How polarimetry may contribute to understand reflective road scenes : theory and applications

Fan Wang

Normandie Université

THÈSE

Pour obtenir le grade de Docteur

opéré par INSA de Rouen

Discipline : *Informatique* **Spécialité :** *Vision par ordinateur*

How polarimetry may contribute to understand reflective road scenes:

Theory and Applications

Présentée et soutenue publiquement par

Fan WANG

Thèse soutenue publiquement le 16 Juin 2016 devant le jury composé de		
M. Olivier COLOT	Professeur, Université de Lille 1	Rapporteur
M. Majdi KHOUDEIR	Professeur, Université de Poitiers	Rapporteur
M. Olivier MOREL	Maître de conférences, Université de Bourgogne	Examinateur
M. Stéphane MOUSSET	Maître de conférences, Université de Rouen	Examinateur
M. Abdelaziz BENSRHAIR	Professeur, INSA de Rouen	Directeur de thèse
Mme. Samia AINOUZ	Maître de conférences, INSA de Rouen	Encadrante

Thèse dirigée par Abdelaziz Bensrhair et Samia Ainouz, laboratoire LITIS

INSA INSTITUT NATIONAL DES SCIENCES APPLIQUÉES ROUEN

SPMii ECOLE DOCTORALE

litis

To my dear husband and parents...

My big thesis is that
although the world looks messy and chaotic,
if you translate it into the world of
numbers and shapes, patterns emerge
and you start to understand
why things are the way they are.
—Marcus du Sautoy

Acknowledgements

My forty-two months' mission has been finished and concluded in this manuscript, with my tears and joys inside. My adventure in science may not be as exciting as the board journal of the Enterprise, but I thank the persons that will have the patience to read it. Also this thesis would not exist without the help of several essential persons.
First of all, I would like to express my gratitude to my Ph.D. director, prof. Abdelaziz Bensrhair for his guidance. Also, I could not have managed to do all this work without INSA de Rouen. Second, I would like express my gratitude to the jury that accepted to review my thesis: prof. Olivier Colot (Université de Lille1), prof. Majdi Khoudeir (Université de Poitiers), Assoc prof. Olivier Morel (Université de Bougogne) and Assoc prof. Stephane Mousset (Université de Rouen). I won't forget the help from Assoc. prof. Caroline Petitjean, for the interesting discussions and corrections. I am also grateful to prof. Stephane Canu, for all the discussions we have shared and I am impressed by your motivation and ideas in scientific research. This list includes my supervisor Samia Ainouz, for her continuous guidance and help for the past years, both professionally and personally.
I would like to thank also my fellow comrades, the PhD students with whom I have shared ideas, anxieties and joy: Alina, Rawia, Dina, Fabien, Yadu, Nadeen, Nawal. Also a special thank to my dear friends, Leslie Lu, Yi Wu, Hongjie Wang, Changwu Huang, Yingchun Wu, Zelong Ma, Yujun Cao, Xiaoshuang Cai, Yahao Chen, Bo jiang, Yuan Liu, Liangjun Chen, that had the patience to listen to me complain during these years.
Last, but not the least, I would like to thank my love, Chunfeng Lian, and my dear parents, who supported me to follow my dreams and gave me continuous encouragement.

Rouen, 15 June 2016 Fan WANG

Abstract

Advance Driver Assistance Systems (ADAS) aim to automate/adapt/enhance transportation systems for safety and better driving. Various research topics are emerged to focus around the ADAS, including the object detection and recognition, image understanding, disparity map estimation etc. The presence of the specular highlights restricts the accuracy of such algorithms, since it covers the original image texture and leads to the lost of information. Light polarization implicitly encodes the object related information, such as the surface direction, material nature, roughness etc. Under the context of ADAS, we are inspired to further inspect the usage of polarization imaging to remove image highlights and analyze the road scenes.

We firstly propose in this thesis to remove the image specularity through polarization by applying a global energy minimization. Polarization information provides a color constraint that reduces the color distortion of the results. The global smoothness assumption further integrates the long range information in the image and produces an improved diffuse image.

We secondly propose to use polarization images as a new feature, since for the road scenes, the high reflection appears only upon certain objects such as cars. Polarization features are applied in image understanding and car detection in two different ways. The experimental results show that, once properly fused with *rgb*-based features, the complementary information provided by the polarization images improve the algorithm accuracy.

We finally test the polarization imaging for depth estimation. A post-aggregation stereo matching method is firstly proposed and validated on a color database. A fusion rule is then proposed to use the polarization imaging as a constraint to the disparity map estimation.

From these applications, we proved the potential and the feasibility to apply polarization imaging in outdoor tasks for ADAS.

Key words: Polarization, image understanding, car detection, fusion, disparity

Résumé

Les systèmes d'aide à la conduite (ADAS) visent à automatiser/ adapter/ améliorer les systèmes de transport pour une meilleure sécurité et une conduite plus sûre. Plusieurs thématiques de recherche traitent des problématiques autour des ADAS, à savoir la détection des obstacles, la reconnaissance de formes, la compréhension des images, la stéréovision, etc. La présence des réflexions spéculaires limite l'efficacité et la précision de ces algorithmes. Elles masquent les textures de l'image originale et contribuent à la perte de l'information utile. La polarisation de la lumière traduit implicitement l'information attachée à l'objet, telle que la direction de la surface, la nature de la matière, sa rugosité etc. Dans le contexte des ADAS, l'imagerie polarimétrique pourrait être utilisée efficacement pour éliminer les réflexions parasites des images et analyser d'une manière précise les scènes routières.

Dans un premier temps, nous proposons dans cette thèse de supprimer les reflexions spéculaires des images via la polarisation en appliquant une minimisation d'énergie globale. L'information polarimétrique fournit une contrainte qui réduit les distorsions couleurs et produit une image diffuse beaucoup plus améliorée.

Nous avons ensuite proposé d'utiliser les images de polarisation comme une caractéristique vu que dans les scènes routières, les hautes réflexions proviennent particulièrement de certains objets telles que les voitures. Les attributs polarimétriques sont utilisés pour la compréhension de la scène et la détection des voitures. Les résultats expérimentaux montrent que, une fois correctement fusionnés avec les caractéristiques couleur, les attributs polarimétriques offrent une information complémentaire qui améliore considérablement les résultats de la détection.

Nous avons enfin testé l'imagerie de polarisation pour l'estimation de la carte de disparité. Une méthode d'appariement est proposée et validée d'abord sur une base de données couleur. Ensuite, Une règle de fusion est proposée afin d'utiliser l'imagerie polarimétrique comme une contrainte pour le calcul de la carte de disparité.

A partir des différents résultats obtenus, nous avons prouvé le potentiel et la faisabilité d'appliquer l'imagerie de polarisation dans différentes applications liées aux systèmes

d'aide à la conduite.

Mots clefs : Polarisation, compréhension de la scène, détection des voitures, fusion, carte de disparité

Contents

List of Figures

List of Tables

Introduction

Road accident is causing 1.24 million death and between 20 and 50 million injuries each year [36]. It is predicted to become the fifth major cause of death by 2030 [42]. Since human error is the main cause for occurrence of accidents, it is crucial to equip the vehicle with safety systems. The Advance Driver Assistance System (ADAS) take proactive steps to prevent an accident before it happens, including obstacle detection and recognition, stereo vision, localization etc. Since *rgb*-image only describes limited information, it is fused with different sensors, such as Radar, Laser or Lidar based sensors, to provide an improved result.

Image polarimetry describes the vibration pattern of the reflected light, which carries important information about the reflection surface, such as material type, object shape etc. It has been applied into various indoor computer vision tasks, including medical imaging [10], object detection [66], 3D surface reconstruction [38] etc. Toward outdoor applications, it has been used as image dehazing [47], visual surveillance enhancing [33] etc. Up to its application in ADAS, it has received much less attentions than other no-conventional sensors.

A common problem in the traditional ADAS imaging system, especially for road scenes, is the presence of the image highlights, since it covers the original information on the images and hence restricts the accuracy of some detection or recognition algorithms. Since polarization images implicitly encode surface-related information, including surface material, orientation and roughness, it can be properly used to remove the image highlight or analyze the scene with the presence of the highlights.

Main objective

Since the different polarization patterns carry important information about the environment, such as the material type, the object shape etc, with proper analysis, the polarization information can be used to enhance the performance of the outdoor

computer vision tasks. Aiming to solve the problem of the specular reflection which leads the computer vision tasks to fail:

1 We firstly proposed to remove the specular reflection to recover the desired scene, since specular and diffuse reflections show different polarization patterns and that polarization images provide complementary color constraints.

2 Towards the context of road scene images, since some special polarization pattern occurs only under certain objects (e.g. cars which have smooth metal surface and transparent windows), we propose to use polarization patterns as features to better detect or recognize these objects.

3 Since the specular highlight and the transparent area on the object make stereo matching result erroneous, we intend to use the polarization information to constraint the stereo matching process and improve the disparity map computation on such areas.

Thesis overview

In the first chapter, after a brief presentation of polarization theory, an introduction of different polarization formalisms is given. The measurement of polarization state through Stokes vector is also presented since it is the only measurable formalism for the outdoor context. Based on the measurement principal, we introduce two frequently used polarization systems, being a manual rotating system and an automatic system. In the end of this chapter, several applications of polarization in computer vision are presented, namely : shape from polarization, image dehazing and specularity removal.

The main contributions of this thesis is presented in three parts : the specularity removal, the polarization imaging used as image features and the stereo matching, where the polarization is used either to remove the image highlights, or to analyze the image in the presence of the highlights. In the first part, we propose an image specularity removal method based on global energy minimization. The global energy function is constructed by an independence assumption and a constraint given by a first approximate assumption.

Since polarization images implicitly encode some surface properties, which provides complementary information to *rgb*-images, in part II we propose to use polarization image as a new image feature for semantic segmentation (chapter 3) and car

detection (chapter 4). The polarization features are extracted and fused in different ways for these two tasks. In semantic segmentation, both HOG, LBP and Lab feature are extracted from both polarization and *rgb*-images and then concatenated into a long feature vector. The joint boosting classifier, as a fusion process, is used to select informative features by iterations. In car detection, a feature selection among various polarization features is firstly performed. Two car detection models are trained separately using the selected polarization image and the *rgb*-image. The final result is given through a proposed fusion rule.

In the part III (chapter 5), We attempt to apply polarization image into stereo matching. A proposed post-aggregation stereo matching method on *rgb*-image to improve the disparity map. A fusion rule is proposed to combine the polarization image and the resulted disparity map.

Finally, a perspective is given in chapter 6 to conclude the limitation and the future work following the realized work through this thesis.

1 Polarization imaging

1.1 Polarization of the light

One of the common properties of all types of waves is polarization. Light is considered as an electromagnetic (EM) wave since the establishment of *"theory of wave optics"* by Fresnel and *"theory of electromagnetism"* by Maxwell. As light radiation is translated through synchronized oscillations of electric and magnetic fields (as seen in Figure 1.1), the polarization of the light can be determined by analyzing only the electric field. Light is considered to be *polarized* when its vibration has a fix directional

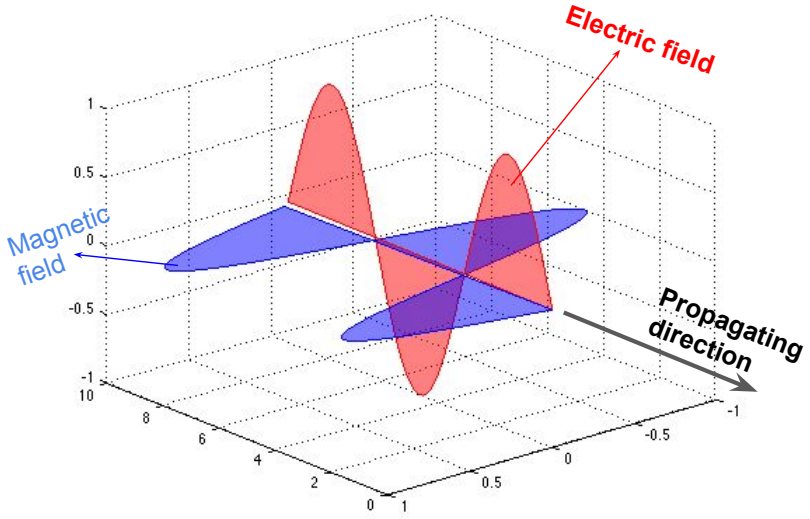


Figure 1.1 – Translation of the wave in both electric and magnetic fields

preference. Otherwise, when the light vibrates randomly in all directions, it is defined

to be *unpolarized*. Most of nature light sources provide unpolarized light, such as the sun, the bulb or a candle. Polarized light is more specifically called linearly polarized, circularly polarized or elliptically polarized according to the behavior of the wave vibration.

In this chapter, we firstly introduce some background of polarization in section 1.1, then we recall several mathematical representations of the polarization. In section 1.2, the principal of the polarization measurement method is described followed by introducing two different systems. Finally, some examples about polarization application are given at the end of this chapter.

1.1.1 Types of polarization

The EM wave $\vec{E}$ can be decomposed according to two orthogonal directions $\vec{E}_x$ and $\vec{E}_y$ as follows:

$$\vec{E} = \vec{E}_x(z,t) + \vec{E}_y(z,t) \tag{1.1}$$

where z is the propagation direction of the light wave and t is the time. The computation of $\vec{E}_x$ and $\vec{E}_y$ are given by

$$\begin{cases} \vec{E}_x(z,t) = \vec{E}_{0x} cos(kz - \omega t + \delta_x)\vec{x} \\ \vec{E}_y(z,t) = \vec{E}_{0y} cos(kz - \omega t + \delta_y)\vec{y} \end{cases}, \tag{1.2}$$

where k is the wave number (if λ is the wave length, k is defined $k = 2\pi/\lambda$), ω is the angular frequency, δ_x and δ_y are the phase angles of $\vec{E}_{0x}$ and $\vec{E}_{0y}$ respectively, $\vec{E}_x(z,t)$ and $\vec{E}_y(z,t)$ are the peak magnitudes in the x and y directions. These two components behave independently in reflection and refraction process. Based on different configurations of their magnitudes and phases, different types of polarization are generated. The most general case of polarization is the elliptical polarization. As shown in Figure 1.2, an elliptical polarization is defined by its orientation ψ and ellipticity χ ($\chi = \frac{b}{a}$), where a and b are respectively the semi-major and semi-minor axes of the ellipse.

From Equation (1.2), the combination of E_x/E_{0x}, E_y/E_{0y} and their dot product results

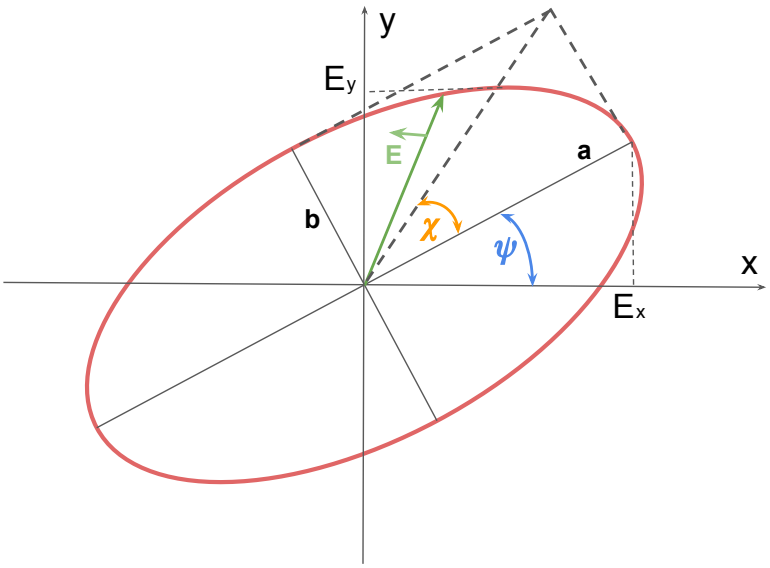


Figure 1.2 – Elliptical polarization

in an elliptical configuration as :

$$(\frac{E_x}{E_{0x}})^2 + (\frac{E_y}{E_{0y}})^2 - 2\frac{E_x}{E_{0x}}\frac{E_y}{E_{0y}}cos\delta = sin^2\delta \tag{1.3}$$

where $\delta = \delta_y - \delta_x$ From this equation, the different polarization states are obviously derived. Indeed, the polarization state is :

- linear when $\vec{E}_x$ and $\vec{E}_y$ have the same phase or have the difference of $\pm\pi$, the electric field of light will vibrate in a fixed plane.
- circular when $\vec{E}_x$ and $\vec{E}_y$ share equal amplitude and have the phase difference of $\pm\pi/2$,
- elliptical otherwise

Figure 1.3 illustrates the different stats of polarization. When a beam of light does not have a consistent pattern of its polarization ellipse, it is said to be *unpolarized*. That is to say δ_x and δ_y varies randomly in Equation (1.2). In contrast, when a beam of light has a constant ellipticity χ and orientation angles ψ, it is *totally (completely) polarized*.

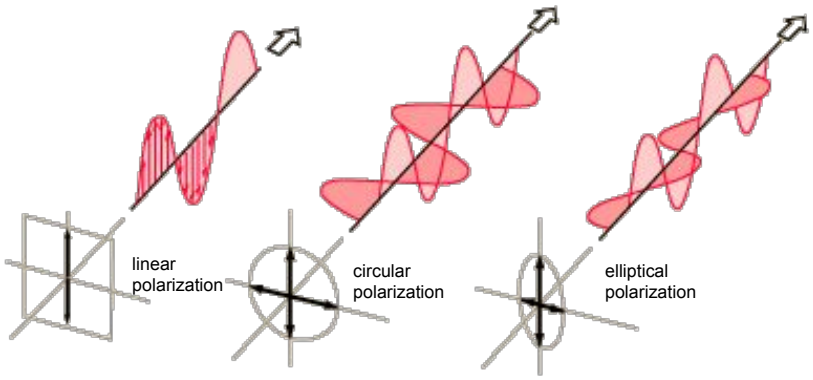


Figure 1.3 – Types of polarization [picture courtesy of Department of physics and astronomy Georgia State University]

When, however, the light is a superposition of a completely polarized component and an unpolarized component, it is known as *Partially polarized* light.

1.1.2 Polarization from reflection

Polarization is usually generated from reflection and scattering, where reflection could be related to diffraction, absorption and refraction. This phenomenon largely occurs in natural environment, but is not visible for human perception except with the involvement of optical instruments.

Light is polarized from reflection when strikes a surface. The polarization state is decided by both the material of the surface and the angle of incidence. From the Snell's law and the reflection law, when the angle of incidence of an unpolarized light equals the angle of Brewster (θ_B defined in Equation (1.4)), the reflected light is totally polarized, otherwise, it is partially polarized.

$$tan\theta_B = \frac{n_2}{n_1} \tag{1.4}$$

where n_1 and n_2 are the refractive indexes of the air and the material respectively. Figure 1.4 illustrates the refractive index as well as the incident angle θ_i, the reflected angle θ_r and the refracted angle θ_t. The relationship among these light parameters is

defined as:

$$\begin{cases} \theta_i = \theta_r \\ n_1 sin\theta_i = n_2 cos\theta_t \end{cases} \tag{1.5}$$

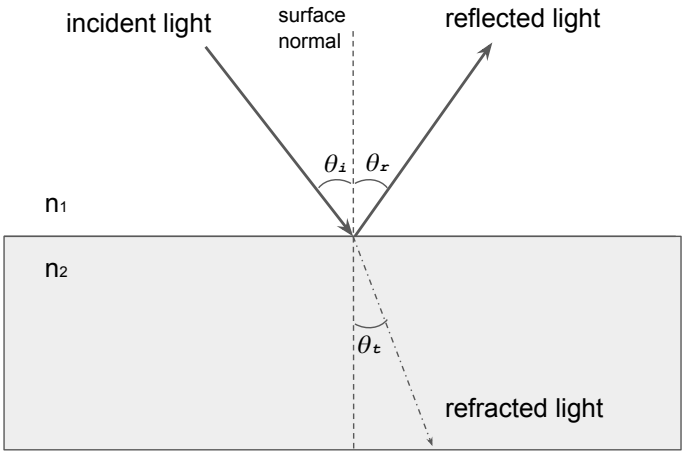


Figure 1.4 – Polarization by reflection

Let $A_\perp$ and $A_\parallel$ be the amplitudes of the incident light, and $R_\perp$ and $R_\parallel$ be the amplitudes of the reflected light. The relationship between the amplitudes of the reflected light and the incident light are described using *the Fresnel coefficients* defined by the following formula :

$$\begin{cases} f_\perp = \frac{R_\perp}{A_\perp} = -\frac{sin(\theta_i - \theta_t)}{sin(\theta_i + \theta_t)} \\ f_\parallel = \frac{R_\parallel}{A_\parallel} = -\frac{tan(\theta_i - \theta_t)}{tan(\theta_i + \theta_t)} \end{cases} \tag{1.6}$$

The reflective coefficient are determined as: $F_\perp = |f_\perp|^2$ and $F_\parallel = |f_\parallel|^2$.

1.2 Polarization formalism

Different mathematical representations are used to analysis light polarization, such as Jones calculus, Stokes vector or Mueller matrix. In this section, we briefly introduce all these different representations.

1.2.1 Jones calculus

Jones calculus is first introduced by R. C. Jones in 1941 to describe the polarized light. More specifically, the polarized light is represented by the complex Jones vector as :

$$V = \begin{bmatrix} V_x \\ V_y \end{bmatrix} = \begin{bmatrix} E_x^{i\phi_x} \\ E_y^{i\phi_y} \end{bmatrix} \tag{1.7}$$

where ϕ_x and ϕ_y are the absolute phases of the component in x and y directions. The intensity of the light is summarized in a 2×2 matrix, named the Jones matrix. It is computed through the product of V with its conjugate transpose.

$$J = V^\dagger V \tag{1.8}$$

The Jones matrix (J) describes a system that for an input incident light V_i, the output reflected light V_o can be defined as:

$$V_o = JV_i \tag{1.9}$$

It is important to note that this formalism is only applicable to totally polarized light.

