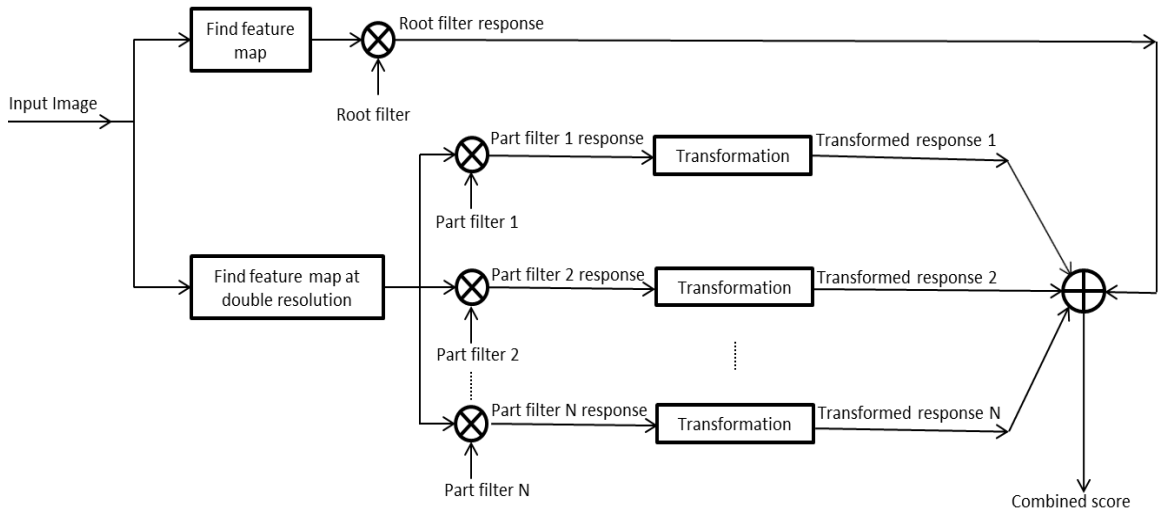
Figure 1.3. DPM detector a one image scale.

So far the DPM model and DPM detection have been discussed but what about the DPM training? The training stage requires having an image dataset that contains annotations for the bounding boxes of the object class of interest. Images containing other classes can be considered as negatives. The first step in the training is to apply Support Vector Machine (SVM) [27] on the positive and negative samples to obtain an initial model that is capable of discriminating between positives and negatives. A detector containing the initial model is then applied on the positive part of the dataset. The resulting detections are warped and then some iterations of SVM are applied on these warped images in which the positions of the root filters are kept latent (hidden) as discussed in details in [20]. The output of these iterations are new bounding boxes that maximizes the performance of the model according to the PASCAL criteria [33]. After that part filters are then initially placed either along the center axis of the root filter or symmetrically around the axis in a manner

16

that maximizes the energy in each root filter. Latent SVM [20] is then used iteratively to enhance the initial filters to obtain a better model.

1.3.3. Integral Channel Features (ICF) using fast feature pyramids

Dollar et al. [19] investigated using integral channel features (ICF) and Adaboost classifiers [28] for pedestrian detection. In ICF, multiple registered channels are extracted from the input image using either linear or nonlinear transformations. Features are then extracted from these channel using integral images [29].

If having a gray scale input image, a simple example for a channel is the image itself while if having an RGB color input image, the channels can be simply the R, G and B channels. More complicated channel examples are histogram of oriented gradients, normalized gradient magnitude, and LUV color channels. For a certain channel, the first-order channel feature is defined as the sum of pixel inside a rectangular region [19], while higher order channel features are computed by performing some operations (e.g. difference) on multiple first-order channel features. It was found in [19] that a detector in which more than a channel are combined can outperform detectors that use individual channels. For example combining histogram of oriented gradients, normalized gradient magnitude, and LUV outperformed using histogram of oriented gradients alone.

For the pedestrian detector to handle different image scales, a feature pyramid is computed at a finite number of scales. Typically 4 to 12 scales per octave are used [21]. This multi-scaling can be computationally expensive. Dollar et al. [21] discussed a modification to the system described in [19] to speed up the computation as they used fast feature pyramids instead of regular feature pyramids used in [19]. Fast feature

17

pyramids are based on computing features at a number of scales lower than that of a typical feature pyramid and then using extrapolation to estimate the features at other scales. Performing extrapolation is faster than computing features at every scale while almost preserving the same performance.

In this work, the fast feature pyramid ICF based pedestrian detector presented in [21] with channels of histogram of oriented gradients, normalized gradient magnitude, and LUV will be used to illustrate the performance of a feature proposed in this work.

1.3.4. Popular pedestrian datasets

In this section, a brief overview will be given on the most popular publically available pedestrian datasets that were used throughout this work.

1.3.4.1 INRIA Person Dataset

INRIA person dataset contains 1804 64×128 images of cropped humans. It was collected as part of work on detection of upright people in images and video [26]. The dataset contains image from GRAZ01 dataset, personal digital image collected over a long period of time and few Google images. Marking in the images is only for upright persons with height >100 pixels.

1.3.4.2. Caltech Pedestrian Dataset

The Caltech Pedestrian Dataset [30, 31] consists of approximately 10 hours of 30Hz video taken from a vehicle driving through regular traffic in an urban environment at a resolution of 640x480. Annotation was done to about 250,000 frames (in 137 approximately minute long segments) resulting in a total of 350,000 bounding boxes and

2300 unique pedestrians annotations. The annotation includes temporal correspondence between bounding boxes and detailed occlusion labels.

1.3.4.3. The PASCAL Visual Object Classes (VOC) Challenge

Everingham et al. [32, 33] introduced the PASCAL visual object classes (VOC) challenge which is a benchmark in visual object category recognition and detection. This challenge was organized annually starting from 2005. The challenge consists of a publicly available dataset of images and annotation, together with standardized evaluation software; and an annual competition and workshop. A new dataset with ground truth annotation was released each year since 2006. The datasets contains several annotated classes among which is the 'person' class. This work will focus on using the VOC2007 dataset [32], since the annotations for both the train set and the test set are publically available.

1.3.5. Performance evaluation using Precision-Recall curves

The performance of a detector will be evaluated using precision-recall curves which are denoted in some of the literature as receiver operating curves (ROC). Available pedestrian datasets contain images and annotation bounding boxes which represent the ground truth for a detection system [20]. When testing a detector, the input to the system is some images and the output is a set of bounding boxes with corresponding scores [20]. These score can be thresholded at different values to plot the precision-recall curve where

$$Precision = \frac{tp}{tp+fp}, \tag{1.25}$$

$$Recall = \frac{tp}{tp+fn}, \tag{1.26}$$

where tp is the number of true positives (correct detections), fp is the number of false positives (unexpected detections), and fn is the number of false negatives (missing detections). At a certain threshold, the precision represents the fraction of the bounding boxes that are correct detections while the recall represents the fraction of the pedestrians in the image that are detected correctly.

Throughout this work, a PASCAL measure has been used to determine the detection rates [33]. If there is an overlap between a detected bounding box and the ground truth bounding box and this overlap is more than 50%, this will be considered as a correct detection, otherwise it is a false positive detection. For a certain ground truth bounding box, if there are more than one overlapping detection bounding boxes, only one of them is counted. A system is scored by the average precision AP of its precision-recall curve across a test set [20].
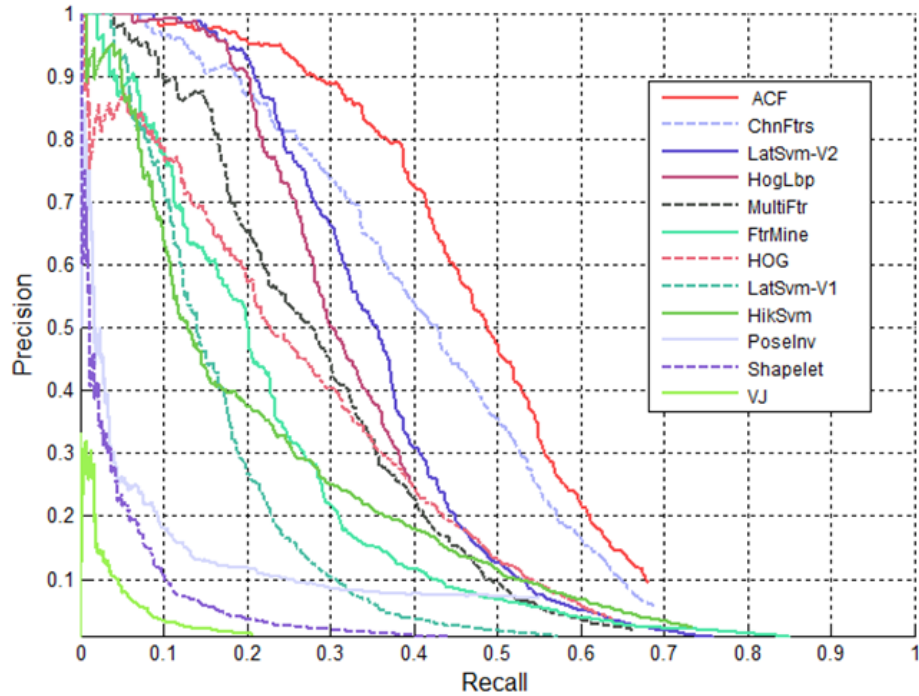
Figure 1.4. State of the art precision-recall curves for Caltech pedestrian dataset.
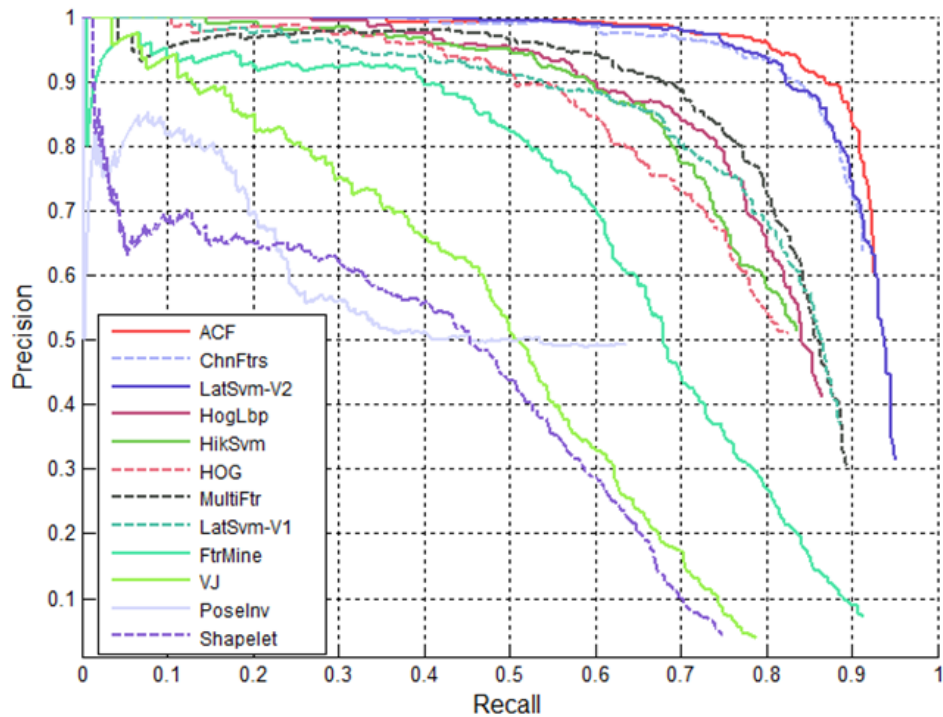
Figure 1.5. State of the art precision-recall curves for INRIA dataset.

Figure 1.4 gives the precision-recall curves of the state of the art pedestrian detectors [19, 21] for the Caltech pedestrian dataset for which the performance is evaluated every 30th frame. Figure 1.5 gives the precision-recall curves of the state of the art pedestrian detectors for the INRIA dataset respectively. The higher the curve, the better the performance.

CHAPTER 2

LITERATURE REVIEW

This chapter gives a review on the literature related to the proposed work. It focuses mainly on shape recovery from polarization, and pedestrian detection for driving assistance.

2.1. Shape recovery from polarization

Light plays an important role in the process of image formation, as an image can be formed when light is reflected from an object in space into an image sensor. In the process of image formation, the object 3D shape is converted into 2D information. The 3D shape information of the object is carried by the light wave propagating from the object to the image sensor, and hence the properties associated with the light can be exploited to recover the 3D shape of the object after the process of image formation. From the prospective of wave propagation, light can be viewed as an electromagnetic wave that is characterized by several properties among which are intensity, spectrum, and polarization. The intensity is a measure of the magnitude of the oscillating electric field associated with the light, the spectrum is the frequency range over which the electric field oscillates, and the polarization is a measure of the direction at which the electric field oscillates in space. Recently, the intensity, spectrum, and polarization properties are being used to develop polarization cameras that are capable not only of measuring the spectrum and magnitude of the electric field associated with the light, but also its

direction in space. While the intensity and the spectrum can give information about the materials, polarization can give information about surface shape, conductivity, refractive indices, and roughness [34]. Polarization imaging has been used for many applications in machine vision [35], remote sensing [34], biomedical imaging [36], industrial control [37] and mud detection [38].

Of concern in this chapter are reported methods for shape recovery based on polarization. Polarization imaging for shape recovery of textureless specular surfaces was proposed in [39]. It was also used to determine the shape for a transparent object [40]. Atkinson and Hancock [18] exploited polarization by surface reflection, using images of smooth objects, to recover surface normals and, hence, height. They focused on dielectric surfaces which diffusely polarize light transmitted from the dielectric body into the air. They modeled the diffuse polarization of the reflection process using a transmitted radiance sinusoid curve and the Fresnel transmission theory. Degree of diffuse polarization was used to estimate the zenith of the surface normals. The azimuth was estimated using the phase of diffuse polarization but up to the ambiguity. Disambiguation was done using zenith angle ranking. The same authors [41] used diffuse polarization in conjunction with shading from two views to estimate the shape. The surface normals were calculated from diffuse polarization for each view independently. The results were fed into a shading algorithm to be enhanced. In [42], the authors used multi-spectral polarization images from a single view to get a robust estimate for the surface normal azimuth in the presence of noise. Along the same line, a method for simultaneous estimation of surface orientation and index of refraction from the spectral variation of the phase of polarization was proposed in [43].

23

This research proposed a direct method for shape recovery from shading and polarization [45]. The observation underpinning the proposed method is that shading information provides one equation on the surface normal (equivalent to two unknowns) at each image pixel. Thus in total, there are twice as many unknowns as the number of equations. However polarization information can double the set of available equations without increasing the number of unknowns, thus rendering the shape recovery problem from both polarization and shading well-posed.

The proposed method utilizes the two sources of information, shading and polarization, for shape recovery. A few methods in literature [41, 42] share the same objective with the proposed method. However the proposed method is novel in several aspects. First, it relies on one view and one spectral imaging band, while other methods try to make use of multiple spectrum bands [42, 43]. As such, the proposed method is considered simpler and needs less expensive imaging setup. Second, the proposed method utilizes simultaneously polarization and shading to find the surface normal, while earlier method [42] uses separately polarization information to estimate the azimuth angle of the normal, and shading to find the zenith angle. However relying only on polarization [18, 42, 43] to obtain the azimuth angle results in an ambiguity in the estimation. To disambiguate the azimuth angle, earlier methods either assumed always convexity on the surface under observation [42, 43] or resorted to a zenith angle ranking procedure [18] which often work only on simple cases. The proposed method solves this ambiguity by making use of the complementary shading and polarization information to estimate the azimuth angle. It is well-know also that traditional shape from shading (SFS) suffers from a convex/concave ambiguity [17, 44], which was shown to be resolved if the attenuation

24

of the illumination due to distance is taken into consideration [46]. Due to the use of polarization information as well, the proposed method can resolve the convex/concave ambiguity of the shading component without having to complicate the illumination model. Thus the method is shown to successfully recover the shape of convex and concave surfaces using the simpler classical assumption of a single distant light source.

2.2. Pedestrian detection for driving assistance

The United States Department of Transportation - Federal Highway Administration (FHWA) [47] reported that each year, pedestrian fatalities comprise about 12 percent of all traffic fatalities and there are approximately 4,000 pedestrian deaths. Moreover, 59,000 pedestrians are injured in roadway crashes. This addresses the problem of pedestrian detection as an important problem in the design of driving assistance systems to achieve a livable community provides safe and convenient transportation choices to all citizens [47]. This section provides a brief survey on the related works to this problem.

Yanwu Xu et al. [48] defined the problem detecting sudden pedestrian crossing. They proposed a video-based driving assistance framework that detects crossing pedestrians before they fully enter the camera view. Their framework consists of three levels; a local level followed by a frame level followed by a video level. In the local level, they used sparse sliding window sampling with a local binary pattern (LBP) difference based motion filter to detect the regions that might contain pedestrians. The output of the local level is then fed for more verification into the frame level which uses a pair of generic pedestrian detectors trained with half-sized pedestrian samples. The output

of the frame level is then fed to the third level to perform appearance and motion based spatiotemporal refinement in order to reduce the false positives.

Olmeda et al. [49] proposed a pedestrian detection system based on discrete features in infrared images. They detected the region of interest by searching for descriptors based on the histogram of the phase congruency orientation which is matched with defined regions of the body of a pedestrian. The detected region of interest is then classified whether to be a pedestrian or a non-pedestrian using a support vector machine (SVM) classifier. Olmeda et al. [49] then compared their approach to the standard HOG algorithm presented in [26].

Kidono et al. [50] proposed a pedestrian recognition system by the fusion between high definition LIDAR and a single camera. They extracted the pedestrian candidates from the two sensors in parallel by SVM classifiers. They used the object information provided by LIDAR efficiently to reduce the computational burden and the number of false positives in the vision-based detection.

Itoh et al. [51] discussed a driver assistance system that informs the driver of the desired steering direction to avoid a collision with a pedestrian in case of emergency. Yun Luo et al. [52] presented a review on pedestrian detection using active night vision sensors operating in the near infrared region of the electromagnetic spectrum and passive night vision sensors operating in the far infrared region of the electromagnetic spectrum. They then proposed a technique to enhance the pedestrian detection systems that use active sensors operating in the near infrared region of the electromagnetic spectrum.

Their technique involves using laser illumination and a camera equipped with a narrow bandpass filter to remove the excess ambient illumination.

Liang Zhao et al. [53] presented an algorithm for detecting pedestrians in a cluttered scene from a pair of moving cameras. They used stereo-based segmentation and neural network-based recognition as foreground objects were first detected through foreground/background segmentation based on stereo vision. Each object is then classified as pedestrian or non-pedestrian by a trained neural network. Inumaru et al. [54] proposed a method for pedestrian detection using a stereo video a voting using Gaussian distribution. They represented the detected object region by HOG features and it was classified if the region contains a human by SVM.

Jayasumana et al. [55] introduced a family of provably positive definite kernels on the Riemannian manifold of Symmetric Positive Definite (SPD) matrices that could be used to design Riemannian extensions of existing kernel-based algorithms, such as SVM and kernel k-means. They used their Riemannian kernel with kernel SVM Multiple Kernel Learning (MKL) for pedestrian detection and tested it on the INRIA person dataset.

Bar-Hillel et al. [56] introduced an approach for learning part-based object detection through feature synthesis (FS). They empirically evaluated their method on the INRIA dataset, the Caltech pedestrian dataset and a dataset of children images with difficult poses.

Levi et al. [57] presented a part-based object detection algorithm with hundreds of parts performing real-time detection which is an accelerated version of the FS method

[56]. Their algorithm uses a coarse-to-fine strategy where a coarse part-based detector that is used to eliminate most image regions is followed by a fine detector to detect the object in the remaining regions. To speed up the coarse level, they developed an approximate nearest neighbor search algorithm, termed "KD-Ferns", which compares each image location to only a subset of the model parts. They evaluated their method on the INRIA dataset and the Caltech pedestrian benchmark which showed to maintain almost the same accuracy performance of the original FS at a speed increase of more than four times.

Suard et al. [58] presented a method for pedestrian detection using stereo infrared images. They used HOG features combined with SVM to build their classifier. Chang et al. [59] also presented a method for pedestrian detection using infrared thermal images. They used HOG features representation combined with adaboost to perform the detection. Their approach contains some preprocessing techniques as foreground-background subtraction, morphology, and aspect ratio filtering.

Ouyang and Wang [60] proposed a probabilistic framework for detecting pedestrians that appear in a group. Single pedestrian detection was performed by the aid of multi-pedestrian detection. Multi-pedestrian detection learning was done using discriminatively trained deformable part models presented in [20] where each single pedestrian is considered as a part. Their framework models the configuration relationship between multi-pedestrian detection and single-pedestrian detectors. They validated their approach using the Caltech, TUD-Brussels and ETH datasets.

Commercially, pedestrian detection systems have been involved in the vehicle industry. BMW [61] uses night vision with dynamic light spot to detect pedestrians in the dark and selectively illuminates them. As soon as unlit pedestrians on a collision course are detected by the remote infrared sensors, the system directs two separately controllable high-performance light spots onto them without any unpleasant dazzling effect. The light very effectively warns drivers and pedestrians of a potentially hazardous situation. On the other hand, Mercedes-Benz Night View Assist PLUS [62] uses invisible infrared beams and a special camera, to provide real-time view of the dark road ahead. It can point out and even shine the headlamps toward pedestrians it detects.

Recently, several approaches for pedestrian detection have been investigated using discriminatively trained part based models [20]. In these approaches, an object detection system is achieved using mixtures of multi-scale deformable part models that are discriminatively trained using support vector machines (SVM) requiring only the knowledge of the box bounding the object in the image. Felzenszwalb et al. [20] investigated using the histogram of oriented gradients (HOG) features introduced by Dalal and Triggs [26]. The basic idea behind the HOG is that the object local appearance and shape can be described using the distribution of the local intensity gradients or edge directions without the need to precisely know the position of the corresponding gradient or edge [26].

Ren and Ramanan [63] replaced the HOG feature in the framework of [26] by histograms of sparse codes (HSC). In their approach, they formed local histograms by computing sparse codes with dictionaries learned using K-SVD which outperformed the HOG based approach.

Jun et al. [64] introduced the local gradient patterns feature (LGP) and the binary histogram of oriented gradients feature (BHOG). The LGP makes the local intensity variations along the edge components robust by assigning one if the gradient of a neighboring pixel is the average of the gradients of the eight neighborhood pixels and zero otherwise. For the BHOG, one is assigned if the values of the histogram bin is greater than the average value of the total histogram binsand zero otherwise. They used their approach for human detection using the INRIA [26] and the Caltech [30, 31] human databases.

Dollar et al. [19] proposed using integral channel features and Adboost classifiers [28] for detecting pedestrians. Their detector involved using a feature pyramid to handle different image scales which can be computationally expensive. In [21, 65] Dollar et al. discussed using fast feature pyramid that uses extrapolation instead of conventional feature pyramid to decrease the computational cost. Benenson et al. [66] reverted the pedestrian detector discussed in [65] to avoid resizing the input images at multiple scales.

CHAPTER 3

RADIATION: EMISSION AND REFLECTION

The surface of any object at a temperature above 0 K emits radiation. The emitted radiation carries information about the object temperature and its surface properties. The theory behind that emission of radiation started by studying the radiation of a blackbody (ideal body that does not exist in nature) then was developed to include non-blackbodies (real surfaces) e.g. the sun, human being, light bulbs, terrain surfaces, etc. If the radiation emitted by a body is incident on a surface and reflected, the properties of the reflected radiation e.g. polarization is affected by the surface and when analyzed can give information about the surface e.g. its shape.

3.1. Blackbody radiation

A blackbody at a temperature T above 0 K continuously absorbs and then diffusely reemits photons (light quanta) of all possible wavelengths (λ ranges from 0 to ∞). It was found that the number of photons emitted per unit wavelength at a certain wavelength is dependent on the temperature T of the body. This is quantitatively clear in Planck's law for blackbody radiation [23].

3.1.1. Planck's Law

Planck's law gives the blackbody radiation spectral intensity $i_{\lambda,b}(\lambda, T, \theta, \varphi)$ [23] which is defined as the emitted power in the direction (θ, φ) at wavelength λ per unit

31

wavelength per unit solid angle per unit area of the body surface projected in the direction of emission when the blackbody is at a temperature T, where θ and φ are the zenith and azimuth angles of the spherical coordinates system respectively. Since a blackbody emits diffusely, $i_{\lambda,b}(\lambda, T, \theta, \varphi)$ will be independent of the direction (θ, φ). Planck's law will be derived below.

For a blackbody with an elemental surface area dA as shown in Figure 3.1, $dA' = dA\cos\theta$ is the projection of a dA in the direction (θ, φ). Hence, from the definition of the blackbody spectral intensity, the power η emitted from dA' into a solid angle $d\omega$ for wavelengths between λ and $\lambda + d\lambda$ at a temperature T is:

$$\eta = i_{\lambda,b}(\lambda, T, \theta, \varphi) \, dA' d\omega d\lambda \, . \tag{3.1}$$
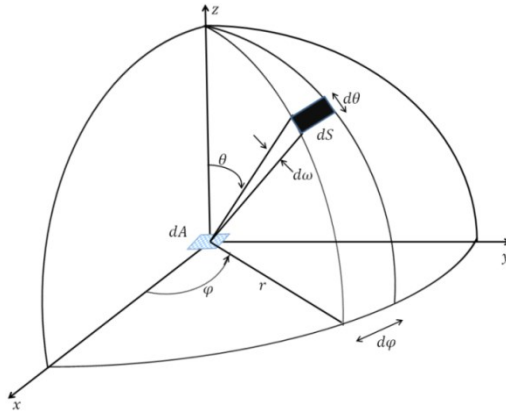
Figure 3.1. Geometry used to derive Planck's law.

Imagine having an infinitesimal cuboid volume of height dl above dA'. A photon traverses that infinitesimal with the speed of light c and thus the time it takes is $\frac{dl}{c}$ Seconds [67]. Hence, the energy enclosed by the infinitesimal volume is $\eta\frac{dl}{c}$ and thus the energy per unit volume between λ and $\lambda + d\lambda$ is:

$$\eta\frac{dl}{c} \times \frac{1}{dl dA'} = \eta\frac{1}{c dA'} = i_{\lambda,b}(\lambda, T, \theta, \varphi) \, d\omega d\lambda \frac{1}{c} \tag{3.2}$$

32

Dividing Equation (3.2) by $d\omega$ gives the energy per unit volume per unit solid angle as $i_{\lambda,b}(\lambda, T, \theta, \varphi) \, d\lambda \frac{1}{c}$ which can be evaluated by multiplying $n_{\lambda}(\lambda)d\lambda$ (Equation (1.6)) by the energy of 1 photon at wavelength λ (Equation (1.2)) as follows:

$$i_{\lambda,b}(\lambda, T, \theta, \varphi) \, d\lambda \frac{1}{c} = n_{\lambda}(\lambda)d\lambda \, h\frac{c}{\lambda}. \tag{3.3}$$

Substituting Equation (1.6) into Equation (3.3) and rearranging gives:

$$i_{\lambda,b}(\lambda, T, \theta, \varphi) = \frac{hc^2}{\lambda e^{\frac{hc}{KT\lambda}} - 1} G_{\lambda}(\lambda). \tag{3.4}$$

Substituting Equation (1.13) into Equation (3.4) gives:

$$i_{\lambda,b}(\lambda, T, \theta, \varphi) = \frac{2hc^2}{\lambda^5} \frac{1}{e^{\frac{hc}{KT\lambda}} - 1}. \tag{3.5}$$

Equation (3.5) is known as Planck's law which gives the emitted power in the direction (θ, φ) at wavelength λ per unit wavelength per unit solid angle per unit area of the body surface projected in the direction of emission when the blackbody is at a temperature T. To find the quantity given in Equation (3.5) but per unit area of the body surface instead of per unit area of the body surface projected in the direction of emission, Equation (3.5) should be multiplied by $\frac{dA'}{dA} = cos\theta$ which gives:

$$e_{\lambda,b}(\lambda, T, \theta, \varphi) = i_{\lambda,b}(\lambda, T, \theta, \varphi)cos\theta = \frac{2hc^2}{\lambda^5} \frac{1}{e^{\frac{hc}{KT\lambda}} - 1} cos\theta, \tag{3.6}$$

where $e_{\lambda,b}(\lambda, T, \theta, \varphi)$ is the emitted power in the direction (θ, φ) at wavelength λ per unit wavelength per unit solid angle per unit area of the body surface when the blackbody is at

a temperature T. $e_{\lambda,b}(\lambda, T, \theta, \varphi)$ is known as the directional spectral emissive power for the blackbody [23].