1.2.2 Wolf's coherency matrix

For the partially polarized light, the parameters E_x,E_y,ϕ_x and ϕ_y may be randomly distributed around an intermediate. To describe the randomness of the wave, an intermediate matrix is introduced, called the coherency matrix. This matrix describes the covariance between the Jones components V_x and V_y of the light beam. It is defined as the temporal means of the product of the Jones vector V_t in an instant t and its conjugate transpose:

$$\Phi = \langle V(t) \otimes V^\dagger(t) \rangle = \begin{bmatrix} \Phi_{xx} & \Phi_{xy} \\ \Phi_{yx} & \Phi_{yy} \end{bmatrix} = \begin{bmatrix} \langle V_1(t)V_1^\dagger(t) \rangle & \langle V_1(t)V_2^\dagger(t) \rangle \\ \langle V_2(t)V_1^\dagger(t) \rangle & \langle V_2(t)V_2^\dagger(t) \rangle \end{bmatrix} \tag{1.10}$$

with:

$$V(t) = \begin{bmatrix} V_1(t) \\ V_2(t) \end{bmatrix} = \begin{bmatrix} E_x(t)^{i\phi(t)} \\ E_y(t)^{i\phi(t)} \end{bmatrix} \tag{1.11}$$

The coherency matrix is Herminia and positively-definite by construction. It fully describes a partial polarized light. The trace of the matrix represents the total intensity of the light I_0.

1.2.3 Stokes parameters

In 1852, Sir Gabriel Stokes proved that the polarized light can be represented by a vector S which contains four measurable parameters. These parameters are obtained as the combination of the coherency matrix components and defined as:

$$\begin{aligned} S_0 &= \langle E_x^2 \rangle + \langle E_y^2 \rangle \\ S_1 &= \langle E_x^2 \rangle - \langle E_y^2 \rangle \\ S_2 &= 2\langle E_x E_y cos(\phi_y - \phi_x) \rangle \\ S_3 &= 2\langle E_x E_y sin(\phi_y - \phi_x) \rangle \end{aligned} \tag{1.12}$$

Stokes parameters are directly computed from the measured intensity through the optical system. In this way, the Stokes parameters could also be defined according to the intensity by the following formula:

$$\begin{aligned} S_0 &= I_{0^o} + I_{90^o} \\ S_1 &= I_{0^o} - I_{90^o} \\ S_2 &= I_{45^o} + I_{-45^o} \\ S_3 &= I_{RCP} + I_{LCP} \end{aligned} \tag{1.13}$$

According to the polarization state of the light, I_{0^o} represent the linear horizontal component of the intensity, I_{90^o} the linear vertical component, I_{45^o} and I_{-45^o} are the linear components oriented at 45^o and -45^o respectively and I_{RCP}, I_{LCP} are the right and the left circular components. The first term S_0 represents the total transmitted intensity transmitted. S_1 represents how much the tendency that the light is horizontally polarized is larger than the tendency that the light is vertically polarized . S_2 measures how much the tendency that the light is polarized in 45^o is larger than at -45^o . S_3 represents how much the tendency that light is right circularly polarized is larger than the tendency that the light is left circularly polarized in (I_{LCP}). These four parameters are grouped in a 4×1 vector, named the Stokes Vector S. The different polarization states can also be described in term of the amplitude and the phase of the light wave as in Table 1.1.

From its construction, the stokes vector satisfy the physical admissibility constraint defined by :

$$\begin{cases} S_0^2 \geqslant S_1^2 + S_2^2 + S_3^2 \\ S_0 > 0 \end{cases} \tag{1.14}$$

For different polarization states, the Stokes vector is observed to be :

- Linearly polarized if $S_3 = 0$
- Circularly polarized if $S_1 = 0, S_2 = 0, S_3 \neq 0$
- Totally polarized if $S_0^2 = S_1^2 + S_2^2 + S_3^2$
- partially polarized if $S_0^2 > S_1^2 + S_2^2 + S_3^2$
- Unpolarized if $S_1 = S_2 = S_3 = 0$

Partially polarized light is a superposition of unpolarized light and totally polarized light. In order to quantify the amount of the polarized part of a partially polarized wave, the *degree of polarization* (ρ) is introduced and defined by the following equation :

$$\rho = \frac{\sqrt{S_1^2 + S_2^2 + S_3^2}}{S_0} \tag{1.15}$$

For the linearly polarized light, the degree of polarization is is simplified so that:

$$\rho = \frac{\sqrt{S_1^2 + S_2^2}}{S_0} \tag{1.16}$$

Note that $0 \leq \rho \leq 1$, and thus:

- $\rho = 0$ for non-polarized light
- $\rho = 1$ for totally polarized light
- $0 < \rho < 1$ for partially polarized light

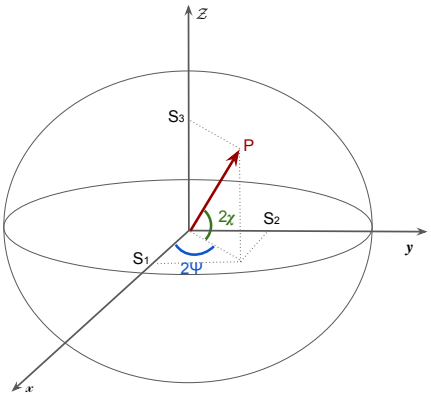


Figure 1.5 – Poincare sphere

The *angle of polarization* φ defines the orientation of the polarized part of the wave light (shown as ψ in Figure 1.2) . It is computed as:

$$\varphi = \frac{1}{2} tan^{-1}(\frac{S_2}{S_1}) \tag{1.17}$$

1.2.4 Poincare sphere

When the Stokes vector is normalized by its first component S_0 ($\bar{S} = (1, \frac{S_1}{S_0}, \frac{S_2}{S_0}, \frac{S_3}{S_0})^T$), it may be represented on a unit sphere called the Poincare sphere (from the name of Henri Poincare). This provides a unique geometric representation of any partially polarized or totally polarized light. Any point $P = (\bar{S_1} = S_1/S_0, \bar{S_2} = S_2/S_0, \bar{S_3} = S_3/S_0)$ on the Poincare sphere could be defined by its latitude 2ψ ($-\pi \leqslant 2\psi \leqslant \pi$) and its longitude 2χ ($-\pi/2 \leqslant 2\chi \leqslant \pi/2$). The coordinates of the point P on the sphere are computed as:

$$P = \begin{cases} \frac{S_1}{S_0} = cos(2\psi)sin(2\chi) \\ \frac{S_2}{S_0} = sin(2\psi)cos(2\chi) \\ \frac{S_3}{S_0} = sin(2\chi) \end{cases} \tag{1.18}$$

In the case of the totally polarized light, its Stokes vector is uniquely found on the surface of the sphere. For an unpolarized light, it can be considered like containing

Table 1.1 – Different polarization states

Polarization State	Amplitude	Phase	Equation	Stokes Parameter
Linear horizontal polarization	$E_{0y} = 0$, $E_0 = E_{0x}$	$\phi = \pm n\pi$, $n \geqslant 0$	$E = E_x + E_y = E_x$, $E_x(z,t) = iE_0 cos(kz - wt)$	$\begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}$
Linear vertical polarization	$E_{0x} = 0$, $E_0 = E_{0y}$	$\phi = \pm n\pi$, $n \geqslant 0$	$E = E_x + E_y = E_y$, $E_y(z,t) = jE_0 cos(kz - wt)$	$\begin{pmatrix} 1 \\ -1 \\ 0 \\ 0 \end{pmatrix}$
Linear polarization at $+45^o$	$E_0 = E_{0x} = E_{0y}$	$\phi = \pm n\pi$, $n \geqslant 0$	$E = E_x + E_y$, $E = (i + j)E_0 cos(kz - wt)$	$\begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix}$
Linear polarization at -45^o	$E_0 = E_{0x} = E_{0y}$	$\phi = \pm n\pi$, $n \geqslant 0$	$E = E_x + E_y$, $E = (i - j)E_0 cos(kz - wt)$	$\begin{pmatrix} 1 \\ 0 \\ -1 \\ 0 \end{pmatrix}$
Right circular polarization	$E_{0x} = E_{0y}$	$\phi = \pm n\pi/2$, $n \geqslant 0$	$E = E_0(icos(kz - wt) + jsin(kz - wt))$	$\begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}$
Left circular polarization	$E_{0x} = E_{0y}$	$\phi = \pm n\pi/2$, $n \geqslant 0$	$E = E_0(icos(kz - wt) - jsin(kz - wt))$	$\begin{pmatrix} 1 \\ 0 \\ 0 \\ -1 \end{pmatrix}$

all other states, so that it is represented by points that uniformly distributed on the whole surface of the sphere. For the case where the light is partially polarized, it is represented by a group of points around P, where P has the maximum probability of the polarization state. The more concentration of the points distributes around P, the more the state tends to be polarized. This behavior is quantified by the degree of polarization.

Each circle of the attitude corresponds to a fixed azimuth angle, and each longitudinal line matches the same ellipticity. The north pole and the south pole represents two orthogonal states, being right circular polarization and left circular polarization respectively. Their ellipticities are equal to ± 1 with azimuth angles equal to infinity. The equator of the sphere represents the linear polarization where the ellipticities are zero.

1.2.5 Mueller matrix

Hans Mueller at 1943 first introduced the Mueller matrix which fully characterizes a polarization element [54]. When the light travels through or reflected by an optical element, its Stokes vector is modified. The reflected output Stokes vector S_o is a function of the incident stokes vector S_i and its interaction with the optical element (modeled by the Mueller matrix M). This relationship between S_o and iS_i is given by :

$$S_o = MS_i \tag{1.19}$$

If a beam of light S_i passes through n optical elements with their respective Mueller matrices being $M_1, M_2, ..., M_n$, the Stokes vector of the output light S_o is:

$$S_o = M_n M_{n-1} ... M_1 S_i \tag{1.20}$$

It is important to note that the stokes vector characterize the reflected light while the Mueller matrix translates the intrinsic properties of the optical element it represents.

1.2.6 Polarizer

A polarizer is an optical element which is able to unequally modify the amplitude of the electric field components E_x and E_y. The character of a polarizer is defined by

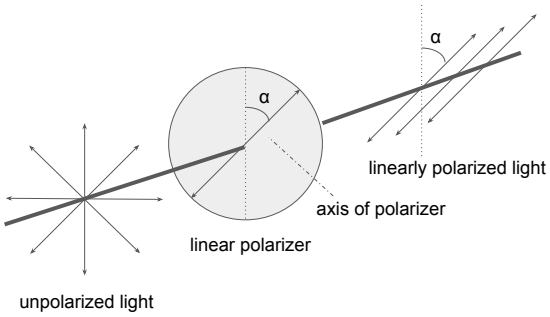


Figure 1.6 – Linear polarizer oriented at α

the state of the polarization of the output light, known as circular polarizer or linear polarizer. For an unpolarized light beam with known wave length, a polarizer can be used to polarize the light beam in the required direction. Note that this process considerably reduces the intensity of the light beam. For an ideal linear polarizer, 50% of the light intensity is transmitted. When an unpolarized light beam passes through two polarizers, the transmitted intensity varies from a maximum for parallel axes and a minimum when the axes are perpendicular. Based on the Malus' law, the transmitted intensity through a polarizer is computed as:

$$I_t = I_o cos^2\alpha \tag{1.21}$$

where I_t is the transmitted intensity, I_o is the original energy and α is the angle between the axis of the polarizer and the axis of the electric field (as shown in Figure 1.6).

The Mueller Matrix of a linear polarizer oriented at an angle α is:

$$M_{pol}(\alpha) = \begin{pmatrix} 1 & cos2\alpha & sin2\alpha & 0 \\ cos2\alpha & cos^2 2\alpha & cos2\alpha sin2\alpha & 0 \\ sin2\alpha & cos2\alpha sin2\alpha & sin^2 2\alpha & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{1.22}$$

Table 1.2 shows some Mueller matrices of different polarizers.

Table 1.2 – Example of optical elements and their corresponding Mueller Matrices

optical elements	Mueller Matrix	optical elements	Mueller Matrix
Horizontal Linear Polarizer	$\frac{1}{2}\begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$	Vertical Linear Polarizer	$\frac{1}{2}\begin{pmatrix} 1 & -1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$
Linear polarizer with a transmission axis at 45^o	$\frac{1}{2}\begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$	Linear polarizer with a transmission axis at -45^o	$\frac{1}{2}\begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$

1.3 Measurement principle

Usually, when an unpolarized light strikes a surface and be polarized through reflection, the reflected beam is linearly polarized and the direction of the polarization is parallel to the plane of the surface. To describe the linearly polarized light, only the first three Stokes parameters are used. The last component is considered to be zero since it refers to the circular part of the wave.

For the linear polarization, the axis of the light is oriented at φ and the degree of polarization is ρ. The Stokes vector could be written as:

$$\begin{aligned} S_0 &= I_{total} \\ S_1 &= \rho I_{total} cos2\varphi \\ S_2 &= \rho I_{total} sin2\varphi \\ S_3 &= 0 \end{aligned} \tag{1.23}$$

In the case of the partially linearly polarized light, ideally, the wave is passed unattenuated when its electric field is aligned with the axis of the polarizer. By rotating the polarizer, the wave is attenuated as a cosine function (Equation (1.21)) as shown in Figure 1.7. The transmitted intensity $I_p(\alpha)$ changes with the orientation of the polarizer α in a manner that:

$$I_p(\alpha) = \frac{1}{2}(S_0 + S_1 cos2\alpha + S_2 sin2\alpha) \tag{1.24}$$

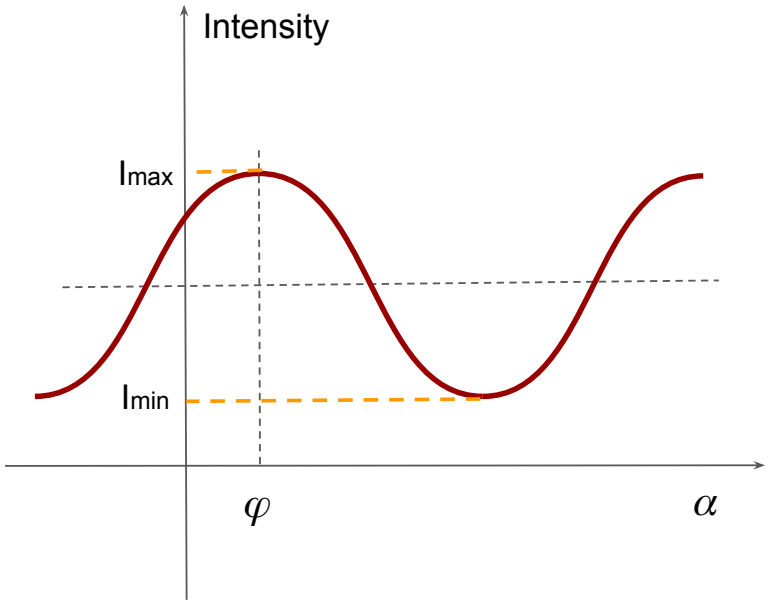


Figure 1.7 – Partial linear polarization

From Equation (1.23) and Equation (1.24), for an arbitrary linear polarization light with the angle of polarization φ, the degree of polarization ρ and the total intensity I, the intensity that passes through a polarizer oriented at α is:

$$I_p(\alpha) = \frac{1}{2} I_{total}(1 + \rho cos(2\alpha - 2\varphi)) \tag{1.25}$$

Consequently, a linear polarized light can be described using three parameters, being (S_0, S_1, S_2) or $(I_{total}, \rho, \varphi)$. To find these three unknown parameters, as introduced in [65], at least three measurements need to be achieved with different orientations of the polarizer. Wolff *et al.*[65] mentioned that $0^o, 45^o, 90^o$ as orientations of the polarizer ensure a better reception of the signal by the camera. From Equation (1.25),

the parameters I, ρ, φ are conveniently computed using:

$$\varphi = \begin{cases} \frac{1}{2} tan^{-1}[\frac{I_0 + I_{90} - 2I_{45}}{I_0 - I_{90}}] + \frac{\pi}{2}, & if\ I_{90} > I_0 \\ \frac{1}{2} tan^{-1}[\frac{I_0 + I_{90} - 2I_{45}}{I_0 - I_{90}}] + \pi, & if I_{90} < I_0\ and\ I_{45} < I_0 \\ \frac{1}{2} tan^{-1}[\frac{I_0 + I_{90} - 2I_{45}}{I_0 - I_{90}}], & if I_{90} < I_0\ and\ I_{45} > I_0 \end{cases} \tag{1.26}$$

$$\rho = \frac{I_{90} - I_0}{(I_{90} + I_0) cos2\varphi}$$

$$I = I_0 + I_{90}$$

1.4 Polarization imaging systems

From the measurement principal introduced in the last section, we introduce here two systems to acquire polarization images : the manual system and the automated system.

1.4.1 Manual rotating polarizer system

To measure the partial linear polarized light, one generally make use of a polarization system which installs a polarizer in front of a CCD camera (as shown in Figure 1.8). Three image are taken at three different positions $(\alpha_1, \alpha_2, \alpha_3)$ of the polarizer. This configuration measures the polarization parameters in a very easy way, but it is very sensitive to the noise, especially on the value of the angle of polarization. Some other methods use much more measurements (e.g. 36 measurements in [46]). The parameters I, ρ and φ are approximated using linear least mean squares. 36 measurements gives more noise-resistant, yet since it is much more time consuming, it is used mostly in laboratory level applications.

1.4.2 System of automated camera

As seen above, for polarization imaging systems, to compute polarization parameters, at least three intensity values are needed for each pixel. This configuration, is suitable only for static scenes, since the user needs some time to turn the orientation of the polarizer manually before taking each of the three images. For moving objects, hanlon *et al.*[23] designed an automated polarization camera as seen in Figure 1.9. In this

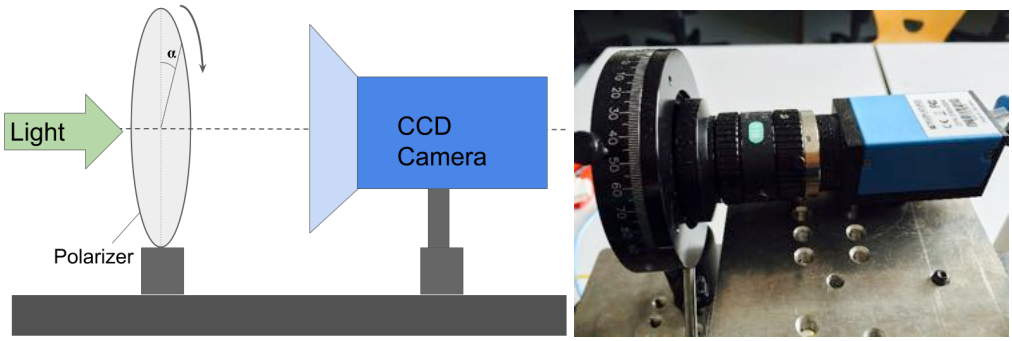


Figure 1.8 – Manual rotating polarizer system

configuration, a prism is replaced with a neutral prismatic beam-splitter made from chronic material. A linear polarizer (small disc Polaroid) is placed in front of each camera tube to distribute a different part of the polarization sensitivity (with directions of polarizers axes at $0^o, 45^o, 90^o$ respectively) for each of the three channels.

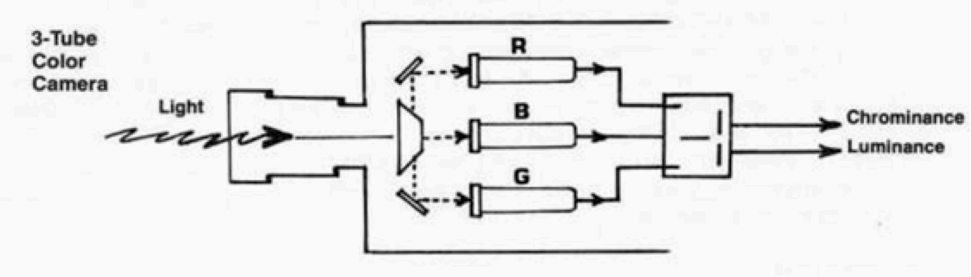


Figure 1.9 – Automated polarization camera setup in [23] [picture courtesy from [23]]

Similar configuration can be found in [17] with a 3-ways beam splitter. This camera consists of two coating surfaces; the first coating surface reflects 30% and transmits 70% of the light. The other coating surface provides a 50% reflectance and transmittance respectively. These combinations are splitting the incoming light into three components with equal spectral and spatial contents (see Figure 1.9). The orientation of the filters are commonly at $0^o, 45^o, 90^o$. This camera is configured with either color or monochrome sensors for each channel (see in Figure 1.10).

Such a polarization system setup enables the outdoor applications of the polarization since for outdoor applications, lots of computer vision problems focus on moving objects. These cameras are also more feasible for embedding on robots or cars.

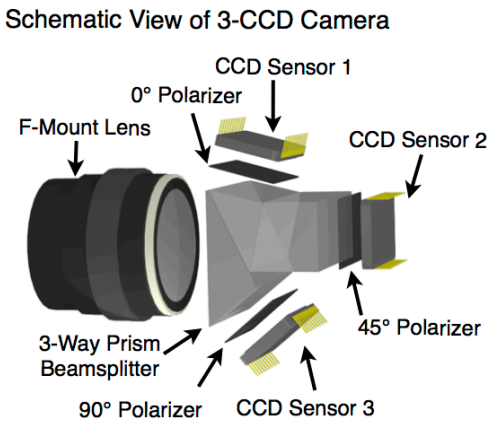


Figure 1.10 – Camera of Flux data FD-1665-P

1.5 Polarization imaging applications

Polarization parameters describe the light polarization states, which is not perceptible for human vision. This extra-information provides important cues for some computer vision problems. For this reason, a lots of polarization methods are applied in solving problems that attract our attention in the context of road scenes analysis which is at the heart of our work.

1.5.1 Shape from polarization

Conventional imaging system can solve shape reconstruction problems through modeling the reflection intensity and the reflection angle, which is computed through exhaustive iterations. While for unpolarized incident light, the polarization state of the reflected light depends on the angle of incidence and the normal of the surface. This information is an important cue for surface reconstruction and thus for shape from polarization.

Wolff *et al.*[66] presented a polarization reflectance model known as the Fresnel reflectance model using Fresnel reflection coefficients. Many researches has been achieved based on this model. In 1999,Rahmann [43] presented a method to estimate the orientation of a flat object and the position of the light source by polarization analysis. Later, Rahmann [44] estimated the shape of specular objects from multiple views.

For transparent object, Saito *et al.*[46] implemented a method to estimate surface orientation of transparent object using polarization highlights. Miyazaki *et al.*[37] resolved two ambiguity of the reconstruction of transparent objects by producing specular reflection over the whole object surface. By comparing the degree of polarization at the corresponding point of each set of polarization data, the unique surface orientation is determined.

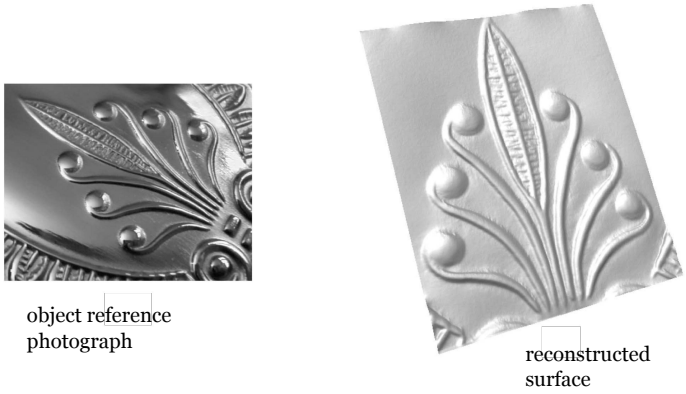


Figure 1.11 – 3D reconstruction of metallic surface [39]

Respect to the metallic surface, Morel *et al.*[38] proposed a shape reconstruction of smooth metallic surfaces using a Fresnel reflectance model with complex index of the refraction. In a later publication [39], the ambiguity about the determination of the azimuth angle from the angle of polarization is resolved by varying the illumination with an active lighting system. The result is shown in Figure 1.11. In 2007, Atkinson [3] combined the polarization and a shading information system for 3D reconstruction from two views of smooth non metallic surfaces. The main idea of Atkinson is to combine shading information to enhance the estimation of the normal of the surface by measuring some statistics on the pixel brightness that depends on the surface orientation.

1.5.2 Defogging

For image defogging, conventional computer vision methods require weather conditions to change between image acquisitions which is obviously not time efficient. However, as the scattering effect polarizes the transmitted light, polarization is used to restore the dehazed image without waiting the change of the weather conditions.

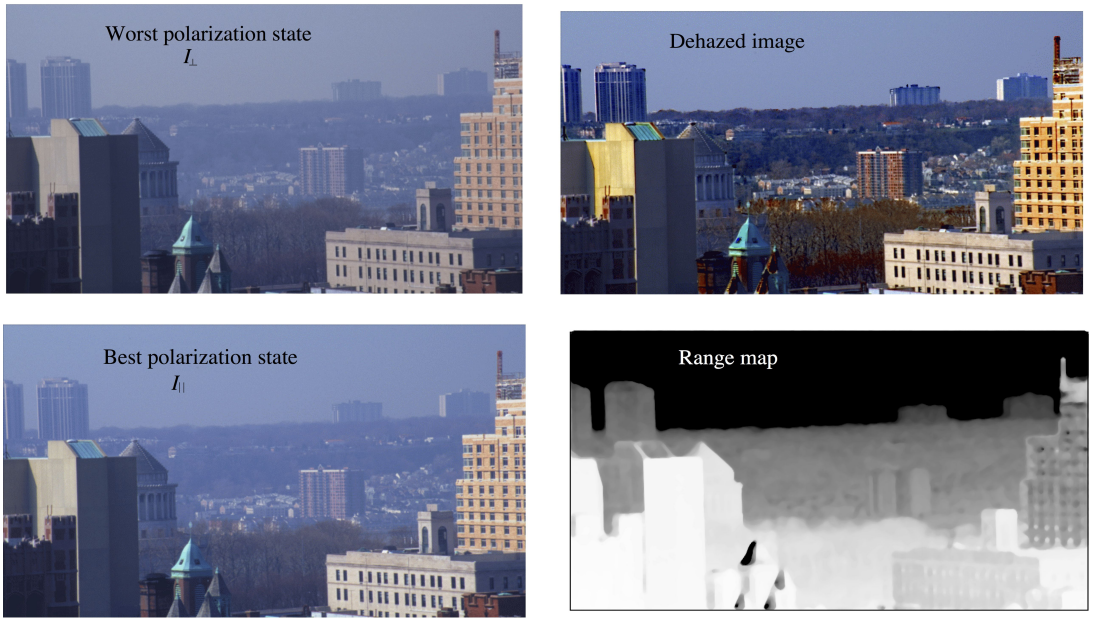


Figure 1.12 – Instant dehazing using polarization [47]

Schechner *et al.*in [47] and [48] used polarization to improve the quality of the image taken in hazy weather. In this approach the polarization effect of the atmospher is modeled and used to recover the dehazed scene. A brute depth map is obtained as a bonus, since the polarization effect increases with the distance of the object.

As seen Figure 1.12, the polarizer is rotated to take two images with the best dehaze orientation ($I_{\perp}$) and the worst dehaze orientation ($I_{\parallel}$). based on the polarization sum ($I_{\parallel} + I_{\perp}$) and difference ($I_{\parallel} - I_{\perp}$), this method computes a partial polarization of the airlight. Its stability depends, however, on the stability of the degree of polarization. This method may be less effective under an overcast sky and even may fail in very dense haze or foggy. Also the airlight (object from very far away, almost horizon) is needed and should be manually labeled.

1.5.3 Reflection separation

Another important application of the polarization is to separate the specular reflection and the diffuse reflection. Specular reflection occurs when the light is totally reflected in the same direction from a smooth surface. In contrast, diffuse reflection is the reflection such that an incident light is reflected at many directions as illustrated in

Figure 1.13. In most computer vision methods, only diffuse reflection is considered, while specular reflection can appears as highlights which covers the image texture and makes erroneous the imaging systems results.

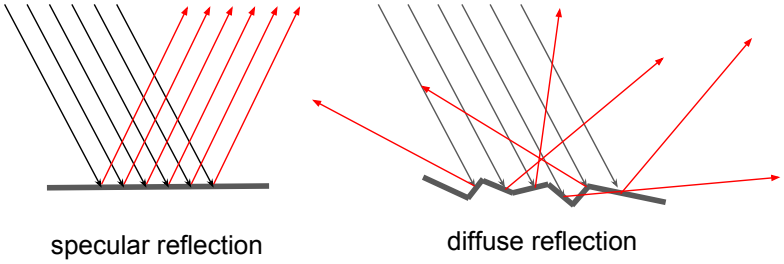


Figure 1.13 – Specular and diffuse reflections

When the light is reflected from a surface, for diffuse reflection, the reflected light is almost unpolarized (except on occluding contours). The polarization effect is almost generated by the specular reflection. Thus, polarization provides important cues to separate the specular and the diffuse reflections.

Wolff *et al.*[66] used gray level polarization images for separating reflection components. This method assumes that the diffuse component and the ratio of Fresnel coefficients [4]are constant, so that the diffuse component is estimated using a large set of image points lying on a pre-segmented highlight region.

Nayar *et al.*[40] proposed to separate reflection using the polarization in conjunction with the color information. The specular highlights are detected by giving a threshold on the degree of polarization. By using the dichromatic model, the specular color is decided on every single point of the image with the variation of the polarizer. Neighboring pixels are used to compute the specular scalar and thus the diffuse-only image is produced.

Kim *et al.*[27] extended the Nayar's work by dividing the color space into a specular line space and a diffuse plane space. The diffuse pixels are picked using a threshold on the intensity variation (which is also related to the degree of polarization) when turning the polarizer. An energy function is proposed to smooth the spatial variation of the specular component. They also proposed to address the probable over-exposure by using an inpainting technique. Umeyama *et al.*[61] assumed that the acquired image is a linear combination of a specular and a diffuse components, under the condition that the light source is far away from the object and that the angle of incidence does not

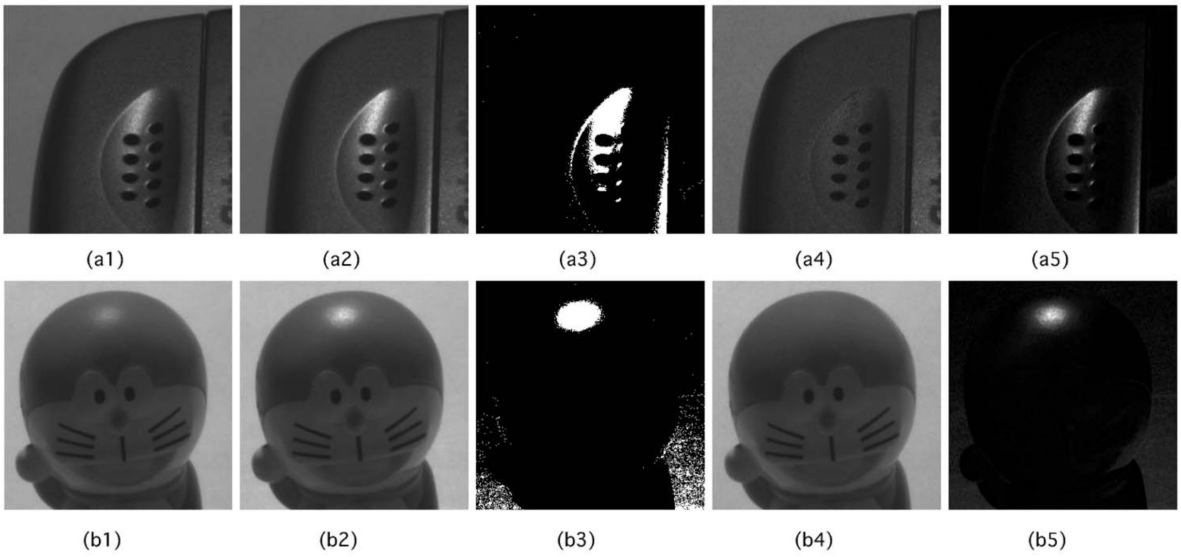


Figure 1.14 – Result of the separation in [61]

change. The combination parameter is reversed by using an independent component analysis (ICA) [25]. In Figure 1.14, we show some results that Umeyama *et al.*achieved. (a1), (a2) and (b1), (b2) are the input images takes by a polarizer oriented at 0^o and 90^o respectively, (a3) and (b3) are the estimated regions where possibly have specular reflection, (a4) and (b4) shows the results of the estimated diffuse components, (a5) and (b5) shows the estimated specular components.

1.6 Conclusion

This chapter gives a short description of the polarization theory, including different types of polarization. We presented several mathematical representations of the polarization including Jones Calculus, coherency matrix, Stokes vector, Poincare sphere and Mueller matrix. A convenient polarization measurement method and the corresponding manual and automated systems are also introduced. Regarding to the important information that polarization provides to different computer vision problems, we introduced several important applications of the polarization in computer vision, such as 3D reconstruction, image defogging and reflection separation. It is worth to note that the separation of specular and diffuse components greatly affects the accuracy of various computer vision algorithms, since specular reflection usually generates highlights and covers the real texture of objects in the image. To this end,

we put our efforts to improve the accuracy of the separation of specular and diffuse reflection to be efficiently applicable to complex road scenes.

Specularity removal Part I

2 Specularity removal using polarization imaging

2.1 Introduction

From the dichromatic reflection model [50], each brightness value in the image is viewed as the sum of two components, being the diffuse part and the specular part. The *diffuse* component arises from body reflection, when the light rays penetrate the body of the object, reflected back to the surface and then into the air. The *specular* reflection, however, is the result of a single surface reflection of incident light rays, where the angle of reflection equals to the angle of incidence [57]. For this reason, specular reflection usually occurs as a concentrated compact lobe in the image. The concentration of the light energy causes strong *highlights* and challenges the robustness of a large variety of vision algorithms. Since many algorithms (e.g. feature extraction, object detection) assume perfect diffuse surfaces and ignore specular reflections, the specularity removal is necessary and can be considered as a pre-processing step to improve the accuracy of such algorithms.

The major difference between specular and diffuse reflections, in the point of view of optics, is that they have different degrees of polarization (DOP). When a beam of an unpolarized light is reflected, the DOP of specular reflection is larger than that of the diffuse reflection for most angles of incidence. As the DOP represents the ratio of the light being polarized, this means that the specular reflection is generally much more polarized than the diffuse reflection [4].

2.1.1 Related work

Traditional methods separate the diffuse and the specular components based on *color-only* images, where the main idea is to find a variable which is independent from the specular component. By estimating this variable for each given pixel, its diffuse component is found.

As a foundation of the color-based methods, Tan *et al.*[58] inspected the specular component via the chromaticity, proved to be independent from the specular component. An additional hue-based segmentation method is required for the multi-colored surfaces. Yang *et al.*[69] extended this work by detecting diffuse pixels in the HSI space, which also requires the hue-based segmentation. The color covariance is defined as a constant variable to recover the diffuse component. Kim *et al.*[28] used the dark channel prior as a pseudo-solution and refined the result through the Maximum A Posteriori (MAP) estimation of the diffuse component. The dark channel prior, however, only works for highly colored surfaces. To avoid extra segmentation, Tan and Ikeuchi [57] proposed another diffuse pixel pick-up method via computing the logarithmic differential between up to four neighboring pixels. The common limitation of the above presented color-based methods, is their high color distortion on the recovered diffuse component[40, 57, 69]. The main reason is that these methods assume that the specular color is constant through-out the whole image.

To better recover the diffuse component, other methods proposed to accomplish the separation using *polarization* [66], since specular and diffuse components hold different DOPs. When rotating the polarizer, the change of the intensity is only related to the specular part, so that the intensity change refers directly to the specular color. With these color constraints, polarization based methods produce more accurate results with less color distortions. Nayar *et al.*[40] firstly constrained the diffuse color on a line in RGB space. The neighboring diffuse-only pixels are used to estimate the diffuse component, providing state-of-the-art locally polarization-based specularity removal results. The specular pixels are detected by simply tresholding the DOP. This term changes not only with different specular portions, but also with different incident angles, different refractive indexes etc. The computation of the DOP involves more than three images, making it largely contaminated by the camera noise. This makes Nayar's method restricted since its computation highly relies on the DOP. The existing literature solves the separation problem, either locally or globally (up to now, only few global methods have been proposed). Local methods, based on the dichromatic

reflection model [50], assume that the intensity of a pixel is a linear combination of its diffuse and specular components. Global-based methods, like in [61] simplified this model into the image level, under the condition that the light source is far away from the object and that the incident angle does not change. In other words, the acquired image is linearly combined by a specular image and a diffuse image with respect to a constant parameter. This parameter is reversed using the Independent Component Analysis (ICA) [25]. These ideal conditions discussed in [61] rarely conform to reality and only a part of the specular reflection component is removed.

With respect to the literature, we come out with two observations: (1) The color-based methods produce heavy color distortions; (2) the local-patch-based methods can only use the information offered by very closing pixels without any consideration of long range cues.

Based on these observations, we proposed a global method using the polarization setup and a local first approximate solution as detailed bellow [63].

2.1.2 Contribution

Our main contribution for the separation problem could be summarized by the following points :

1. Inspired from [61], we assume that the acquired image is linearly combined by a diffuse and specular reflection images. However, we depart from the use of a fixed coefficient and instead investigate the benefit of using a spatially varying coefficient, which generalizes the model proposed in [61] to better conform to the reality.

2. Based on these assumptions, a global energy function is constructed to better consider the long range information, and to produce more accurate and robust result than local-patch-based methods. The optimum solution is found by applying the graph-cut scheme [6].

3. Apart from the independency assumption, a first approximate solution is computed as a supplementary constraint. We propose to compute a more reliable approximate solution by combining the logarithm differential specular detection method as in [57] and the specular-to-diffuse mechanism as in [40].

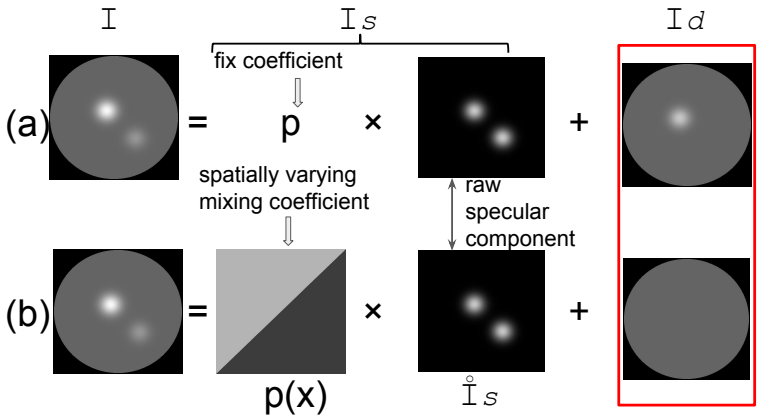


Figure 2.1 – An example of specularity removal process on synthetic image. (a) fixed coefficient as assumed by Umeyama *et al.*[61], where the specular component is not fully removed; (b) spatially varying coefficient as assumed in our approach, which more effectively removes the specular reflection.

4. A histogram-based criterion is proposed in the experimental part to quantitatively evaluate the results. The proposed method is compared with two well-known separation algorithms : Nayar's polarization setup [40] and Umeyama's method [61].

2.2 Polarization parameters

Light reflected from an object is generally partially linearly polarized. It is fully described using three parameters [2]: *light magnitude I*, *degree of polarization* ρ, *angle of polarization* φ. These parameters are measured with the polarization imaging system described in Figure 1.8 and computed using Equation (1.26). The terms I_{max} and I_{min} define the maximum and minimum intensity for each pixel as shown in the curve in Figure 1.7, and can be conveniently achieved with :

$$I = I_{max} + I_{min}, \quad \rho = \frac{I_{max} - I_{min}}{I_{max} + I_{min}} \tag{2.1}$$

2.3 Problem formulation

From the dichromatic reflectance model [29], a beam of light is a linear combination of the diffuse and the specular components. Umeyama *et al.*[61] simplified this model to the image level by assuming that the acquired image I is the sum of a diffuse image I_d and a specular image I_s, where I_s is defined as a combination of a raw specular image $\mathring{I}_s$ with a fixed mixing coefficient p (as shown in Figure 2.1 (a)). The image I is related to its components by the following formula:

$$\begin{cases} I = I_d + I_s \\ I_s = p\mathring{I}_s \end{cases}. \tag{2.2}$$

The diffuse and the specular reflection images are assumed to be probabilistically independent. The optimum p is found by minimizing the Mutual Information (MI) [**?**] between I_d and I_s to ensure their maximum independency. The raw specular image $\mathring{I}_s$ is defined as:

$$\mathring{I}_s = I_{max} - I_{min}, \tag{2.3}$$

where I_{max} and I_{min} are the maximum and the minimum intensities described in Equation (2.1).

This method globally models the diffuse and specular separation problem. However, it suffers from the following limitations:

- In real applications, the assumption that the incident angle on each pixel keeps constant is not always valid. Fixing the mixing coefficient over the whole image is inapplicable.
- Umeyama *et al.*[61] minimized the MI at the image level since the mixing coefficient is assumed to be constant over the whole image. However, the computation of MI does not suite for spatially varying mixing coefficient, since the sum of MI needs to be minimized at each local patch, which is too much time consuming (about 25 minutes for a 240×320 image).

To handle these limitations, we propose to reformulate the specularity removal problem by the following four points:

(1) The mixing coefficient is assumed to be spatially varying as shown in Figure 2.1(b). The computed diffuse component is shown in the lower part of the red square. The specular reflection is more effectively removed than that of Figure 2.1(a).

(2) We assume as in [61] that I_d and I_s are probabilistically independent. The goal is to minimize their similarity.

(3) Another similarity measurement is applied instead of MI which is much more time-efficient and produces competitive results.

(4) Since only maximizing the independency is not enough to produce the best solution, a first approximate solution is brought as a constraint to ensure the reliability of the final solution.

2.4 Global energy function

As global methods can better integrate the long range cue via the global smoothness assumption, they can produce more accurate and robust result than local-patch-based methods. For this purpose, we propose a global energy function, which is composed of a data term and a smoothness term.

As mentioned above, in this chapter we assume that the mixing coefficient is spatially varying. Equation (2.2) is transformed to:

$$\begin{cases} I(x) = I_d(x) + I_s(x) \\ I_s(x) = p(x)\mathring{I}_s(x) \end{cases} \tag{2.4}$$

where the $p(x)$ is the Local Mixing Coefficient (LMC). Our objective is to find an optimum $p(x)$ for each pixel x. As $\mathring{I}_s(x)$ is computed using Equation (2.3), once the $p(x)$ is found, its corresponding diffuse $I_d(x)$ and specular components $I_s(x)$ are computed using Equation (2.4).