3.1.2. Wien's displacement law

If the blackbody radiation spectral intensity $i_{\lambda,b}(\lambda, T, \theta, \varphi)$ is plotted versus wavelength λ for various temperatures T, it can be observed that the peak of $i_{\lambda,b}(\lambda, T, \theta, \varphi)$ moves to the right when the temperature T of the body is decreased. This means that as the temperature decreases, the wavelength at which the peak occurs increases. This is clear when comparing $i_{\lambda,b}(\lambda, 6000 \text{ K}, \theta, \varphi)$ in Figure 3.2 with $i_{\lambda,b}(\lambda, 200 \text{ K}, \theta, \varphi)$ in Figure 3.3. At $T = 6000$ K, the peak occurred in the visible region of the electromagnetic spectrum while at $T = 200$ K, the peak occurred in the infrared region of the electromagnetic spectrum. From this, it is concluded that at relatively low temperatures, the blackbody radiation is concentrated in the infrared region while at relatively high temperatures, the radiation is concentrated in the visible region. A good example for this is the comparison between the sun and the human if they are roughly considered, for now, as blackbodies. The sun whose temperature is around 5800 K peaks at 0.5 μm which is in the green visible part of the electromagnetic spectrum [11]. On the other hand, a human whose temperature is around 310 K peaks at 9.345 μm which is in the infrared region of the electromagnetic spectrum. Wien [11] related the wavelength of the peak λ_{max} to the temperature of the blackbody T in his displacement law. Wien's displacement law states that the product of the wavelength at which the peak of $i_{\lambda,b}(\lambda, T, \theta, \varphi)$ occurs and the temperature T is a universal constant equal to 2.897×10^{-3} m K as follows:

$$\lambda_{max} T = 2.897 \times 10^{-3} \text{ m K.} \tag{3.7}$$

Another observation is that the peak of the blackbody radiation spectral intensity $i_{\lambda,b}(\lambda, T, \theta, \varphi)$ moves downwards when decreasing the blackbody temperature. This means that the radiation power emitted from blackbodies of high temperature is higher than that emitted from blackbodies at low temperatures. This is clear when comparing the radiation power emitted from the sun to that emitted from a human if they are roughly considered again, for now, as blackbodies.

3.1.3. Stefan-Boltzmann law

So far, this chapter derived the blackbody radiation spectral intensity $i_{\lambda,b}(\lambda, T, \theta, \varphi)$ and directional spectral emissive power $e_{\lambda,b}(\lambda, T, \theta, \varphi)$ but the total power emitted by the blackbody at all wavelengths and all directions per unit area of its surface was not computed. In order to do this, consider the hemisphere shown in Figure 3.1 around the blackbody elemental area dA. The power emitted per unit surface area in the direction (θ, φ) in a solid angle $d\omega$ for wavelengths between λ and $\lambda + d\lambda$ is given as:

$$de_b = e_{\lambda,b}(\lambda, T, \theta, \varphi) \, d\omega \, d\lambda. \tag{3.8}$$

Substituting for $e_{\lambda,b}(\lambda, T, \theta, \varphi)$ by its value from Equation (3.6) and for $d\omega$ by the elemental solid angle in the spherical coordinates system which is $sin\theta d\theta d\varphi$ gives:

$$de_b = \frac{2hc^2}{\lambda^5} \frac{1}{e^{\frac{hc}{KT\lambda}} - 1} cos\theta sin\theta d\theta d\varphi d\lambda. \tag{3.9}$$

The radiation power per unit area of the source surface emitted in a hemisphere is found by integrating Equation (3.9) for $\theta \in [0, \frac{\pi}{2}]$, $\varphi \in [0, 2\pi]$ and $\lambda \in [0, \infty[$ as:

35

$$e_b = \int_{\lambda=0}^{\infty} \int_{\varphi=0}^{2\pi} \int_{\theta=0}^{\frac{\pi}{2}} \frac{2hc^2}{\lambda^5} \frac{1}{e^{\frac{hc}{KT\lambda}}-1} cos\theta sin\theta d\theta d\varphi d\lambda. \qquad (3.10)$$
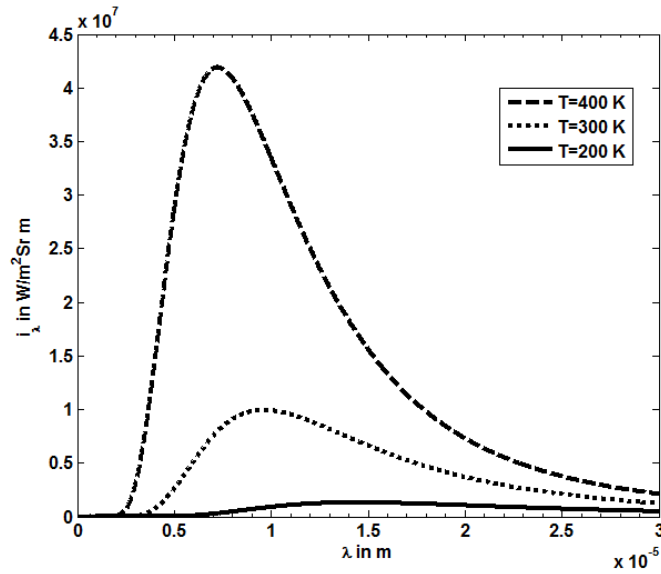
Figure 3.2. Planck's law at low temperatures.
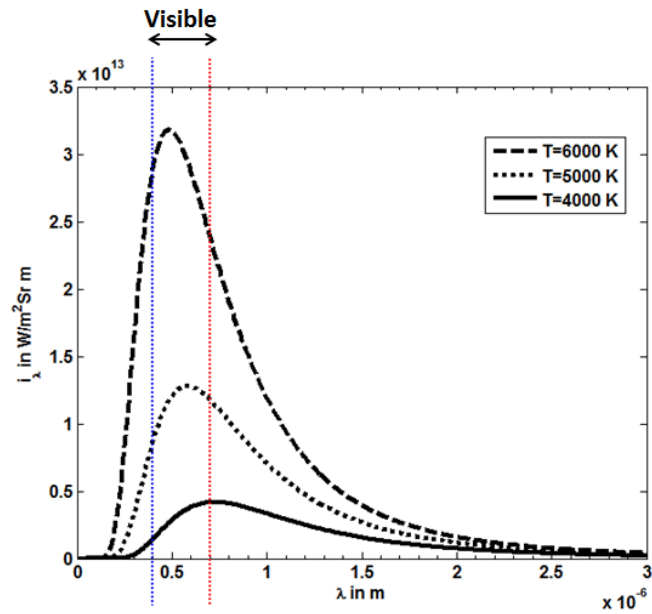
Figure 3.3. Planck's law at high temperatures.

Simplifying Equation (3.10) gives:

36

$$e_b = 2hc^2\pi \int_{\lambda=0}^{\infty} \frac{1}{\lambda^5} \frac{1}{e^{\frac{hc}{KT\lambda}}-1} d\lambda. \tag{3.11}$$

Now define η_1 as:

$$\eta_1 = \int_{\lambda=0}^{\infty} \frac{1}{\lambda^5} \frac{1}{e^{\frac{hc}{KT\lambda}}-1} d\lambda. \tag{3.12}$$

To solve for η_1, a dummy variable s is used such that:

$$s = \frac{hc}{KT\lambda}. \tag{3.13}$$

From Equation (3.13),

$$\lambda = \frac{hc}{KTs}, \tag{3.14}$$

and

$$d\lambda = -\frac{hc}{KT} \frac{1}{s^2} ds. \tag{3.15}$$

Also $\lambda = [0:\infty[$ corresponds to $s =]\infty:0]$.

The minus sign in Equation (3.15) can be used to change the direction of integration, thus substituting Equation (3.14) and Equation (3.15) into Equation (3.12) gives:

$$\eta_1 = \int_{s=0}^{\infty} \frac{K^5 T^5 s^5}{h^5 c^5} \frac{1}{e^s-1} \frac{hc}{KTs^2} ds. \tag{3.16}$$

Simplifying Equation (3.16) gives:

$$\eta_1 = \frac{K^4 T^4}{h^4 c^4} \int_{s=0}^{\infty} \frac{s^3}{e^s-1} ds. \tag{3.17}$$

From integration tables, $\int_{s=0}^{\infty} \frac{s^3}{e^s-1} ds = \frac{\pi^4}{15}$. Substituting into Equation (3.17) gives:

$$\eta_1 = \frac{\pi^4 K^4 T^4}{15 h^4 c^4}. \tag{3.18}$$

37

Substituting Equation (3.12) and Equation (3.18) into Equation (3.11) gives:

$$e_b = \frac{2\pi^5 K^4 T^4}{15 h^3 c^2}. \tag{3.19}$$

Now define $\sigma = \frac{2\pi^5 K^4 T^4}{15 h^3 c^2} = 5.67 \times 10^{-8}\ \mathrm{W\,m^{-2}K^{-4}}$, hence

$$e_b = \sigma T^4. \tag{3.20}$$

σ is known as Stephan-Boltzmann constant and Equation (3.20) is known as Stephan-Boltzmann law for blackbody radiation which indicates that the power of the radiation emitted in a hemisphere from a blackbody at a temperature T per unit area of the body surface is proportional to T^4. Stephan-Boltzmann law is illustrated in Figure 3.4 as e_b is plotted versus T for $T\epsilon[0{:}\,700]$ Kelvins.
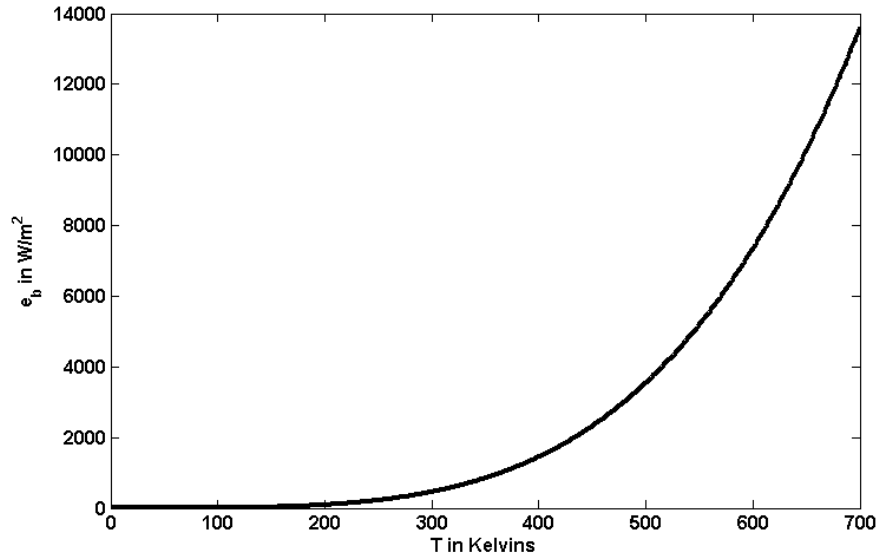
Figure 3.4. Stephan-Boltzmann law for blackbody radiation.

3.1.4. Thermal imaging of a blackbody

The temperature dependence of the emitted radiation as clear in Equation (3.5), and Equation (3.20) helped in arising the concept of thermal imaging and thermal cameras.

38

In thermal camera, an element of a plane sensor records the radiation power emitted by a blackbody as seen by the sensor element over a certain limited band of wavelengths.
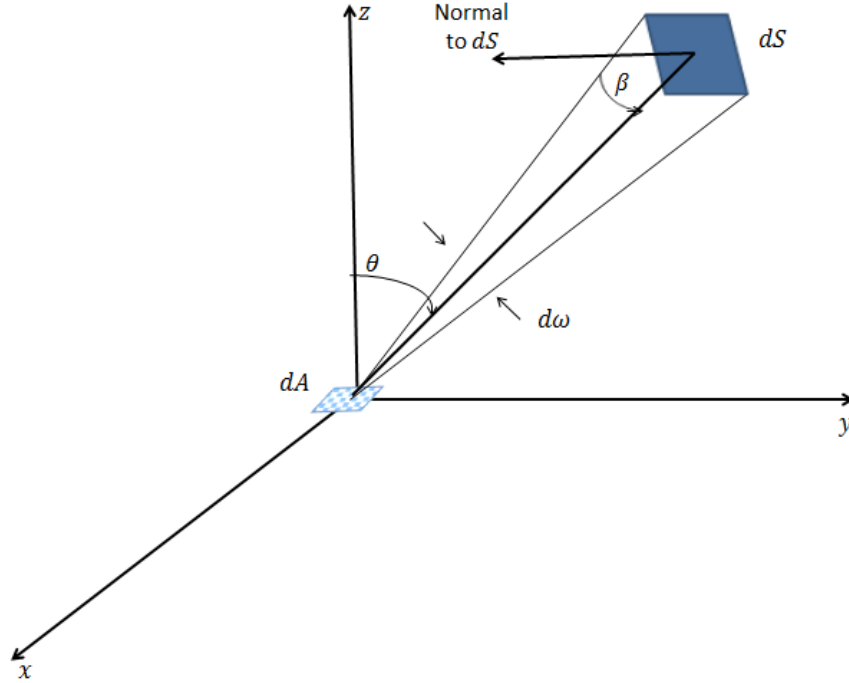
Figure 3.5. Geometry used for image formation.

Assume that the camera sensor is band limited over $\lambda \in [\lambda_1, \lambda_2]$ and dS is the elemental area of the sensor element surface. Moreover, r is the line joining the centers of ds and dA which is the elemental surface area of the blackbody. The normal to dA makes an angle θ with r while the normal to dS makes an angle β with r. Also $d\omega$ is the solid angle subtended by dS as seen by dA [68] where

$$d\omega = \frac{dS\cos\beta}{r^2}.$$ (3.21)

The power intercepted by dS between λ and $\lambda + d\lambda$ due to dA is

39

$$dP_{sensor\ element} = e_{\lambda,b}(\lambda, T, \theta, \varphi)dAd\omega d\lambda. \qquad (3.22)$$

The total power seen by the camera sensor element over its bandwidth is found by integrating A over the whole blackbody surface, ω over the sensor element and λ over $[\lambda_1, \lambda_2]$ as:

$$P_{sensor\ element,b} = \int_{\lambda=\lambda_1}^{\lambda_2} \int_{\substack{sensor \\ element}} \int_{object} e_{\lambda,b}(\lambda, T, \theta, \varphi)dAd\omega d\lambda. \qquad (3.23)$$

In the previous equation, the camera sensor was treated ideally as a blackbody with a limited absorption bandwidth and lens usage effect was not discussed. Also the transmittance of the atmosphere through which the radiation is transmitted from the object to be imaged to the camera sensor is assumed to be unity. In other words, it is assumed that no radiation is absorbed or reflected by the atmosphere.

From Equation (3.23), it is clear that a camera sensor element gets light from all points it sees from the blackbody surface. However, practically each sensor element is required to record only the signal coming from a specific single point (very small area) on the blackbody surface. In order to do this, a lens is placed between the blackbody to be imaged and the camera sensor. Equation (3.23) can still be used with lens but taking into account that A is the surface area of a very small region on the blackbody surface (single point).

So far, blackbodies have been discussed. However a blackbody does not exist practically in nature. Non-blackbody contains other aspects that should be taken into account. Non-blackbodies will be discussed below and compared to blackbodies.

3.2. Radiation from real surfaces

Practically, the behavior of a surface of a real body (non-blackbody) deviates to some extent from that of a blackbody [23]. A real body absorbs and reemits radiation over a limited band of wavelength as opposed to a blackbody that absorbs and reemits radiation over all possible wavelengths ($\lambda \epsilon [0, \infty[$). Moreover, for a real body, the absorption and the emission of radiation can be dependent on the direction of radiation incidence or emission [23] which is a non-diffuse behavior as opposed to a blackbody whose absorption and emission of radiation is diffuse. Also since the surface of a real body does not absorb the entire radiation incident upon it, it reflects back some of the radiation as opposed to a blackbody surface which reflects nothing. As a result, a real body emits radiation less than that emitted by a blackbody of the same temperature. This can be quantitatively described using spectral directional emissivity.

3.2.1. Spectral directional emissivity

The spectral directional emissivity $\varepsilon_{\lambda,\theta,\varphi}(\lambda, T, \theta, \varphi)$ of a real surface (non-blackbody) is defined as the ratio of its radiation spectral intensity in the direction (θ, φ) at wavelength λ and temperature T to radiation spectral intensity of a blackbody at the same temperature and wavelength. Thus the spectral directional emissivity $\varepsilon_{\lambda,\theta,\varphi}(\lambda, T, \theta, \varphi)$ of a non-blackbody can be written as [68]:

$$\varepsilon_{\lambda,\theta,\varphi}(\lambda, T, \theta, \varphi) = \frac{i_{\lambda,nb}(\lambda,T,\theta,\varphi)}{i_{\lambda,b}(\lambda,T,\theta,\varphi)}, \tag{3.24}$$

where $i_{\lambda,nb}(\lambda, T, \theta, \varphi)$ is the non-blackbody radiation spectral intensity which is defined as the radiation emitted power in the direction (θ, φ) at wavelength λ per unit wavelength per unit solid angle per unit area of the body surface projected in the direction of

emission when that non-blackbody is at a temperature T. For a certain wavelength λ_0 and a certain direction (θ_0, φ_0) at temperature T, $\varepsilon_{\lambda,\theta,\varphi}(\lambda_0, T, \theta_0, \varphi_0)$ takes a value between 0 and 1.

If the spectral directional emissivity $\varepsilon_{\lambda,\theta,\varphi}(\lambda, T, \theta, \varphi)$ of a certain surface is independent of the direction (θ, φ), then that surface is called a diffuse surface. Also for a certain surface for a certain direction (θ_0, φ_0), if $\varepsilon_{\lambda,\theta,\varphi}(\lambda, T, \theta_0, \varphi_0)$ is independent of the wavelength λ, then the surface is known as a gray surface in the direction (θ_0, φ_0) [23]. If $\varepsilon_{\lambda,\theta,\varphi}(\lambda, T, \theta_0, \varphi_0)$ is independent of both the wavelength λ, and the direction (θ, φ), then the surface is a gray diffuse surface. To illustrate the effect of emissivity, the radiation spectral intensity $i_{\lambda,nb}(\lambda, T, \theta, \varphi)$ for gray diffuse surfaces with different emissivity values at temperature $T = 300\ Kelvins$ is plotted in Figure 3.6.
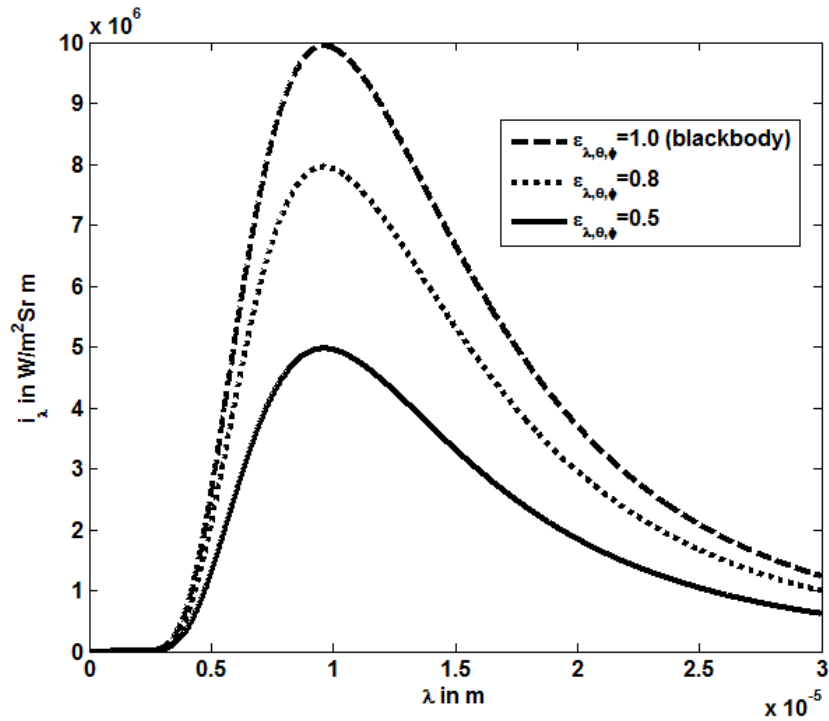
Figure 3.6. Radiation spectral intensity $i_{\lambda,\text{nb}}(\lambda, T, \theta, \varphi)$ for gray diffuse surfaces with different emissivity values at temperature T=300 Kelvins.

This chapter defined the directional spectral emissive power for the blackbody $e_{\lambda,b}(\lambda, T, \theta, \varphi)$ in Equation (3.6). Now, it will define the directional spectral emissive power of a real surface $e_{\lambda,nb}(\lambda, T, \theta, \varphi)$ as the emitted power in the direction (θ, φ) at wavelength λ per unit wavelength per unit solid angle per unit area of the body surface when the body is at a temperature T. $e_{\lambda,nb}(\lambda, T, \theta, \varphi)$ is related to $i_{\lambda,nb}(\lambda, T, \theta, \varphi)$ by:

$$e_{\lambda,nb}(\lambda, T, \theta, \varphi) = i_{\lambda,nb}(\lambda, T, \theta, \varphi)cos\theta. \tag{3.25}$$

Substituting Equation (3.24) into Equation (3.25) gives:

$$e_{\lambda,nb}(\lambda, T, \theta, \varphi) = \varepsilon_{\lambda,\theta,\varphi}(\lambda, T, \theta, \varphi)i_{\lambda,b}(\lambda, T, \theta, \varphi)cos\theta. \tag{3.26}$$

Substituting Equation (3.5) into Equation (3.26) gives:

$$e_{\lambda,nb}(\lambda, T, \theta, \varphi) = \varepsilon_{\lambda,\theta,\varphi}(\lambda, T, \theta, \varphi)\frac{2hc^2}{\lambda^5}\frac{1}{e^{\frac{hc}{KT\lambda}}-1}cos\theta. \tag{3.27}$$

3.2.2. Stefan-Boltzmann law for real surfaces

As done for a blackbody, it is intended to find the total power emitted in a hemisphere by a non-blackbody at all wavelengths and all directions per unit area of its surface. This quantity will be denoted as e_{nb}. In order to do this, consider a hemisphere around the non-blackbody elemental area dA. The power emitted per unit surface area in the direction (θ, φ) in a solid angle $d\omega$ for wavelengths between λ and $\lambda + d\lambda$ is:

$$de_{nb} = e_{\lambda,nb}(\lambda, T, \theta, \varphi)\, d\omega\, d\lambda. \tag{3.28}$$

Now substituting for $e_{\lambda,nb}(\lambda, T, \theta, \varphi)$ by its value from Equation (3.27) and for $d\omega$ by the elemental solid angle in the spherical coordinates system which is $sin\theta d\theta d\varphi$ gives:

$$de_{nb} = \varepsilon_{\lambda,\theta,\varphi}(\lambda,T,\theta,\varphi)\frac{2hc^2}{\lambda^5}\frac{1}{e^{\frac{hc}{KT\lambda}}-1}cos\theta sin\theta d\theta d\varphi d\lambda. \tag{3.29}$$

The radiation power per unit area of the source surface emitted in a hemisphere is found by integrating Equation (3.29) over $\in [0,\frac{\pi}{2}]$, $\varphi \in [0,2\pi]$ and $\lambda \in [0,\infty[$ as:

$$e_{nb} = \int_{\lambda=0}^{\infty}\int_{\varphi=0}^{2\pi}\int_{\theta=0}^{\frac{\pi}{2}}\varepsilon_{\lambda,\theta,\varphi}(\lambda,T,\theta,\varphi)\frac{2hc^2}{\lambda^5}\frac{1}{e^{\frac{hc}{KT\lambda}}-1}cos\theta sin\theta d\theta d\varphi d\lambda. \tag{3.30}$$

To be able to compare e_{nb} for a non-blackbody to e_b for a blackbody, e_{nb} and e_b will be related by:

$$e_{nb} = \varepsilon_{total}e_b, \tag{3.31}$$

where ε_{total} is the total emissivity of the non-blackbody over all wavelengths and all directions. ε_{total} can be found by comparing Equation (3.30) and Equation (3.31) as:

$$\varepsilon_{total} = \frac{\int_{\lambda=0}^{\infty}\int_{\varphi=0}^{2\pi}\int_{\theta=0}^{\frac{\pi}{2}}\varepsilon_{\lambda,\theta,\varphi}(\lambda,T,\theta,\varphi)\frac{2hc^2}{\lambda^5}\frac{1}{e^{\frac{hc}{KT\lambda}}-1}cos\theta sin\theta d\theta d\varphi d\lambda}{e_b}. \tag{3.32}$$

Substituting Equation (3.20) into Equation (3.32) gives:

$$\varepsilon_{total} = \frac{\int_{\lambda=0}^{\infty}\int_{\varphi=0}^{2\pi}\int_{\theta=0}^{\frac{\pi}{2}}\varepsilon_{\lambda,\theta,\varphi}(\lambda,T,\theta,\varphi)\frac{2hc^2}{\lambda^5}\frac{1}{e^{\frac{hc}{KT\lambda}}-1}cos\theta sin\theta d\theta d\varphi d\lambda}{\sigma T^4}, \tag{3.33}$$

where ε_{total} gives the ratio of the radiation power emitted by a non-blackbody in a hemisphere at all wavelengths and temperature T per unit area of the body surface to that emitted by a blackbody at the same temperature. Substituting Equation (3.20) into Equation(3.31) gives the Stefan-Boltzmann law for a real surface (non-blackbody surface) as:

$$e_{nb} = \varepsilon_{total}\sigma T^4. \tag{3.34}$$

Stefan-Boltzmann law is plotted for different values of ε_{total} in Figure 3.7.
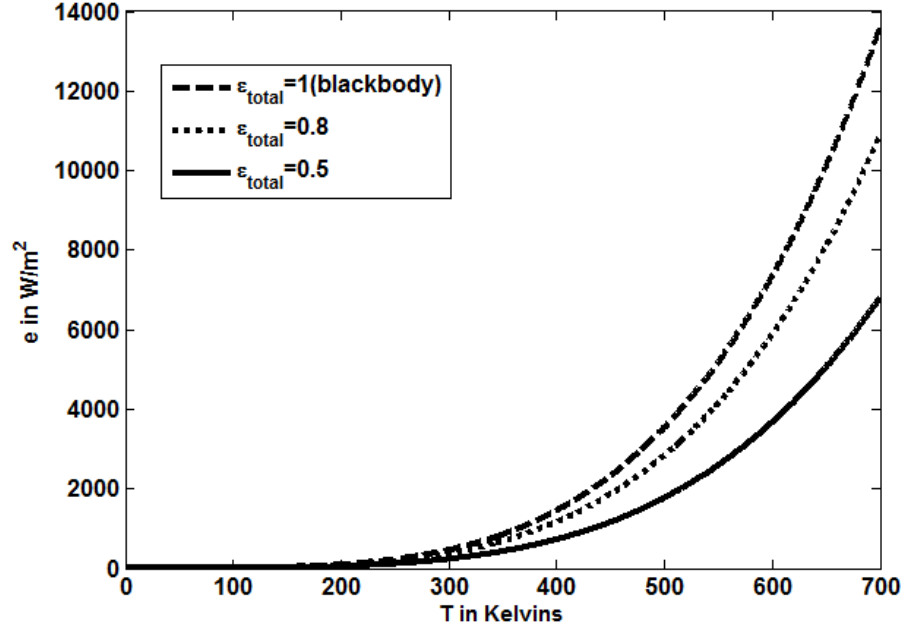
Figure 3.7. Stefan-Boltzmann law for surfaces with different emissivity ε_{total}.