The optimum $p(x)$ should be found by solving the following constraint optimization

problem:

$$\begin{aligned} &\underset{p(x)}{\operatorname{argmin}} \left[\sum_{x} \Phi(p(x)) + \lambda_1 \sum_{x} \sum_{y \in \mathcal{N}(x)} \Psi(p(x), p(y))] \right] \\ &\quad s.t.\ 0 \leq p \leq \tilde{p} \end{aligned} \tag{2.5}$$

where $\Phi(p(x))$ is the data term, $\Psi(p(x), p(y))$ is the smoothness term where $y \in \mathcal{N}(x)$ and $\mathcal{N}(x)$ is the 4-connected neighborhood of x. Term λ_1 is a hyper-parameter which balance the data and the smoothness terms. Scalar $\tilde{p}(x)$ is the upper-boundary so that the $I_d(x)$ is kept positive since it defines an intensity.

2.4.1 Data term

The data term $\Phi(p(x))$ contains two parts: a patch-based dissimilarity (independency) measurement $C_{DC}(p(x))$, and a pixel-wise constraint part $D(p(x))$ related to the first approximate solution:

$$\Phi(p(x)) = -\lambda_2 C_{DC}(p(x)) + D(p(x)) \tag{2.6}$$

where λ_2 is an empirically chosen hyper-parameter. Term $C_{DC}(p(x))$ is used to maximize the independency between the diffuse and the specular images. However, since only maximizing $C_{DC}(p(x))$ is not enough to guarantee the best solution, the constraint part $D(p(x))$ is brought into the data term. This constraint measures the distance between the final solution and a first approximate solution, thus ensuring the reliability of the result. In addition, this first solution used as the initialization to the optimization process also improves the time efficiency. By minimizing $\Phi(p(x))$, the optimum $p(x)$ is found through the trade-off between maximizing $C_{DC}(p(x))$ and minimizing $D(p(x))$.

Dissimilarity measurement

Since using the mutual information for a patch centered at each pixel is highly time consuming, instead, we apply the DIFF-census cost function ([35]) to measure the dissimilarity of the diffuse and the specular images. DIFF-census is used in [35] to optimize the disparity map. It is known to be resistant to noises and color distortions.

The advantage of this cost function is that it gives similar results as compared to mutual information and is less time consuming.

Given a pixel location x and an arbitrary $p(x)$ $(0 \leq p(x) \leq \tilde{p}(x))$, the independency between its corresponding diffuse component $I_d(x)$ and specular component $I_s(x)$ are computed as:

$$\begin{aligned} C_{DC}(p(x)) &= DIFFcensus(I_d(x), I_s(x)) \\ &= \rho(C_{census}(\bar{I}_d(x), \bar{I}_s(x)), \lambda_{census}) \\ &+ \rho(C_{DIFF}(\bar{I}_d(x), \bar{I}_s(x)), \lambda_{DIFF}) \end{aligned} \tag{2.7}$$

where $\bar{I}_d(x)$ and $\bar{I}_s(x)$ are the $n \times m$ patches $\mathcal{W}$ centered at x with an arbitrary size (the size of $\mathcal{W}$ is chosen to be 5×5 by default in our experimentation). λ_{DIFF} and λ_{census} are the hyper-parameters to balance the two parts, we use $\lambda_{DIFF} = 55$ and $\lambda census = 95$ as suggested in [35]. More specifically, in Equation (2.7), we have:

$$\begin{cases} C_{census}(\bar{I}_d(x), \bar{I}_s(x)) = H(CT(\bar{I}_d(x)), CT(\bar{I}_s(x))) \\ C_{DIFF}(\bar{I}_d(x), \bar{I}_s(x)) = \dfrac{|DIFF(\bar{I}_d(x)) - DIFF(\bar{I}_s(x))|}{n \times m} \end{cases} \tag{2.8}$$

where the $n \times m$ is the same size of the patch used in Equation (2.7), the $CT(\cdot)$ is the *Census Transform*, $H(\cdot)$ is the *hamming distance*. For more information about census transform and hamming distance, please refer to [71]. $DIFF(\cdot)$ is computed as:

$$\begin{cases} DIFF(\bar{I}_d(x)) = \sum_{y \in \bar{I}_d(x)} (|I_d(x) - I_d(y)|) \\ DIFF(\bar{I}_s(x)) = \sum_{y \in \bar{I}_s(x)} (|I_s(x) - I_s(y)|) \end{cases} \tag{2.9}$$

DIFF-census makes a trade-off between classical Census Transform (CT) and Sum of Absolute Difference (SAD) by introducing the hyper-parameters $\lambda DIFF$ and λ_{census}, which provides an improved result for the disparity map computation problem[35].

Constraint part

Lack of constraint, the optimum solution provided by the independency assumption alone over minimize the energy function, and does not provide a good separation result. We put a constraint to limit the final solution $p(x)$ to be close to the approximate

solution $p_{init}(x)$. The constraint is given by measuring their euclidean distance:

$$D(p(x)) = \sqrt{|p^2(x) - p^2_{init}(x)|} \tag{2.10}$$

The first approximate solution $p_{init}(x)$ is found by combining the logarithm differential specular detection method proposed by [57] and the specular-to-diffuse mechanism in [40]. The computation of $p_{init}(x)$ will be detailed in Section 2.4.3.

2.4.2 Smoothness term

We apply the smoothness term among the 4-connected neighborhood $\mathcal{N}(x)$ of the pixel x. To keep the original texture of the image, a color discontinuity detection based on thresholding the RGB values suggested in [57] is applied. Let Th_R and Th_G be small threshold values (this value is adjusted in our experiment according to the input images, $Th_R = Th_G = 0.005$ as in [57]), and indices r and g be respectively the red and the green channels in the RGB space. The following rule is applied to judge if a pixel sits on color discontinuous or not:

$$(\delta_r > Th_R \; and \; \delta_g > Th_G) \begin{cases} true\colon \; color - discontinuity \\ false\colon \; otherwise \end{cases} \tag{2.11}$$

where

$$\begin{cases} \delta_r(x) = \sigma_r(x) - \sigma_r(x-1) \\ \delta_g(x) = \sigma_g(x) - \sigma_g(x-1) \end{cases} \tag{2.12}$$

and

$$\begin{cases} \sigma_r = \dfrac{I_r}{I_r + I_g + I_b} \\ \sigma_g = \dfrac{I_g}{I_r + I_g + I_b} \end{cases} \tag{2.13}$$

The chromaticity changes are computed between neighboring pixels on both red channel and green channel (no need to include the blue channel since $\sigma_r + \sigma_g + \sigma_b = 1$). When change on two channels are both larger than the threshold, the pixel is referred

to as the color discontinuity. With the known color discontinuity, for $y \in \mathcal{N}(x)$, the smoothness term is defined as:

$$\Psi(p(x), p(y)) = \begin{cases} 0, & color\ discontinuity \\ \sqrt{p(x)^2 - p(y)^2}, & otherwise \end{cases} \tag{2.14}$$

2.4.3 First approximate solution

The first approximate solution $p_{init}(x)$ provides supplementary constraint to the data term to improve the accuracy of the model, as well as a sub-optimal initial solution to improve the efficiency of the optimal process.

To locally obtain an approximate solution, one usually follows two steps: *i*) diffuse region detection : detecting the pixels upon which only the diffuse component exists and the specular component is almost zero; *ii*) specular-to-diffuse mechanism : for the pixel where the specular component appears, inferring its diffuse component according to its neighborhood information.

Polarization-based methods, in step *i*), usually apply a simple thresholding on the DOP as in [40], which is unstable caused by the noise, and inconvenient as the threshold largely varies from a scene to another one. For the step *ii*), the color-based methods, because of the absence of the lightening information, estimate the specular color for the whole image, which is unreliable. Polarization images, contradictory, provides the specular color constraint conveniently.

To obtain a more stable first approximate solution, and handle the limitation of each step, we propose to combine the above two steps i) and ii) but coming from two different articles: step *i*) as proposed by [57], namely the logarithm differential, and step *ii*), the specular-to-diffuse mechanism as in [40].

Logarithm differential

The logarithm differential is a 'diffuse verification' process, which verifies each pixel if it is a diffuse-only pixel or not. This process is based on the assumption that for diffuse-only pixels who share the same chromaticity, their maximum chromaticity is only related with the brightness. That is to say, when their maximum chromaticity is shifted to a same value, their brightness should be the same. The *chromaticity* is

defined as the normalized RGB, being:

$$\sigma(x) = \frac{I(x)}{I_r(x) + I_g(x) + I_b(x)} \tag{2.15}$$

where $\sigma = \{\sigma_r, \sigma_g, \sigma_b\}$. And the *maximum chromaticity* is defined as :

$$\tilde{\sigma}(x) = \frac{max(I_r(x), I_g(x), I_b(x))}{I_r(x) + I_g(x) + I_b(x)} \tag{2.16}$$

For the whole image, the $I(x)$ is scaled so that the maximum chromaticity $\tilde{\sigma}(x)$ is turned to an arbitrary constant value (we set $\tilde{\sigma}(x) = 0.5$ as suggested by [57]). After this process, a new image $\boldsymbol{I'}$ is generated. Based on the above assumption, if two neighboring pixels are both diffuse pixels, their intensity logarithm differential should be zero. Otherwise, if they are not on the color discontinuity, they should be both specular pixels. The logarithm differential $\delta(x)$ is computed as:

$$\delta(x) = [log(I(x)) - log(I(x+1))] - [log(I'(x)) - log(I'(x+1))] \tag{2.17}$$

where

$$\delta(x) \begin{cases} = 0, \ diffuse-only \\ \neq 0, \ specular \ or \ color \ discontinuity \end{cases} \tag{2.18}$$

The ambiguity between specular and color discontinuity is suppressed via the color discontinuity detection described in Equation (2.11).

Specular-to-diffuse mechanism

By applying the logarithm differential process, all pixels have been verified if they are diffuse-only pixels, or they have both diffuse and specular components. The specular-to-diffuse mechanism is then applied upon the pixels who have both diffuse and specular components. For the purpose of convenience, we name these pixels *specular pixels*.

Nayar *et al.*[40] showed the diffuse and specular components as color vectors on RGB space as in Figure 2.2. The point D defines the diffuse-only intensity of a pixel x, its diffuse color vector is represented as $I_d(x)$. The $I_s(x) = p(x)\mathring{I}_s(x)$ is the color vector of its specular component. When a pixel contains both specular and diffuse components,

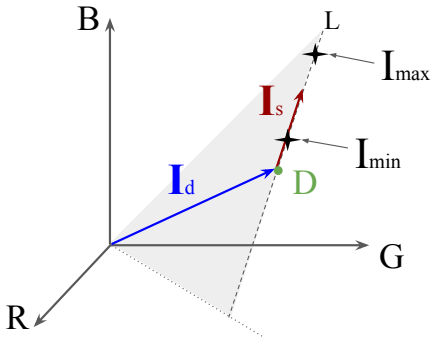


Figure 2.2 – Diffuse and specular components in RGB space

the intensity acquired by the camera should be the sum of I_d and I_s, which lies on the line L. Theoretically, since only the specular component I_s is polarized [66], by rotating the polarizer, $I(x)$ only varies along the line L. This line is the color constraint that can be obtained only through polarization. Line L is determined using the I_{max} and I_{min} through Equation (2.1). Since the specular component is rarely fully polarized, rotating the polarizer is not able to completely remove the specular component.

To get the point D in Figure 2.2 which is the diffuse-only component, the neighboring diffuse-only pixels are used. For $y_i \in \mathcal{N}_x$, where $\mathcal{N}_x$ is the neighborhood of x, D is estimated by applying the following rules:

- Each y_i provides a corresponding estimation D_y^i by computing the projection of $I(y)$ on the line L.
- $I(y)$ is used if it is a diffuse-only pixel. Alternatively, it can be used if its diffuse component has already been computed.
- $I(y)$ is used if it lies close to the gray plane in Figure 2.2 (the plane defined by I_d and L). This is verified by computing the angle between $I(y)$ and this gray plane.

After having all estimations D_y^i, the final D is found by the mean of all estimations.

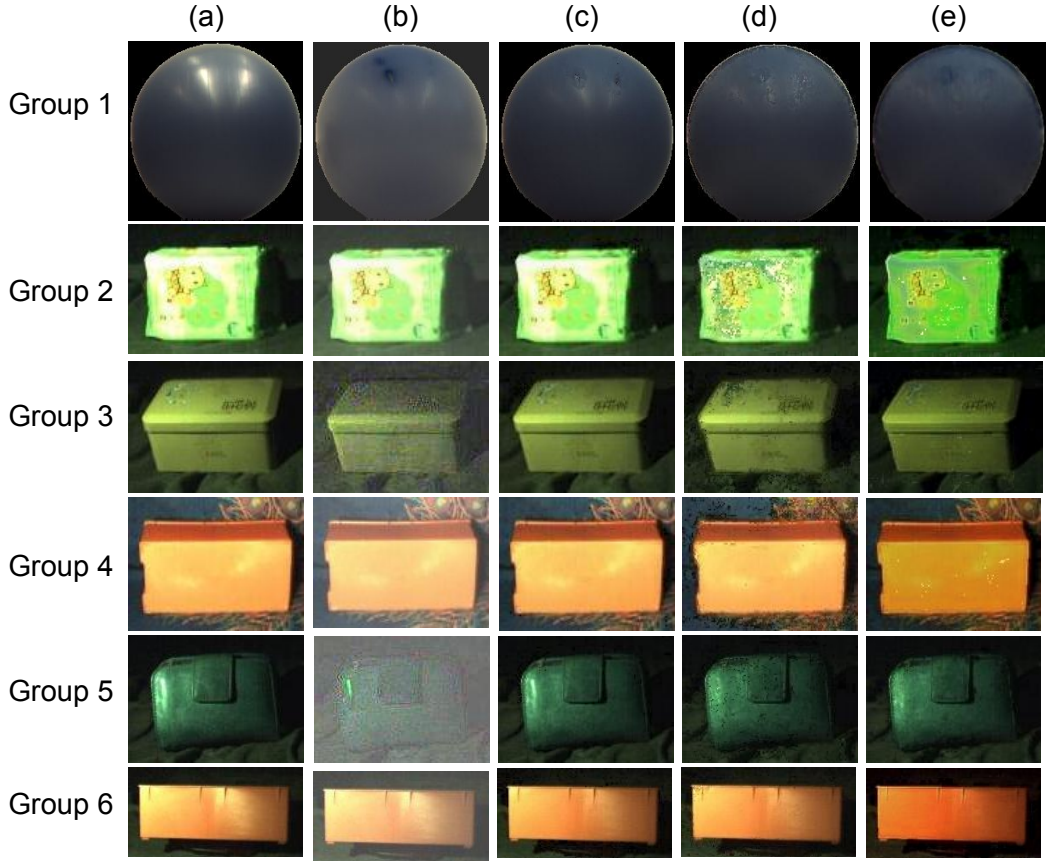


Figure 2.3 – (a) original image (I_0); (b) results of Umeyama's method; (c) results of Nayar's method; (d) first approximate solutions; (e) results of the proposed method.

2.5 Experimentation

2.5.1 Data and parameters

In the experiments, we compared the proposed approach with two well-known methods in the literature: Nayar's method [40] which provides the state-of-the-art local specular and diffuse separation using polarization and Umeyama's method which is a global polarization-based algorithm. The energy function was solved through graph-cut with 4-connected neighbors using the gco_v3.0 library [6], [30], [5]. The problem of optimizing $p(x)$ was formulated as a global labeling problem, where labels are from 0 to 255. The hyper-parameters were chosen as $\lambda_1 = 5$ (in Equation (2.5)) and $\lambda_2 = 1.5$ (in Equation (2.6)) for our experiments. It is worth to note that the specularity removal results are stable in the range of 3.5 to 7 for λ_1 and 1.3 to 1.7 for λ_2. The

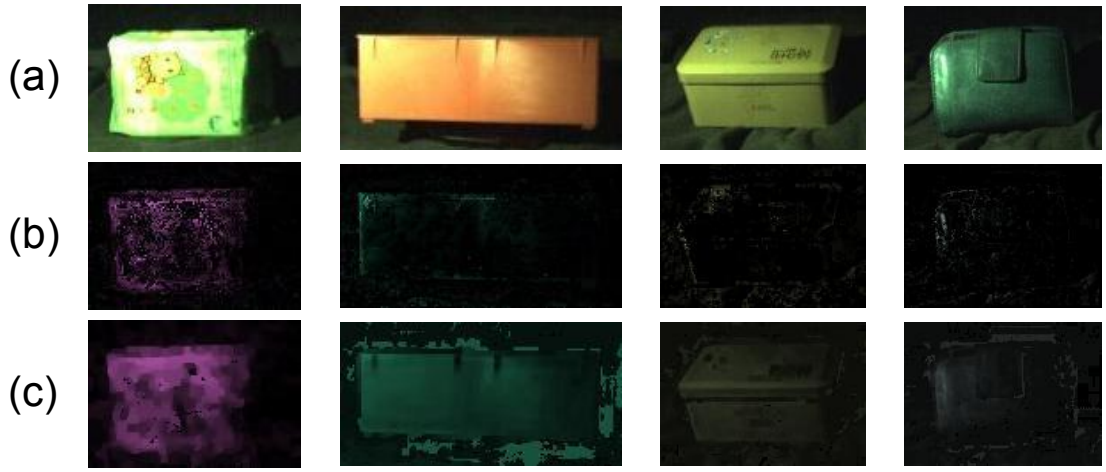


Figure 2.4 – (a) original image (I_0); (b) the first approximate solution P_{init} (c) the final solution P by the proposed method

algorithm was implemented on Matlab 2012a and C++ platform. The running time for an image of 240×320 in around 10 seconds including computing the first approximate solution, data term and the optimization process.

2.5.2 Visually evaluation

For visually evaluation, we show the diffuse component resulted by six groups of images (Figure 2.3 (a), where the I_0 is used) from four methods, being Umeyama's method [61] (Figure 2.3 (b)), Nayar's method [40] (Figure 2.3 (c)), our first approximate solution (Figure 2.3 (d)) and the proposed final solution (Figure 2.3 (e)). It can be observed that our proposed method produces the best specularity-removal result. Umeyama's method removes only a small part of the specular component, with a reduction on the contrast. The reason is that the assumption of uniform incident angle made by Umeyama does not hold in real situations. It also proves that the independency assumption alone is not able to yield a good result. The first approximate solution that we computed better detects the specular region, so that it produces similar but slightly better results that the Nayar's method that only removes part of the specularity. The proposed method produces the best specular removal method, thanks to the independency assumption, and the constraint given by the first approximate solution.

We also show some examples of specular components on four groups of images

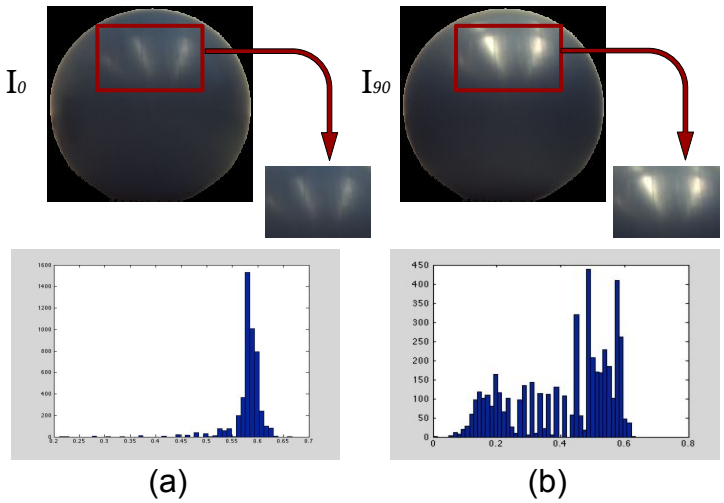


Figure 2.5 – (a) histogram of hue with I_0 (weak specular reflection); (b) histogram of hue with I_{90} (strong specular reflection)

(Figure 2.4 (a)) with their first approximate solution P_{init} (Figure 2.4 (b)) and their final solution P (Figure 2.4 (c)). It can be observed that on the first approximate solution, only a part of the specular component is figured out, while the global method can more completely estimate the specular component. It is worth to note that the global method is not only a smooth effect on the approximate solution, this should be the benefit from the independency assumption that we made on the data term.

2.5.3 Quantitative evaluation

As far as we know, there is no existing public polarization-based benchmark. The proposed approach was evaluated on four images taken with our polarization device composed by a polarizer[1] and a CCD camera[2].

In the literature, only visual comparison of results from different methods is given, however, no numerical evaluation is presented [40, 57, 68]. The reason is that, firstly, a ground truth is not always accessible; Secondly, the ground truth is acquired usually under an extreme dark illumination and thus not usable for error computation.

[1] http://www.edmundoptics.fr/optomechanics/optical-mounts/polarizer-prism-mounts/rotary-optic-mount/1978/

[2] http://www.theimagingsource.com/en_US/products/cameras/gige-cmos-ccd-color/dfk33gv024/

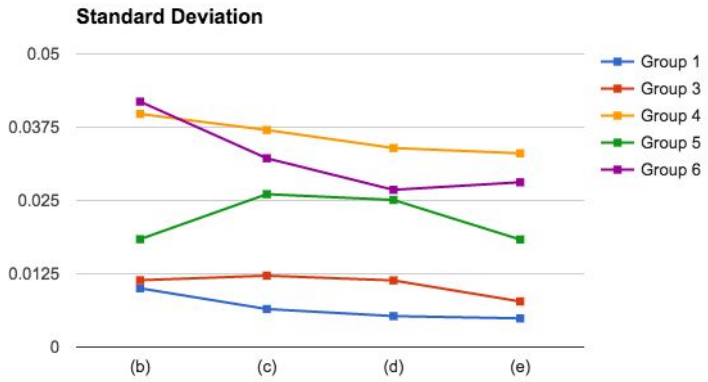


Figure 2.6 – SD computed from different group of images, where (b) is the Umeyama's method, (c) is the Nayar's method, (d) is the first approximate solution and (e) is the proposed global method

For the numerical assessment, we propose to evaluate the specularity removal results using the Standard Deviation (SD, σ) of the histogram distribution. To explain this criterion, we take an object with uniform color (uniform hue) as in Figure 2.5. I_0 is taken with a polarizer positioned at 0^o and I_{90} at 90^o. These images are analyzed in the HSV space, since in this color space, the chromaticity is straightforwardly presented as hue. From Figure 2.5, it can be observed that, with weak specular reflections, the histogram of hue (Figure 2.5 (a)) is more centralized than that with strong specular reflections (Figure 2.5 (b)).

This observation inspired us to use the SD of the histogram on hue to evaluate the result. The smaller SD is, the more successful the specularity is removed. Note that this criterion is applicable only for images where the specular reflection does not cover the majority of the image, and that the texture of the original image is relatively simple.

We show the evaluation on SD in Figure 2.6 with images from Figure 2.3, with the same compared methods on the x-axis. It can be noted that our proposed method also achieve the best results as been observed in the last subsection, and followed by the first approximate solution and the Nayar's method [40]. The Umeyama's method [61] always produces the largest SD, except on group 5, where a very large over whitening effect is presented and leads to its relatively small SD value. However, from both visual and quantitative evaluations, we can conclude that our proposed method still

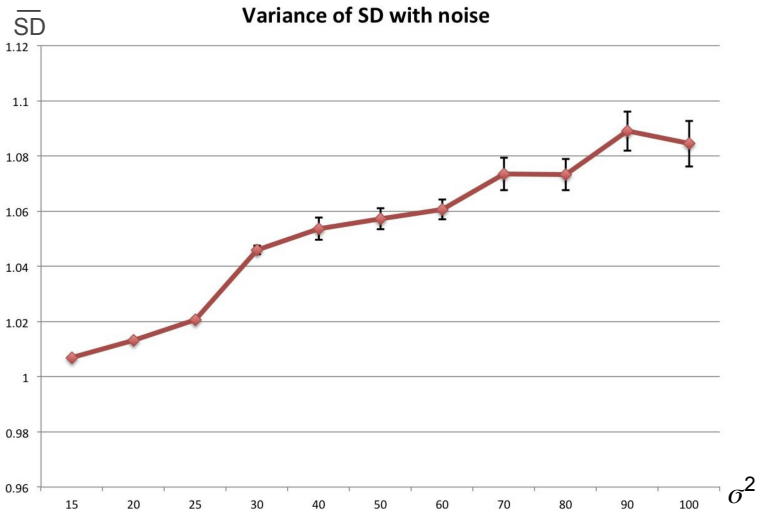


Figure 2.7 – The performance of the SD by adding white Gaussian noise. the x-axis is the variation of the added noise

produces the best specularity removal results.

2.5.4 Robustness analysis

Local-based methods are usually based on the DOP. Since the DOP is computed from at least three images, it is largely contaminated by the noise. This motivates us to analysis the performance of our method with respect to different levels of noise.

White the Gaussian noises, with the zero mean and the same σ is added to I_0, I_{45} and I_{90}. The SD of the result of each group of images are computed correspondingly. Let SD_0 be the SD of the result without noise, $\overline{SD}$ is normalized as $\overline{SD} = SD/SD_0$. It is straightforward that when $\overline{SD}$ is near to 1, the result is not largely influenced by the noise. The mean (red point) and the variance (vertical bound) of $\overline{SD}$ over the groups of Figure 2.3 are computed and shown in Figure 2.7. The x-axis is the $\sigma^2 = \{15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100\}$ of the different noises that we added. For $\sigma^2 < 15$, little variation on SD can be noticed. It can be seen that even by adding a noise with $\sigma^2 = 25$ (which is considerable), the change of the SD still keeps very small (nearly 5%). That is to say, our method keeps stable for noise with $\sigma^2 \leq 25$. The noise with $\sigma^2 > 25$ rarely appears even in real applications thanks to the improved the camera quality. According to the above analysis, we may say that our method is robust

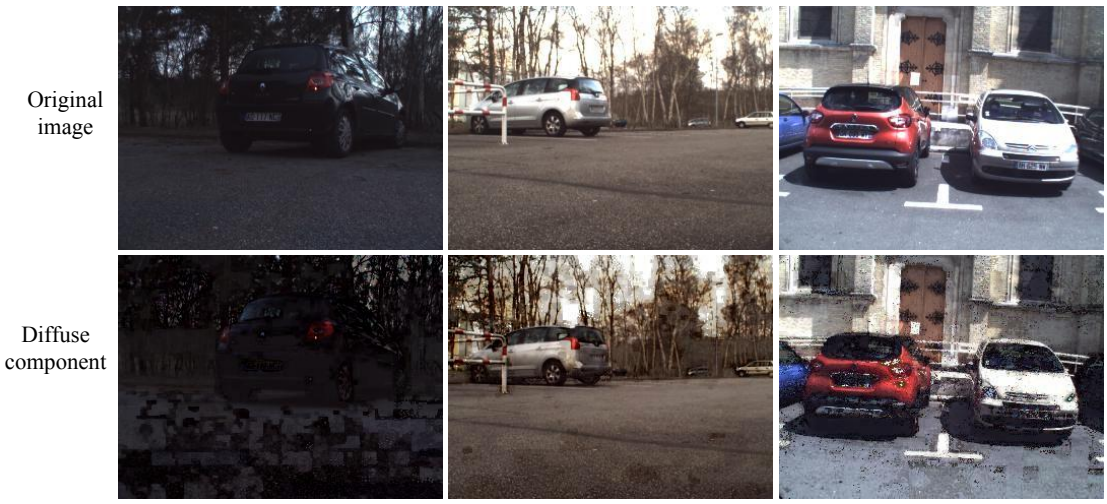


Figure 2.8 – The specularity removal result on outdoor images.

to noise.

2.6 Discussion and future work

The specularity-removal method can be applied as a pre-processing for applications in which the specularity reflection largely influences the results, such as medical image, stereo vision, edge detection etc.

In this chapter, we proposed a polarization-based global energy minimization approach to remove the specular component from images. This method is based on an independency assumption, with constraints given by a first approximate solution. Polarization information is used as a color constraint which largely reduce the color distortion produced by the traditional color-based methods. The robustness analysis also shows that the proposed method is stable for camera noises, which is quite problematic for classical local methods.

As a limitation, although the proposed technique outperforms other state-of-the-art methods, it still have limited performance for outdoor scenes as shown in Figure 2.8. The light from the sky is judged as specular reflection and partially removed by our algorithm, since the skylight are partially polarized under the cloudy days. The estimated diffuse images are too noisy and thus not reliable enough to further apply other computer vision tasks. There is still a long way to go and more efforts to be made

to finally apply the specular and diffuse separation for outdoor applications.

As another limitation, similar to all the specularity removal methods, this method is defined to work with the specular component which varies inside the camera sensor range (0-255). Once either of the color channel goes out of the range, the chromaticity information is permanently lost. In that case, the diffuse component is hardly recovered by the specularity removal method. As the future work, we suggest to detect the image over-saturation using polarization, and then apply the inpainting methods to tackle these regions. These techniques can efficiently recover the lost information through the smoothness assumption over texture, color or other features.

Part II Polarization used as image features

Polarization used as image feature

For unpolarized incident light, the polarization state of its reflection is related to the refraction index, the microfacet and the angle of incidence. More generally, the polarization state is related to the surface material, the surface roughness and the surface geometric structure.

As described in the last chapter, polarization-based specularity removal methods are able to better recover the diffuse reflection of the image than color-based methods. This application can be considered as a pre-processing step to remove the specular highlight and to better apply other image processing algorithms. This process stays, however, still in an indoor application level. More effort should be made to handle the extremely complex illuminating condition for outdoor scenes.

Other than adapting the specularity removal method for outdoor scenarios, it is also advantageous to use polarization information as image features. The polarization state, namely polarization features, provides the hint to the object surface since it is highly related to various surface properties. It can be combined with classical image features, which describe the object shape, like HOG, texton etc, to provide complementary information that traditional color-based image does not possess.

Under outdoor conditions, the light source can not be strictly controlled. The stokes vector is used in this case to describe the polarization state. Other polarization characters, such as the Degree Of polarization (DOP) and the Angle Of Polarization (AOP) are also used.

In this part, we proposed to combine polarization and color features in two different ways for two different applications: semantic segmentation and vehicle detection. In both applications, polarization features provide complementary information that can not be perceived through color-only images.

Semantic segmentation gives a meaningful class label to each pixel in the image. It enables the intelligent device to understand the scene, and has received sufficient attention during recent years. It is an important goal for various artificial intelligent tasks.

Vehicle detection by computer vision system is another research interest. The high velocity of the car make it a highly dangerous obstacle. The car detection has broad applications such as autonomous driving and obstacle avoidance. The output of the vehicle detection can also be used as an input for road scene semantic segmentation.

Based on these two applications, we finally prove that polarization can provide useful features for outdoor road scene analysis.

The following of this part is shown in two topics, which present the proposed semantic segmentation and car detection method correspondingly. In each part, we give the background of the application, the detail of the proposed method, and the experimental results on our self-collected database. In the end of this chapter, we discuss about the advantages and limitations of the two proposed methods.

3 Semantic segmentation

3.1 Introduction

Semantic segmentation, which is also known as scene/image parsing or image understanding, aims to divide an image into predefined meaningful non-overlapped regions (e.g. car, grass, road, etc). As an important task in intelligent vehicle (IV) applications, it's ultimate goal is to equip IV with the ability to understand the surrounding environment. Other IV tasks, such as pedestrian detection, obstacle detection or road surface estimation, could benefit from semantic segmentation.

The substantial development of image classification, object detection, and superpixel segmentation in the past few years have boosted the research in the supervised scene parsing. However, the challenges ranging from feature representation to model design and optimization are still not fully resolved. Up to feature extraction, most methods extract features from RGB or gray level images. Since local low-level features are sensitive to perspective variations, researchers tried to combine some other information with RGB image to give a better performance, such as RGB-D image [22], and geometry information [59] etc. In another aspect, some special illumination cases, such as reflective surface (too bright) or dark shaded surface, would appear to cover real texture or feature information, hence limiting the algorithm performance. Considering this limitation, we adopt polarization image as a new information source to improve the classification result.

Light is polarized once it is reflected from a surface. As has been stated before, the light polarization properties are related to different surface material, surface geometry

structure, the roughness of the surface etc. So that these characteristics are coded implicitly in the light polarization state. In this point of view, polarization attributes can provide description of some surface features that can not be offered by color images. It is worth to know that, these attributes are still kept distinguishable under high reflection or in shadow areas, where the color-image based methods fail to produce reliable results.

In computer vision, there are many indoor polarization applications under ideal lighting conditions since early 1990s, such as surface modeling, shape recovery and reflectance analysis. However, not much outdoor applications have been realized. The reason is that the outdoor incident and reflect light are extremely complex. To the best of our knowledge, no work in the literature has applied polarization in semantic segmentation, this is the first work which attempts to utilize polarization information as features for outdoor image processing applications.

In this chapter, we propose to combine the polarization images (resulted from polarization state of each pixel) with the color images to improve the accuracy of image semantic segmentation. The combination method, more specifically, is through the HOG, LBP and LAB features that are extracted on both the polarization images and the color images independently. These features are concatenated and feed into a joint boosting classifier, a feature selection based classifier known for its facility to integral new sources of features. In the training process, the classifier randomly selects different polarization features and color features from the input space to produce the polarization-based semantic segmentation results. In comparison, we repeat the same algorithm, which extracts the HOG, LBP and LAB features on, however, only color images. After training another joint boosting classifier, the color-based semantic segmentation results are given. The comparison shows that the accuracy of the semantic segmentation is improved thanks to the included polarization features.

3.2 Background

As very classical methods in image parsing, bottom-up semantic segmentation methods usually pursue the following pipelines[73]: 1) Grouping nearby pixels to image patches according to the local homogeneity. For this step, there exists methods like K-means, mean shift, Simple Linear Iterative Clustering(SLIC) [1], normalized-cut [52] etc; 2) Extracting local features, e.g., HOG, LBP, texture or curvature, from each patch;

3) Feeding the extracted features and hand-labeled ground truth to a classification model to produce a compatible score for each training sample; 4) Applying the test samples to the trained classifier. After all these steps, a brute semantic segmentation result is obtained on the test image. To give a more coherent result, state-of-the-art methods commonly perform a global optimization based on Markov Random Field (MRF) or Conditional Random Field (CRF), in which training result acts as the unary term, while the pairwise term is defined over 4 or 8-connected neighborhood.

Numerous methods accomplish the semantic segmentation through these four steps. Superpixel is frequently used as it is more natural and efficient in representation than pixel level features. Because the latter one is ambiguous and sensitive to noise. There are some recent efficient and good performing superpixel generation methods as TurboPixel [32], SLIC superpixels [1] etc. For local features, scale-invariant feature transform (SIFT) feature [34], histogram of gradient (HOG) feature [9], local binary patterns (LBP) [41] and textons [53] are widely used. [53] proposed to use texture layout filters with textons to represent the local texture layout information. Tighe *et al.*[59] tried to combine global and local features to nonparametrically classify over retrieved similar exemplars. Recently, deep convolutional neural network learned features [15] have been applied to replace the hand craft features which achieved promising performances.

Torralba *et al.*[60] proposed an efficient Joint Boosting classifier based on sharing features between different classes. This classifier fits well the semantic segmentation problem, as it enables to classify large scale of different object classes, and to describe many different views of the object. Shotton *et al.*[53] applied the texture layout filter in this classifier, and proposed an optimized search instead of the original best search to accelerate the computation. Costea and Nedvschi [7] used multi-features (HOG, LBP and Color feature) to replace the texture features in the texture layout filter, and applied the original version of Joint Boosting as in [60] at over 50 FPS. In this chapter, we still use the same features as that used in [7]. While, the difference is that the feature extraction procedure is applied in the superpixel level, instead of sampling pixels from image grid. Additionally, we include polarization features into classification in our algorithm. We employ the Joint Boosting classifier used in [53], along with an optimal strategy to reduce threshold searching space.

3.3 Polarization applied on semantic segmentation

In this section, we describe the application of polarization with the texture layout filter [53]. The texture layout filter describes the feature distribution and integrates the spatial context hint together. It is combined with the Joint Boosting classifier which is known by its efficiency in multi-class classification and its facility to integrate new sources of features (e.g., polarization features in our case).

3.3.1 Local feature description

Shotton *et al.*[53] computed texton features at each individual pixel. A 17 dimensional descriptor vector (texton), obtained by 17 individual linear filters applied over the image, was provided. Costea *et al.*[7] combined three descriptor types as they declared that the combined use of multiple descriptor types for boosting can significantly improve the classification accuracy [31]. HOG and LBP features were computed from gray scale images with Gaussian smooth of $\sigma = 0.25$, and also computed from color features on RGB images with Gaussian smooth of $\sigma = 1.0$. The 24-dimensional HOG feature, 24-dimensional LBP feature and 3-dimensional color feature were extracted correspondingly. We denote *HLC feature* as the combination of HOG, LBP and Color features. The same configuration in [7] was applied since these parameters are optimized to fit the real-time application.

3.3.2 Polarization feature

In the image acquisition, the manually system of Figure 1.8 shown in Chapter 1 is used. The angles (θ_i) are chosen to be $[0^o, 45^o, 90^0]$ as recommended by [64]. By turning the polarization filter according to the marked degree, one image is taken with each θ_i. Since polarization images are even more noisy than color images, a Gaussian smoothing is applied, with $\sigma = 1.5$ which is larger than $\sigma = 1.0$ used for color features, before further computation. The DOP and AOP are densely computed for every pixel, producing DOP image and AOP images as shown in Figure 3.1. Regarding DOP and AOP as gray scale images, HOG and LBP features are extracted separately from them, which produces four polar-features: *DOP-HOG, DOP-LBP, AOP-HOG* and *AOP-LBP*. We take each polar-feature to combine with the HLC features separately, so as to make comparison and find the best polar-feature.

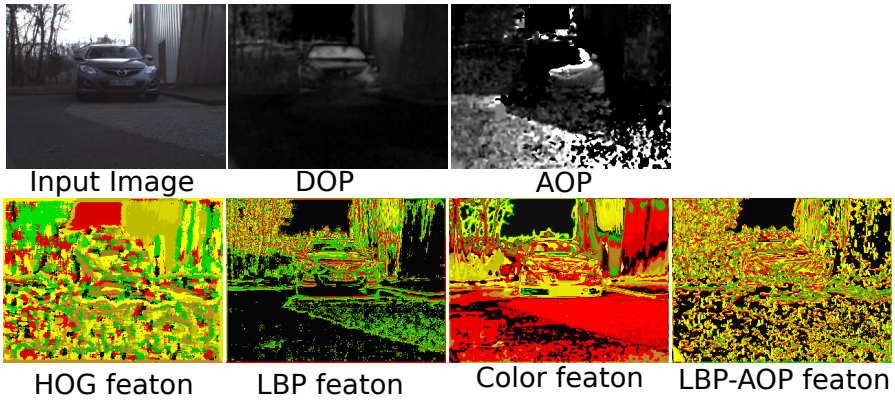


Figure 3.1 – Example of an input image with its DOP and AOP images, and the second row shows its featon image obtained through feature quantization.

3.3.3 Codebook visualization

Feature vector quantization is used to generate a codebook. A large set of descriptor vectors are sampled and then trained using K-means clustering. The resulting centroids are recorded as visual words of the code book. Once we have the codebook, a new feature vector can be matched to the most similar visual words of the codebook. Euclidean distance is used to measure the similarity. By matching densely the local descriptors from one image, a codebook map is then generated. Figure 3.1 illustrates the codebook maps of an image from HLC combined with AOP-LBP feature. This process reduces the high dimensional feature vector to a single discrete value and the post-processing becomes more efficient. The quantized feature vector are invariant to small changes as similar descriptor vectors tend to be attached to the same visual word.

We use 50 visual words for the codebook as in [7], which is less than usual size in the literature that varies from 100 to 4000. Since that we don't have a large data set while we also have multi-feature descriptors, and outdoor scene are less complex both in terms of variety or textures. In every experiment, we have 3 features as HLC features plus one polar feature. we train 50 visual words for every feature, and get 200 visual words in total. After having all the visual words, every pixel in the image is matched with the most similar visual word. With this process, we get four images which are named featon images(similar to the texton map in [53]), with each featon image corresponds to a feature.

3.3.4 Feature-layout-filters

The texton features and the texture-layout-filters are based on the texture features in [53]. Rather than use the texture features, we extract the HOG, LBP and LAB features to keep the algorithm be more noise resistant, and to describe both the shape and texture features. As the texton features and the texture-layout-filters are all applied on HOG, LBP and LAB features, we call them the *featon* and the *feature-layout-filters* to avoid ambiguity, but the formation of the filters is kept the same as in [53]. The main idea of this filtering is that a pixel is classified based on visual word counts in specific regions around the pixel. Each feature-layout filter is a pair (r, f) of an image region r and a featon f ($f = 1...50$).

A set of R which contains $N = 50$ rectangles is randomly generated, such that their top-left and bottom-right corners lie inside a region of 200×200 pixels. These rectangles are defined using the relative coordinate of the pixel, means that giving a rectangle $r_n, n \in N$ and pixel i, the position of the rectangle turns to be $r + i$. As recommended in [53], any configuration could be used which covering over half of the image size. The feature response at the location i is the proportion of pixels under the offset region $r + i$ that have featon index f using (Equation (3.1)).

$$v_{[r,f]}(i) = \frac{1}{area(r)} \sum_{j \in (r+i)} [T_j = f]. \tag{3.1}$$

The featon map is separated into $4 \times 50 = 200$ channels. For each channel, the integral image [8] T_i is used to efficiently compute in R over the whole image. This process will result in a 10000 possible classification features.

3.3.5 Joint-Boosting Classifier

Shotton *et al.*[53] proposed an adapted version of Joint Boost algorithm from [60], which is more efficient in computation and proved to give comparable result as [60]. The algorithm iteratively selects discriminant featon-layout filters as 'weak learner', which is shared between a set of classes C. More specifically, for the m^{th} training turn, for the c^{th} class and on the most dominant feature dimension i, a weak learner $h_i^m(c)$

is learned:

$$h_i^m(c) = \begin{cases} a[v_{[r,t]}(i) > \theta] + b & if c \in C, \\ k^c & otherwise, \end{cases} \tag{3.2}$$

Where $v_{[r,t](i)}$ is the feature computed from Equation (3.1), a, b, θ and k^c are the parameters that given through the training process.

After M turns, a 'strong' classifier is added up by the trained the weak learners as $H(c,i) = \sum_{m=1}^{M} h_i^m(c)$. To reduce the computational cost, exhaustive feature search for each weak learner $h_i^m(c)$ is replaced by random feature selection. Thus, in each turn, the algorithm examines only a randomly chosen fraction $\zeta \ll 1$ of the possible features. It was stated that the randomization not only speeds up learning, but also improves generalization by preventing over fitting to the training data [53].