3.2.3. Thermal imaging of a non-blackbody

Assume that the sensor of the thermal camera used to image a real surface is band limited over wavelengths $\lambda \in [\lambda_1, \lambda_2]$, Equation (3.23) can be modified for a non-blackbody to be:

$$P_{sensor\ element,nb} = \int_{\lambda=\lambda_1}^{\lambda_2} \int_{\substack{sensor \\ element}} \int_{object} e_{\lambda,nb}(\lambda, T, \theta, \varphi)dA d\omega d\lambda, \tag{3.35}$$

where $P_{sensor\ element,nb}$ is the total power seen by the camera sensor element over its bandwidth ($\lambda \in [\lambda_1, \lambda_2]$).

Substituting Equation (3.27) into Equation (3.35) gives:

45

$$P_{sensor\ element,nb} = \int_{\lambda=\lambda_1}^{\lambda_2} \int_{\substack{sensor \\ element}} \int_{object} \varepsilon_{\lambda,\theta,\varphi}(\lambda,T,\theta,\varphi) \frac{2hc^2}{\lambda^5} \frac{1}{e^{\frac{hc}{KT\lambda}}-1} cos\theta dAd\omega d\lambda. \quad (3.36)$$

Practically, the object to be imaged can be considered gray over the camera bandwidth ($\lambda \in [\lambda_1, \lambda_2]$), and diffuse over the solid angle subtended by the camera sensor element, and thus $\varepsilon_{\lambda,\theta,\varphi}(\lambda,T,\theta,\varphi)$ can be considered constant. Now define $\varepsilon_{\lambda,\theta,\varphi}(\lambda,T,\theta,\varphi)$ for $\lambda \in [\lambda_1, \lambda_2]$ and under diffuse emission assumed in the equation as $\varepsilon(T)_{\lambda_1 \to \lambda_2}$, then

$$P_{sensor\ element,nb} = \varepsilon(T)_{\lambda_1 \to \lambda_2} \int_{\lambda=\lambda_1}^{\lambda_2} \int_{\substack{sensor \\ element}} \int_{object} \frac{2hc^2}{\lambda^5} \frac{1}{e^{\frac{hc}{KT\lambda}}-1} cos\theta dAd\omega d\lambda. \quad (3.37)$$

Comparing Equation (3.23) and Equation (3.37),

$$P_{sensor\ element,nb} = \varepsilon(T)_{\lambda_1 \to \lambda_2} P_{sensor\ element,b}. \quad (3.38)$$

In the previous equation, $\varepsilon(T)_{\lambda_1 \to \lambda_2}$ for a specific material for a specific camera band $\lambda \in [\lambda_1, \lambda_2]$ can be found from emissivity tables.

Based on Equation (3.38), two objects made of different materials can be distinguished at the same temperature using their $\varepsilon(T)_{\lambda_1 \to \lambda_2}$.

3.3. Reflection and absorption of radiation

For the surface of a real body (non-blackbody), when radiation is incident upon it, it can be reflected, absorbed or transmitted through the body. The reflected fraction is called reflectivity ρ, the absorbed fraction is called absorptivity α, and the transmitted fraction is called the transmissivity τ. The reflectivity, the absorptivity, and the transmissivity are related by [23]

$$\alpha + \rho + \tau = 1. \quad (3.39)$$

For opaque bodies, the transmissivity τ is zero, thus Equation (3.39) is reduced to:

$$\alpha + \rho = 1. \tag{3.40}$$

At thermal equilibrium where the body is maintained at constant temperature, all of the absorbed part will be reemitted in the same in the same direction it was absorbed at (θ, φ). This makes the directional spectral emissivity $\varepsilon_{\lambda,\theta,\varphi}(\lambda, T, \theta, \varphi)$ equal to the directional spectral absorptivity $\alpha_{\lambda,\theta,\varphi}(\lambda, T, \theta, \varphi)$. This is given by Kirchhoff's law as [68]:

$$\varepsilon_{\lambda,\theta,\varphi}(\lambda, T, \theta, \varphi) = \alpha_{\lambda,\theta,\varphi}(\lambda, T, \theta, \varphi). \tag{3.41}$$

From Equation (3.41), it is clear that dealing with emissivity is the same as dealing with absorptivity if the body is at thermal equilibrium. Since the previous sections discussed emissivity and the emitted part of radiation in details, the next section will focus on the reflected part of the radiation. Throughout this context, it will be dealt with the reflected radiation classically either as rays in describing shape from shading or as electromagnetic waves in understanding the concept of polarization imaging that will be used in object shape recovery.

3.4. Intensity images and shape from shading

Considering the classical geometric optical model, a simple image formation model is the lambertian model [17]. In the lambertian image formation model, the grey level intensity of each pixel of the image depends on the light source direction and the surface normal at that pixel. Also in lambertain image formation model, the surface is assumed to reflect light in all direction (diffuse). According to these common assumptions, the image intensity at any point is given by:

47

$$I = \eta N \cdot L, \tag{3.42}$$

where η is the surface albedo, N is the surface normal, and L is the light source direction. The surface normal can be written in terms of the surface zenith angle θ and azimuth angle γ as $N=[\cos\gamma \sin\theta, \sin\gamma \sin\theta, \cos\theta]^T$, while the light source direction is written as $L=[l_1, l_2, l_3]^T$. Substituting into (3.42) gives:

$$I = \eta(l_1\cos\gamma \sin\theta + l_2\sin\gamma \sin\theta + l_3\cos\theta). \tag{3.43}$$

Shape from shading aims at recovering an object shape from the gradual variation of the shading in the image [17]. The recovered shape can be expressed in many ways among which are [17] the depth $Z(x,y)$, the surface normal (n_x, n_y, n_z), or (θ, γ), where θ is the surface normal zenith angle and γ is the surface normal azimuth angle.

Shading information provides one equation for the surface normal (equivalent to two unknowns) at each image pixel. Thus in total, there are twice as many unknowns as the number of equations. This can cause ambiguity in the recovered shape. However polarization information can double the set of available equations without increasing the number of unknowns, thus rendering the shape recovery problem from both polarization and shading well-posed. The basics of polarization imaging are discussed briefly in the following section considering the electromagnetic nature of radiation.

3.5. Polarization imaging

The primary physical quantities associated with an optical field (electromagnetic wave) are the intensity, wavelength, coherence, and polarization. Conventional panchromatic cameras measure the intensity of optical radiation over some wave band of interest. Spectral imagers measure the intensity in a number of wave bands, which can

range from one or two (three is common for a color camera) through multispectral systems that measure of the order of 10 spectral channels to hyperspectral systems that may measure 300 spectral channels or more. Spectral sensors tend to give information about the distribution of material components in a scene. Polarimetry seeks to measure information about the vector nature of the optical field across a scene. While spectral information can help in understanding materials, polarization information can help in understanding surface features, shape, shading, and roughness. Polarization provides information that is largely uncorrelated with spectral and intensity images, and thus has the potential to enhance many fields of optical metrology. The elementary definitions and equations for polarization imaging can be found in [34]. [34] also discusses different types of polarimeters (1D, 2D, 3D, Full-Stokes, and active imaging polarimeters). [34] also compares different architectures for imaging polarimeters (Rotating element, Division of amplitude, Division of aperture, and Division of focal plane).

3.5.1. Describing the polarization state of the illuminating light and the objects in the scene

Polarization is a general descriptor of light and contains information about reflecting objects that traditional intensity-based sensors ignore. All the possible states of polarization can be represented in one vector known as the Stokes vector S [69], named after George Gabriel Stokes, who introduced it in 1852. The Stokes vector is a shorthand notation that describes the polarization of light through four components [69]. An object in a scene can be described by a Mueller matrix. The Mueller matrix is a 4×4 matrix with real-valued elements. The Mueller matrix M for a polarization altering device is defined as the matrix which transforms an incident Stokes vector S_i into the exiting

49

(reflected, transmitted, or scattered) Stokes vector S. Each element of the incident S_i is related to the four elements of S by the elements of M [70] as:

$$S = MS_i. \tag{3.44}$$

Thus, Stokes vectors describe the light and Mueller matrices describe the scene. Moreover, knowing the incident and reflected light Stokes vectors at each point (pixel) in a scene, the Muller matrix at that pixel can be estimated and analyzed to get information about that pixel. If E_x and E_y represent the electric field component of the light reflected form a scene in the x and y directions respectively, then the Stokes vector S is given as [69]:

$$S = \begin{bmatrix} S_0 \\ S_1 \\ S_2 \\ S_3 \end{bmatrix} = \begin{bmatrix} < |E_x|^2 + |E_y|^2 > \\ < |E_x|^2 - |E_y|^2 > \\ 2Re < E_x E_y{}^* > \\ -2Im < E_x E_y{}^* > \end{bmatrix}, \tag{3.45}$$

where $< \cdot >$ is the time averaging operator [5].

S_3 is the difference between the left and right circular polarization components which is not being of interest of this context. This context will focus only on S_0, S_1 and S_2 which represent the linear Linear Stokes parameters and can be measured experimentally using a rotating linear polarizer mounted in front of a CCD camera as [69]:

$$\begin{bmatrix} S_0 \\ S_1 \\ S_2 \end{bmatrix} = \begin{bmatrix} I_0 + I_{90} \\ I_0 - I_{90} \\ I_{45} - I_{135} \end{bmatrix}, \tag{3.46}$$

where I_0, I_{45}, I_{90} and I_{135} are respectively the intensities received by the CCD sensor for polarizer orientation of 0°, 45°, 90° and 135°.

These angles are measured with respect to a given reference. An alternative way to represent polarization information in a scene is by using a polarization image (I, β, φ) where I is the intensity, β is the degree of linear polarization and φ is the angle of polarization (φ). $I, \beta \ and \ \varphi$ are related to S_0, S_1 and S_2 by [69]

$$I = S_0, \tag{3.47}$$

$$\beta = \frac{\sqrt{S_1^2 + S_2^2}}{S_0^2}, \tag{3.48}$$

$$\varphi = \frac{1}{2} \arctan(\frac{S_2}{S_1}). \tag{3.49}$$

Atkinson and Hancock [18] used I_0, I_{45}, and I_{90} only to calculate $I, \beta \ and \ \varphi$ as:

$$I = I_0 + I_{90}. \tag{3.50}$$

$$\varphi = \begin{cases} 0.5tan^{-1}\left(\frac{I_0 + I_{90} - 2I_{45}}{I_{90} - I_0}\right) & if \ I_{90} < I_0 < I_{45} \\ 0.5tan^{-1}\left(\frac{I_0 + I_{90} - 2I_{45}}{I_{90} - I_0}\right) + 180^o & if \ I_{90} < I_0 \ and \ I_{45} < I_0 \\ 0.5tan^{-1}\left(\frac{I_0 + I_{90} - 2I_{45}}{I_{90} - I_0}\right) + 90^o & otherwise, \end{cases} \tag{3.51}$$

$$\beta = \frac{I_{90} - I_0}{(I_{90} + I_0)cos2\varphi}. \tag{3.52}$$

3.5.2. When is polarimetric imaging preferable to classical intensity imaging for target detection?

In target detection applications, the relevant efficiency criterion is contrast (or discrimination ability) between a target and a background. [71] assumed that the scene consists of two regions with different Mueller matrices; the target region and the

51

background region. They discussed the settings of the polarimetric imager for which the scene appears with maximal contrast. They investigated four settings; purely polarized illumumination/purely polarized imaging, purely polarized illumination/ intensity imaging, purely unpolarized illumination/ purely polarized imaging, and purely unpolarized illumination/ intensity imaging. They reached the conclusion that it is always preferable to purely polarize the illumination, or not to polarize it at all. There is never interest in partially polarizing it. Similarly, for each type of illumination, there is a "turning point" at which polarization imaging becomes preferable to intensity imaging. Interestingly, this turning point is "sharp," in the sense that there is no interest of having a "partially" polarized imager; it must be totally depolarized (i.e., intensity imaging), or totally polarized. Also they found that the optimal setup depends on the scene (Mueller matrices) and on the noise perturbing the scene. Also it is worth mentioning that [72] defined the contrast between regions with different polarimetric properties under coherent illumination (such as Laser). They showed that the performances of maximum likelihood-based detection and segmentation algorithms are bijective functions of this contrast parameter which makes it possible to characterize the performance of such algorithms by simply specifying the value of the contrast parameter. [73] presented a polarization reflectance model known as the Fresnel reflectance model. Their reflectance model accurately predicts the magnitudes of polarization components of reflected light. They demonstrated the capabilities of their model to segment material surfaces according to varying levels of relative electrical conductivity, in particular distinguishing dielectrics which are nonconducting and metals which are highly conductive. They discussed polarization based methods that can provide cues for distinguishing different intensity-

edge types arising from intrinsic light-dark or color variations, intensity edges caused by specularities, and intensity edges caused by occluding contours where the viewing direction becomes nearly orthogonal to surface normals. Analysis of reflected polarization components was also shown to enable the separation of diffuse and specular components of reflection, unobscuring intrinsic surface detail saturated by specular glare. Finally they addressed polarization based methods used for constraining surface normals.

3.5.3. Shape from diffuse polarization

If the zenith and azimuth angles previously described are estimated, shape recovery can be done. The degree of linear polarization β [18] is one-to-one related to zenith angle for diffuse surfaces by

$$\beta = \frac{\left(n_t - \frac{1}{n_t}\right)^2 sin^2\theta}{2 + 2n_t^2 - \left(n_t - \frac{1}{n_t}\right)^2 sin^2\theta + 4cos\theta\sqrt{n_t^2 - sin^2\theta}}, \tag{3.53}$$

where n_t is the refractive index. If β is found from Equation (3.52), the zenith angle θ can be directly found from Equation (3.53). For the azimuth angle γ, it is related to the angle of polarization φ [18] but up to the ambiguity as:

$$\gamma = \varphi \ \ or \ \ \gamma = \varphi \pm 180^o. \tag{3.54}$$

This work presented a method for removing the ambiguity in γ [45]. Knowing θ and γ, shape recovery can be done. This will be discussed in detail in chapter 5.

CHAPTER 4

EQUIPMENT

This chapter discusses the equipment involved in the proposed research. The equipment used contains mainly two robotic vehicles (ATRV2 and ATRVmini), a polarization camera (SALSA), and a thermal camera (FLIR SC655).

4.1. ATRV2

Figure 4.1. ATRV2.

The ATRV2 shown in Figure 4.1 is a Linux based robotic vehicle platform equipped with a stereo pair of cameras and pan-tilt units which can be used for object detection and depth estimation. The ATRV2 is equipped with other depth estimation devices such as a LADAR and a Kinect which could enhance the shape recovery if based only on stereo vision. In addition to the depth estimation devices, the ATRV2 is equipped

also with an electronic compass and a crossbow IMU400CC-200 modules (accelerometers and gyroscopes) that give estimates for the yaw, pitch and roll of the robotic vehicle which can be used to compensate for any errors arising from the system. Furthermore, the ATRV2 is equipped with a wireless communications system that enables it to communicate with other robotic platforms. The drivers and interfaces to all modules of the ATRV2 were written in c/c++ making it easy to be integrated with available robotic development packages. The software and the hardware of the ATRV2 were upgraded during this work. The features and upgraded status of the ATRV2 will be discussed briefly below.

4.1.1. Features

- 4-Wheel drive (4 motors)

- 1.5 m/sec

- Skid steering

- Climbs hills up to 35 degrees

- Payload: 220 lbs.

- 4-6 hour battery

- 12 sonar sensors

- Laser range detector (LADAR)

- Two 2.4 GHz Quad core Pc's

- GPS

- Stereo imaging

- Wireless Ethernet enabled

- 4 emergency stop buttons

4.1.2. Internal Structure

Figure 4.2 is a top-view block diagram for the ATRV2 when the upper cover is open. The figure shows that the ATRV contains two Pcs; alaa PC, and chuck PC. Most of the upgrade was done on those PCs. The upgrade involves the hardware and the software.

4.1.2.1. Hardware Configuration for alaa PC
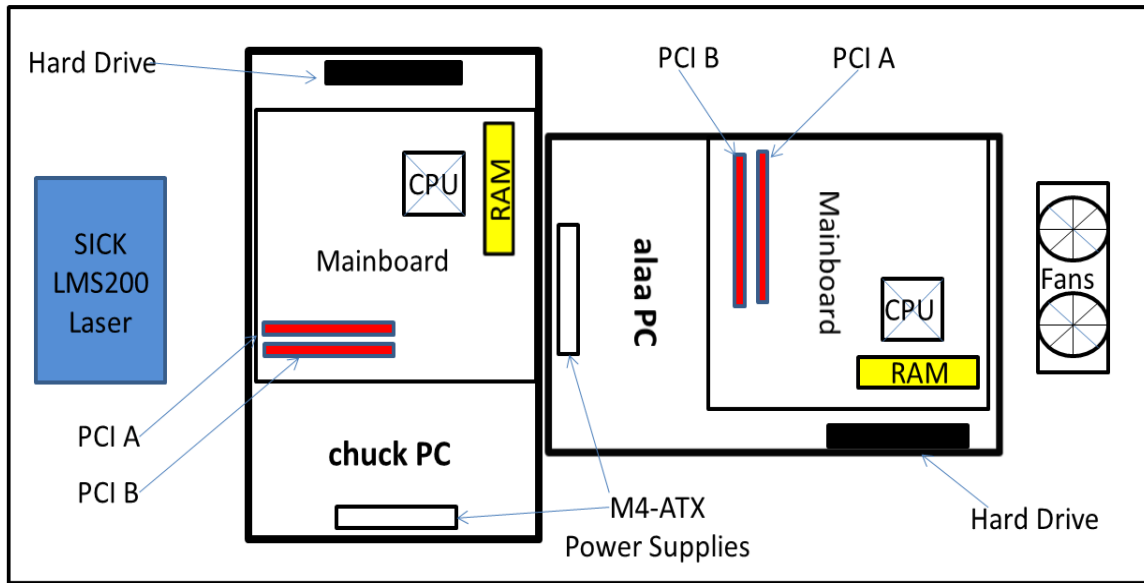
Figure 4.2. Top-view schematic diagram for the ATRV2 when the upper cover is open.

- The mainboard is a BIOSTAR TA890GXE.

- 8 GB (4x2GB modules) of DDR3 1333MHz CORSAIR memory are installed.

- The CPU is an AMD Athlon X4 610e with 2.4 GHz True Quad-Core design and 2.0 MB total cache.

- A 640 GB Western Digital Hard Drive is installed. The drive is SATA/64 MB cache.

- A serial RJ-12 card is installed in PCI A slot. This card provides 8 ports of numbers ttyR0-ttyR7 (up to down).

56

- Another RJ-12 card is installed in PCI B slot. This card provides 8 ports of numbers ttyR8-ttyR15 (up to down).

- A SICK LMS200 laser module is connected to port ttyR8

- A PTU-D46 Pan-Tilt controller module is connected to port ttyR9. This module was made by Directed Perception. It serves the ATRV2 left camera.

- Another PTU-D46 Pan-Tilt controller module is connected to port ttyR10. It serves the ATRV2 right camera.

- A G12 GPS OEM board is connected to port ttyR11. (Now GPS port A connected to ttyR6 and port B to ttyR7)

- A Crossbow IMU400CC-200 module is connected to port ttyR12.

- A doubletalk module is connected to port ttyR13.

- An electronic compass is connected to port ttyR14.

- The ATRV2 rflex is connected to port ttyR15.

4.1.2.2. Hardware Configuration for chuck PC

- The mainboard is a BIOSTAR TA890GXE.

- 8 GB (4x2GB modules) of DDR3 1333MHz CORSAIR memory are installed.

- The CPU is an AMD Athlon X4 610e with 2.4 GHz True Quad-Core design and 2.0 MB total cache.

- A 640 GB Western Digital Hard Drive is installed. The drive is SATA/64 MB cache.

- A Bt878 video capture card is installed in PCI A slot. This card is made by Brooktree Corporation.

- Another Bt878 video capture card is installed in PCI B slot.

- A SONY XC-555 CCD color video camera module is connected to terminal B of the Bt878 video capture card installed in PCI A slot.

- Another SONY XC-555 CCD color video camera module is connected to terminal B of the Bt878 video capture card installed in PCI B slot.

4.1.2.3. Software Configuration for alaa PC

The packages installed on alaa PC are:

- Mandriva 2010.2 64 bits as the operating system.

- Player 3.0.1

- Stage 3.2.2

- Opencv 2.3.0 [74]

4.1.2.4. Software Configuration for chuck PC

The packages installed on chuck PC are:

- Mandriva 2010.2 64 bits as the operating system.

- Player 3.0.1

- Stage 3.2.2

- Opencv 2.3.0

4.2. ATRVmini

The ATRVmini shown in Figure 4.3 is a Linux based robotic vehicle platform equipped with a camera and a zoom-pan-tilt unit which can be used for object detection. The ATRVmini is equipped also with an electronic compass and a crossbow IMU400CC-200 modules (accelerometers and gyroscopes) that give estimates for the yaw, pitch and roll of the robotic vehicle which can be used to compensate for any errors arising from

the system. Furthermore, the ATRVmini is equipped with a wireless communications system that enables it to communicate with other robotic platforms. The drivers and interfaces to all modules of the ATRVmini were written in c/c++ making it easy to be integrated with available robotic development packages. The software and the hardware of the ATRVmini was upgraded at the CVIP laboratory. The features and upgraded status of the ATRVmini will be discussed briefly below.

4.2.1. Features

- 4-Wheel drive (4 motors)
- 1.5 m/sec

Figure 4.3. ATRVmini.

- Skid steering
- 3-6 hour battery
- 16 sonar sensors
- Laser range detector
- Quad-core 1.6 GHz processor

- Wireless Ethernet enabled

- Pan/ Tilt/ Zoom unit

- 2 emergency stop buttons

- USB speakers

4.2.2. Software Configuration for ATRVmini PC

The packages installed on ATRVmini PC are:

- Mandriva 2010.2 64 bits as the operating system.

- Player 3.0.1

- Stage 3.2.2

- Opencv 2.3.0

4.3. SALSA polarization camera

The primary physical quantities associated with light when dealt with classically as an electromagnetic wave are the intensity, wavelength, coherence, and polarization. Conventional panchromatic cameras measure the intensity of optical radiation over some wave band of interest. Spectral imagers measure the intensity in a number of wave bands, which can range from one or two (three is common for a color camera) through multispectral systems that measure of the order of 10 spectral channels to hyperspectral systems that may measure 300 spectral channels or more. Spectral sensors tend to give information about the distribution of material components in a scene. Polarimetry seeks to measure information about the vector nature of the optical field across a scene. While the spectral information can explain materials, polarization information can explain surface features, shape, shading, and roughness. Polarization tends to provide information

that is largely uncorrelated with spectral and intensity images, and thus has the potential to enhance many fields of optical metrology. The elementary definitions and equations for polarization imaging can be found in [34]. [34] also discusses different types of polarimeters (1D, 2D, 3D, Full-Stokes, and active imaging polarimeters). [34] also compares different architectures for imaging polarimeters (Rotating element, Division of amplitude, Division of aperture, and Division of focal plane). To get a 3D linear Stokes imaging polarimeter, a motorized rotating linear polarizer can be mounted in front of a CCD camera that captures four images with the polarizer oriented at -45º, 0º, 45º and 90º measured from a given reference angle. Combining these four images gives the polarization image provided that these images are temporally and spatially registered. Sometimes, the movement of the linear polarizer can cause the images to be unregistered.

Bossa Nova Tech [75] used a liquid crystal modulator in their SALSA linear Stokes (3D) polarization camera (Figure 4.4) to rotate the polarization direction instead of using a rotating polarizer. This makes the captured four images spatially registered.

Figure 4.4. Salsa polarization camera.

Later on it developed full-Stokes polarization camera that is capable of giving information about the circular polarization aspect. However in the experiments done throughout this work, its SALSA linear Stokes polarization camera was used. The SALSA camera has the following specifications [75]:

- Video format: Digital IEEE-1394, 12 bits monochrome, 782×582 pixels.

- Acquisition interface: IEEE-1394.

- Synchronization interface: USB.

- Camera size: $4'' \times 4'' \times 6''$.

- Lenses: Standard Nikon F mount lenses.

- Frame rate at maximum resolution and 12 bits mode: 8.75 polarization image/sec.

- Frame rate at 320×240 resolution and 12 bits mode: 28 polarization image/sec.

- Spectral range: 520-550 nm.

4.4. FLIR LWIR SC655 thermal camera

The concept of thermal imaging arose from studying the phenomenon of Blackbody radiation [22]. As discussed before, a blackbody at a temperature T above 0 K continuously absorbs and then diffusely reemits photons (light quanta) of all possible wavelengths λ (λ ranges from 0 to ∞). It was found that the number of photons emitted per unit wavelength at a certain wavelength is dependent on T. This is clear in Planck's law which gives the radiation spectral intensity of a blackbody at temperature [23]. And from Wien displacement law [24], the dominant emitted wavelength is in the visible light range for an extremely high temperature body (thousands of kelvins). As the body

temperature gets lower, the dominant wavelength moves towards the infrared regions (NIR, SWIR, MWIR, LWIR).

For a non-blackbody, beside the phenomenon of temperature based emission of photons, there is also reflection of photons incident from an external visible light source. Since the emitted photons carry information about the body temperature, succeeding in separating that portion of photons from the reflected portion gives the temperature of the body. Fortunately, at low temperature (few hundreds of kelvins), the dominant wavelength of the emitted photons lies in the infrared region, thus they are band separated from the reflected visible photons. Based on this, thermal infrared cameras can give the temperature of a body.

In this work, an SC655 LWIR thermal camera developed by FLIR [76] will be used. The FLIR LWIR SC655 thermal camera is shown in Figure 4.5 and has the

Figure 4.5. FLIR SC655 thermal camera.

following specifications [76]:

- Resolution: 640×480 pixels.
- Frame rate: Up to 200 frame/sec.

- Spectral range: 7.5-14.0 μm.

- Dynamic range: 14 bit

- Temperature range: -20°C to 150°C

- Acquisition interface: Gigabit Ethernet and USB.

- Thermal sensitivity: 50 mK.

- Lens: 41.3 mm focal lens

CHAPTER 5

UNAMBIGUOUS SHAPE FROM POLARIZATION AND SHADING

Polarization imaging can give information about surface shape. It has been used for shape recovery, but with convex/concave reconstruction ambiguity. In this chapter, a direct method utilizing shading and polarization for shape recovery without the need for nonlinear optimization routines is presented. The proposed method utilizes simultaneously polarization and shading to find the surface normal, thus eliminating the reconstruction ambiguity. The albedo of the surface is also estimated in an EM-like fashion. Several experiments on synthetic and real datasets are reported to evaluate the proposed method. The method consistently outperforms a well-known method based on polarization information alone.

5.1. Introduction

Light plays an important role in the process of image formation, as an image is formed when light is reflected from an object in space into an image sensor. In the process of image formation, the object 3D shape is converted into 2D information. The 3D shape information of the object is carried by the light wave propagating from the object to the image sensor, and hence the properties associated with the light to recover the 3D shape of the object can be exploited after the process of image formation. From the prospective of wave propagation, light is an electromagnetic wave that is characterized by several properties among which are intensity, spectrum, and

polarization. The intensity is a measure of the magnitude of the oscillating electric field associated with the light, the spectrum is the frequency range over which the electric field oscillates, and the polarization is a measure of the direction at which the electric field oscillates in space. Recently, the intensity, spectrum, and polarization properties are being used to develop polarization cameras that are capable not only of measuring the spectrum and magnitude of the electric field associated with the light, but also its direction in space. While the intensity and the spectrum can give information about the materials, polarization can give information about surface shape, conductivity, refractive indices, and roughness [34]. Polarization imaging has been used for many applications in machine vision [35], remote sensing [34], biomedical imaging [36] and industrial control [37].

Of concern in this chapter are reported methods for shape recovery based on polarization. Polarization imaging for shape recovery of textureless specular surfaces was proposed in [39]. It was also used to determine the shape for a transparent object [40]. Atkinson and Hancock [18] exploited polarization by surface reflection, using images of smooth objects, to recover surface normals and, hence, height. They focused on dielectric surfaces which diffusely polarize light transmitted from the dielectric body into the air. They modeled the diffuse polarization of the reflection process using a transmitted radiance sinusoid curve and the Fresnel transmission theory. Degree of diffuse polarization was used to estimate the zenith of the surface normals. The azimuth was estimated using the phase of diffuse polarization but up to the ambiguity. Disambiguation was done using zenith angle ranking. The same authors [41] used diffuse polarization in conjunction with shading from two views to estimate the shape. The surface normals

were calculated from diffuse polarization for each view independently. The results were fed into a shading algorithm to be enhanced. In [42], the authors used multi-spectral polarization images from a single view to get a robust estimate for the surface normal azimuth in the presence of noise. Along the same line, a method for simultaneous estimation of surface orientation and index of refraction from the spectral variation of the phase of polarization was proposed in [43].

The observation underpinning this chapter is that shading information provides one equation on the surface normal (equivalent to two unknowns) at each image pixel. Thus in total, there are twice as many unknowns as the number of equations. However polarization information can double the set of available equations without increasing the number of unknowns, thus rendering the shape recovery problem from both polarization and shading well-posed.

In this chapter, a method utilizing the two sources of information, shading and polarization, is proposed for shape recovery. A few methods in literature [41, 42] share the same objective with the proposed method. However the proposed method is novel in several aspects. First, the proposed method relies on one view and one spectral imaging band, while other methods make use of multiple spectrum bands [42, 43]. As such, the proposed method is considered simpler and needs less expensive imaging setup. Second, the proposed method utilizes simultaneously polarization and shading to find the surface normal, while earlier method [42] uses separately polarization information to estimate the azimuth angle of the normal, and shading to find the zenith angle. However relying only on polarization [18, 42, 43] to obtain the azimuth angle results in an ambiguity in the estimation. To disambiguate the azimuth angle, earlier methods either assumed always

convexity on the surface under observation [42, 43] or resorted to a zenith angle ranking procedure [18] which often work only on simple cases. The proposed method solves this ambiguity by making use of the complementary shading and polarization information to estimate the azimuth angle. It is well-know also that traditional shape from shading (SFS) suffers from a convex/concave ambiguity [17, 44], which was shown to be resolved if the attenuation of the illumination due to distance is taken into consideration [46]. Due to the use of polarization information as well, the proposed method can resolve the convex/concave ambiguity of the shading component without having to complicate the illumination model. Thus the method is shown to successfully recover the shape of convex and concave surfaces using the simpler classical assumption of a single distant light source. Moreover, the proposed method uses fast and direct computation without the need for nonlinear optimization, which may lead to a local minimum unless a good initialization is provided.

Another important characteristic of the proposed method is that the albedo of the surface is also estimated within the same framework. Once the surface normal are estimated from the polarization and shading information using an initial estimate [77] of the surface albedo, a better least-mean-square estimate for the albedo is readily found. The new revised albedo estimate is then used to compute the surface normals. This process may be iterated until stability is reached. Viewed in this way the proposed algorithm has a feature reminiscent of the EM algorithm. The surface normals may be regarded as hidden or missing data that must be recovered from the observed image brightness and polarization. In the expectation-step, the surface albedo is computed. The

maximization step is concerned with finding the revised surface normal directions that minimize the image brightness and polarization error.

The rest of this chapter is organized as follows. Section 5.2 derives the surface normals from the image shading and polarization information. Section 5.3 explains how the surface albedo is estimated within a complete algorithm for shape recovery with Albedo estimation. Several experimental results on synthetic and real data are reported in Section 5.4.

5.2. Surface Normal from Polarization and Shading

In this section, information from both polarization and shading is used for a direct method to estimate surface normals. This is motivated by the following observation. For an *NxN* image, there are *2N²* unknowns (account for two angles of the surface normal per pixel). Shading information provides *N²* equations, which makes shape recovery from only shading information ill-posed. However, polarization can provide additional *N²* equations, which increases the total number of available equations to *2N²* (enough to solve for all unknowns).

In the setting for this problem, a polarization imaging camera that is composed from an intensity CCD camera with a linear polarizer mounted in front of it is considered. For a given pixel at any polarizer orientation θ_p measured with respect to a given reference, the intensity of the reflected light reaching the camera sensor I_{θ_p} follows the transmitted radiance sinusoid given by [18, 42]

$$I_{\theta_p} = \frac{I_{max} + I_{min}}{2} + \frac{I_{max} - I_{min}}{2} \cos(2\theta_p - 2\gamma), \qquad (5.1)$$

where I_{max} and I_{min} are the maximum and minimum intensities of the light reaching the camera sensor while rotating the polarizer in front of the CCD camera, respectively, with γ being the azimuth angle of the normal to the surface of the object reflecting light onto the camera. Assuming diffuse reflection where the light penetrates the object and is refracted at the object-air interface back towards the camera, a polarization image can be formed of three components, the intensity I, the degree of diffuse polarization β, and the angle of polarization φ, all of which can be found in terms of I_{θ_p} at θ_p equal to 0°, 45°, and 90° using the following equations [18]:

$$\varphi = \begin{cases} 0.5tan^{-1}\left(\frac{I_0+I_{90}-2I_{45}}{I_{90}-I_0}\right) & if\ I_{90} < I_0 < I_{45} \\ 0.5tan^{-1}\left(\frac{I_0+I_{90}-2I_{45}}{I_{90}-I_0}\right) + 180^o & if\ I_{90} < I_0\ and\ \ I_{45} < I_0 \\ 0.5tan^{-1}\left(\frac{I_0+I_{90}-2I_{45}}{I_{90}-I_0}\right) + 90^o & otherwise, \end{cases} \qquad (5.2)$$

$$I = I_0 + I_{90}. \qquad (5.3)$$

$$\beta = \frac{I_{90}-I_0}{(I_{90}+I_0)cos2\varphi}. \qquad (5.4)$$

The degree of diffuse polarization β was found [18] in terms of the zenith of the surface normal and the refractive index n_t of the object material as:

$$\beta = \frac{\left(n_t-\frac{1}{n_t}\right)^2 sin^2\theta}{2+2n_t^2-\left(n_t-\frac{1}{n_t}\right)^2 sin^2\theta+4cos\theta\sqrt{n_t^2-sin^2\theta}}, \qquad (5.5)$$

where θ is the zenith angle of the object surface normal. The refractive index falls between 1.3 and 1.6 for most dielectrics, and its exact value is not a critical requirement for shape recovery [18]. However its exact value for a specific material can be found from some available lookup tables or measured experimentally [18].

Knowing I_0, I_{45}, and I_{90}, the degree of diffuse polarization β can be calculated. And from (5.5), since the relationship between the degree of diffuse polarization and the surface normal zenith is one to one, the surface normal zenith θ can be directly estimated. For the surface normal azimuth γ, it can be estimated from the angle of polarization φ [18]

$$\gamma = \varphi \ \ or \ \ \gamma = \varphi \pm 180^o. \tag{5.6}$$

However this introduces ambiguity in the estimation. To resolve this, the information in the shading component is used. The model used for forming the image I is assumed to be the Lambertian model as Lambertian surfaces have diffuse reflectance only. Moreover, the orthographic projection is also considered. According to these common assumptions, the image intensity at any point is given by:

$$I = \eta \boldsymbol{N} \cdot \boldsymbol{L}, \tag{5.7}$$

where η is the surface albedo, $\boldsymbol{N}$ is the surface normal, and $\boldsymbol{L}$ is the light source direction. The surface normal can be written in terms of the surface zenith angle θ and azimuth angle γ as $\boldsymbol{N}=[\cos\gamma \ \sin\theta, \ \sin\gamma \ \sin\theta, \ \cos\theta]^{\mathrm{T}}$, while the light source direction is written as $\boldsymbol{L}=[l_1, l_2, l_3]^{\mathrm{T}}$. Substituting into (5.7) gives

$$I = \eta(l_1\cos\gamma \ \sin\theta + l_2\sin\gamma \ \sin\theta + l_3\cos\theta). \tag{5.8}$$

Assuming that the light source direction is known, there is some known estimate for the surface albedo η, and if θ is obtained from (5.5), then the last equation becomes having only one unknown, the azimuth angle γ. Equation (5.8) can be written in the form

$$a \cos\gamma + b \sin\gamma = c, \tag{5.9}$$

where $a = \eta l_1 sin\theta$, $b = \eta l_2 sin\theta$, $c = I - \eta l_3 cos\theta$. Solving (5.9) gives two possible solutions, γ_1 and γ_2

$$\gamma_{1,2} = \cos^{-1}(\frac{ac \pm b\sqrt{a^2+b^2-c^2}}{a^2+b^2}). \qquad (5.10)$$

This equation has real solutions only when $a^2 + b^2 \geq c^2$.

Thus there are two candidates for γ from the shading information, Equation (5.10), and two candidates provided by the phase of polarization, Equation (5.6). These multiple possible solutions give rise to the ambiguity found in previous approaches [18, 42, 43] in recovering the surface normals. In order to resolve this ambiguity, the two sources of information are exploited. The common solution, or the best approximation solution for (5.6) and (5.10), is sought as follows. Define the sets $R_1 = \{\gamma_1, \gamma_2\}$, and $R_2 = \{\varphi, \varphi + 180^o\}$. Let R_m denotes either of those groups. Define the distance of a number r to the set R_m as $d(r, R_m) = min_{j=1,...,n_m}|r - r_{mj}|$ where r_{mj} is the j-th member of R_m and n_m is the set size. Hence, the problem can be formulated as finding the real number r which minimizes

$$\Sigma_{m=1}^2 d(r, R_m) = \Sigma_{m=1}^2 min_{j=1,...,n_m}|r - r_{mj}|. \qquad (5.11)$$

To solve this problem, a simple practical solution [78] is adopted. The function $d(r, R_m)$ is piecewise linear and changes its shape only at r_j and $0.5(r_j + r_{j+1})$. Thus to find the global minimum of the sum $\Sigma_{m=1}^2 d(r, R_m)$, it is sufficient to evaluate it at all r_{mj} and $0.5(r_j + r_{j+1})$ and to pick out r at which the minimum value occurs. Thus by using this approach, the best estimate of the azimuth angle γ can be found without iterative numerical methods.

Another important component in the proposed method is imposing smoothness prior on the recovered surface normal. Indexing the surface normals according to their pixel locations, suppose that $\Gamma_{i,j} = \{\boldsymbol{N}(k,l)|(k,l) \in W_{i,j}\}$ is the set of surface normals in the neighborhood $W_{i,j}$ of the pixel (i,j). To ensure the smoothness of the recovered surface, the normal at each pixel is set to the average normal over the neighborhood. That is,

$$\widehat{\boldsymbol{N}}(i,j) = \frac{1}{|\Gamma_{i,j}|}\sum_{(k,l)\in W_{i,j}} \boldsymbol{N}(k,l). \tag{5.12}$$

Instead of using the averaging operation to impose smoothness, one can also use the median of the normals over the neighborhood $W_{i,j}$ of the pixel.

Having estimated the zenith and the azimuth of the surface normal for all points, the surface shape can be obtained using the shapelets algorithm [79].

5.3. Complete Shape Recovery Algorithm

The previous section assumed that there is a known estimate of the surface albedo. This is important to make use of Equation (5.8). In this section, an iterative EM-like algorithm is introduced to generalize the method described in Section 5.2 to be used in case of unknown albedo. Initially, a rough estimate for the albedo η can be found using [77, 80]

$$\eta_o = \frac{\sqrt{6\pi^2 <I^2> - 48 <I>^2}}{\pi}, \tag{5.13}$$

where $<I>$ is the average of the given image intensity. The initial estimate of the albedo η is then fed into the method described in Section 5.2 to find an estimate of the

zenith angle θ and azimuth angle γ of the surface normal at each point. Then the surface normals are subject to the smoothing step (5.12). Since the initial albedo estimate is often not very accurate, the recovered surface normals will not be precise either. However from the obtained normals, one can seek a better estimate of the surface albedo that minimize the image brightness error

$$\hat{\eta} = \arg\min_{\eta} \sum_{k \in I} (I_k - \eta \boldsymbol{N}_k . \boldsymbol{L})^2. \tag{5.14}$$

The solution can be obtained by setting the derivative of (5.14) with respect to η to zero, yielding

$$\hat{\eta} = \frac{\sum_{k \in I} I_k (\boldsymbol{N}_k . \boldsymbol{L})}{\sum_{k \in I} (\boldsymbol{N}_k . \boldsymbol{L})^2}. \tag{5.15}$$

This new revised albedo estimate is then used to compute the surface normals as described in Section 2. This process may be iterated until no further significant change in the estimated albedo or the surface normal is attained. In the experimentations, convergence is typically reached in a few (2-6) iterations. Viewed in this way, this procedure proceeds in an EM algorithm fashion. The surface normals may be regarded as hidden or missing data that must be recovered from the observed image brightness and polarization. In the expectation-step, the surface albedo is computed. The maximization step is concerned with finding the revised surface normal directions that minimize the shading and polarization error. From the final found surface normals, the surface shape using the shapelets algorithm [79]. Algorithm 5.1 summaries the complete shape recovery algorithm from shading and polarization information.

It is important to stress that the method assumes that the light source direction is known. This can be realized during the data acquisition procedure. Alternatively,

especially when working with already-captured datasets, it can be estimated using one of the well-documented methods for determining it automatically [77, 81, 82].

Algorithm 5.1: A Complete Algorithm for Shape Recovery with Albedo Estimation

Input: Intensity image *I*, degree of diffuse polarization β, the angle of polarization φ , and light source direction L.

1. Find an initial estimate for the albedo η_o using (5.13).
2. **Repeat**

    2.1   Find the surface normals at each point as described in Section 5.2.
    2.2   Smooth the obtained surface normals using (5.12).
    2.3   Revise the albedo estimate using (5.15)

    Until almost no change in the albedo estimate and the surface normals

3. Find the surface shape using the shapelets algorithm [79].

5.4. Experimental Results

The performance of the proposed method is evaluated using simulations and real data. The former is used to assess the method accuracy versus various noise levels and for comparison with other methods. The real data experiments are used to demonstrate the method practical utility. The proposed approach is compared with the method of Atkinson and Hancock [18] which relies on polarization information. Other approaches that use both polarization and shading [41, 42] or other related methods [43] use multiple

spectrum bands [41, 42] or multiple views [43] are thus not compared with as the proposed approach uses only one spectrum band and one view.

5.4.1. Simulations

This section begins by showing the results for applying the proposed method on a synthetically generated polarization image for a vase. The 3D height Z of the vase is generated using the equation [83]

$$Z(x,y) = \sqrt{f(y)^2 - x^2} ,$$
(5.16)

where $f(y) = 0.15 - 0.1y(6y + 1)^2(y - 1)^2(3y - 2), \quad -0.5 \leq x \leq 0.5, and\ 0 \leq y \leq 1$

From the height map, the polarization images are generated. Then the images are corrupted with zero-mean Gaussian noise with varying standard deviation σ from 0 to 10 in steps of 1. Figure 5.1(a-c) shows the synthetically generated vase images (intensity, phase of polarization, and degree of diffuse polarization,) at noise standard deviation σ =0. Figure 5.1(d) shows the ground truth 3D surface. At noise standard deviation σ =0, a very good reconstructed surface using the proposed method is shown in Figure 5.1(e) while the surface reconstructed using Atkinson and Hancock [18] is shown in Figure 5.1(f). At noise standard deviation σ =5, the generated polarization image components are shown in Figure 5.2(a-c). The reconstructed surface using the proposed method is shown in Figure 5.2(d) while the surface reconstructed using Atkinson and Hancock [18] is shown in Figure 5.2(e). It is clear that the proposed method is closer to the true surface shape than Atkinson and Hancock [18], where their method suffers due to noise.

The evolution of the albedo estimate versus number of iterations of the solution using the proposed method is shown for noise standard deviation $\sigma = 0$ in Figure 5.3(a), and for noise standard deviation $\sigma = 5$ in Figure 5.3(b). At noise standard deviation $\sigma = 0$, the albedo estimator converges to the ground truth value after 2 iterations while it converges closer to the ground truth value after 3 iterations at noise standard deviation $\sigma = 5$. The evolution curves of the average root mean square error (RMSE) of the estimated azimuth angle, and the estimated zenith angle versus number of iterations at noise standard deviation $\sigma = 0$ are shown in Figure 5.3(c-d), and at noise standard deviation $\sigma = 5$ are shown in Figure 5.3(e-f). It is clear that after two iterations, the average RMSE reaches a steady state value that is much lower than the initial state.

Figure 5.4 compares the accuracy of the proposed method to Atkinson and Hancock method [18] quantitatively by evaluating the average RMSE in the estimated azimuth and zenith angles for each method versus noise. It is clear that the proposed method gives significantly lower error than [18]. For example at noise standard deviation $\sigma = 7$, the average RMSE in the estimated azimuth for the proposed method was 0.6073 while for [18] was 1.143. Also the average RMSE in the estimated zenith for the proposed method was 0.1021 while for [18] was 0.1282. From the values in Figure 5.4, it is clear that the zenith angles are recovered more accurately than the azimuth angles.

Figure 5.5(a-c) shows the synthetically generated images of a volcano (intensity, phase of polarization, and degree of diffuse polarization,) at noise standard deviation $\sigma = 0$. Figure 5.5(d) shows the ground truth 3D surface. This surface is used because it includes convex and concave parts. At noise standard deviation $\sigma = 0$, the reconstructed surface using the proposed method is shown in Figure 5.5(e) while the surface

reconstructed using Atkinson and Hancock [18] is shown in Figure 5.5(f). At noise standard deviation σ =10, the generated polarization image components are shown in Figure 5.6(a-c). The reconstructed surface using the proposed method is shown in Figure 5.6(d) while the surface reconstructed using Atkinson and Hancock [18] is shown in Figure 5.6(e).
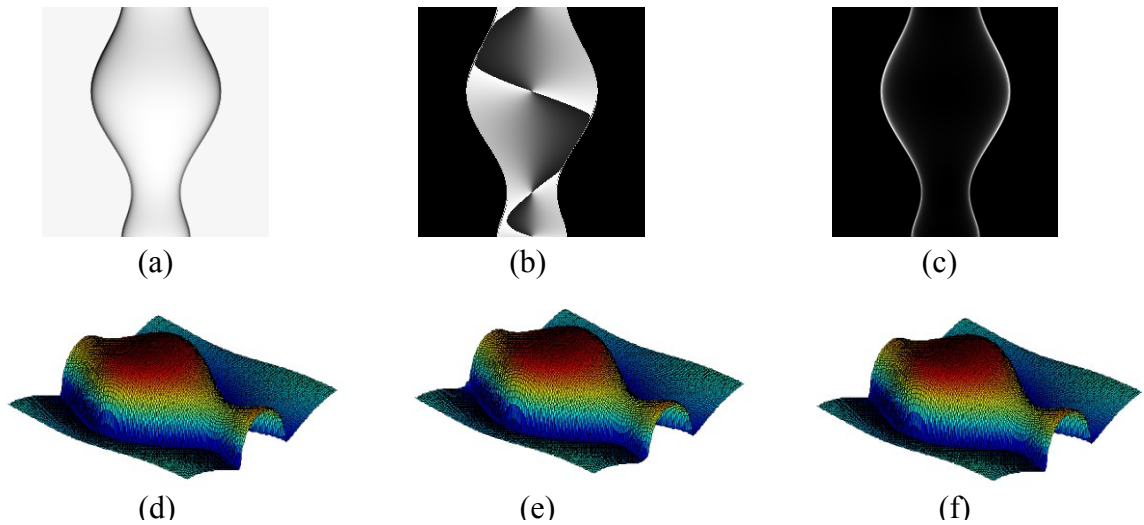

(a)   (b)   (c)

(d)   (e)   (f)

Figure 5.1. Vase Experiment at noise σ = 0: (a) Intensity, (b) Angle of polarization, (c) Degree of polarization, (d) Ground truth surface, (e) Reconstructed surface using proposed method. (f) Reconstructed surface using Atkinson and Hancock [18] method.

(a)   (b)   (c)

(d)   (e)

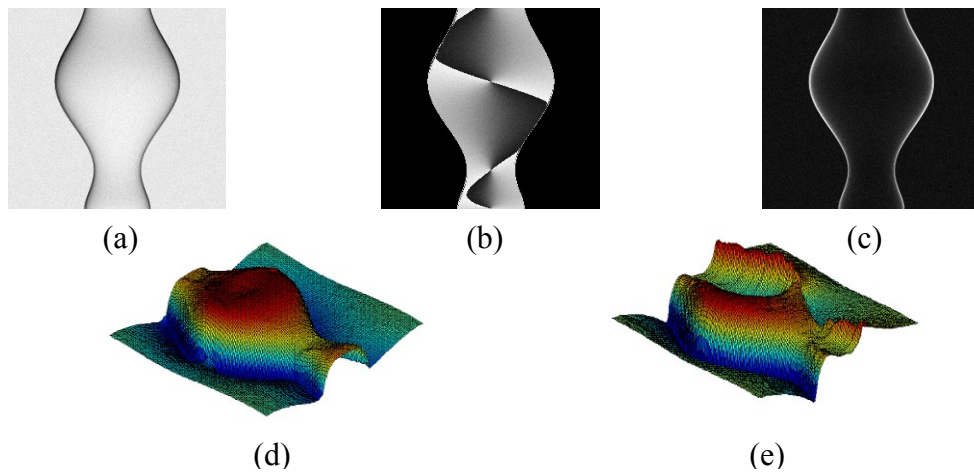**Figure 5.2. Vase Experiment at noise σ = 5: (a) Intensity, (b) Angle of polarization, (c) Degree of polarization, (d) Reconstructed surface using proposed method. (e) Reconstructed surface using Atkinson and Hancock [18] method.**

It is clear that the proposed method is more accurate than that of Atkinson and Hancock [18]. Although the curvature of the volcano surface changes between concavity and convexity, the proposed method succeeded to well discriminate between the concave and convex regions. In contrast, the method in [18] has troubles to resolve this ambiguity.

For the volcano surface, Figure 5.7 compares the accuracy of the proposed method to Atkinson and Hancock method [18] quantitatively by evaluating the average RMSE in the estimated azimuth and zenith angles for each method versus noise. It is clear that the proposed method gives lower error than [18]. For example at noise standard deviation σ=5, the average RMSE in the estimated azimuth for the proposed method was 0.3792 while for [18] was 1.312. Also the average RMSE in the estimated zenith for the proposed method was 0.04482 while for [18] was 0.05365.

The evolution of the albedo estimate versus number of iterations of the solution using the proposed method is shown for noise standard deviation σ =0 in Figure 5.8(a) and noise standard deviation σ =10 in Figure 5.8(b). At noise standard deviation σ =0, the albedo estimator converges to the ground truth value after 4 iterations while it converges to a value not far away from the ground truth value after 5 iterations at noise standard deviation σ =10.