3.3.6 Efficient application

Concerning the optimization over $\theta \in \Theta$, where Θ is all the possible values of θ, Shotton *et al.*proposed to carefully use the histograms to give the thresholded sums necessary for the post-computation. This process might be more efficient but still involves an over-all computation in Θ. Inferred from this, we apply here another optimization process that appears to reduce the searching space of θ. The discrete set Θ contains values of $v_{[r,f]}(i)$ for all training samples, since they are all possible values of θ. It is separated into 20 bins, each bin is weighted by $\omega_i^c z_i^c$ and ω_i^c separately, and summed up to get the histogram value over these 20 bins denoted by H_1 and H_2. The difference of $|H_1 - H_2|$ is then computed to get the three bins with the most variation. Since this difference represents the amount that label changes between +1 and -1, we search in each decision stump to find a threshold which best separate +1 and -1 labels. And this threshold most probably lies in the bins which occur the strongest variation over labels. As this algorithm serves for real-time IV application, this process will largely reduce the time consuming, with a compromise of slightly decent of accuracy.

In [53], a center pixel subsampling was performed over 3×3 or 5×5 grid to reduce training samples. We also apply a SLIC superpixel feature sampling, in which by having the 10000 possible features, a gravity centroid over the feature space is computed to represent the feature of the superpixel instead of the geometry centroid. This will obtain a better boundary result since superpixel attached well the image boundaries.

Figure 3.2 – Polarization image acquired, the first row was taken at dusk, and the second row was taken at daytime.

Table 3.1 – Over-all classificaiton accuracy using polarization

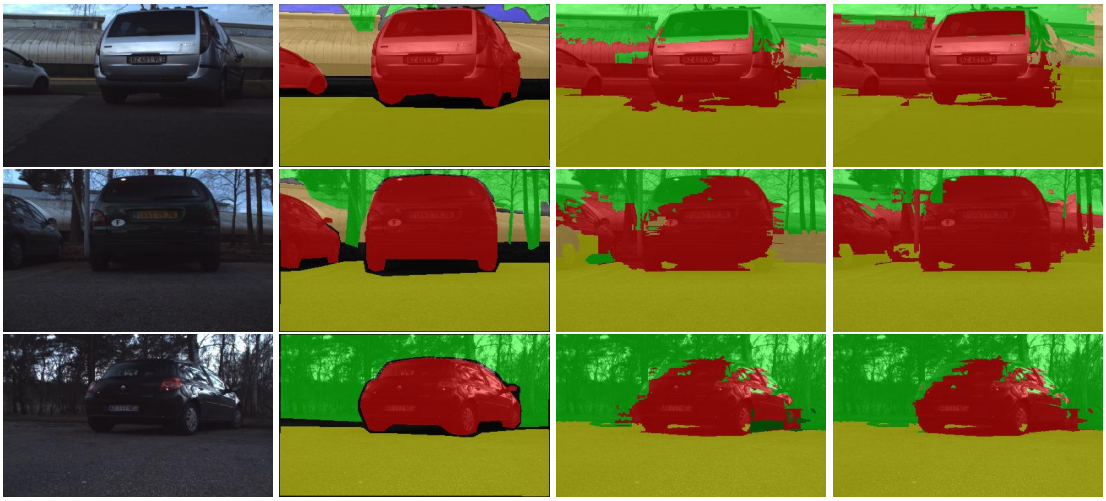
Data	HLC	HOG-DOP	LBP-DOP	HOG-AOP	LBP-AOP
Day-set	11.34	11.27	10.74	11.27	10.97
Dusk-set	13.36	11.35	**10.21**	11.26	10.73

Also, by using a gravity centroid instead of geometry center pixel avoids to select noisy samples which provides more robust features.

3.4 Experiment

3.4.1 Data set

The experiment was applied on our polar-image data sets which contain 21 images at 320 × 240 pixels. The Day-set includes 10 images and the Dusk-set 11 images (an example is shown in Figure 3.2). The Dusk-set uses 6 images for training and 5 images for testing, while the Day-set uses 6 images for training and 4 images for testing. These images were labeled using LableME [45]. We defined 6 classes being car, road, tree, sky, building, grass. Pixels which do not correspond to any of these classes are referred to as void class (see points in black in column (b) of Figure 3.3). After the superpixel segmentation, The label of the superpixel was assigned to the label describing the majority of the pixels it contains. When training the classifier, superpixels which



(a) Input image (b) Ground truth (c) HLC results (d) HLC+LBP-DOP results

Figure 3.3 – Semantic segmentation result over the Dusk-set. Each row corresponds to an image scene. The column (a) shows original color images with its hand labeled ground truth on column (b). Column (c) shows the classification result using HLC and column (d) shows the HLC combined with LBP-DOP feature

correspond to a void class was taken off from the training data. Up to error evaluation, testing samples with void class were also disregarded.

3.4.2 Comparison Day-set and Dusk-set

For each image, 200 SLIC superpixels were extracted with compactness of 20. For the classification, we applied a sampling rate of $\zeta = 0.01$ using 500 boosting rounds. Other methods as [7, 53] ran till 5000 boosting rounds using sampling rate $\zeta = 0.003$. To our case, we applied only 500 boosting rounds, since the classification error converges in 300-400 boosting rounds since our data set is relatively small. We used a sampling rate $\zeta = 0.01$ to examine more feature every turn to make sure that after 500 turns, every feature has got enough comparison.

The over-all error accuracy is shown in Table 3.1. Remaind that *HOG-DOP* stands for combination of HOG-DOP feature with HLC feature, similarly in *LBP-DOP*, *HOG-AOP* and *LBP-AOP*.

Using only the HLC features, Day-set occurs smaller error rate (11.34%) than Dusk-set (13.36%). Using the same algorithm, even Dusk-set got more images in training, its

performance still fall behind the result in Day-set. This is probably due to the poor illumination conditions for the Dusk-set. Paying attention to the performance of polarization, it is not surprise to found that by adding polarization feature, result over Day-set doesn't vary a lot with an improvement of 0.6%, while the Dusk-set polarization features improves the result up to 3.14%. At Day-time, lights acquired by polarization imaging system are of a reflection chaos, since they are from various objects as second or even third hand reflection. While in the Dusk-set, when the illumination goes done and complex reflection appears to be less noticeable, the acquired lights' polarization properties become more coherent and less noisy. Thus, for the Dusk-set, polarization features give good performances. This shows than polarization may potentially provide a solution for optimizing the scene understanding ability for intelligent devices under poor illumination. This should be demonstrated with a larger data set and faster implementation.

From Table 3.1, for the Dusk-set every polarization feature does improve the classification result. It is noticed that from both of the data sets, LBP based polarization features give better performances than HOG-based performances. The best polarization feature is found with LBP-DOP feature.

The comparison between HLC feature and LBP-DOP combined feature is shown in Figure 3.3. We can observe, in the result of HLC feature on the first and second scenes, the miss-classification of the back window of the car. In the first scene, it might be influenced by a light reflection while in the second scene it could be the transparency of the window. These problems are all better resolved using polarization features. In the third image scene, for the HLC result, a small part of the car which has very dark intensity is miss-classified as grass, while in the polarization result, this problem is addressed, and also better contours are got through polarization.

When observing the building part in the first and second scenes, it is found that the building is merged by the car class. The reason could be that the building in the image is covered with some slightly reflecting materials, thus its polarization features would go similar with metal surface. We can also observe from the second and third scenes, over the right part of the car, that polarization features do not contribute to improve the result. As these two surfaces are of sharp angles with image plane, we supposed that polarization feature works better to the surfaces within small angles between the image plane, however, further experiment need to be done.

3.5 Discussion

In this chapter, we proposed a method to apply polarization image on semantic segmentation. The HOG, LBP and LAB features are extracted from polarization images, being DOP and AOP. These features are concatenated with the color-based features as the input of the joint boosting classifier. This classifier is used since it adapts well to combine different features, since it is principally a feature-selection based classifier. In this way, the polarization-based feature is automatically fused with the color-based features.

In the experimentation, the comparison between the polarization-based method and the color-only method is carried out, as well as the comparison of the results between the day-set and the dusk-set. The experiments shows that the polarization improves the result of the scene understanding, especially for the dusk set. This implies that polarization might be a potential solution for intelligent devices under poor illumination.

As the result of this chapter is given on a small data set, it is regarded as a first attempt to apply polarization onto semantic segmentation. The method should be applied on a larger data set, which also needs faster implementation (such as implementation on GPU) to produce more general and efficient results.

4 Car detection

4.1 Introduction

For an intelligent device, under the context of the road scene, car appears to be one of the most frequently observed yet dangerous object. Car detection has a broad of applications such as autonomous driving or obstacle avoidance. It is a challenging problem due to the large structural and appearance variations. Ubiquitous occlusions further increase intra-class variations.

Dalal et al. [9] proposed a single filter based on Histogram of Oriented Gradients (HOG) features to represent an object. The model is a single filter that slides throughout the image. A score is computed with each position and scale. Based on this effective feature, Felzenswalb et al. [16] proposed Deformable Part Models (DPM). The object is modeled by several deformable parts to better handle the object appearance variation, and to integrate the intra-class variations. Specifically for car detection, Wu *et al.*[67] proposed a reconfigurable hierarchical and-or model to integrate the context and the occlusion patterns, in which the DPM is utilized as the deformable feature. The view point (e.g., frontal, rear, left side or right side) of the car is estimated as a by product of this method.

Classical color-based methods, including those mentioned above, extract shape like features from the *rgb*-information of the object. As these features are formulated to be rotation and scale invariant, they produce acceptable and stable detection results. However, it should be noticed that the *rgb*-information does not fully describe an object, since some other pieces of information are not available, such as the geometry

structure of the object, the roughness of the surface, the material of the surface etc. These pieces of information are not perceptible by conventional imaging devices, while they are conveniently provided by the polarization imaging systems.

As mentioned in the last chapter, the polarization state of the reflected light is considered as a physical feature which is highly related to the local patch of the image. Polarization features of road scenes are too much noisy, since their computation involves the acquisition of at least three images under complex outdoor illumination. A fusion scheme with *rgb*-based features could make such features more stable and produce improved and comparable results.

In this chapter, we propose to use polarization features as a complementary information to improve the color-based vehicle detection results. To our knowledge, this is the first work in the literature that attempts to use polarization-based features to outdoor object detection.

Firstly, a feature selection is performed to select the most informative one among the five studied polarization features. This process indicates that the Angle Of Polarization (AOP) is found to be the most informative one. An AOP-based DPM detector (polar-based model) and a color-based DPM detector (color-based model) are then trained independently. With the different score maps produced by the two models, a fusion rule is proposed which takes the polar-based model as a confirmation of the color-based one to produce the final detection bounding boxes. As another contribution, we published a road scene polarization data set, which is the first public available polarization data set to our knowledge. The experiment is performed on this data set, and proved that taking the complementary information provided by the polarization feature, the false alarm (false bounding box) is largely reduced, and the detection accuracy is improved.

4.2 Feature extraction

In this section, we firstly introduce the selection process of the polarization features, where the AOP is found to be the most informative feature. The formulation of the HOG-based feature is introduced as proposed by [16]. This feature is extracted on both AOP and color images and is used to train the AOP-based model and the color-based model independently. The final result is obtained via fusing the results provided by

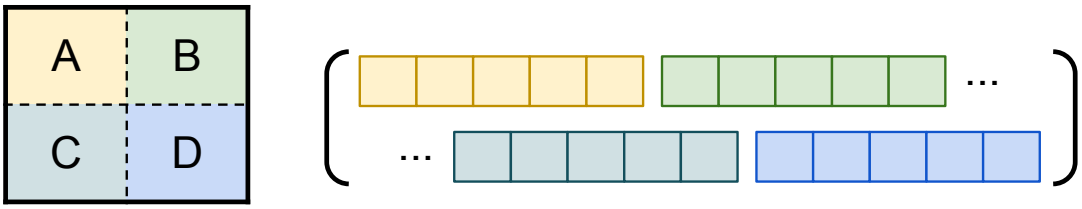


Figure 4.1 – The formulation of the polarization feature for the selection.

the two models as presented in section 4.5.

4.2.1 Polarization feature selection

Basically, the polarization state of a beam light is described by its Stokes vector. The DOP and the AOP are extracted from this physical Vector. Because of the noisy nature of the polarization parameters, a feature selection is required to find the most relevant polar-based feature to train the detection model. This feature selection procedure is a simple but effect, that we roughly train by the Dalal-Trigg detector [9] by replacing the HOG feature by the polarization features.

To train the Dalal-Trigg detector, similar to the original HOG features, the polarization features are extracted based on blocks and cells. More specifically, a 8×8 block is divided into four 4×4 cells, namely A, B, C, D as shown in Figure 4.1. For each pixel in the cell, a feature vector is extracted, which contains the 3-dimensional Stokes vector, the DOP and the AOP, and is described as $[s_0, s_1, s_2, \rho, \varphi]$. The feature vector of a cell is represented by the mean feature vector of all the pixels inside the cell. Each cell holds a 5-dimensional feature vector. The feature vector of the block is then a concatenation of the feature vectors from the four cells as shown in Figure 4.1.

This 20−dimensional features ϕ_{20-d} of all the examples are used to train the Dalal-trigg detector $f \cdot \phi_{20-d}$ to get the filter f, which indicates the weight of each feature dimension. Larger weight means the better relevancy of the corresponding feature. We take the feature which corresponds to the largest weight for further feature extraction.

By applying the feature selection process presented above, the AOP is finally selected to be the most informative feature with respect to the car detection purpose.

Original image | Angle of Polarization

Figure 4.2 – The original image and its corresponding AOP.

4.2.2 Angle of polarization

The AOP refers to the direction of the polarization of the reflected light. It is determined by the angle of the incident light (generally for outdoor applications, the incident light is assumed to be unpolarized), the surface orientation of the object and the material of the object. For rough surface, as the surface orientations of neighboring pixels change a lot, the AOP changes between neighboring pixels in a irregularly way. For smooth surface, however, the AOP changes smoothly and continuously. The AOP can be extracted densely from each pixel of the image, thus forming an image called the AOP image.

Up to road scenes, especially for car detection tasks, the AOP on the car surfaces changes gradually according to the surface geometry structure, while it is noisy for other objects. As shown in Figure 4.2, the AOP image on the tree area is highly noisy, and on the road it is better but still much noisy than on the car. It can be observed that the AOP image describes the geometry structure of the car, which is even more clear than that from the color image.

We can conclude that, the AOP describes the geometry structure that cannot be captured from the color image. In addition, it highlights the area of the car among the noise. For this reason, the fusion between the result of the AOP and those of the color images eliminates the false detection and improve the detection accuracy.

4.2.3 HOG-based feature

Felzenszwal *et al.*[16] proposed a fast Analytic Dimensionally Reduction (ADR) on the HOG feature [9] which performs faster than both the HOG feature and the Principal Component Analysis (PCA) (on 36-dimensional HOG feature). The ADR is based on an observation of the PCA on a large number of images (PASCAL VOC 2007 [14]). From the eigenvectors results from PCA (Figure 4.3),it can be observed that only the first 11 eigenvectors have considerable eigenvalues. Standardly, to reduce the feature dimension, each feature vector should be projected to get a 11-dimensional feature vector. However, as can be observed from Figure 4.3, the structure of the first 11 eigenvectors is quite uniform : either their columns or their rows are equals. The author is then inspired to project the HOG feature to its 9 column values and 4 row values. This projection process is obviously more time efficient than the PCA, simultaneously, it reduces the feature vector from 36-dimensional to 13-dimensional. More specifically, the projection process is formed as follows: Let $S = \{s_1, \ldots, s_9\} \cup T = \{t_1, \ldots, t_4\}$, where

$$s_k(i,j) = \begin{cases} 1 \;\; if \;\; j = k \\ 0 \;\; otherwise \end{cases} \quad , \quad t_k(i,j) = \begin{cases} 1 \;\; if \;\; i = k \\ 0 \;\; otherwise \end{cases} \quad , \tag{4.1}$$

each element of the ADR is computed by taking the dot product of the 36-dimentional HOG feature with each s_k and t_k, which finally forms a 13-dimensional ADR feature vector. This process is much faster than the PCA and the projection on eigenvectors and gives competitive results.

The author repeated this process on a contrast-sensitive version 72-dimensional HOG feature [9] by applying only the 18-dimensional version s_k as in Equation (4.1). The final ADR feature is a 31-dimensional feature vector which is the concatenation of the 13-dimensional feature from the contrast-insensitive HOG feature and 18-dimensional feature from the contrast-sensitive HOG feature.

With the process described above, the HOG-based feature is extracted on both the color image and the AOP image. They are used independently to train a color-detector and a polar-detector. The final result is given by fusing the result given by the two models.

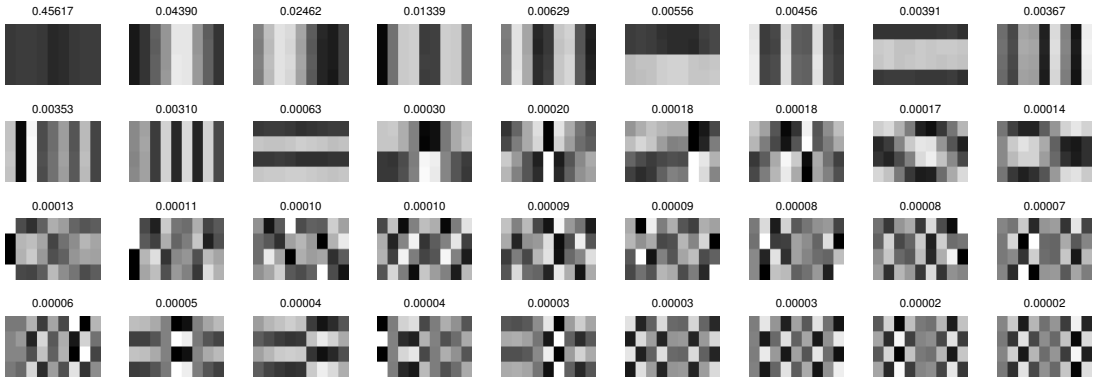


Figure 4.3 – PCA on a large number of HOG features. We have 36 eigenvectors with their corresponding eigenvalues written on the top. [image courtesy from [16]]

4.3 Deformable Part Models

In this section, we describe the DPM model proposed by [16]. Similar to Dalal-Triggs detector [9], DPM is a filter-based detector applied throughout a feature map to get a detection score. A feature map is an array that each entry is a $d-$dimensional feature vector computed from the corresponding image location. Let Φ be the feature map, with $\phi(x, y)$ corresponds to the entry at the location (x, y). The filter f shares the same dimension as the feature map, and the detection score is computed as the dot product $f \cdot \phi(x, y)$. The detection score for a given location is proportional to the possibility of the existence of an object on that location.

The DPM detector M is a mixture of n independent models, being $M = \{m_1, m_2, \ldots, m_n\}$. Each model contains a root filter f_r, and s part filters such as $m_i = \{f_r, p_1, p_2, \ldots, p_s\}$, where $1 \leq i \leq n$. We show in Figure 4.4 an example of the structure of a DPM detector M, who contains three mixture models, each one constructed by a root filter and six part filters.

4.3.1 Mixture of models

The intra-class variation is a common challenge for object detection algorithms. For example, in the car detection, the frontal, rear, left or right views of the car are so different and one model can hardly cover all of them. The mixture of models is constructed so that each model refers to a particular view, thus making the detector much more tolerant to the intra-class variations. The number of models n is a user

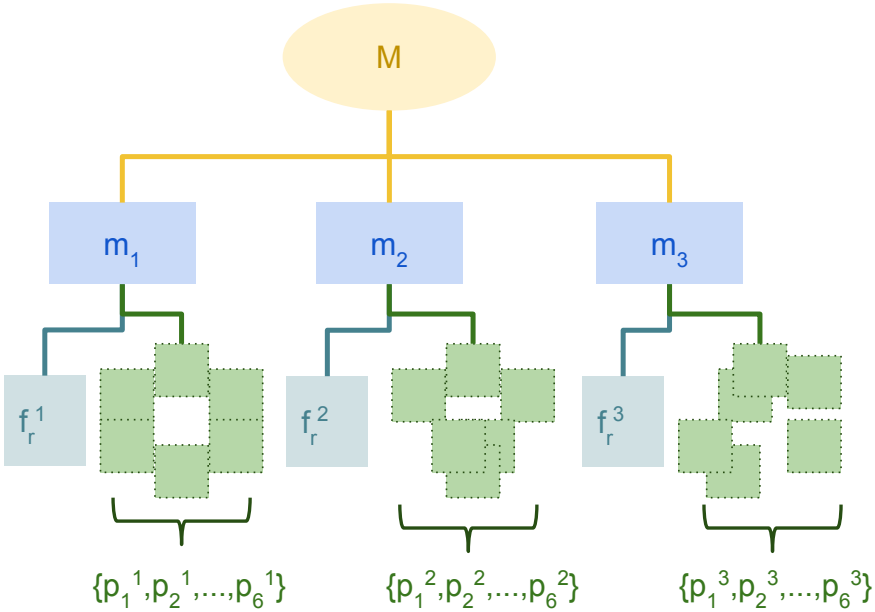


Figure 4.4 – An example of the structure of a DPM detector with three mixture models, and each model contains a root filter and six part filters.

input parameter. For car detection, we obviously use $n = 4$ since the view of the car roughly divided to four views.

For the mixture of models, during the training process, each positive example is clustered so that it feeds only one of the mixture models. For general object detection, the aspect ratio $r = w/h$ is used for the clustering, with w and h be respectively the width and height of the bounding box. During the detection, an object hypothesis is applied on each model to get a score, and the maximum score is used as the final one.