The evolution curves of the average root mean square error (RMSE) of the estimated azimuth angle, and the estimated zenith angle versus number of iterations at noise standard deviation σ =0 are shown in Figure 5.8(c-d), and at noise standard deviation σ =10 are shown in Figure 5.8(e-f). It is clear that after 4 iterations, the average RMSE reaches a steady state value that is much lower than initial state.
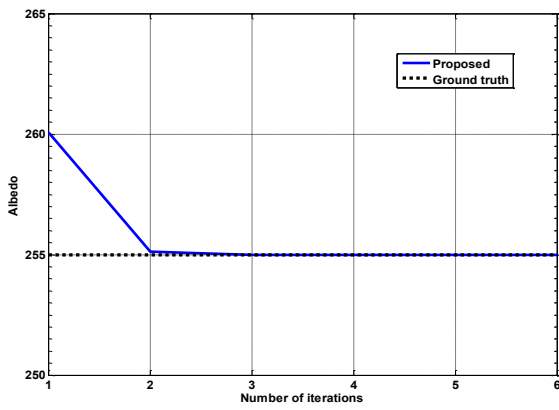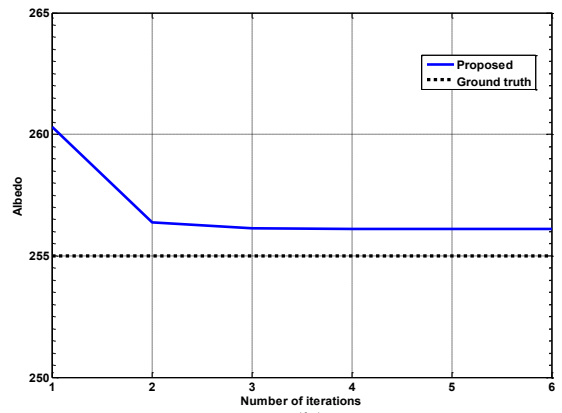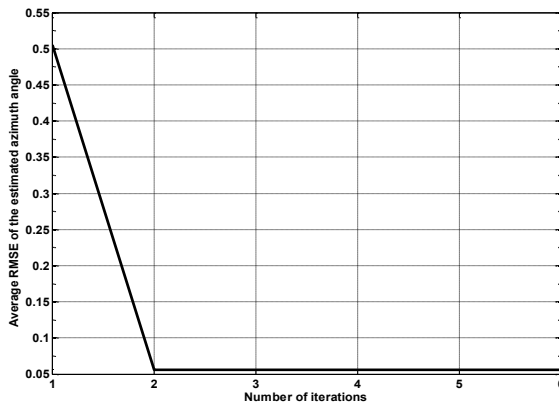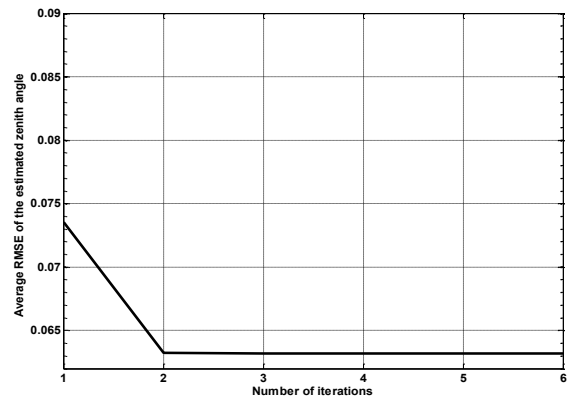
Figure 5.3. Vase Experiment: (a) Albedo reaches steady state after 2 iterations at noise σ=0, (b) Albedo reaches steady state after 3 iterations at noise σ =5, (c) Average RMSE for the estimated azimuth vs. number of iterations at noise σ=0 , (d) Average RMSE for the estimated zenith vs. number of iterations at noise σ=0, (e) Average RMSE for the estimated azimuth vs. number of iterations at noise σ=5 , (f) Average RMSE for the estimated zenith vs. number of iterations at noise σ=5.
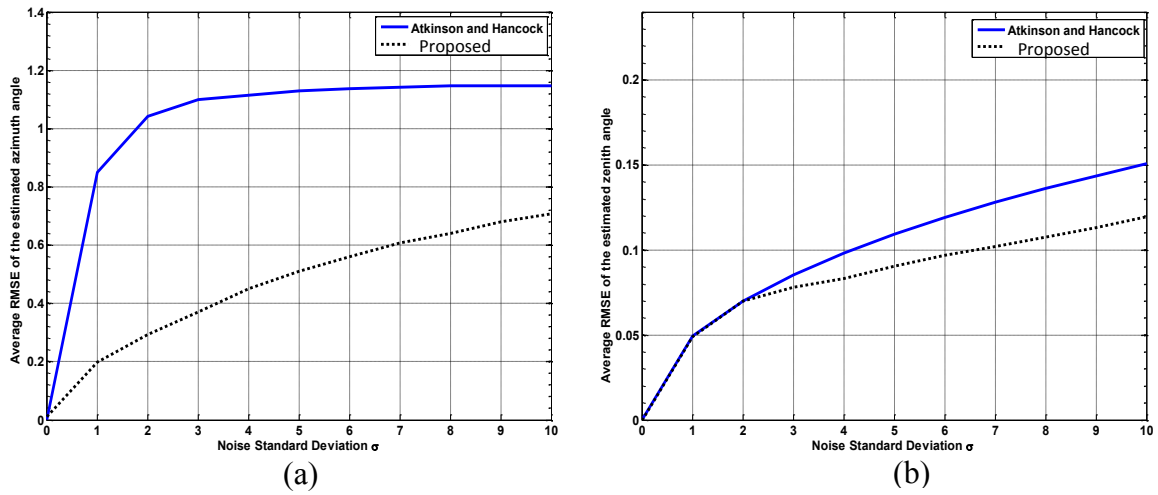
Figure 5.4. Evaluation curves for Vase Experiment: (a) Comparing the azimuth average RMSE of proposed method and Atkinson and Hancock vs. noise σ [18] method, (b) Comparing the zenith average RMSE of proposed method and Atkinson and Hancock vs. noise σ [18] method.

Figure 5.5. Volcano Experiment at noise σ = 0: (a) Intensity, (b) Angle of polarization, (c) Degree of polarization, (d) Ground truth surface, (e) Reconstructed surface using proposed method. (f) Reconstructed surface using Atkinson and Hancock [18] method.

Figure 5.6. Volcano Experiment at noise σ = 10: (a) Intensity, (b) Angle of polarization, (c) Degree of polarization, (d) Reconstructed surface using proposed method. (e) Reconstructed surface using Atkinson and Hancock [18] method.
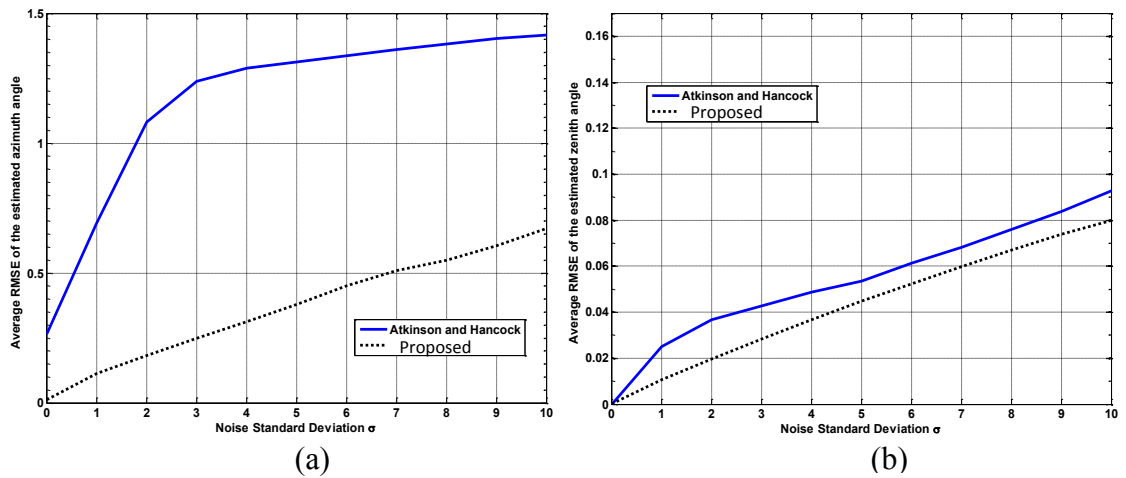
Figure 5.7. Evaluation curves for Volcano Experiment: (a) Comparing the azimuth average RMSE of proposed method and Atkinson and Hancock [18] method vs. noise σ, (b) Comparing the zenith average RMSE of proposed method and Atkinson and Hancock [18] method vs. noise σ.
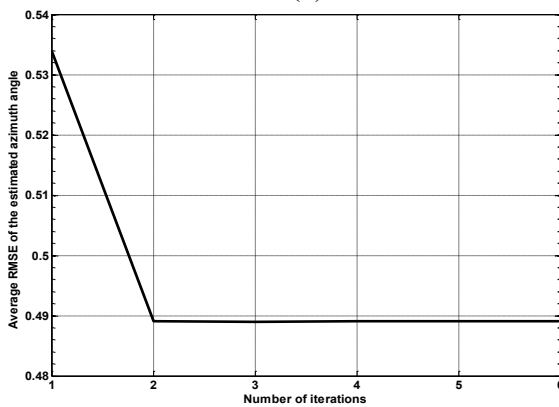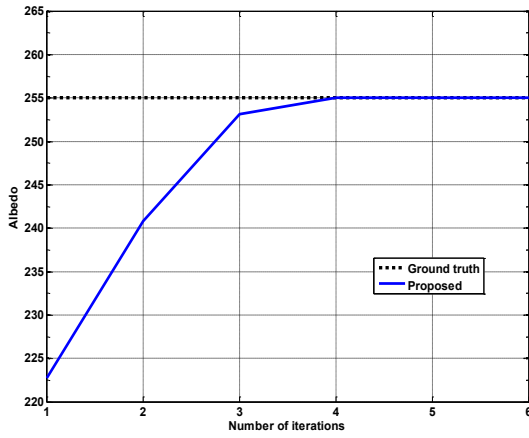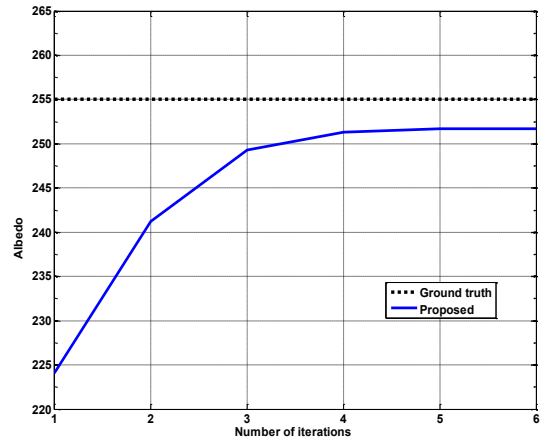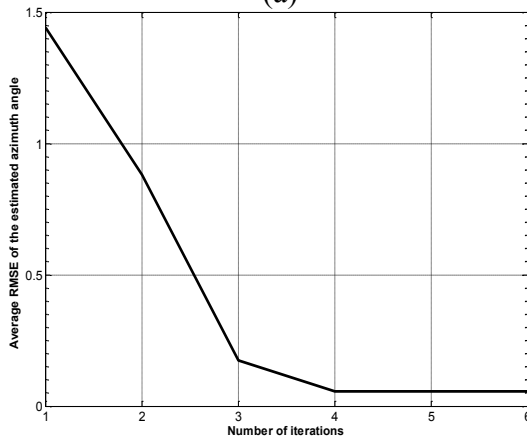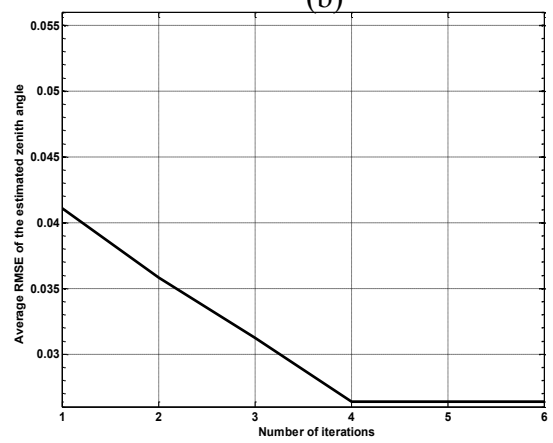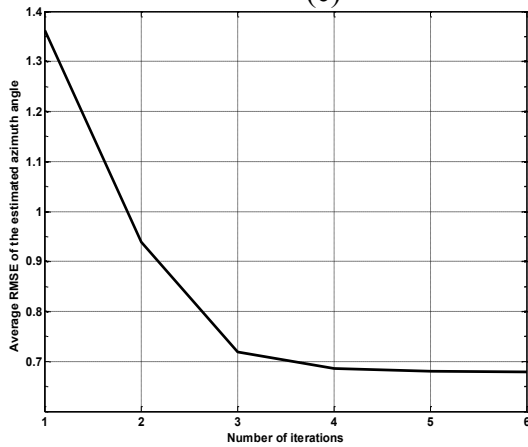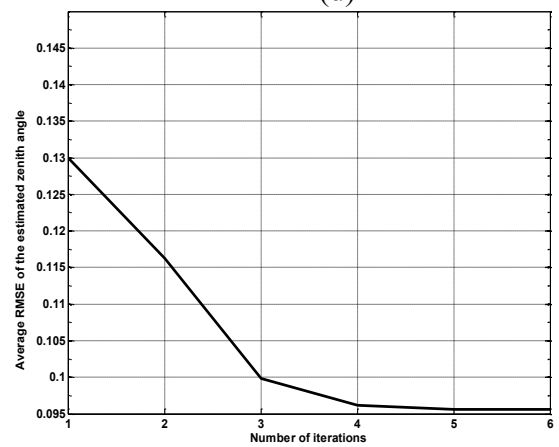
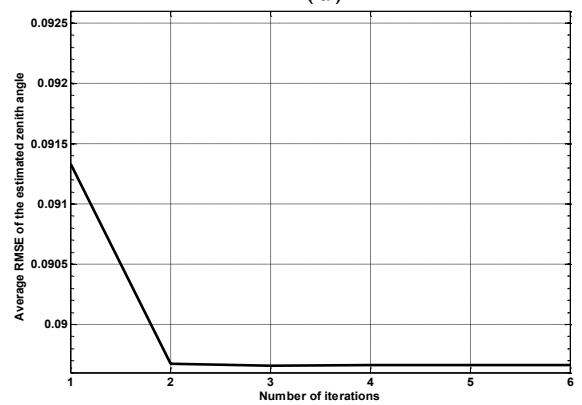**Figure 5.8. Volcano Experiment: (a) Albedo reaches steady state after 4 iterations at noise σ=0, (b) Albedo reaches steady state after 5 iterations at noise σ =10, (c) Average RMSE for the estimated azimuth vs. number of iterations at noise σ=0 , (d) Average RMSE for the estimated zenith vs. number of iterations at noise σ=0, (e) Average RMSE for the estimated azimuth vs number of iterations at noise σ=10 , (f) Average RMSE for the estimated zenith vs. number of iterations at noise σ=10.**
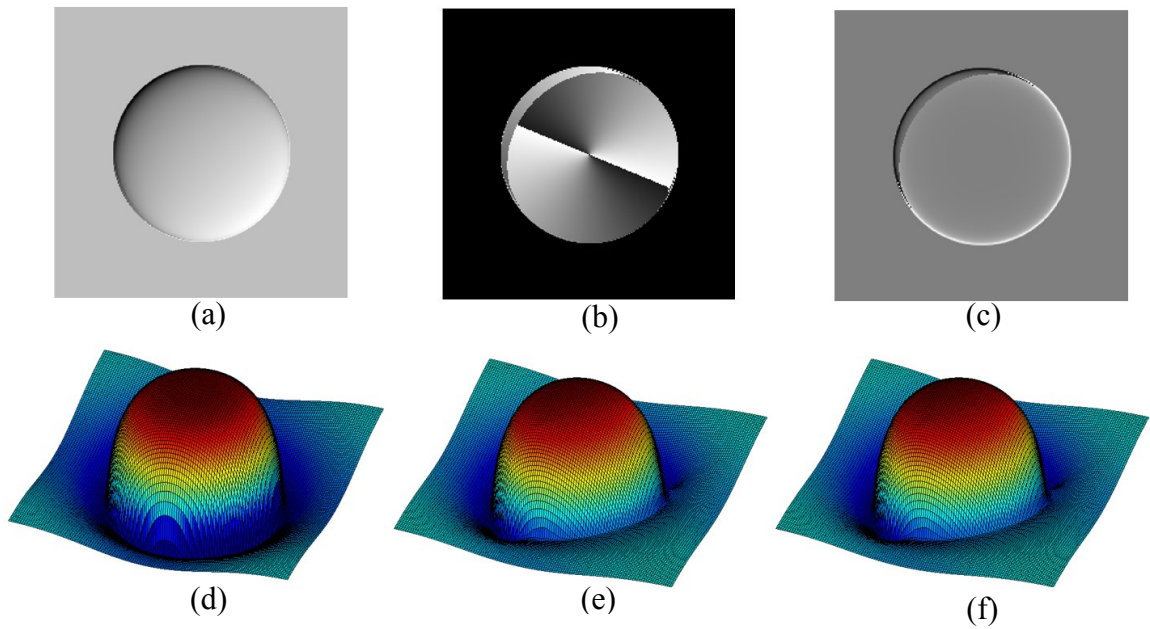
Figure 5.9. Convex hemisphere Experiment at noise σ = 0: (a) Intensity, (b) Angle of polarization, (c) Degree of polarization, (d) Ground truth surface, (e) Reconstructed surface using proposed method. (f) Reconstructed surface using Atkinson and Hancock [18] method.
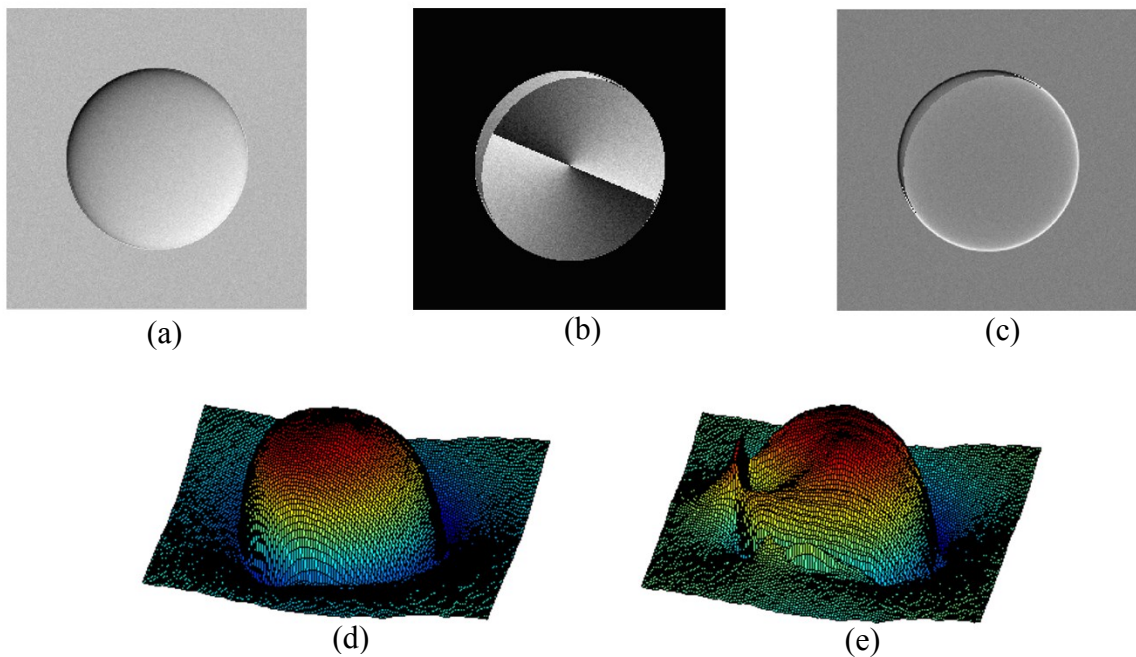
Figure 5.10. Convex hemisphere Experiment at noise σ = 5: (a) Intensity, (b) Angle of polarization, (c) Degree of polarization, (d) Reconstructed surface using proposed method. (e) Reconstructed surface using Atkinson and Hancock [18] method.
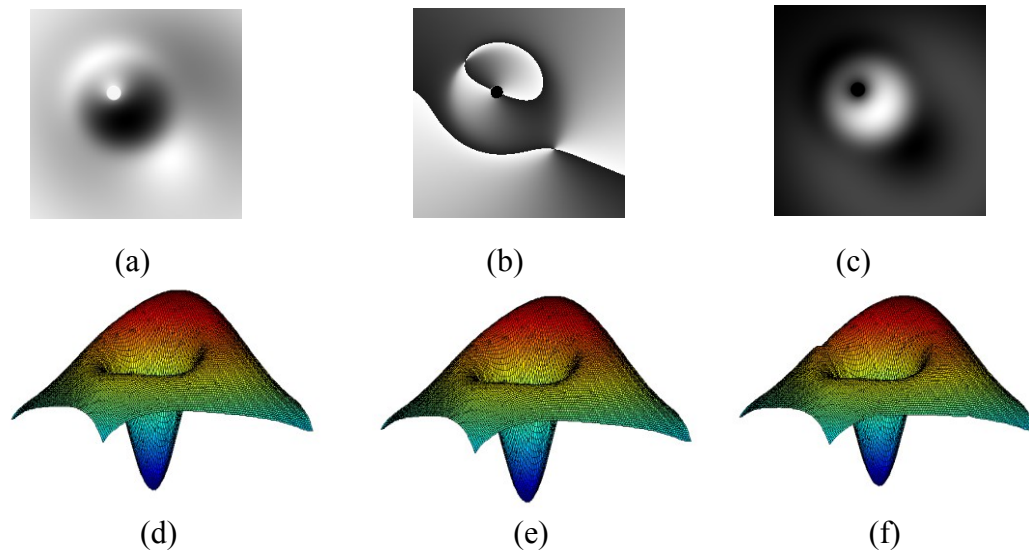
(a) (b) (c)

(d) (e) (f)
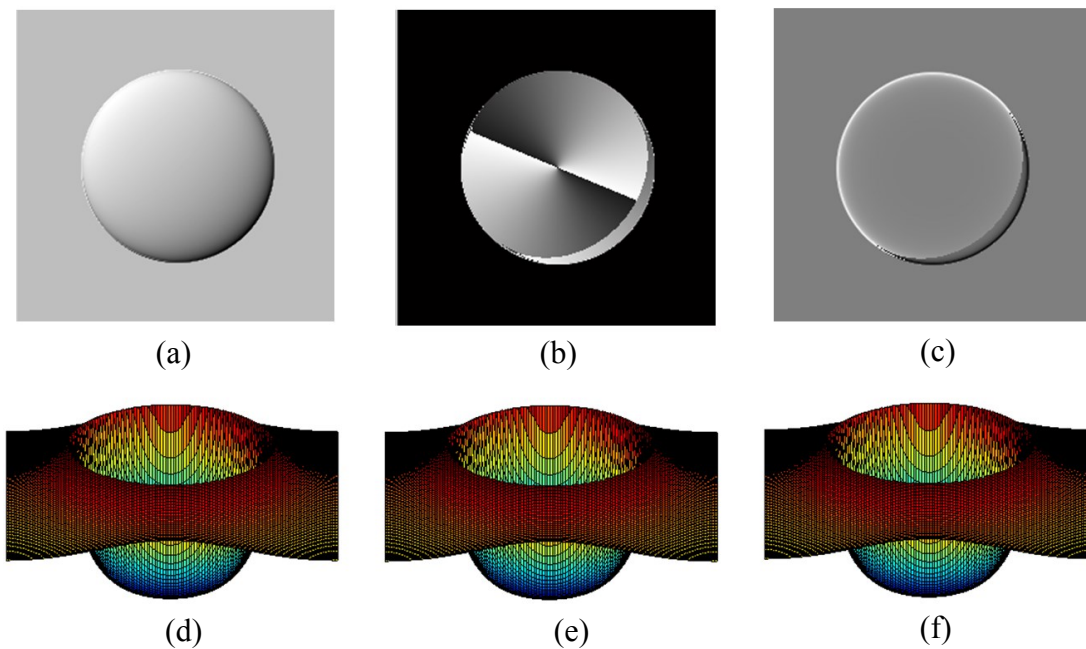
Figure 5.11. Concave hemisphere Experiment at noise σ = 0: (a) Intensity, (b) Angle of polarization, (c) Degree of polarization, (d) Ground truth surface, (e) Reconstructed surface using proposed method. (f) Reconstructed surface using Atkinson and Hancock [18] method.
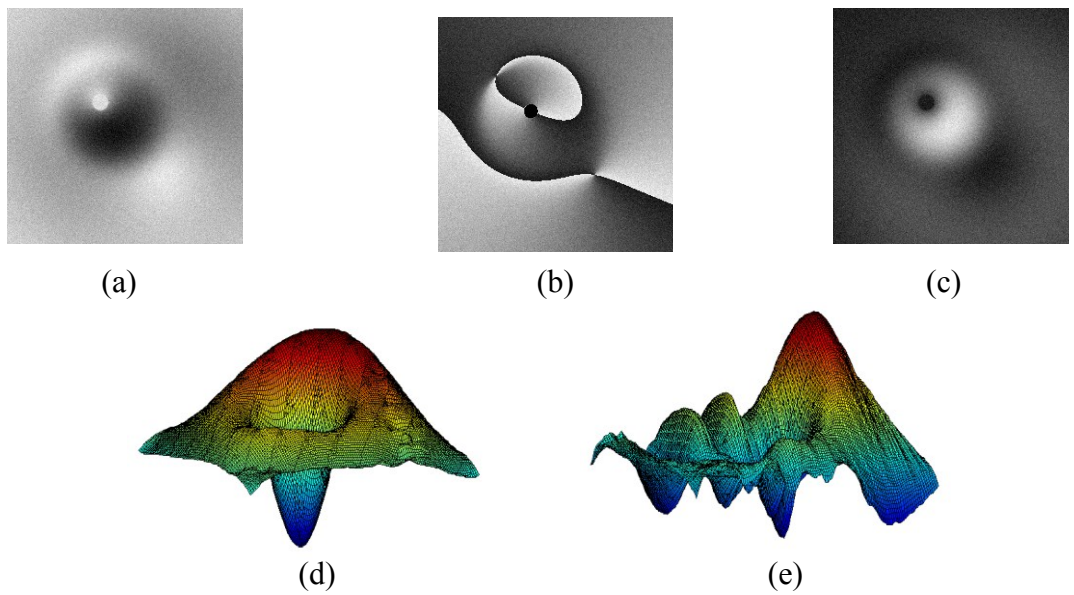

(a) (b) (c)

(d) (e)
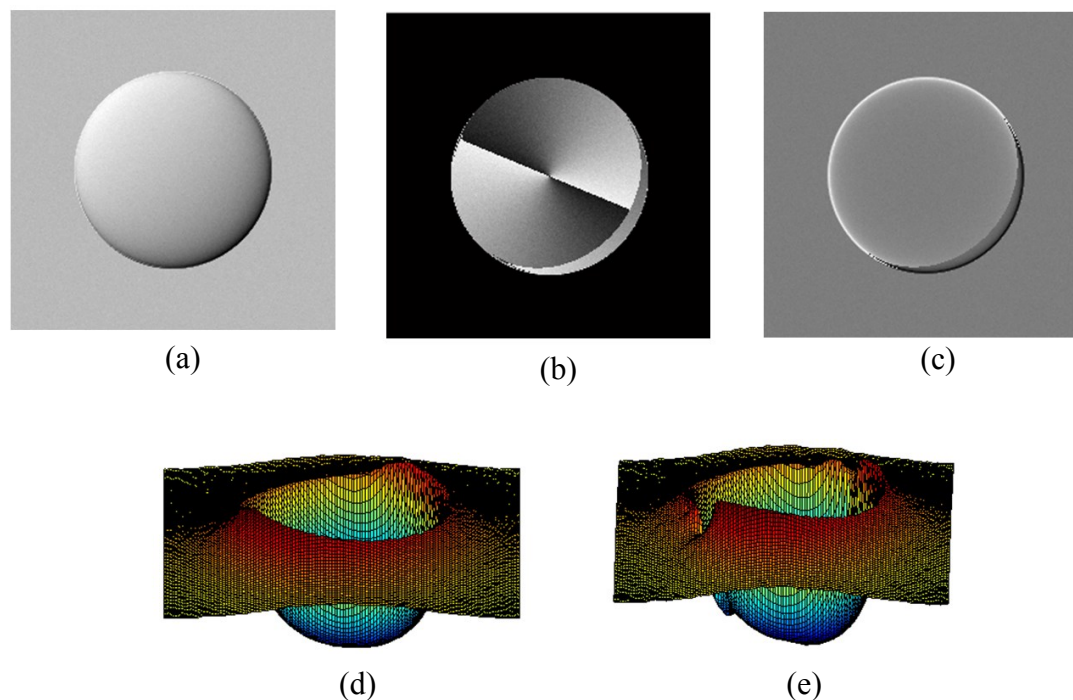
Figure 5.12. Concave hemisphere Experiment at noise σ = 5: (a) Intensity, (b) Angle of polarization, (c) Degree of polarization, (d) Reconstructed surface using proposed method. (e) Reconstructed surface using Atkinson and Hancock [18] method.
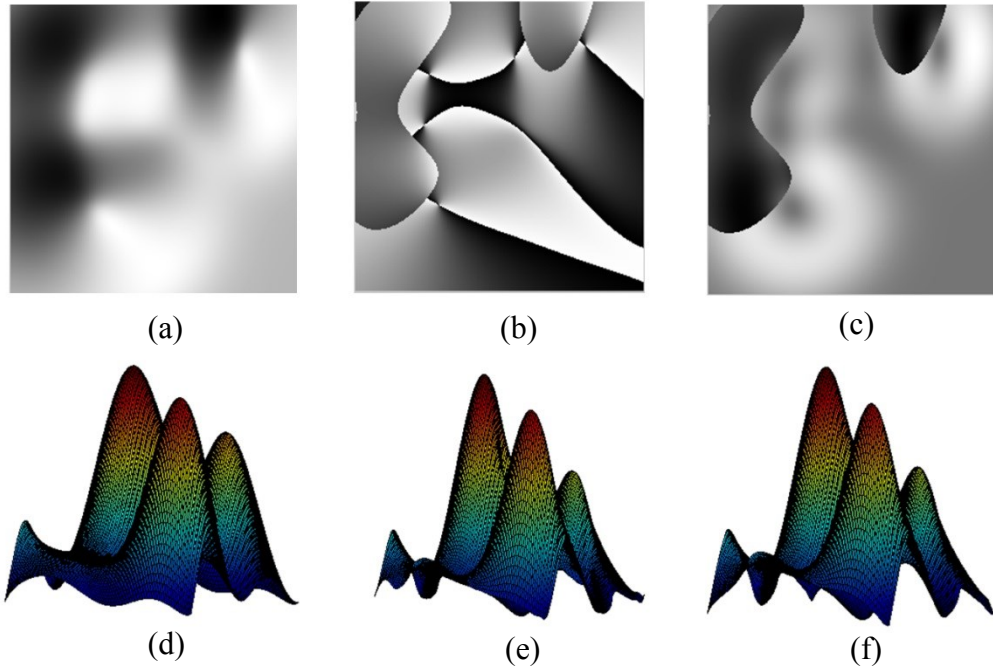
Figure 5.13. Mountains Experiment at noise σ = 0: (a) Intensity, (b) Angle of polarization, (c) Degree of polarization, (d) Ground truth surface, (e) Reconstructed surface using proposed method. (f) Reconstructed surface using Atkinson and Hancock [18] method.
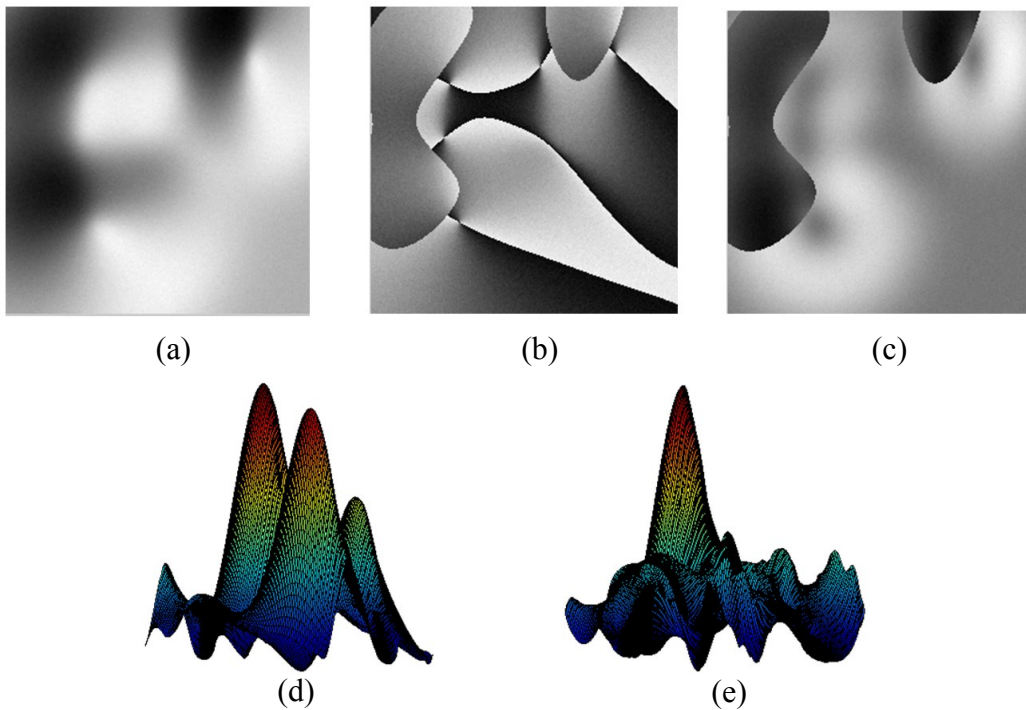
Figure 5.14. Mountainss Experiment at noise σ = 5: (a) Intensity, (b) Angle of polarization, (c) Degree of polarization, (d) Reconstructed surface using proposed method. (e) Reconstructed surface using Atkinson and Hancock [18] method.

86

Figure 5.9 through Figure 5.14 illustrates the proposed approach for three more synthetic surfaces; a convex hemisphere, a concave hemisphere and synthetic mountains.

5.4.2. Real Data

The proposed method was applied on a polarization image taken for a face mask made of plastic using the SALSA polarization camera under short term loan agreement.

The polarization image and the reconstructed surface using the proposed method for the face mask are shown in Figure 5.15. The reconstructed surface is fairly good. It is not perfect though since some of the image parts contain specular reflection which violates the model of diffuse polarization that is used in this chapter. However, the polarized component of the polarization image can be used to segment parts with specularity as specular reflection highly polarize light which opposes diffuse reflection that is low polarized.

(a)   (b)   (c)

(d)   (e)
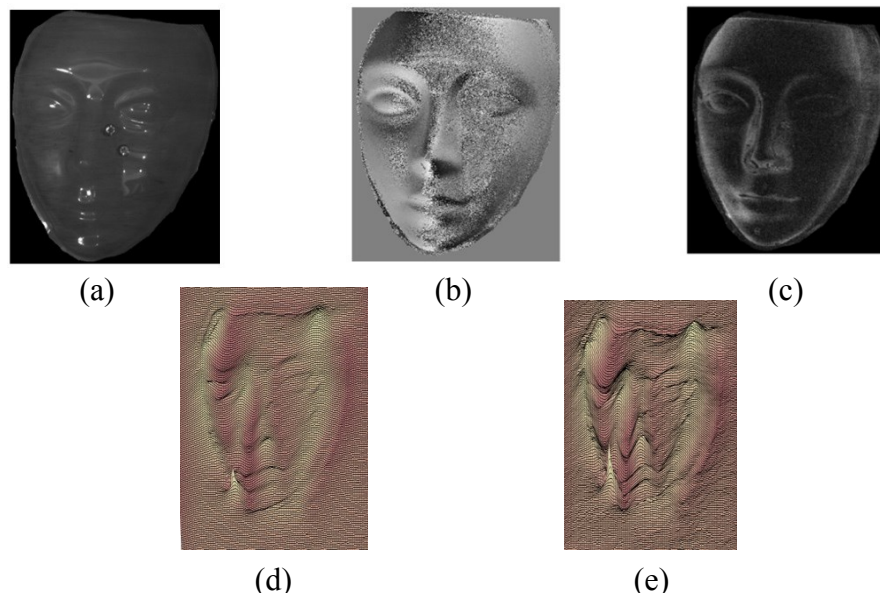
Figure 5.15. Plastic face mask experiment: (a) Intensity, (b) Phase of polarization, (c) Degree of diffuse polarization, (d) Reconstructed surface using the proposed method, (e) Reconstructed surface using Atkinson and Hancock [18] method.

CHAPTER 6

AUTOMATIC ANNOTATION FOR THERMAL IMAGING
PEDESTRIAN DATASETS FOR NIGHT DRIVING ASSISTANCE

Pedestrian detection is an important concern in the design of driving assistance systems that can reduce accidents and save lives. Annotated pedestrian datasets are essential for training a classifier that can be used for pedestrian detection. Many annotated visible pedestrian datasets are publically available. Similar annotated thermal datasets which are essential for training night pedestrian detectors are rare. Since manual annotation for large datasets is time consuming, in this chapter an automatic alternative for constructing an annotated thermal imaging pedestrian dataset is proposed [85]. This is done by transferring detections from registered visible images simultaneously captured at day-time where pedestrian detection is well developed in visible images. Histogram of Oriented Gradients (HOG) features are extracted from the constructed dataset and then fed to a discriminatively trained deformable part based classifier. The resulting classifier was tested for night driving assistance and succeeded to detect pedestrians even in the situations where visible imaging pedestrian detectors failed because of low light or glare of oncoming traffic.

6.1. Introduction

Visible images have been used successfully for pedestrian detection during day-time, but during night-time it suffers from the low light of night scenes and the glare from headlights of the oncoming traffic. On the other hand, thermal images can solve these

night vision problems as illustrated in Figure 6.1. Although the literature contains many annotated visible pedestrian datasets, it lacks annotated thermal datasets which is a necessary but time consuming component for training a classifier. The proposed system captures simultaneous visible and thermal images then registers them so that day-time visible images can be used to automatically annotate the thermal images. The proposed approach used discriminatively trained deformable part models presented in [20] to detect pedestrians in the day-time visible images and then transferred the bounding boxes to the corresponding thermal images and constructed from that an annotated pedestrian thermal dataset. This dataset was used to train a discriminatively trained part based models classifier using HOG features [26]. For the discriminatively trained part based framework [20], an object detection system is achieved using mixtures of multi-scale deformable part models that are discriminatively trained using support vector machines (SVM) requiring only the knowledge of the box bounding the object in the image. The obtained

Figure 6.1. An illustrative example of night problems. The upper sequence shows a pedestrian crossing the street using thermal images. The lower sequence shows the same scene using visible images where the pedestrian is very difficult to detect because of both the glare from the headlights of oncoming traffic and the low light of the night scene.

classifier is used to detect pedestrians in night thermal images when visible images fail. One advantage of this framework is that it makes use of the well-developed datasets of visible images to easily build similar thermal datasets.

6.2. System Description

This chapter proposes a driving assistance system that uses visible and thermal images to detect pedestrians in order to alert the driver in case of emergency. The system registers both visible and thermal images so that the pedestrian is approximately in the same location in both images. This allowed using detections from visible images taken during day-time to automatically annotate corresponding thermal images; which are then used to train a classifier for pedestrian detection on thermal images that can be used at night when visible detector is not reliable. This section describes how the two images are registered and then describes the HOG features extraction from the thermal images to train a discriminatively trained part based models classifier [20] that can detect pedestrians in thermal images.

6.2.1. Registration of Visible and Thermal Images

In the setup, the visible and thermal cameras are aligned horizontally and mounted on a vehicle close to each other as shown in Figure 6.2. The field of view of the visible camera is larger than the field of view of the thermal camera, which results in the scene of the thermal image being completely contained in the visible image. Spatial registration of the visible image (target) is needed in order to be aligned with the thermal image (reference). To do this rigid registration, similarity transformation which can include rotation, scaling and translation is used.

Figure 6.2. Setup used for data collection. A thermal camera and a visible camera are mounted on a vehicle as illustrated on the left part of the image. The right part of the image zooms on the cameras.

To calculate the transformation, a calibration step is used. It needs to be done only once after the system is installed where two control points are manually selected in the visible image with the corresponding points being also selected in the thermal image as shown in Figure 6.3. After that, all the visible images are directly transformed using the computed transformation to be registered with the corresponding thermal images. This results in the pedestrians locations to be approximately the same in both images.

6.2.2. Feature Extraction

For the training stage of the thermal imaging classifier, the features are extracted using the Histogram of Oriented Gradients (HOG) presented by Dalal and Triggs [26] where the basic idea behind the HOG is that the object local appearance and shape can be described using the distribution of the local intensity gradients or edge directions without the need to precisely know the position of the corresponding gradient or edge. Positive and negative samples are automatically generated as described earlier and then resized to 80x40 pixels. Each resized sample is then divided into non-overlapping cells of size 8x8
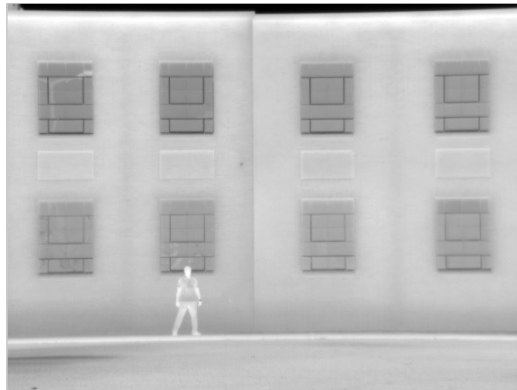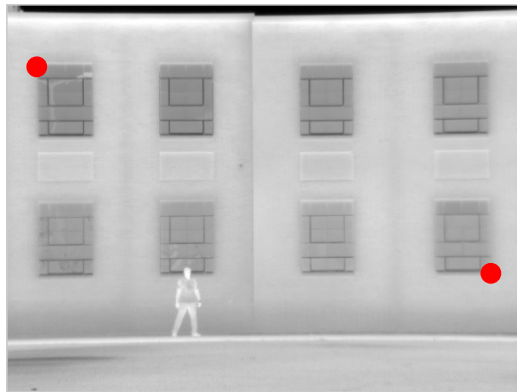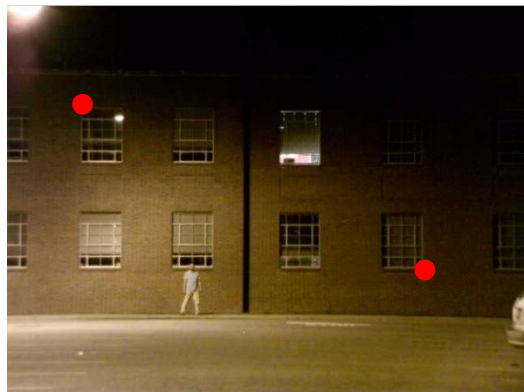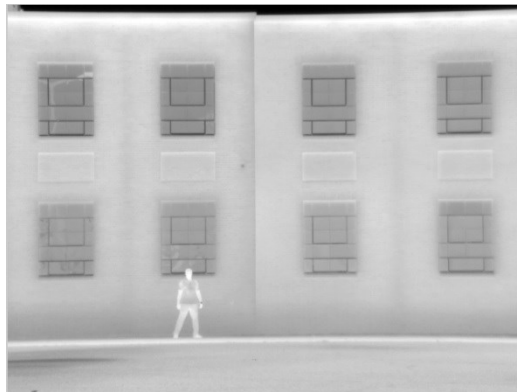
Figure 6.3. Registration of visible and thermal images: (a) Captured thermal image (reference). (b) Captured visible image (target). (c) Marking 2 control points manually in the thermal image. (d) Marking the corresponding points in the visible image. (e) and (f) illustrates the output of the registration process.

pixels resulting in 10x5 cells. For each cell, the gradient orientation histograms are computed by quantizing each pixel into one of undirected bins according to the gradient magnitude. After that, neighborhood normalization is performed for each cell over 2x2 cell blocks using four different normalization factors from the four blocks that contain the cell. These 4 histogram vectors are then concatenated to make a feature vector of dimension 36 per cell. For the case of 10x5 cells, the final feature vector length will be 1800. The system used the VLFeat implementation of the HOG features which is written in C for efficiency [84]. These features were extracted from a dataset of positive and negative samples to train a discriminatively trained part based models classifier [20] as illustrated in Figure 6.4. In a test image, to detect pedestrians with different sizes HOG features are computed on sliding windows over a standard image pyramid.

<div align="center">(a)   (b)   (c)   (d)</div>
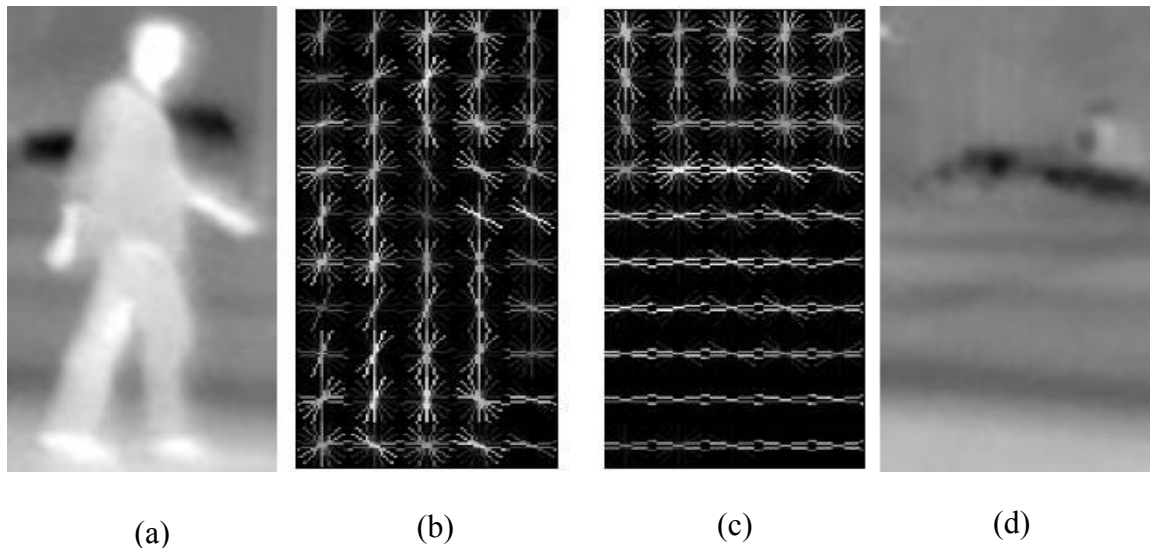
Figure 6.4. HOG features extraction (a) 80x40 positive sample. (b) An iconic representation of HOG features of (a). (c) and (d) same as (b) and (a) but for a negative sample.

6.3. Experimental Results

In the experiments, thermal images were captured by the FLIR LWIR thermal camera described in chapter 4. This thermal camera was mounted on a vehicle along with a visible camera that captures images at a resolution also of 640×480. Both the thermal and visible cameras are capable of capturing simultaneous frames that can be simply registered in the proposed approach.

| (a) | (b) | (c) |
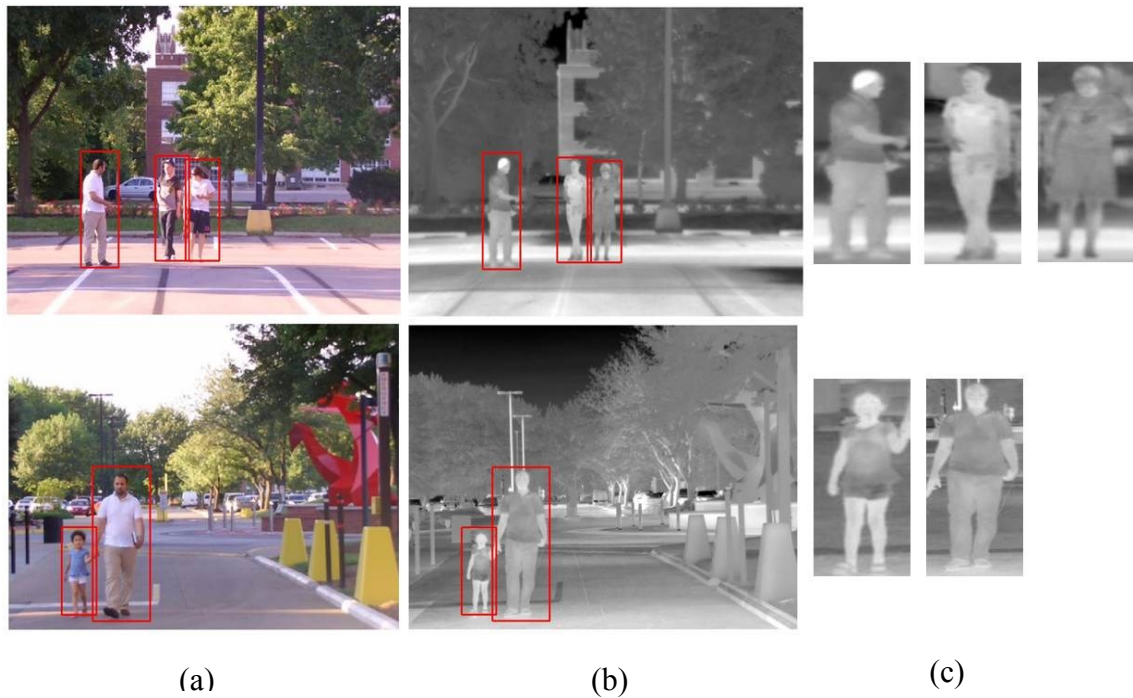
Figure 6.5. Automatic pedestrian annotation in thermal day-time image. (a) Pedestrian detection in visible images (b) Corresponding bounding boxes on thermal images (c) Cropped pedestrians from thermal images resized to 80x40 pixels to be used for training.

In order to obtain thermal data for training a classifier to be used for driving assistance at night, simultaneous thermal and visible images were captured while driving the vehicle during the day-time. A day-time visible image is clear enough to easily detect whether a pedestrian exists in the image or not. The obtained images were captured at different seasons of the year with different weather conditions to make the detector robust

at different weather conditions. The day-time visible and thermal images are spatially registered as described above. The visible images are fed into an off shelf visible pedestrian detector [20] in order to find the bounding boxes for the detected pedestrians. These bounding boxes are automatically applied to the registered thermal images as shown in Figure 6.5. The contents of these bounding boxes are considered the positive samples of the classifier where the negative samples were cropped from thermal images that do not contain pedestrians. In the experiments 1000 positive samples and 2000 negative samples were used for training purpose. The obtained samples are then resized to 80x40 for which the HOG features are extracted. The computed HOG features are then fed to train discriminatively trained part based models classifier. The training was done for different postures. This is well handled using the discriminatively trained deformable part based framework. The obtained classifier is ready now to be used at night.

It is worth mentioning that the pedestrian's height appears smaller in the thermal images as its distance from the camera increases depending on the camera focal length. Using image pyramids handles detection at different distances. The FLIR thermal camera that was used has a focal length of 41.3 mm for which 1 m height at a distance of 10 m from the camera corresponds to about 240 pixels in the image.

As mentioned previously, visible cameras can suffer from the low light of night scenes and the glare from headlights of the oncoming traffic. The visible camera can be saturated by an intensive light source which makes it difficult to extract information from it as shown in Figure 6.6(a) for instance. Although the information is messed up in the visible image, it is still clear in the thermal image that is captured at a different spectral band as shown in Figure 6.6(b) where the pedestrian existing in the image is correctly

detected. The human eye of the driver is similar in the spectral properties to a visible camera so it is likely that the driver sitting behind the wheel will face the same problem as the visible camera which might affect the safety of the pedestrians in case of not using a night vision driving assistance system. Sample results for the obtained thermal imaging pedestrian detector at different weather conditions are shown in Figure 6.7 where rectangular bounding boxes are drawn around the detected pedestrians. The first image row was captured in warm weather while the second row in hot weather. The third image row was captured during cold weather where the left image was captured during a snow event and the right image was captured after a rain event. It was noticed during the experiments that the performance of the detector increases as the temperature difference between the pedestrians and the background increases.

(a)   (b)

Figure 6.6. Night problems of visible image due to headlight glare. (a) A pedestrian is completely hidden by the glare in the visible image. (b) Corresponding thermal image with the pedestrian detected correctly using the obtained system.

In order to numerically evaluate the performance of the proposed automatic annotation method, the training is repeated but this time with manual annotation for the thermal images; hence the automatically annotated trained detector can be compared with

the manually annotated trained detector. When testing a detector, the input to the system is a group of images and the output is a set of bounding boxes with corresponding scores [20]. Scores can be thresholded at different values to plot the precision-recall curve where

$$Precision = \frac{tp}{tp+fp}. \qquad (6.1)$$

$$Recall = \frac{tp}{tp+fn}, \qquad (6.2)$$

where tp is the number of true positives (correct detections), fp is the number of false positives (unexpected detections), and fn is the number of false negatives (missing detections). At a certain threshold, the precision represents the fraction of the bounding boxes that are correct detections while the recall represents the fraction of the pedestrians in the image that are detected correctly.

Throughout this work, a PASCAL measure has been used to determine the detection rates [33]. If there is an overlap between a detected bounding box and the ground truth bounding box and this overlap is more than 50%, this is considered a correct detection, otherwise it is a false positive detection. For a certain ground truth bounding box, if there are more than one overlapping detection bounding boxes, only one of them is counted. Figure 6.8 shows the precision-recall curves comparing the proposed automatic annotation method with the manual annotation method (used as a reference to compare with). Although the curve of the proposed method is lower than that of the manual annotation method by an average precision AP=0.988%, the results of both methods are very close which indicates that automatic annotation can almost achieve the performance of the manual annotation while being less time consuming. Moreover, the experiment was repeated using a Support Vector Machine (SVM) classifier with HOG

features in a manner similar to that used in [58] in order to compare the performance of SVM classifiers with discriminatively trained part based models classifiers. The precision-recall curve of the discriminatively trained part based approach is higher than that of the SVM approach by an average precision AP=3.8730% as shown in Figure 6.9. Thus the part based approach outperformed the SVM approach.

A FLIR LWIR thermal camera has been mounted on the previously mentioned robotic vehicle; the ATRV2 as shown in Figure 6.10. The achieved night vision pedestrian detector was implemented on the ATRV2. Recall that the ATRV2 is a Linux based robotic vehicle platform equipped with cameras for night and two 2.4 GHz Quad core computers for processing.

Figure 6.7. Some results for the obtained thermal imaging pedestrian detector at different weather conditions.
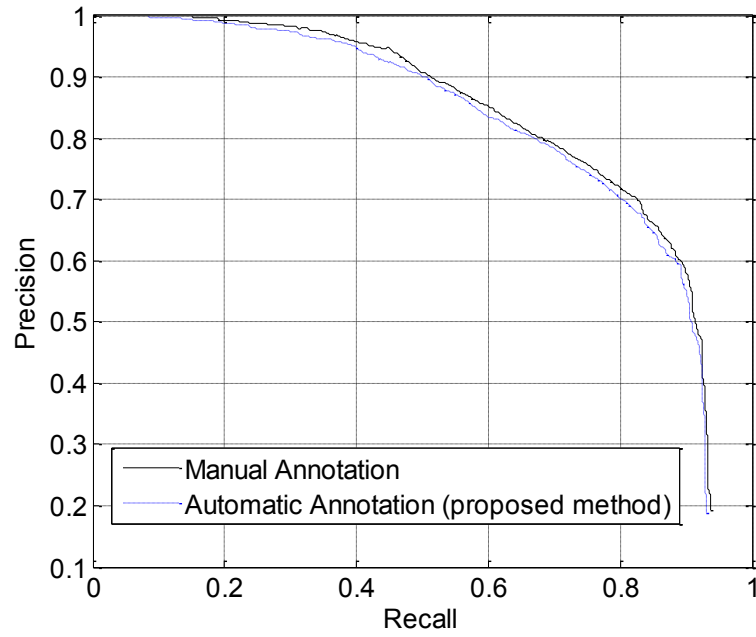
Figure 6.8. Precision-recall curves comparing automatic annotation (proposed work) with manual annotation (reference).
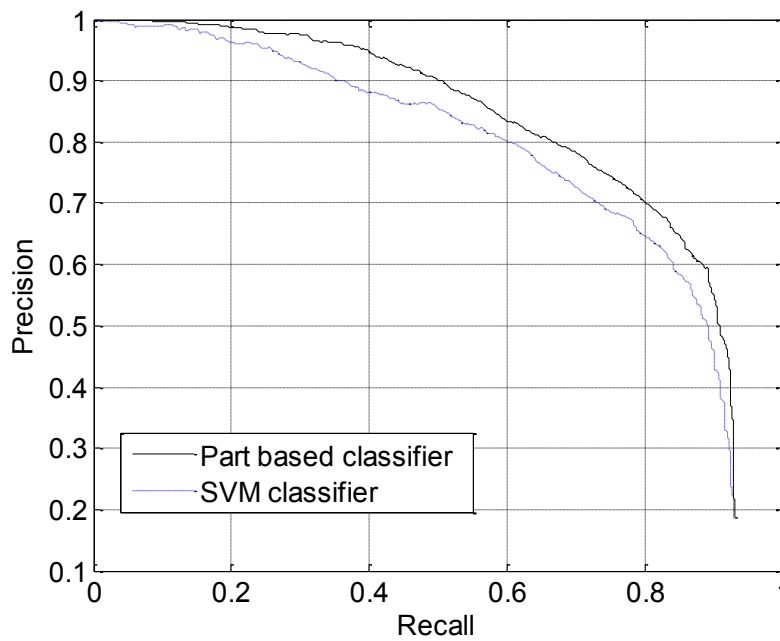
Figure 6.9. Precision-recall curves comparing the performance of an SVM classifier with discriminatively trained part based models classifier, both using the proposed automatic annotation method.