4.3.2 Root filter and part filters

Here we describe only one of the model in the mixture of models since every model shares the same structure. Let m be one model among the mixture of models. It is defined by a root filter that approximately covers an entire object and part filters that cover smaller parts of the object. Figure 4.5 gives an illustration of such a model in a feature pyramid. The root filter location defines a detection window. The part filters are placed several levels down in the pyramid so that it is the twice resolution of the features in the root filter level. It should be noted that using higher resolution features

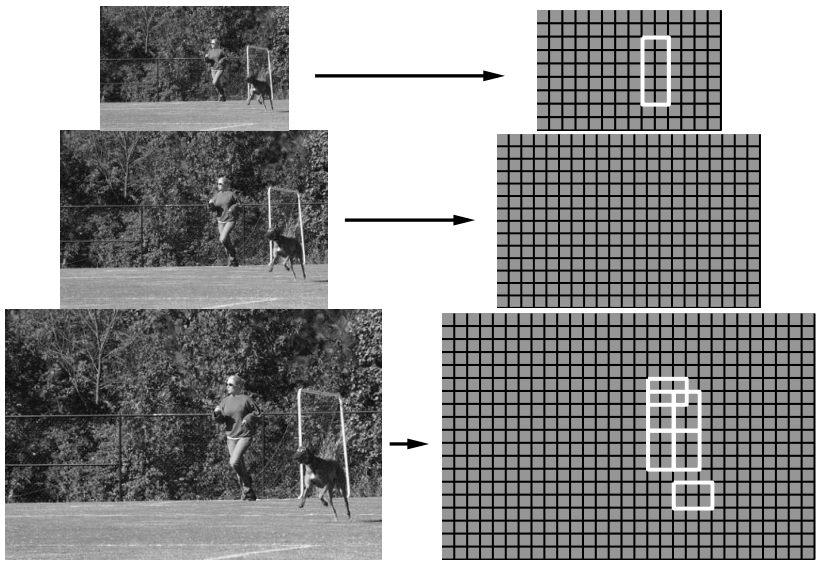


Figure 4.5 – An instantiation of a person model within a feature pyramid. The part filters are placed at twice the spatial resolution of the level of the root filter. [image courtesy from [16]]

for defining part filters is essential to perform a good detection. The reason is that the root filter is able to capture coarse resolution edges while the part filters can capture details.

The model m with s parts is formally defined by $(s+1)$ elements as $(f_r, p_1, \ldots, p_s)$, with a bias term b used to balance the score of each model. Each part filter p_i is defined by three elements (f_i, v_i, d_i), where f_i is the filter for the i-th part, v_i is a two dimensional vector that defines the anchor position of p_i with respect to the root position, and d_i is a four dimensional vector defining a deformation cost for each possible placement of the part with respect to the anchor position.

An object hypothesis should specify the location z_i of each filter in the feature pyramid, where $z_r = (x_r, y_r, l_r)$ defines the location of the root filter, and $z_i = (x_i, y_i, l_i)$ is the location of i-th part filters defined with respect to z_r. The score of this hypothesis is computed as the score of each filter at the defined location, minus a deformation cost for each part filter. The deformation cost for each p_i is a quadratic function which measures the distance between the learned anchor position v_i and the defined location z_i plus a bias. In this way, the model softly constrains the part filter to move

in a short distance with respect to the learned anchor position v_i. Note that the position of the part filter is always defined relatively to the root filter position. The bias is a parameter that balance the score from each model. Finally, let x be an object hypothesis, the score is computed as:

$$score(f_r, p_1, \ldots, p_n) = \sum_{i=0}^{n} f_i \cdot \phi(x, z_i) - \sum_{i=1}^{n} d_i \varphi(dx_i, dy_i) + b, \tag{4.2}$$

where

$$(dx_i, dy_i) = (x_i, y_i) - (2(x_r, y_r) + v_i) \tag{4.3}$$

gives the displacement of the i-th part relative to the anchor position v_i in the twice resolution of the root filter, and

$$\varphi_d(dx, dy) = (dx, dy, dx^2, dy^2) \tag{4.4}$$

describes the deformation (distance) as a feature vector. Term d_i in the part filter specifies the coefficient of this distance which is figured out in the learning process.

4.4 Training and prediction

In the last section, we described the structure of the DPM detector and how the score is computed based on a learned model. In this section, the classifier and the learning process proposed by [16] will be described. More detail about this method is described in appendix A.

4.4.1 Latent SVM

Consider a classifier that scores an example x with a function of the form

$$c_f(x) = \max_{z \in Z(x)} f \cdot \Phi(x, z). \tag{4.5}$$

Here f is the model filter, z is the latent values which define the optimum positions of the part filters and $Z(x)$ defines all possible values of z for an example x. For computation, f is the concatenation of the root filter and part filters which forms a row vector, and $\Phi(x, z)$ is also the concatenation of the feature vectors extracted ϕ_i at

the locations of the root and part filters defined by z correspondingly.

Since the latent variable should be figured out in the training process, the classical SVM does not straightforwardly solve this problem. Felzenszwalb *et al.*[16] proposed a Latent SVM (LSVM) classifier to train the DPM detector. Similar to classical SVM, f is trained from labeled examples $D = (< x_1, y_1 >, \ldots, < x_n, y_n >)$, where $y_i \in \{-1, 1\}$, by minimizing the objective function:

$$L_D(f) = \frac{1}{2}||f||^2 + \lambda \sum_{i=1}^{n} max(0, 1 - y_i c_f(x_i)), \tag{4.6}$$

where $max(0, 1 - y_i c_f(x_i))$ is the standard hinge loss and the scalar λ controls the relative weight of the regularization term.

4.4.2 Two-stage learning for LSVM

Let Z_{pos} be a latent value for each positive example in the training set D. An auxiliary objective function $L_D(f, Z_{pos}) = L_{Z_{pos}}(f)$ is derived by restricting the latent value to Z_{pos}. Under this case, we can minimize $L_{Z_{pos}}(f)$ as

$$L_{Z_{pos}}(f) = \min_{Z_{pos}} L_D(f, Z_{pos}). \tag{4.7}$$

Holding this idea, the function $L_D(f, Z_{pos})$ is minimized iteratively in two stages as:
1) Relabel positive examples: optimize $L_D(f, Z_{pos})$ over Z_{pos} by selecting the highest scoring latent value Z_i for each positive example,

$$Z_i = argmax_{Z \in Z(x_i)} f \cdot \Phi(x_i, Z).$$

2) Optimize filters: optimize $L_D(f, Z_{pos})$ over f by solving the convex optimization problem defined by $L_{Z_{pos}}(f)$ using a stochastic gradient descent algorithm [16], which is detailed in appendix A.

Both steps always minimize or maintain the value of $L_D(f, Z_{pos})$ which converges to a strong local optimum. That is to say, a careful initialization of f is necessary to train a good model. This will be introduced later with the model initialization.

4.4.3 Data-mining hard examples

As a common problem of object detection, one usually have a very large number of negative examples (10^5 negative examples can be generated from on single image). It is unfeasible to consider all negative examples simultaneously. A standard solution is to gather only part of negative examples as *hard negative* through a data-mining process.

The hard instances $H(f,D)$ and easy instances $E(f,D)$are defined relatively on f as follows:

$$\begin{aligned} H(f,D) &= \{\langle x,y\rangle \in D | y \cdot c_f(x) < 1\}, \\ E(f,D) &= \{\langle x,y\rangle \in D | y \cdot c_f(x) > 1\}. \end{aligned} \tag{4.8}$$

More specifically, $H(f,D)$ are the examples in D that are miss-classified or inside the margin of the classifier f, and $E(f,D)$ are the examples in D that are correctly classified and outside the margin.

At each iteration of the two-stage learning process, new negative examples (at least one) are pushed into the training cache. The cache still stores the hard example from the last iterations, and easy examples are removed. By following this data-mining process, the system requirement and computation time are largely reduced.

4.4.4 Model initialization

The two-stage learning algorithm for LSVM may possibly converge to the local minima and thus sensitive to initialization. The model is initialized through two steps described below.

Step 1. Initializing Root filters: As for a mixture of n models, positive examples are splitted to n groups so that each model uses only one of the groups. The split of the examples can be get from one input from the data annotation such as the view of the object. Otherwise the aspect ratio of the bounding box is sorted and used as a simple indicator of the object view. The root filter f_r for each model is trained by a standard SVM without latent information as in [9]. These models with only root filters are merged into a mixture model and retrained on the full data sets.

Step 2. Initializing Part Filters: The number of part filters is a user input parameter. In

[16], a 6-part filters configuration is used. The initial location of the part filter is set by greedily placing parts to cover high-energy regions of the root filter. Once a part is placed, energy of the covered portion of the root filter is set to zero, and then we look for the next highest-energy region, until all parts are chosen. Remind that the part filter is initialized in the level with twice the resolution of the root filter.

4.4.5 Training procedure

Algorithm 1 Training procedure

Data: Positive examples $\{(I_1, B_1), \ldots, (I_n, B_n)\}$
Negative images N
Initial model f
Result: New model f

$Cache_n := \phi$ % empty negative cache
for $relable := 1$ **to** $num - relabel$ **do**
    $Cache_p := \phi$ % empty positive cache
    % step 1, maximize score to get Z_{pos}
    for $i := 1$ **to** n **do**
        Add highest scoring Z_{pos} for each(I_i, B_i) to Positive Cache $Cache_p$
    end for
    % step 2, optimize f as well as data-mining hard negatives
    for $datamin := 1$ **to** $num - datamine$ **do**
        for $j := 1$ **to** m **do**
            if $|Cache_n| > memory - limit$ **then** break
                Add new negative examples to $Cache_n$
            end if
            Optimize f through `gradient-descent` on $(Cache_p \cup Cache_n)$
            Remove easy examples
        end for
    end for
end for

For object detection, the training examples are given by positive bounding boxes B defined on an image I, being a set of pairs (I, B), and a set of Background images N. The training procedure of the DPM detector can be described by the pseudo code shown in Algorithm 1. As described in the last subsections, this process mainly goes through the initialization, searching Z_{pos} by maximizing the score, optimizing f for the given Z_{pos} while selecting the hard negative examples by data mining until the model converges.

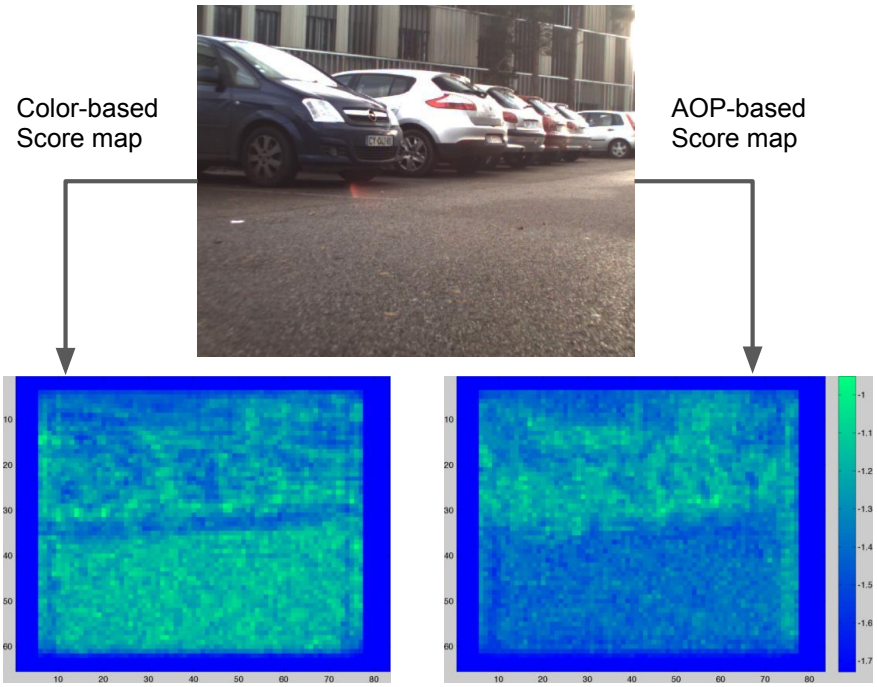


Figure 4.6 – The score map given by color-based and AOP-based methods.

4.5 Fusion rule

As polarization provides complementary information that is not accessible by color images, the fusion of polarization and color result improves the result provided by color alone. In this section, we propose a fusion rule to improve the detection result by taking the polarization result as a confirmation for the color one.

Following the training process, we obtain a polar-model M_p using AOP image and a color-model M_c using color image. These models are also called the DPM detectors, which detect a car in an image by using the score of the model. Since each location is scored by the model, a score map is formulated for each image. The higher score a location gets, the more possible that a car appears at this location. A result without fusion is given by simply thresholding the score map by a trained threshold.

The score map given by M_c and M_p are shown with the original image in Figure 4.6. It can be seen that on the color-based score map, both the car area and the road area get high scores, which consequently generates false positive. Distinctively, on the score map given by AOP image, only the car area gets high scores. This demonstrates the

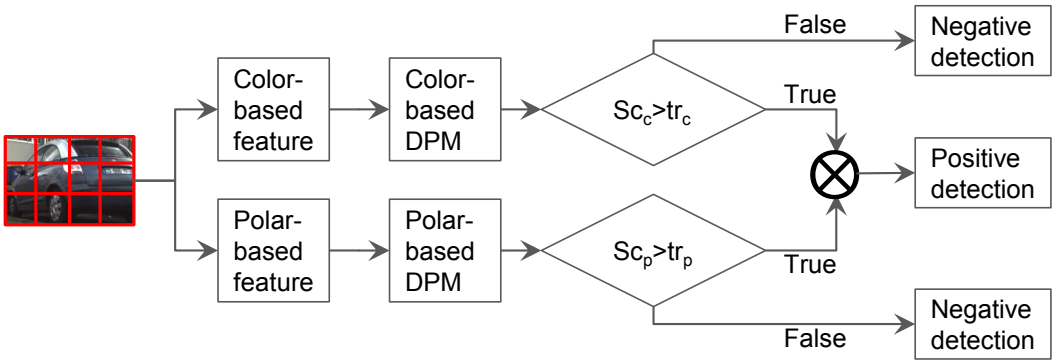


Figure 4.7 – The fusion rule.

idea that we can reduce the false detection by taking AOP results, too. However, the AOP image based model M_p alone does not provides satisfying result. The reason is that AOP image describe the geometry structure of the car in a very rough and noisy way, this information alone is not strong enough to fully describe a car.

This consideration on AOP inspired us to propose a fusion rule which uses the polar-based model M_p as a confirmation of the color-based model M_c. This rule is described in Figure 4.7, Sc_c and Sc_p are the score maps given by M_p and M_c respectively, and tr_c and tr_p are the trained thresholds. The idea is that, the score location, on which M_c produces high scores, are then confirmed by the score of M_p. If both M_c and M_p produces high score, this location is then considered as an object, and vice versa.

The use of M_p as a confirmation serve to exclude the false detection provided by the color-only method, thus improving the detection accuracy.

4.6 Experimentation

4.6.1 Data set and configuration

Since there does not yet exist a public polarization-based data set, our experiment is implemented on a self-collected data set [1]. It contains 153 scenes, with three images I_0, I_{45} and I_{90} for each scene taken by the corresponding polarizer angles. All the cars in the image are labeled by bounding boxes. We divided the data set so that there are 115 scenes are used for training and 38 scenes used for model evaluation. The

[1] The database can be found at http://pagesperso.litislab.fr/fwang/fichiers/

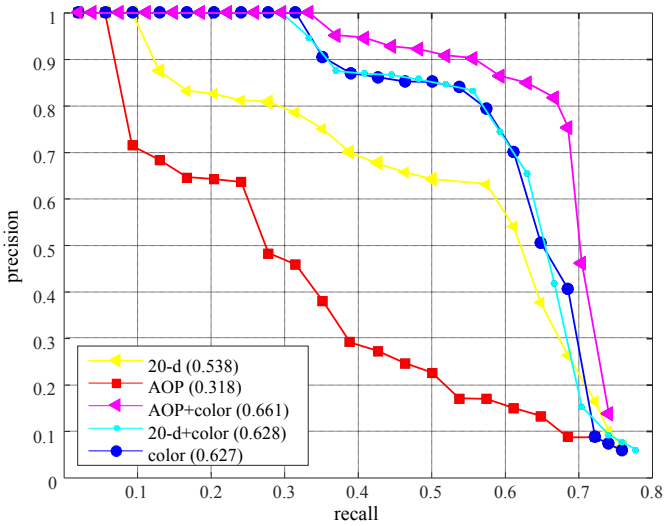


Figure 4.8 – Precision recall curve on the test set

parameter λ in Equation (4.6) is set to be 0.002 as proposed by [16].

4.6.2 Results

As a standard evaluation, the detected bounding boxes on the test set is assessed by the evaluation scheme provided by the PASCAL VOC data set [14].

The Precision-Recall Curve (PRC) is firstly computed and shown in Figure 4.8. The PRC of the AOP-based model, the color-based model and the fusion of AOP with color are shown in the figure. Moreover, we train a model with the 20-dimensional polarization feature (the feature described in 4.2.1) and also fuse this model with the color based model. The results are also shown in Figure 4.8 noted by $20-d$ and $20-d+polar$ correspondingly. The Average Precision (AP) of each curve of Figure 4.8 is shown in Table 4.1.

It can be observed from both Figure 4.8 and Table 4.1 that the fusion of AOP with $color$ produces the best result. Following by the fusion of $20-d+color$ which performs only slightly better than the color only one. The $20-d$ or the AOP alone does not produce comparable results.

This result confirms that AOP alone is not stable enough to produce good results. However, as the most informative polarization feature, once properly fused with the

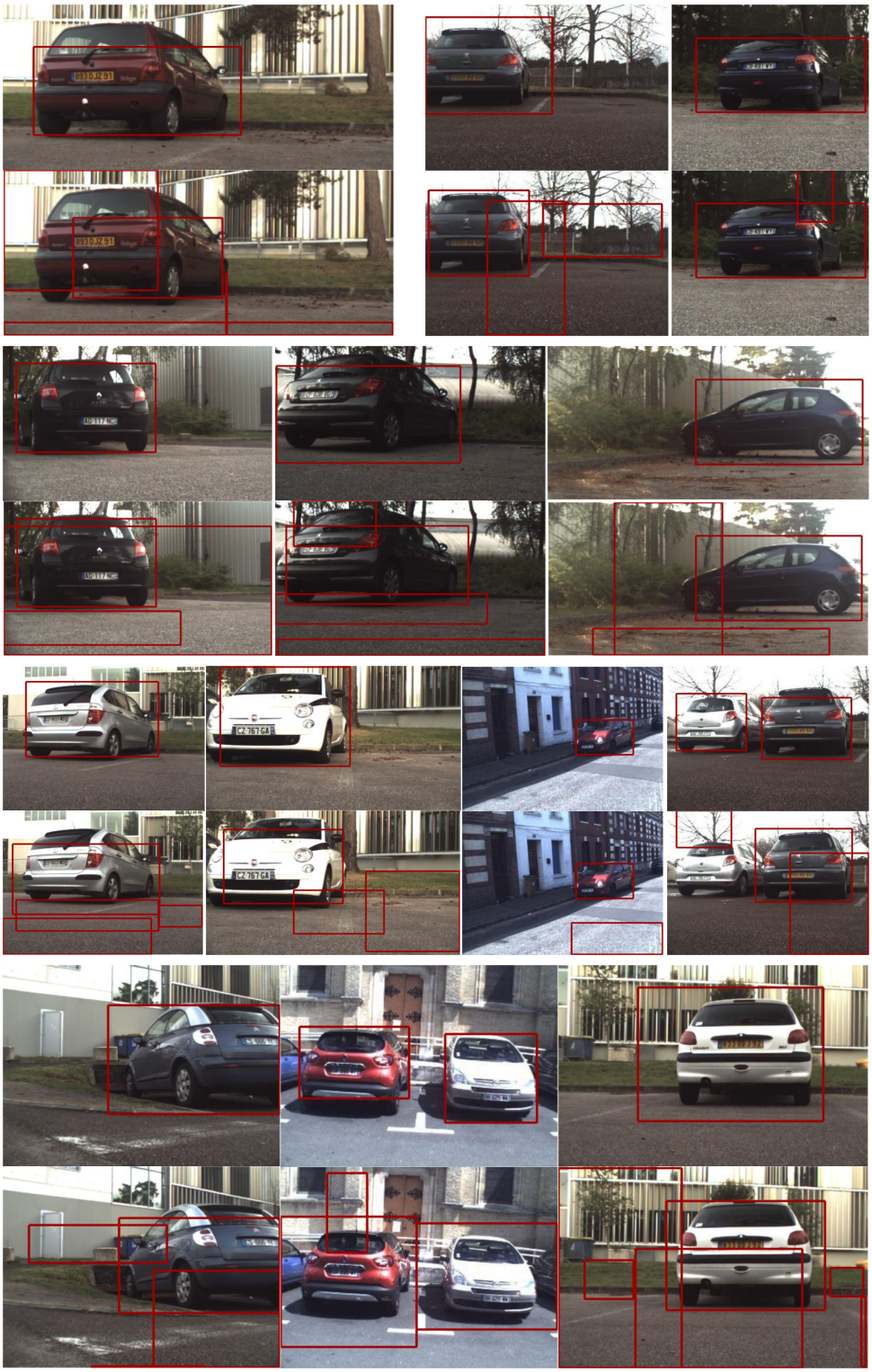


Figure 4.9 – Detection results. In each example, the first row refers to the polar-based model and the second row is the color-only one.

Table 4.1 – The Average Precision of corresponding methods of Figure 4.8

source	AOP	$20-d$	$20-d+color$	color	AOP+color
AP (%)	31.8	53.8	62.8	62.7	**66.1**

color model, it provides complementary information which improves the AP by 3.4%. The $20-d$ performs more stable than the AOP, however, with all the redundant features and the more complex model, it almost does not improve the color-based result (0.1% can be even neglected). According to the above analysis, it can be concluded that our proposed pre-selection method is valid. By using the selected feature (AOP) and the proposed fusion rule, the polarization feature provides useful information which improves the color-based method.