Figure 6.10. The ATRV2 with the FLIR LWIR thermal camera mounted on it.

CHAPTER 7

MIXED PARTIAL DERIVATIVE BASED HISTOGRAM OF ORIENTED
GRADIENTS FOR PEDESTRIAN DETECTION

The Histogram of Oriented Gradients (HOG) has proved to be a robust feature when used with pedestrian detectors. This chapter proposes a new feature based on HOG to be used with the discriminatively trained part based framework for pedestrian detection [87] and the integral channel features based framework for pedestrian detection. The proposed method is based on computing the image mixed partial derivatives to be used to redefine the gradients of some pixels and to reweigh the vote at all pixels with respect to the original HOG. The proposed approach was tested on the PASCAL2007, INRIA and Caltech person datasets and proved to have an outstanding performance.

7.1. Introduction

Recently, several approaches for pedestrian detection e.g. discriminatively trained deformable part based models (DPM) [20] and integral channel features (ICF) [19, 21] have been investigating HOG introduced by Dalal and Triggs [26] for feature description. For the DPM approach introduced by Felzenszwalb et al. [20], an object detection system is achieved using mixtures of multi-scale deformable part models that are discriminatively trained using support vector machines (SVM) requiring only the knowledge of the box bounding the object in the image. Ren and Ramanan [63] replaced the HOG feature in the framework of [20] by histograms of sparse codes (HSC). In their approach, they formed local histograms by computing sparse codes with dictionaries

101

learned using K-SVD which outperformed the HOG based approach. Prioletti et al. [86] presented a two-stage pedestrian detection system. Their system used a Haar cascade classifier to extract candidate which are then fed into a part-based HOG classifier [20] to lower the number of false positives.

Dollar et al. [19] investigated using integral channel features (ICF) and Adaboost classifiers [28] for pedestrian detection. In ICF, multiple registered channels are extracted from the input image using either linear or nonlinear transformations. Features are then extracted from these channel using integral images [29]. Dollar et al. [21] then discussed a modification to the system described in [19] to speed up the computation as they used fast feature pyramids instead of regular feature pyramids used in [19]. Combining histogram of oriented gradients, normalized gradient magnitude, and LUV channels gave good performance [21].

Jun et al. [64] introduced the local gradient patterns feature (LGP) and the binary histogram of oriented gradients feature (BHOG). The LGP makes the local intensity variations along the edge components robust by assigning one if the gradient of a neighboring pixel is greater than the average of the gradients of the eight neighborhood pixels and zero otherwise. For the BHOG, one is assigned if the value of the histogram bin is greater than the average value of the total histogram bins and zero otherwise. They used their approach for human detection using the INRIA [26] and the Caltech [30, 31] human databases.

This chapter proposes a new feature based on HOG to be used for human detection. The proposed feature is based on computing the image mixed partial derivatives to be used to redefine the gradients of some pixels and to reweigh the vote at

102

all pixels with respect to the original HOG. The mixed partial derivative can be interpreted as the rate of change of the slope in the x direction while moving into the y direction or vice versa which carries information different than that of the derivative in a single direction as x or y.

7.2. Histogram of oriented gradients (HOG)

HOG was presented by Dalal and Triggs [26]. This chapter will focus only on their rectangular version of the HOG. The rectangular HOG was based on locally calculating normalized histograms of image intensity gradient orientations, as the local object shape and appearance can be well characterized using these histograms without need to precisely know the position of the image intensity gradients. To extract the HOG from a window of a colored RGB image, the gradients in the x and y directions are calculated for each color channel separately with the best performance achieved when using 1-D $[-1, 0, 1]$ masks for gradient computation [26]. The resulting gradients are used to compute the gradient orientation and magnitude for each color separately at all pixels and then selecting the values at the color channel corresponding to the highest gradient magnitude at a certain pixel to be fed to the next stage of the HOG computation. In the next step, the image window is divided into smaller square patches called cells [26]. An orientation histogram is computed for each cell as follows.

Consider a certain cell; each pixel contributes a weighted vote for a gradient orientation histogram depending on the orientation of the corresponding gradient. The orientation bins of the histogram are either evenly spaced over 0-180 (for unsigned gradients HOG) or 0-360 (for signed gradients HOG). The vote at a certain pixel is a function of the gradient magnitude where the best results are achieved when taking the

vote equal to the gradient magnitude at the pixel under consideration [26]. After that, a normalization process is done [26] by grouping cells into larger overlapping blocks where each block is normalized separately. Using 9 histogram orientation bins spaced over 0-180 worked well for human detection [26].

Felzenszwalb et al. [20] implemented the HOG features by combining unsigned gradient HOG (9 bins spaced over 0°-180°) and signed gradient HOG (18 bins spaced over 0°-360°) in a single feature vector. This leads to a 31-dimensional vector, 27 of them corresponds to these orientations and the remaining 4 captures the overall gradient energy in the four neighboring blocks. This work will use the HOG implemented by [20] as a reference and will build on it.

7.3. Mixed partial derivative based HOG (proposed work)

The proposed method is based on computing the mixed partial derivative of the image window under consideration where the resulting values are used to redefine the x and y components of the gradient of some pixels and to reweigh the vote at all pixels with respect to the original HOG [20, 26].

Let $I_c(x, y)$ denote the image intensity at pixel (x, y) for color channel c, where $c \in \{R, G, B\}$ taking into account that R, G, and B refers to the red, green and blue color channels respectively. Although there are many ways to compute the gradients, using 1-D $[-1, 0, 1]$ masks works best as suggested in [26]. Thus for pixel (x, y) of color channel c as shown in Figure 1, the gradient in the x direction $I_{c_x}(x, y)$ and the gradient in the y direction $I_{c_y}(x, y)$ are calculated as follows [26]:

$$I_{c_x}(x, y) = N_{23} - N_{21}. \tag{7.1}$$

$$I_{c_y}(x, y) = N_{32} - N_{12}. \tag{7.2}$$

$I_c(x-1, y-1)$ $=N_{11}$	$I_c(x, y-1)$ $=N_{12}$	$I_c(x+1, y-1)$ $=N_{13}$
$I_c(x-1, y)$ $=N_{21}$	$I_c(x, y)$ $=N_{22}$	$I_c(x+1, y)$ $=N_{23}$
$I_c(x-1, y+1)$ $=N_{31}$	$I_c(x, y+1)$ $=N_{32}$	$I_c(x+1, y+1)$ $=N_{33}$

Figure 7.1. 8-connected neighborhood pixels for $I_c(x, y)$.

The mixed partial derivative for pixel (x, y) of color channel c $I_{xy}(x, y)$ is computed as:

$$I_{c_{xy}}(x, y) = (N_{33} - N_{31}) - (N_{13} - N_{11}). \tag{7.3}$$

Although using 1-D [-1,0,1] masks for computing gradients in the x, and y directions works best [26], it was found through this work that recomputing the gradient values at some pixels using pixel values of the neighbors added some improvement. The pixels at which the x and y gradients to be recomputed are based on their original x and y gradients and their mixed partial derivatives. It was clear that some pixels have low values for the absolute values of both the x and y components of the gradient although they have a high value of the mixed partial derivative. As an extreme example, at a given pixel (x,y), $I_{c_x}(x, y) = I_{c_y}(x, y) = 0$ while $I_{c_{xy}}(x, y) \neq 0$. This gives unspecificity when calculating $(\frac{I_{c_y}(x,y)}{I_{c_x}(x,y)})$ to compute the corresponding gradient orientation.

105

To remove this unspecificity at this pixel, the x and y components of the gradient are recomputed using the intensity values at the neighbor pixels. It was found through this work that this recomputation improves the performance not only for the extreme case when $I_{c_x}(x,y) = I_{c_y}(x,y) = 0$, but also when $\left|I_{c_x}(x,y)\right| < T_x$ and $\left|I_{c_y}(x,y)\right| < T_y$ while $\left|I_{c_{xy}}(x,y)\right| > T_{xy}$, where T_x, T_y and T_{xy} are positive thresholding values. The recomputation is done as follows:

$$I_{c_x}(x,y) = \begin{cases} \frac{N_{33} + N_{13}}{2} - \frac{N_{31} + N_{11}}{2} & \text{when } \left|I_{c_{xy}}(x,y)\right| > T_{xy}, \left|I_{c_y}(x,y)\right| < T_y \text{ and } \left|I_{c_x}(x,y)\right| < T_x \\ \\ N_{23} - N_{21} & \text{otherwise} \end{cases} \tag{7.4}$$

$$I_{c_y}(x,y) = \begin{cases} \frac{N_{33} + N_{31}}{2} - \frac{N_{13} + N_{11}}{2} & \text{when } \left|I_{c_{xy}}(x,y)\right| > T_{xy}, \left|I_{c_y}(x,y)\right| < T_y \text{ and } \left|I_{c_x}(x,y)\right| < T_x \\ \\ N_{32} - N_{12} & \text{otherwise} \end{cases} \tag{7.5}$$

For each pixel (x,y) for color channel c, the gradient magnitude $r_c(x,y)$ and the gradient orientation $\theta_c(x,y)$ are calculated as follows:

$$r_c(x,y) = \sqrt{I_{c_x}^2(x,y) + I_{c_y}^2(x,y)}. \tag{7.6}$$

$$\theta_c(x,y) = \text{atan}\left(\frac{I_{c_y}(x,y)}{I_{c_x}(x,y)}\right), \tag{7.7}$$

where atan() is the inverse tangent function. The resulting gradient magnitude and the corresponding mixed partial derivative are used to calculate a voting function $v_c(x,y)$ for channel c at pixel (x,y) as follows:

$$v_c(x,y) = \begin{cases} r_c(x,y) \left| I_{c_{xy}}(x,y) \right|^n & when \ \left| I_{c_{xy}}(x,y) \right| > T_{xy}, \\ \\ 0 & otherwise \end{cases} \qquad (7.8)$$

where n is a non-negative constant. For each pixel (x,y), c corresponding to the highest $v_c(x,y)$ is picked to be used for an orientation histogram computation. In the following discussion, $v(x,y)$ will denote the value of the largest $v_c(x,y)$ at pixel (x,y) resulting in having a vote array that has number of columns and rows equal to that of the original image window. After computing $v(x,y)$, the vote array is divided into smaller square patches called cells [26]. The size of a cell will be denoted by w_{cell}.

The next step is to compute the histogram for each cell. Consider a certain cell; each pixel contributes a weighted vote $v(x,y)$ for a gradient orientation histogram depending on the orientation of the corresponding gradient. The orientation bins of the histogram are once evenly spaced over 0°-180° using 9 bins and once evenly spaced over 0°-360° using 18 bins and the resulting values are combined in a manner similar to that presented in [20], where the feature vector of each cell is normalized in neighboring square blocks of four cells [26]. In practice, this leads to a 31-dimensional final feature vector.

7.4. Experimental results

In the first part of the experimental results, the proposed feature was augmented to the DPM framework [20] and the performance was compared to the original feature used by [20]. The performance of the proposed method is evaluated on the PASCAL2007 [32], INRIA [26] and Caltech [30, 31] publically available person datasets using discriminatively trained part based framework with two components [20]. The cell size

w_{cell} is taken as 8×8 pixels. These pedestrian datasets contain images and annotation bounding boxes which represent the ground truth for a detection system [20]. When testing a detector, the input to the system is some images and the output is a set of bounding boxes with corresponding scores [20]. These scores can be thresholded at different values to plot the precision-recall curve where

$$Precision = \frac{tp}{tp+fp}. \tag{7.9}$$

$$Recall = \frac{tp}{tp+fn}, \tag{7.10}$$

where tp is the number of true positives (correct detections), fp is the number of false positives (unexpected detections), and fn is the number of false negatives (missing detections). At a certain threshold, the precision represents the fraction of the bounding boxes that are correct detections while the recall represents the fraction of the pedestrians in the image that are detected correctly.

Throughout this work, a PASCAL measure has been used to determine the detection rates [33]. If there is an overlap between a detected bounding box and the ground truth bounding box and this overlap is more than 50%, this will be considered as a correct detection, otherwise it is a false positive detection. For a certain ground truth bounding box, if there are more than one overlapping detection bounding boxes, only one of them is counted.

Several values of the thresholding constants in Equation (7.4) and Equation (7.5) were tried and some values that gave good results were selected. The selected values are

$T_x = 0.9$, $T_y = 0.9$, and $T_{xy} = 0.9$. Moreover several values for n in Equation (7.8) were tried. Good results were obtained when choosing $n = 0.125$.

Figure 7.2 shows the precision-recall curves comparing the proposed method with [20] using the PASCAL2007 person dataset. The curve of the proposed method is higher by an average precision AP=0.8109% which means better performance. Figure 7.3 shows the precision-recall curves comparing the proposed method with [20] using the INRIA person dataset. The curve of the proposed method is higher by an average precision AP=0.0582% which means better performance. Figure 7.4 shows the precision-recall curves comparing the proposed method with [20] using the Caltech person dataset. Half of the Caltech dataset was used for training and the other half for testing using a skip of 3 while considering pedestrian of height ≥100 pixels. The curve of the proposed method is higher by an average precision AP=0.8190% which means better performance. For the three datasets, the proposed approach results outperform [20].

The proposed method was implemented on the two robotic vehicles mentioned in chapter 4; the ATRV2 and the ATRVmini. Figure 7.5 shows sample detections obtained by the implemented system.

Figure 7.6 shows an inclined image form the PASCAL2007 person dataset. The proposed method succeeded to have an overlap with the ground truth more than 50% and thus was considered as a correct detection while focusing on the subject more than [20]. Since inclined images could be captured by a vehicle moving on an inclined surface, succeeding in detecting pedestrians in these inclined images can be valuable to alert the driver and save lives.
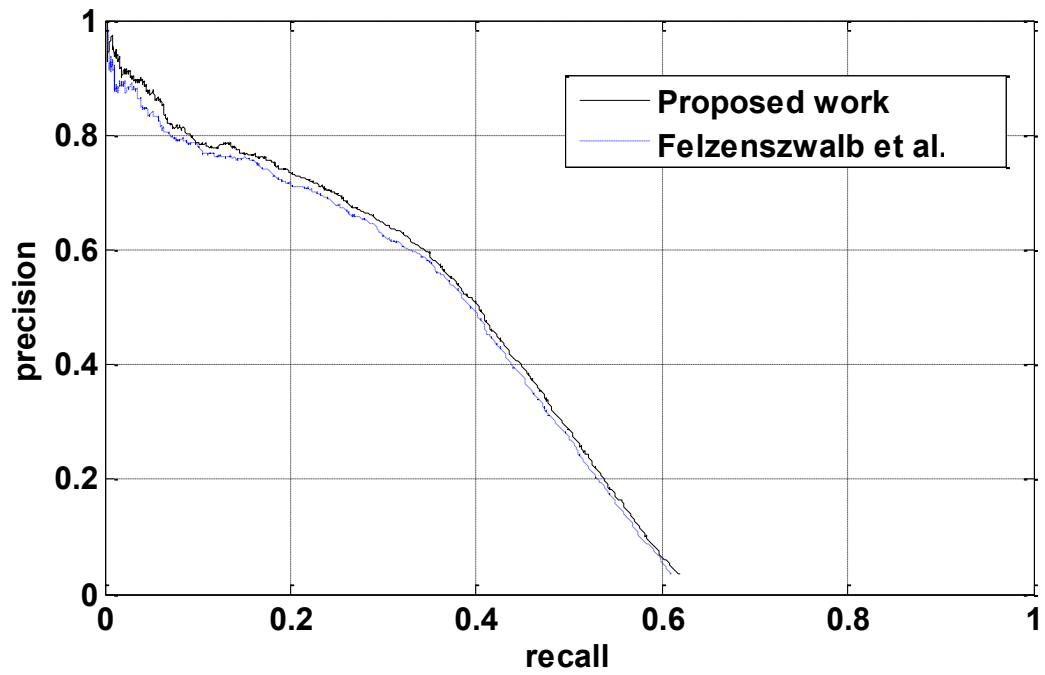
Figure 7.2. Precision-recall curves for PASCAL2007 person dataset using DPM.
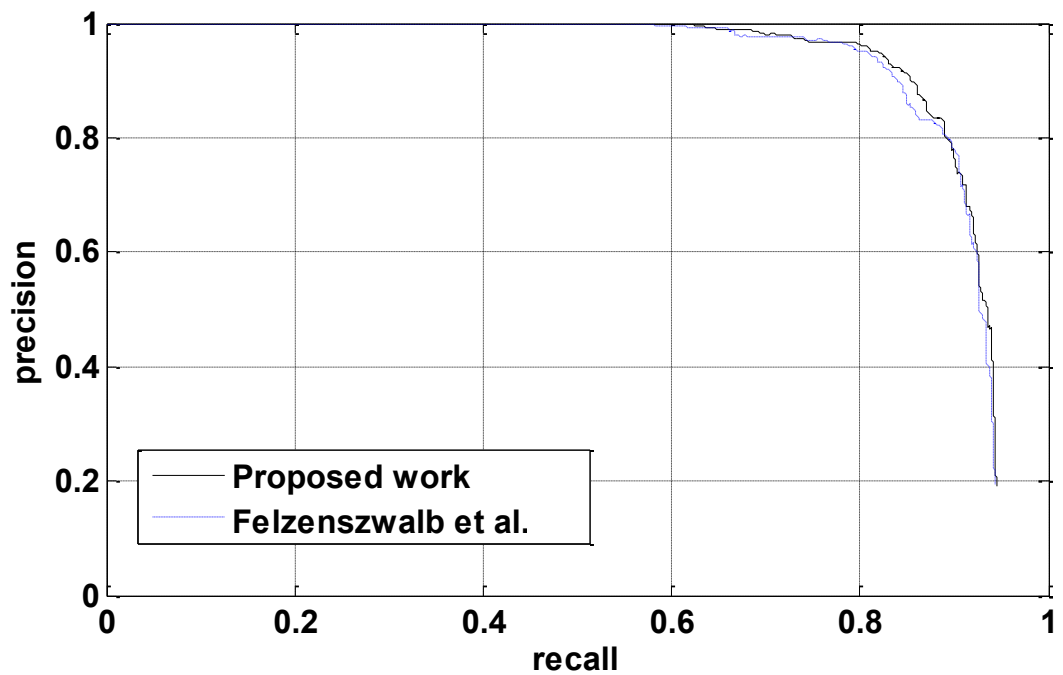
Figure 7.3. Precision-recall curves for INRIA person dataset using DPM.
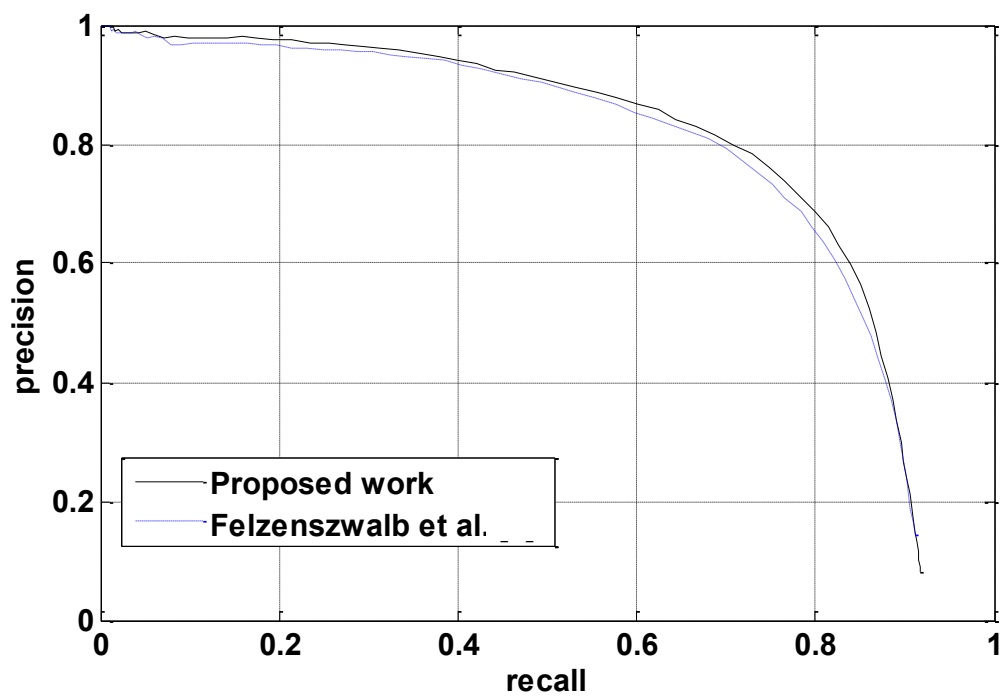
Figure 7.4. Precision-recall curves for Caltech person dataset using DPM.
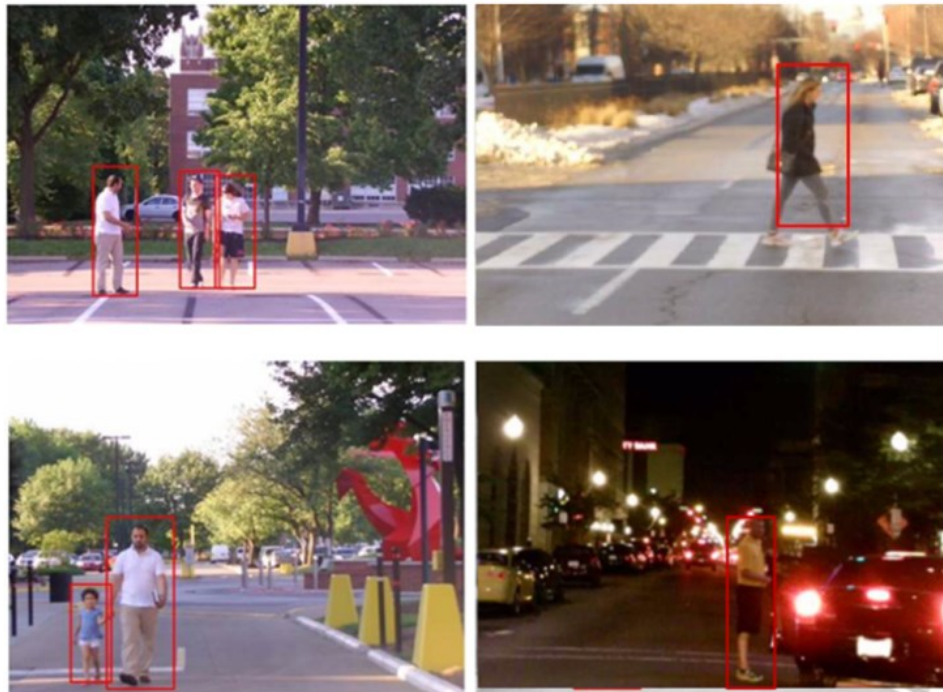
Figure 7.5. Sample detections of the proposed method.

(a)	(b)	(c)
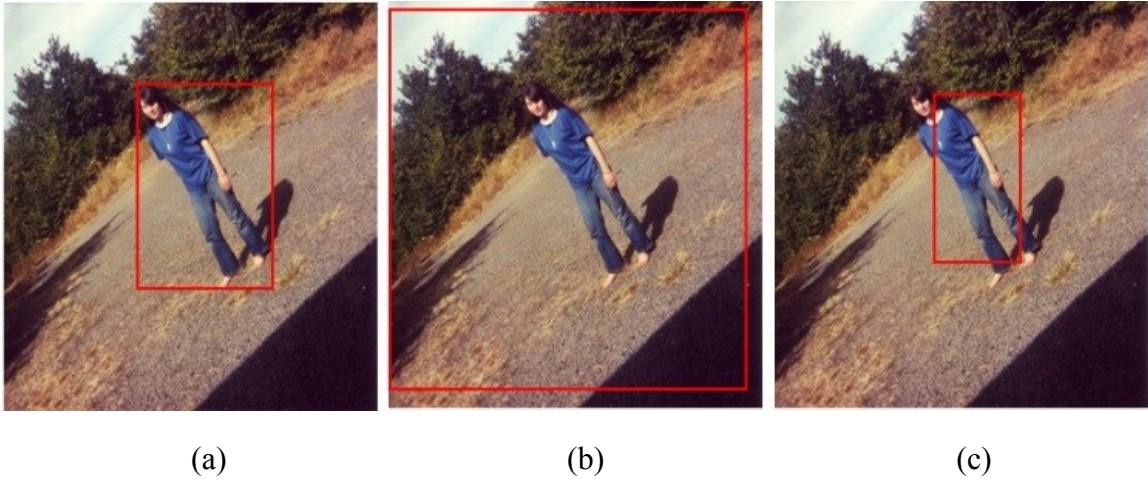
Figure 7.6. Inclined image from the PASCAL2007 person dataset: (a) Bounding box showing the ground truth. (b) Bounding box showing the detection obtained by Felzenszwalb et al. [20]. (c) Bounding box showing the detection obtained by the proposed method which has an overlap with the ground truth more than 50%.
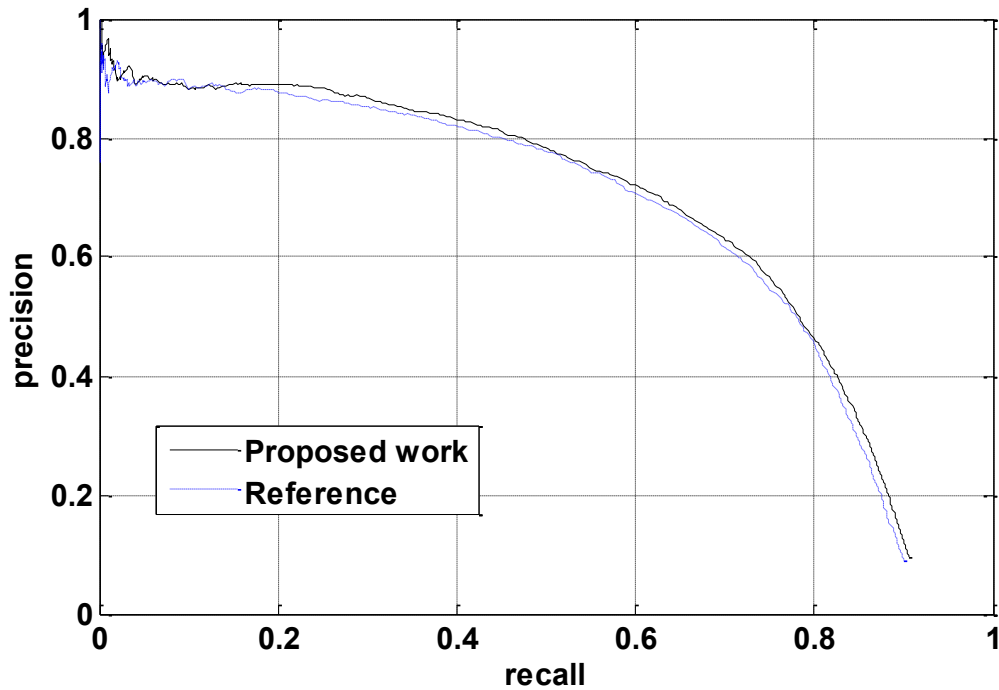
Figure 7.7. Precision-recall curves for Caltech person dataset using ICF.

In the second part of the experimental results, the proposed feature was tested with another pedestrian detection framework which is the fast feature pyramid ICF based pedestrian detector presented in [21]. The experiment was performed twice: (1) once

using channels of Felzenszwalb et al. histogram of oriented gradients [20] (reference to compare with), normalized gradient magnitude, and LUV and (2) once using mixed partial derivative based histogram of oriented gradients (proposed), normalized gradient magnitude, and LUV. To speed up the computation, the unsigned gradient parts of the feature vectors were only considered. Caltech pedestrian dataset was used in the evaluation. Half of the Caltech dataset was used for training and the other half for testing using a skip of 3 while considering pedestrian of height ≥100 pixels. PASCAL measure has been used to determine the detection rates [33]. Figure 7.7 shows the precision-recall curves comparing the proposed method with the reference method. The curve of the proposed method is higher by an average precision AP=1.8117% which indicates a better performance.

CHAPTER 8

CONCLUSIONS AND FUTURE WORK

Electromagnetic radiation contains lots of information that can be used by a machine e.g. a smart robot to make a decision and perform an action. Several types of images can be formed from radiation e.g. polarization images, thermal images and visible images. This work made use of these three types of images to aid a robotic vehicle in the collection of information from a scene for the purpose of driving assistance.

In chapter 5, a method utilizing shading and polarization for shape recovery was proposed. The proposed method is novel in several aspects. Compared to some multiple spectrum bands-based methods [42, 43] and multi-view approaches [41], only one view and one spectral imaging band is utilized. As such, the proposed method is considered simpler and needs less expensive imaging setup. Second, the proposed method utilizes simultaneously polarization and shading to find the surface normal. As such, the proposed method is able to resolve the inherent ambiguity in estimating the azimuth angle from only the polarization information. The method also resolves the convex/concave ambiguity of the traditional SFS methods without having to complicate the illumination model. Another important characteristic of the proposed method is that the albedo of the surface is also estimated in an EM-like fashion.

The experimental results have shown that the proposed method successfully recovered the surface normals at various noise levels. The method consistently

outperformed a well-known method based on polarization information [18], and the experiments clearly demonstrated that the method converges quickly in a few iterations.

Knowing whether the reconstructed surface is concave or convex is an important issue in many robotic applications e.g. an autonomous robotic vehicle can be able to discriminate between a ditch (negative obstacle) and a hill (positive obstacle). Proper discrimination leads to a proper decision and a proper action that helps the robot to complete its mission safely. The proposed approach assumes uniform albedo for the surface under consideration. For the future, the work may be extended to handle non-uniform albedo surfaces.

In chapter 6, the problem of night vision pedestrian detection was tackled which is important in the design of driving assistance systems to achieve a livable community. An annotated thermal imaging pedestrian dataset was automatically constructed using an available visible imaging pedestrian detector. This was done by registering the visible and thermal images such that pedestrians are approximately in the same place in both of them. The constructed dataset was obtained at day-time where the HOG [26] features were extracted for positive and negative samples and then used to train a discriminatively trained part based models classifier [20]. This system was mounted on a robotic vehicle and tested at night-time. The classifier showed good results in detecting pedestrians even in difficult situations as in the case of very low light or the presence of glare from oncoming traffic that can cause image saturation with visible imaging pedestrian detectors. For the future work, it is proposed to investigate using the proposed approach with pedestrian detection frameworks other than the discriminatively trained deformable part based models. Also the occlusion problem of pedestrians could be studied.

In chapter 7, a new feature based on HOG was proposed to be used with the discriminatively trained deformable part based framework for pedestrian detection [20] and the integral channel features framework for pedestrian detection [21]. The proposed feature is based on computing the image mixed partial derivatives to be used to redefine the gradients of some pixels and to reweigh the vote at all pixels with respect to the original HOG. The mixed partial derivative can be interpreted as the rate of change of the slope in the x direction while moving into the y direction or vice versa which carries information different than that of the derivative in a single direction as x or y.

The proposed system was tested on three publically available pedestrian datasets and the performance in all cases showed to be equivalent or better than existing approaches. For the future work, it is proposed to investigate using the discussed approach with other pedestrian detection frameworks.

REFERENCES

[1] J. Hollas. *Modern Spectroscopy*. WILEY, 2004.

[2] R. Cipolla, S. Battiato, and G. Farinella .*Computer Vision: Detection, Recognition and Reconstruction*. Springer, 2010.

[3] S. Lipson, H. Lipson and D. Tannhauser. *Optical Physics*. Third edition. Cambridge University Press, 1995.

[4] E. Hecht. *Optics*. Second edition. Addison-Wesley Publishing Company, 1987.

[5] C. Brosseau. *Fundamentals of Polarized Light, A Statistical Optics Approach*. John Wiley & Sons, 1998.

[6] M. Scully and M. Suhail Zubairy. *Quantum Optics*. Cambridge University Press, 1997.

[7] D. Ghosh, N. Ghosh and P. Haldar. *Engineering Physics*. University Science Press, 2008.

[8] W. Heisenberg. Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik. *Z. Phys*. 43, pp.172-198. 1927.

[9] P. Würfel and U. Würfel . *Physics of Solar Cells: From Basic Principles to Advanced Concepts.* Wiley, 2008.

[10] S. Eliezer, A. Ghatak, and H. Hora. *Fundamentals Of Equations Of State*. World scientific publishing Co. 2002.

[11] B. Soffer and D. Lynch. Some paradoxes, errors, and resolutions concerning the spectral optimization of human vision. *Am. J. Phys*., 67, pp.946–953. 1999.

[12] T. Guenault. *Statistical Physics*. second edition. Springer, 2007.

[13] Yan Li, Thomas X. Wu, and Shin-Tson Wu. Design Optimization of Reflective Polarizers for LCD Backlight Recycling. *Journal of Display Technology*, vol.5, no.8, pp.335-340, August 2009.

[14] Zhibing Ge, Xinyu Zhu, and Shin-Tson Wu. A transflective liquid crystal display using an internal wire grid polarizer. *Journal of Display Technology*, vol.2, no.2, pp.102-105, June 2006.

[15] A. Jha. *A Textbook Of Applied Physics.* Prentice Hall, 2011.

[16] A. EL-Barkouky, A. Mahmoud, J. Graham, and A. Farag. An Interactive Educational Drawing System Using A Humanoid Robot and Light Polarization. *ICIP*, 2013.

[17] Ruo Zhang, P-S Tsai, J.E. Cryer, and M.Shah, Shape-from-shading: a survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.21, no.8, pp.690-706, August 1999.

[18] G. Atkinson and E. Hancock. Recovery of Surface Orientation From Diffuse Polarization. *IEEE Transactions on Image Processing*, Vol. 15, no. 6, June 2006.

[19] P. Dollar, Z. Tu, P. Perona and S. Belongie. Integral Channel Features. *Proceedings of the British Machine Conference,* pp.91.1-91.11. BMVA Press, September 2009.

[20] P. Felzenszwalb, R. Girshick, D. McAllester, and D. Ramanan. Object Detection with Discriminatively Trained Part Based Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(9), pp.1627-1645, 2010.

[21] P. Dollár, R. Appel, S. Belongie, and P. Perona. Fast Feature Pyramids for Object Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, to appear, 2014

[22] A. Mahmoud, A. EL-Barkouky, H. Farag, J. Graham, and A. Farag. A Non-invasive Method for Measuring Blood Flow Rate in Superficial Veins from a Single Thermal Image. *PBVS CVPR workshop*, 2013.

[23] J. Robert Mahan. *Radiation Heat Transfer, A Statistical Approach*. WILEY, 2002.

[24] Lianxi Ma, Junjun Yang and Jiacai Nie. Two forms of Wien's displacement law *.Lat. Am. J. Phys*. Educ. Vol. 3, No. 3, September 2009.

[25] P. Felzenszwalb and D. Huttenlocher. Pictorial structures for object recognition. *International Journal of Computer Vision*, vol. 61, no. 1, 2005.

[26] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 886-893, June 2005.

[27] C. J. C. Burges. A Tutorial on Support Vector Machines for Pattern Recognition. *Data Mining and Knowledge Discovery*, Vol. 2, No. 2, pp.121-167, 1998.

[28] J. Friedman, T. Hastie, and R. Tibshirani. Additive logistic regression: a statistical view of boosting. *The Annals of Statistics*, vol. 38, no. 2, pp.337–374, 2000.

[29] P. Viola and M. Jones. Robust real-time object detection. *IJCV*, 57(2), pp.137154, 2004.

[30] P. Dollar, C. Wojek, B. Schiele, and P. Perona. Pedestrian detection: A benchmark. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.304-311, June 2009.

[31] P. Dollar, C. Wojek, B. Schiele, and P. Perona. Pedestrian Detection: An Evaluation of the State of the Art. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.34, no.4, pp.743-761, April 2012.

[32] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The PASCAL Visual Object Classes Challenge 2007 (VOC2007) Results. http://www.pascalnetwork. org/challenges/VOC/voc2007/workshop.

[33] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The Pascal Visual Object Classes (VOC) challenge. *Int. J. Comput*. Vis., vol. 88, no. 2, pp.303–338, June 2010.

[34] J. Scott Tyo, Dennis L. Goldstein, David B. Chenault, and Joseph A. Shaw. Review of passive imaging polarimetry for remote sensing applications. *J. Opt. Soc. Am*. A, Vol. 45, No. 22, August 2006.

[35] L. Wolff. Polarization vision: a new sensory approach to image understanding. *Image & Vision Comp*., vol. 15, pp.81-93, 1997.

[36] Yu.A. Vashpanov, Jung-Young Son, and Kae-Dal Kwack. Multispectral images in polarized light for medical applications. *IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, 2008. MFI 2008., vol., no., pp.86-89, 20-22 August 2008.

[37] O. Morel, C. Stolz, F. Meriaudeau, and P. Gorria. Active lighting applied to three-dimensional reconstruction of specular metallic surfaces by polarization imaging. *Appl. Opt*. 45(17), pp.4062-4068, 2006.

[38] A. Rankin and L. Matthies. Passive Sensor Evaluation for Unmanned Ground Vehicle Mud Detection. *Journal of Field Robotics*, 27(4), pp.473-491, 2010.

[39] S. Rahmann, and N. Canterakis. Reconstruction of specular surfaces using polarization imaging. *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2001)*, vol.1, no., pp.I-149- I-155 vol.1, December 2001.

[40] D. Miyazaki, M. Kagesawa, and K. Ikecuhi. Transparent Surface Modeling from a Pair of Polarization Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.26, No.1, pp.73-82, January 2004.

[41] G. Atkinson, and E. Hancock. Shape Estimation Using Polarization and Shading from Two Views. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.29, no.11, pp.2001-2017, November 2007.

[42] Cong Phuoc Huynh, A Robles-Kelly, and E. Hancock. Robust Shape from Polarisation and Shading. *20th International Conference on Pattern Recognition (ICPR),* vol., no., pp .810-813, August 2010.

[43] Cong Phuoc Huynh, A. Robles-Kelly, and E. Hancock. Shape and refractive index recovery from single-view polarisation images. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.1229-1236, June 2010.

[44] B. Horn. *Robot Vision*. McGraw-Hill, 1986.

[45] A. Mahmoud, M. El-Melegy and A. Farag. Direct Method for Shape Recovery from Polarization and Shading. *IEEE Int. conf. on Image Processing (ICIP)*, pp.1769-1772 September 2012.

[46] E. Prados, and O. Faugeras. Shape from shading: a well-posed problem?. *Computer Vision and Pattern Recognition (CVPR)*, pp.870- 877, June 2005.

[47] http://safety.fhwa.dot.gov/ped_bike/

[48] Yanwu Xu, Dong Xu, S. Lin, T. Han, Xianbin Cao, and Xuelong Li. Detection of Sudden Pedestrian Crossings for Driving Assistance Systems*, IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol.42, no.3, pp.729,739, 2012.

[49] D. Olmeda, J. Armingol, and A. de la Escalera. Discrete features for rapid pedestrian detection in infrared images. *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS),* vol., no., pp.3067-3072, 7-12 October 2012.

[50] K. Kidono, T. Naito, and J. Miura. Reliable pedestrian recognition combining high-definition LIDAR and vision data. *2012 15th International IEEE Conference on Intelligent Transportation Systems (ITSC),* vol., no., pp.1783-1788, 16-19 September 2012.

[51] M. Itoh, T. Inagaki, and H. Tanaka. Haptic steering direction guidance for pedestrian-vehicle collision avoidance. *2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC),* vol., no., pp.3327-3332, 14-17 October 2012.

[52] Yun Luo J. Remillard, and D. Hoetzer. Pedestrian detection in near-infrared night vision system. *Intelligent Vehicles Symposium (IV), 2010 IEEE* , vol., no., pp.51-58, June 2010.

[53] Liang Zhao, and C. Thorpe. Stereo- and neural network-based pedestrian detection. *IEEE Transactions on Intelligent Transportation Systems,* vol.1, no.3, pp.148-154, Sep 2000.

[54] K. Inumaru, Joo Kooi Tan, H. Hyoungseop, and S. Ishikawa. Detecting pedestrians from a video by stereo cameras. *ICCAS-SICE, 2009* , vol., no., pp.663-666, August 2009.

[55] S. Jayasumana, R. Hartley, M. Salzmann, Hongdong Li, and M. Harandi. Kernel Methods on the Riemannian Manifold of Symmetric Positive Definite Matrices. *IEEE Conference on Computer Vision and Pattern Recognition*, pp.73-80, 2013.

[56] A. Bar-Hillel, D. Levi, E. Krupka, and C. Goldberg. Part-based feature synthesis for human detection. *11th European Conference on Computer Vision (ECCV),* pp.127-142, 2010.

[57] D. Levi, S. Silberstein, and A. Bar-Hillel. Fast multiple-part based object detection using KD-Ferns. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.947-954, June 2013.

[58] F. Suard, A. Rakotomamonjy, A. Bensrhair, and A. Broggi. Pedestrian Detection using Infrared images and Histograms of Oriented Gradients. *IEEE Intelligent Vehicles Symposium*, pp.206-212, 2006.

[59] Shyang-Lih Chang, Fu-Tzu Yang, Wen-Po Wu, Yu-An Cho, and Sei-Wang Chen. Nighttime pedestrian detection using thermal imaging based on HOG feature. *2011 International Conference on System Science and Engineering (ICSSE)*, pp.694-698, June 2011.

[60] Wanli Ouyang, and Xiaogang Wang. Single-Pedestrian Detection aided by Multi-pedestrian Detection. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* pp.3198-3205, June 2013.

[61] www.bmw.com

[62] www.mbusa.com

[63] Xiaofeng Ren, and D. Ramanan. Histograms of Sparse Codes for Object Detection. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.3246-3253, June 2013.

[64] Bongjin Jun, Inho Choi, Daijin Kim. Local Transform Features and Hybridization for Accurate Face and Human Detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* , vol.35, no.6, pp.1423-1436, June 2013.

[65] P. Dollar, S. Belongie, and P. Perona. The fastest pedestrian detector in the west. *BMVC*, 2010.

[66] R. Benenson, M. Mathias, R. Timofte, and L. Van Gool. Pedestrian detection at 100 frames per second. *2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), vol., no.,* pp.2903-2910, June 2012.

[67] S. Chandrasekhar. *Radiative Transfer*. Dover Publications, Inc., 1960.

[68] F. Inropera and D. DeWitt. *Introduction to Heat Transfer*. Second Edition. WILEY, 1990.

[69] A. Andreou and Z. Kalayjian. Polarization Imaging: Principles and Integrated Polarimeters. *IEEE Sensors Journal*, Vol. 2, No. 6, pp.556-576, December 2002.

[70] R. Chipman. *HANDBOOK OF OPTICS*. College of Optical Sciences University of Arizona Tucson, Arizona.

[71] F. Goudail, and J. Scott Tyo. When is polarimetric imaging preferable to intensity imaging for target detection?. *J. Opt. Soc. Am. A*, Vol. 28, No. 1, pp.46-53, January 2011.

[72] F. Goudail, P. Refregier. Contrast definition for optical coherent polarimetric images. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol.26, no.7, pp.947-951, July 2004.

[73] L. Wolff, and T. Boult. Constraining object features using a polarization reflectance model. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol.13, no.7, pp.635-657, July 1991.

[74] http://opencv.willowgarage.com

[75] http://www.bossanovatech.com/polarization_imaging.htm

[76] http://www.flir.com/US/

[77] Qinfen Zheng, and R. Chellappa. Estimation of Illuminant Direction, Albedo, and Shape from Shading. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.13, no.7, pp.680-702, July 1991.

[78] P. Strum. Self-calibration of a moving zoom-lens camera by pre-calibration. *Image and Vision Computing*, 15, pp. 583-589, 1997.

[79] P. Kovesi. Shapelets correlated with surface normals produce surfaces. *Tenth IEEE International Conference on Computer Vision (ICCV)*, vol.2, no., pp.994-1001, October 2005.

[80] E. Trucco and A. Verri. *Introductory Techniques for 3-D computer Vision*, Prentice Hall, 1998.

[81] A. Pentland. Finding the illuminant direction. *Journal of the Optical Society of America*, vol. 72, issue 4, p.448, 1982.

[82] Yufei Zhang, Yee-Hong Yang. Illuminant direction determination for multiple light sources. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. vol.1, pp.269-276, 2000.

[83] S. Yuen, Y. Tsui, Y. Leung, and R. Chen. Fast marching method for shape from shading under perspective projection. *In Proceedings of VIIP'02*, pp.584–589, September 2002.

[84] A. Vedaldi, B. Fulkerson. VLFeat: An open and portable library of computer vision algorithms. http://www.vlfeat.org/, 2008.

[85] A. Mahmoud, A. EL-Barkouky, J. Graham, and Aly Farag. Pedestrian Detection Using Thermal Imaging for Night Driving Assistance. Accepted for publication *in International Journal of Multimedia Technology (IJMT)*, 2014.

[86] A. Prioletti, A. Mogelmose, P. Grisleri, M. Trivedi, A. Broggi, and T. Moeslund. Part-Based Pedestrian Detection and Feature-Based Tracking for Driver Assistance: Real-Time, Robust Algorithms, and Evaluation. *IEEE Transactions on Intelligent Transportation Systems*, vol.14, no.3, pp.1346-1359, September 2013.

[87] A. Mahmoud, A. EL-Barkouky, J. Graham, and A. Farag. Pedestrian detection using mixed partial derivative based histogram of oriented gradients. Accepted for publication in *2014 IEEE International Conference on Image Processing (ICIP)*, 2014.

CURRICULUM VITAE

Contact Information

<u>Name:</u>
Ali Mahmoud

<u>Address:</u>
Department of Electrical and Computer Engineering
University of Louisville
Computer Vision and Image Processing Laboratory
Lutz Hall, Room #006
2211 South Brook Street
Louisville, KY 40292, USA

<u>E-mail:</u>
ali.mahmoud@louisville.edu

<u>Phone:</u>
(502) 852-6130

Education

• Currently, a PhD candidate at the ECE department of University of Louisville with GPA 4.0
• Degree of M.Sc in Electrical Engineering (Electronics and Communications) obtained in September 2009 from Faculty of Engineering, Alexandria University, Egypt. Thesis entitled "Channel Estimation in Orthogonal Frequency Division Multiplexing".

• Degree of B.Sc in Electrical Engineering (Electronics and Communications) obtained in June 2005 with an overall grade Distinction with degree of Honor from Faculty of Engineering, Alexandria University, Egypt. Graduation Project entitled "Multimedia Mobile Communications Systems".

• High School Education obtained from El Nasr Boys English School (E.B.S) with a percentage of 98.3%.

Employment History

• Graduate Research Assistant at the robotics section of the Computer Vision and Image Processing Laboratory (CVIP Lab.), ECE department at University of Louisville, Kentucky, United States, from 11th of January 2010 till now.

• Assistant Lecturer at the Department of Electrical Engineering, Faculty of Engineering, Alexandria University (From the 19th of October 2009 till the 28th of December 2009).

• Demonstrator at the Department of Electrical Engineering, Faculty of Engineering, Alexandria University (From the 30th of April 2006 to the 18th of October 2009).

Programming Experience

• C/C++ (gcc/g++ with Linux systems and Microsoft Visual Studio with Windows systems).

• Linux (Mandriva, Ubuntu, Mageia, and Gentoo).

• Familiar with Makefiles and Cmake.

• Converting C/C++ codes between Linux and Windows systems.

• Familiar with parallel processing in C++ using OpenMP.

• Some experience with Python.

• Player robot server which is free software that enables research in robot and sensor systems.

•OpenCV

• MATLAB
• Cross compiling between a Linux system and the humanoid robot manufactured by Aldebaran.

Robotics, Sensors, and Electronics Experience

• Refurbished the robotic vehicle platform of the CVIP lab (ATRV2) as he upgraded its hardware and built its Linux based software. The refurbishment process involved soldering, module wiring and networking. The drivers and interfaces to all platform modules were written in C/C++. He developed an algorithm for pedestrian detection on that robotic platform which is an important problem in driving assistance.

• Contributed to using the humanoid robot NAO to develop an interactive educational drawing system using a humanoid robot and light polarization. He was responsible mainly for the programming (C/C++ on a Linux system) and cross compiling the code to have a stand-alone system on the robot NAO. This work won the first place in 5th annual

graduate research symposium poster session and the second place in the E-Expo 2014. This system was used with normal kids and is being currently tested with kids with autism.

• Setup a network connecting all the robotic platforms of the CVIP lab (ATRV2, ATRVmini, NAO), so now all the robotic platforms can interface with each other.

• Thermal Imaging as he developed an algorithm for measuring a fluid flow rate from a single thermal image. Also, he used thermal imaging for pedestrian detection at night.

• Polarization Imaging as he developed a direct method for shape recovery from polarization and shading making use of the direction of the electric field accompanying the Light.

• Sonars and LADARs.

• Visible cameras and Kinect.

• Electronic compass, Inertial Measurement Unit, GPS, and pan/tilt units.

• Familiar with most of the modules of NAO manufactured by Aldebaran (visual and audio sensors, motors, sonar, etc.)

• Added to the BOSS face recognition system at the CVIP Lab a hardware based module for capturing simultaneous stereo pairs of thermal images.

• During the period from the 30th of April 2006 till the 28th of December 2009, he assisted in setting up some laboratories at the Department of Electrical Engineering, Faculty of Engineering, Alexandria University (Solid State Electronics, Logic Design, Optical Devices, Advanced Communications Systems, Analog Communications, Microprocessors Interfacing, Acoustics, and Electrical Measurements).

Publications and Training

• Ali Mahmoud, Ahmed EL-Barkouky, James Graham, and Aly Farag, "Pedestrian detection using mixed partial derivative based histogram of oriented gradients", Accepted for publication in 2014 IEEE International Conference on Image Processing (ICIP), 2014.

• Ali Mahmoud, Ahmed EL-Barkouky, James Graham, and Aly Farag, "Pedestrian Detection Using Thermal Imaging for Night Driving Assistance", Accepted for publication in International Journal of Multimedia Technology (IJMT), 2014.

• Ahmed EL-Barkouky, Ahmed Shalby, Ali Mahmoud, and Aly Farag, "Selective Part Models for Detecting Partially Occluded Faces in the Wild", Accepted for publication in 2014 IEEE International Conference on Image Processing (ICIP), 2014.

• Ali Mahmoud, Ahmed EL-Barkouky, Heba Farag, James Graham, and Aly Farag, "A Non-invasive Method for Measuring Blood Flow Rate in Superficial Veins from a Single

Thermal Image", 2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp.354,359, June 2013

• Ahmed EL-Barkouky, Ali Mahmoud, James Graham, and Aly Farag, "An Interactive Educational Drawing System Using a Humanoid Robot and Light Polarization", 2013 IEEE International Conference on Image Processing (ICIP), Sept. 2013.

• Ali Mahmoud, Moumen El-Melegy and Aly Farag, "Direct Method for Shape Recovery from Polarization and Shading", Proc. of 2012 IEEE International Conference on Image Processing (ICIP), pp.1769-1772, 2012.

• Eslam Mostafa, Aly Farag, Ahmed Shalaby, Asem Ali, Travis Gault, and Ali Mahmoud, "Long Term Facial Parts Tracking in Thermal Imaging for Uncooperative Emotion Recognition", The IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS 2013), Sept. 29 – Oct. 2, 2013.

• Ahmed Shalaby, Ali Mahmoud, Eslam Mostafa, Ahmed Abdoulmalek, and Aly A. Farag, "Segmentation framework of vertebral body using 2D-PCA", STEM, 2012.

• Ahmed Shalaby, Ali Mahmoud, Ahmed Abdoulmalek, and Aly A. Farag, "Registration of Human Foot using X-ray and CT images", STEM, 2012.

• Ali Mahmoud, Aly Farag, Aly Abdelrahim, and Said Elnoubi, "A channel estimation technique in OFDM systems using no pilots," IEEE International Conference on Multimedia Technology (ICMT), pp.6680 - 6683, July 2011.

• Aly Abdelrahim, Mostafa Abdelrahman, Ali Mahmoud, and Aly Farag, "Image retrieval based on content and image compression," IEEE International Conference on Multimedia Technology (ICMT), pp.6696 - 6703 , July 2011.

• Ali Mahmoud and Said Elnoubi " Channel Estimation in OFDM Systems using Deterministic Modeling for Stochastic Frequency Selective Mobile Radio Channels" at the 26th National Radio Science Conference, Egypt, March 2009.

• Attended a training and earned a certificate in thermal imaging at ITC center in May 2012 entitled "An introduction to infrared thermography for research and testing"

• Earned a certificate in thermal imaging entitled "Thermography Basics" in May 2012.

• Earned a certificate in thermal imaging entitled "WEB-SW30 Getting Started with R&D Thermography Systems" in May 2012.

• IT training at Vodafone Egypt, July 2004.

Awards

Electrical Engineering Outstanding Graduate Student Award, ECE department, University Of Louisville, April 22, 2014.

- Martha and Frank Diebold Award in recognition of exceptional work to inspire engineering for change, University Of Louisville, April 22, 2014.
- 2nd Place in Engineering Exposition 2014 Student Research Competition, Graduate Category, University Of Louisville, March 22, 2014.
- 5th Annual Graduate Research Symposium Poster Presentation Award (1st Place), University Of Louisville, April 23, 2013.
- 5th Annual Graduate Research Symposium Oral Presentation Award (2nd Place), University Of Louisville, April 23, 2013.