To evaluate the improvement provided by the fusion of $AOP+color$, we compare the detected bounding box by $AOP+color$ method and the color-only method. This comparison is shown in Figure 4.9, where for each example, the $AOP+color$ results is shown on the first row and the color-only results on the second row. It is worth to note that, after fusing the two different sources of information via the proposed method, while keeping the true positive bounding boxes, the false detection are effectively removed.

These results demonstrates that by properly choosing the polarization feature and fuse it with the color model, the detection result is improved.

Figure 4.10 – Main limitations

4.7 Discussion, limitation and perspectives

In this chapter, we proposed to use the polarization features to improve the car detection accuracy. A polarization feature selection is applied to look for the most informative polarization attribute and to keep the simplicity of the model. The AOP feature (AOP image) was found to be the most informative feature, and was used to train a DPM detector. We then proposed a fusion rule to fuse this detector with another DPM detector that trained using color-only images. The final detection results were evaluated by the same evaluation scheme used in [14]. The detection results validated the proposed method, and shown that our method largely reduced the false detection rate and enhance the robustness of the model.

The proposed method demonstrates that polarization features can provide useful information for car detection. The false alarm is largely reduced by a simple but effective fusion rule. However, as can be seen in Figure 4.10, the limitation is that the polarization-based feature is not formed to generate new true positive detection boxes. The reason why we did not take new detection from the polar-based model in the fusion rule, is that this model is not robust enough, and that the new detection boxes it provides are usually incorrect.

To deal with this limitation, as a future work, the polarization feature should be properly fused with the color one inside the model and inside the training loop. A stable model which properly integrates the color-based feature and the polar-based feature might able to both reduce the false detection and produce new true positive bounding boxes as well.

Stereo matching Part III

5 Post-aggregation for stereo matching based on Dempster-Shafer theory

5.1 Introduction

As an increasingly important issue in computer vision, stereo matching allows a system to perceive the $3D$ knowledge of the environment by producing the disparity map from input stereo images. The disparity map allow us to estimate the distances to the objects, upon which further techniques such as obstacle/pedestrian detection, recognition, classification can be realized.

The presence of the specular highlights in the environment (e.g. urban road scenario) restricts the accuracy of the estimated disparity map, since highlight totally covers the original image texture, causing in the lost of information. In addition, the position of the specular highlight also changes from different views.

In this chapter, we study if polarization can be used to improve the accuracy of the stereo matching results, especially in presence of highlights. We firstly proposed a post-aggregation step which improves the disparity map on color-based images [62]. We then proposed a fusion rule to constraint the color stereo matching result using the polarization images.

5.2 Related work

In recent years, local stereo matching methods have advanced due to their efficiency. The key problem of the local methods is to find a stable cost function and an appropriate window with reasonable size for cost aggregation. As feasible cost functions,

there exists, for instance, the sum of squared-differences (SSD), the sum of absolute-difference (SAD), Census-transform [71], Diff-Census [35]. In order to obtain a robust aggregation strategy, various aggregation windows are constructed, such as adaptive window [26], adaptive weighted support window [70] or cross zone aggregation [72].

When comes to reality, the matching cost computed by the intensity or the correlation of pixels cannot be considered to be fully reliable even if they use a robust cost function. When it is not possible to obtain a precise measurement from the matching cost, some probabilistic methods are used by considering the risk and reliability of the information source (e.g. belief propagation [56]).

A common assumption in classical probabilistic methods (Bayesian probabilities for instance), is that all probabilities that satisfy specific properties must add to one [49]. In practice however, it could be questionable when we don't have a complete knowledge of an event. For instance, in stereo matching, the corresponding points do not always exist, or hardly a reasonable corresponding pixel can be chosen because of radiometric distortions or untextured areas.

We seek to remedy these problems by employing Dempster-Shafer Theory (DST) [11], which is also known as belief function theory or evidence theory. Due to its ability to handle situations with nonspecific ambiguous or conflicting information, the DST, in comparison with the Bayesian theory, introduces the Frame of Discernment (FoD) which models explicitly the uncertainty. The value of FoD indicates the probability that we cannot decide to which class the pixel belongs. Fortunately, this concept matches with situations where there is an occlusion or disparity conflicts in the neighboring pixels.

Some previous work of DST for stereo matching has been done as in [21]. Firstly the author defined a 1×6 feature vector for each pixel. Then two categories of pixels are identified namely true matches and false matches by prior knowledge. For each pixel, authors compute its similarities to each categories to make a decision.

In this chapter, we propose another approach of using DST: the disparity searching problem is regarded as a classification problem, in which each class represents a disparity. For instance, if the disparity range is 0 to 127, there will be 128 classes. Firstly, Diff-Census cost function [35] with fix window aggregation is employed to initialize the mass function (probability for every class at each pixel). Secondly, a cross zone window is used for the DST strategy. This method could be used as a

post-aggregation method to replace the winner-takes-all(WTA) strategy.

The proposed post-aggregation method is designed for two considerations: i) It is observed that some disparity refinement is performed only after having the disparity map. In this post-aggregation step, the aggregated cost is kept and processed in each disparity to make a better final decision; ii) Cross zone window consists in finding for each pixel a cross support zone given by a threshold of the intensity difference. Yet this threshold, in the presence of noisy pixels, could to be restrictive for the support zone. Also a fix window will usually result in a foreground flattening. We combined these two windows to reduce both their disadvantages.

The proposed method is assessed firstly on a well known stereo benchmark (KITTI, [18–20]) to demonstrate its feasibility on color-based images. This result is further fused with the Degree Of Polarization (DOP) and compared with the color-based results to discuss if the fact of using polarization in stereo matching produce a better disparity map or not.

5.3 Dempster-Shafer theory

The DST is an effective uncertain reasoning framework [12, 55]. Let the *frame of discernment* be a finite set of all possible values $\Omega = \{\omega_1, \ldots, \omega_c\}$ for X (the Ω is the set of all disparity in the disparity range, and X is a pixel). In the framework of DST, based on a given source of uncertain information or partial knowledge, the *mass function m* is used to quantify the evidence about the value of X (the disparity of X), m is defined from the powerset of Ω, namely 2^{Ω}, to [0,1], such that

$$\sum_{A \subseteq \Omega} m(A) = 1 \tag{5.1}$$

where $m(A)$ is the subjective probability (*mass of belief*) that supports $X \in A$. Each subset of A is named a focal set if and only if $m(A)$ is nonzero. And when $m(\phi) = 0$ (ϕ refers to the empty set), the mass function m is said to be normalized.

The corresponding *belief* ($Bel(A)$) and *plausibility* ($Pl(A)$) functions based on the mass function could be defined respectively for each subset $A \subseteq \Omega$ as

$$Bel(A) = \sum_{B \subseteq A} m(B) \tag{5.2}$$

$$Pl(A) = \sum_{B \cap A \neq \emptyset} m(B) \tag{5.3}$$

where $Bel(A)$ represents the probability that evidence supports $X \in A$, and accordingly, $Pl(A)$ represents the probability that evidence does not contradict the fact that $X \in A$.

When different and independent sources of uncertain evidence about X are introduced, they can be combined using the Dempster's combination rule. For instance, let m_1 and m_2 be two independent mass functions, the combination of them can be written as

$$(m_1 \oplus m_2)(A) = \frac{1}{1-K} \sum_{B \cap C = A} m_1(B) m_2(C) \tag{5.4}$$

For all subset $A \neq \emptyset$ and $(m_1 \oplus m_2)(\emptyset) = 0$, where

$$K = \sum_{B \cap C = \emptyset} m_1(B) m_2(C) \tag{5.5}$$

can be viewed as a measurement of the *degree of conflict* between the evidence offered by m_1 and m_2. The pignistic probability distribution for each singletons can be computed with the newly produced mass function. Or, the corresponding value of the belief function or the plausibility function can be alternative ways for the final decision about X.

5.4 Post-aggregation algorithm using DST

Local stereo matching usually seeks for a better way to use the information provided by the neighborhood. The developed method combines together a fix window aggregation followed by a DST cross-zone post-aggregation. The combination of the two kind of windows reduces the disadvantages of both the fix window and the cross zone window. To further inspect the information provided in each disparity map, the post-aggregation step keeps and processes the aggregated cost in each disparity to make the final decision. Based on this idea, our algorithm is detailed in the following six steps:

1- Obtaining matching cost: To have a more robust result, a window aggregation is used to compute the matching cost. A fix window is used for this pre-step. For every pixel p in the image, the matching cost (after aggregation) can be represented by a vector $cost(p)$, as

$$cost(p) = \{c_p^1, \ldots, c_p^n\} \tag{5.6}$$

Here c_p^i refers to the cost corresponding to each disparity $\{d_i\}_{i=\overline{1,n}}$, where n is the maximum disparity.

2- Cost-mass transformation: the mass is initialized by transforming the matching cost obtained in the first step. To build up the transformation function for pixel p, the nominator is set to $c_p^{max} - c_p^i + c_p^{min}$ to keep the result in the realm of $[c_p^{min}, c_p^{max}]$. The denominator is then written as $n \times (c_p^{min} + c_p^{max}) - \sum_{i=1}^{n} c_p^i$ to normalize the mass in each disparity. Here c_p^{min} and c_p^{max} are respectively the maximum and the minimum costs in $cost(p)$. For this pixel, the mass of belief attached to each possible disparity $d_i \in D = \{d_1, \ldots, d_n\}$ is computed as :

$$m(\{d_i\}) = \frac{c_p^{max} - c_p^i + c_p^{min}}{n \times (c_p^{min} + c_p^{max}) - \sum_{i=1}^{n} c_p^i} \tag{5.7}$$

Hence, the subjective probability of each possible disparity d_i is inversely proportional to the corresponding cost c_p^i, and $\sum_{i=1}^{n} m(\{d_i\}) = 1$.

3- Cross zone computation: a specific window should be computed for each pixel as proposed by [72]. The idea behind is to construct a cross region for each pixel. For this, it is necessary to find only four pixels, corresponding to the end of the four arms: up, down, left and right (Figure 5.1.a). Then, in order to construct a region of various shapes, for each pixel that lies on the vertical arm, the horizontal arm will give the region boundaries for the specific row (Figure 5.1.b, Figure 5.1.c).

More specifically, given a pixel p, its directional arms (left, right, up or down) are found by applying the following rules:

$D_c(p, p_a) < \tau$. The intensity difference D_c between the pixel p and an arm pixel p_a should be less than a given threshold τ, where $D_c(p, p_a) = max|I(p) - I(p_a)|$, $I(p)$ refers to the intensity of the pixel p.

$D_s(p, p_a) < L$. The Euclidean distance D_s between p and p_a is limited by the maxi-

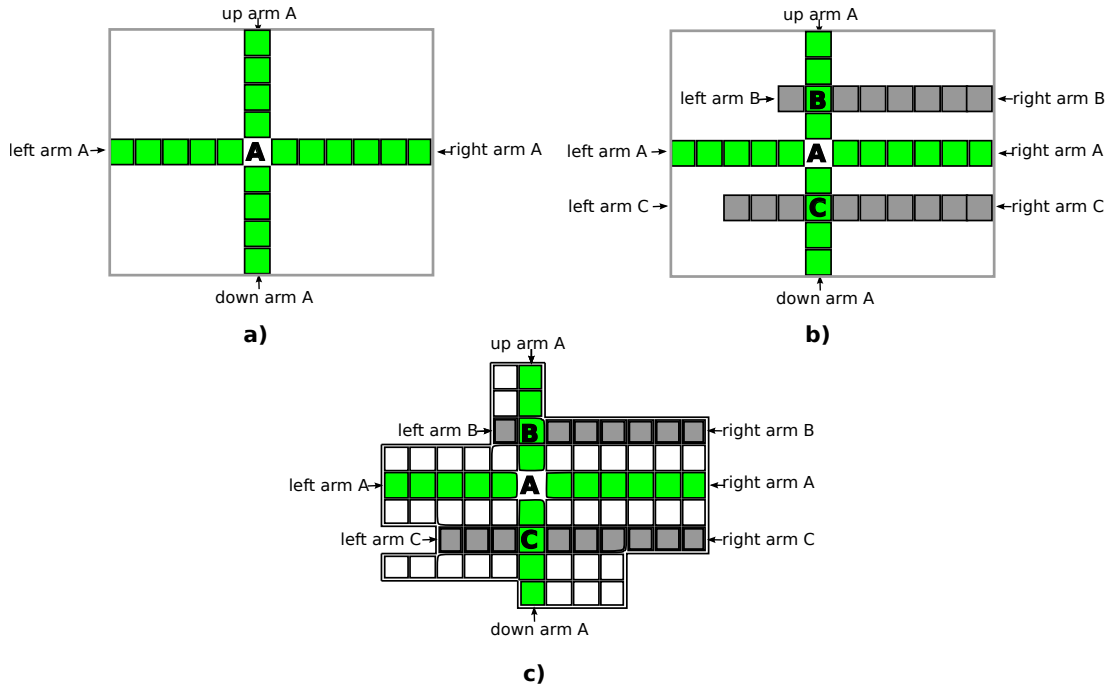


Figure 5.1 – Cross region construction: a) For each pixel four arms are chosen based on some color and distance restrictions; b),c) The cross region of a pixel is constructed by taking for each pixel situated on the vertical arm, its horizontal arm limits.

mum length threshold L.

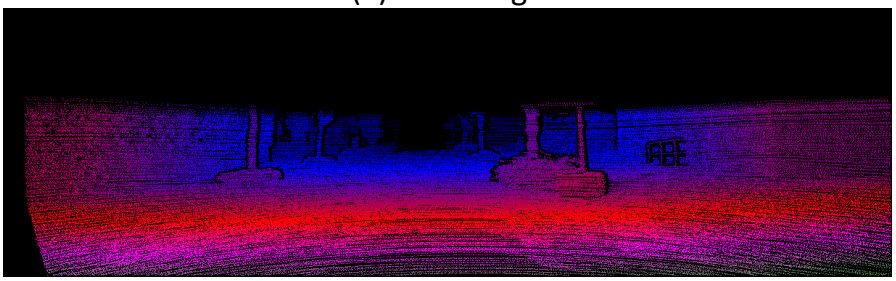
4- Cross zone mass discounting: Let the neighborhood of the pixel p in the cross zone be $\Gamma = \{\tau_1, \ldots, \tau_m\}$. Each pixel $\tau_p \in \Gamma$ can be regarded as a piece of evidence to help reasoning the actual disparity of p. As done in k-NN evidence theory [13, 51], the evidence offered by τ_p is partially reliable, and its influence is inversely proportional to the spatial distance between τ_p and p (denoted by $\omega_{p,q}$). Under this assumption, the evidence provided by the neighbor pixel can be quantified as a mass function:

$$m_{p,q}(\{d_i\}) = \alpha \cdot e^{-\gamma \omega_{p,q}^2} \cdot m_q(\{d_i\}), \quad \forall i = 1, \ldots, n \tag{5.8}$$

where α and γ are the arbitrary parameters. Additionally, according to the fact that the evidence provided by each neighbor pixel is not of 100% certainty, the rest mass of belief in on the FoD as follows:

$$m_{p,q}(D) = 1 - \sum_{i=1}^{n} m_{p,q}(\{d_i\}) \tag{5.9}$$

(a) Left image



(b) groud truth disparity

Figure 5.2 – (a) The original left image, (b) The ground-truth disparity given by KITTI.

The value of $m_{p,q}(D)$ represents, based on the evidence from τ_p, how much the probability that current pixel p doesn't belong to any particular disparity, meaning possibly an occlusion, a radiometric distortion or untextured area. By training the value of α and γ , a reasonable value on FoD will be given.

5- Cross zone mass fusion: this step is taken after getting the knowledge of each neighbor's evidence in the last step. The evidence will be fused together using the Dempster's Combination rule stated in Equation (5.4). The knowledge provided by each evidence is focused on the singletons or the whole FoD and can be computed succinctly as below:

$$M_p(\{d_i\}) = \frac{1}{Q}\prod_{q=1}^{m}\left[m_{p,q}(\{d_i\}) + m_{p,q}(D)\right] - \prod_{q=1}^{m} m_{p,q}(D) \tag{5.10}$$

$$M_p(D) = \frac{1}{Q}\cdot\prod_{q=1}^{m} m_{p,q}(D) \tag{5.11}$$

$$Q = \sum_{i=1}^{n} \prod_{q=1}^{m} [m_{p,q}(d_i) + m_{p,q}(D)] - (n-1) \cdot \prod_{q=1}^{m} m_{p,q}(D) \tag{5.12}$$

In which $M_p(\{d_i\})$ represents the subjective probability of d_i, $M_p(D)$ represents the ignorant degree, and Q represents the conflict degree.

6- Decision making: the final mass of each singleton $\{d_i\}_{i=\overline{1,n}}$ and FoD D in the image describe for each pixel p the probability for every possible disparity and respectively the probability for 'cannot decide the disparity'. This better can describe the uncertainty of the pixel p. The final decision will be made by computing the belief function or plausibility function for each disparity $\{d_i\}_{i=\overline{1,n}}$ which are represented as

$$Bel_p(\{d_i\}) = M_p(\{d_i\}) \tag{5.13}$$

and

$$Pl_p(\{d_i\}) = M_p(\{d_i\}) + M_p(D) \tag{5.14}$$

Comparing the results from Equation (5.13) and Equation (5.14), the disparity of the pixel p could be determined conveniently by searching for the maximum mass of belief.

5.5 Fusion rule

Polarization attributes, such as DOP and AOP, are densely calculated for each pixel in the image. Influenced by the complex illumination condition and the noise from three different images, the polarization images are too noisy to be used directly for stereo matching.

From the post-aggregation algorithm described in the last section, the belief $Bel_p(\{d_i\})$ of the pixel p on different disparities is computed, where n is the disparity range. The result of the color-based image is generated by simply choosing the disparity which corresponds to the maximum belief.

In this section, we propose a fusion rule, which use polarization image as a constraint when generating the disparity map. In the fusion rule, the belief Bel_p is used as

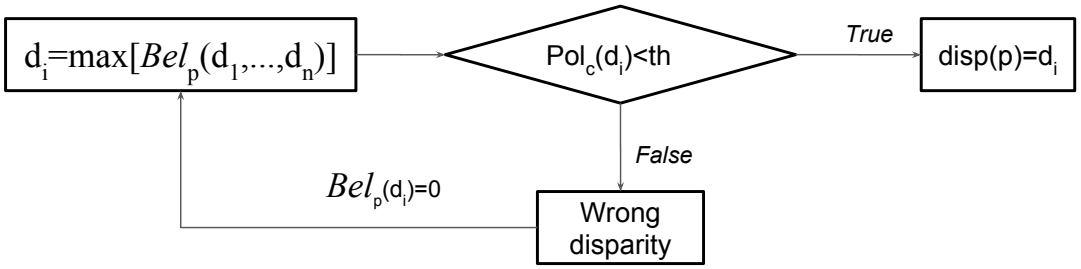


Figure 5.3 – Fusion rule using polarization.

Table 5.1 – Mean DE with different cost functions

cost functions	fix window WTA(%)	cross zone WTA(%)	fix window DST(%)
SAD	29.73	34.83	24.21
Census	15.26	22.71	10.90
ADCensus	16.95	16.11	14.82
DiffCensus	14.16	12.57	**10.66**

an input: for the disparity d_i which gains the maximum belief, its corresponding polarization cost $Pol_c(d_i)$ is computed. Taking the DOP image as an example the cost $Pol_c(d_i)$ is equal to $|DOP_L(p) - DOP_R(p)|$. If the polarization cost is larger than a threshold th, the disparity is neglected, and a new disparity is chosen by setting $Bel(d_i) = 0$ (th is a hyper-parameter that is empirically chose according to the data set). This fusion rule is further illustrated in Figure 5.3.

Table 5.2 – Mean DE results from KITTI data set

Error	Out-Noc	Out-All	Avg-Noc	Avg-All
2 pixels	14.77 %	16.53 %	2.5 px	3.1 px
3 pixels	**9.39** %	11.15 %	2.5 px	3.1 px
4 pixels	7.48 %	9.21 %	2.5 px	3.1 px
5 pixels	6.48 %	8.18 %	2.5 px	3.1 px

5.6 Experimental results

5.6.1 Results on color-based images

Four cost functions are tested in the experimental part: SAD as a standard and widely used cost function, Census as a well-known one, ADCensus as top-ranked in middlebury data set and DiffCensus as a stable one for KITTI data set. These cost functions are tested with fix window aggregation and our DST cross zone post-aggregation strategy. As comparison, every cost function is combined with the fix window aggregation and cross zone aggregation in both using WTA strategy. In this experiment, the belief function in Equation (5.13) is used to compute the disparity (the term of FoD in Equation (5.9) is not yet applied to predict the occlusion). The size of the fix window is 11×25 and the arm boundary of the cross zone is 9×20 with the 'intensity difference' threshold of 14. The arm boundary and intensity threshold are adjusted accordingly, otherwise the cross zone is computed exactly as in [72].

The result of mean disparity error (MDE) on the KITTI training set is shown in Table 5.1 with the resulting disparity map shown in Figure 5.4. As can be seen, the group using DST cross zone has smaller error than the other two groups. The best result is of 10.66% using DiffCensus cost function, fix window aggregation and our method and of 14.16% and 12.57% respectively for the other two groups with the DiffCensus cost function. The best case in tested on the KITTI data set, in which the ground truth is not provided. To have a mean DE of testing set, the resulting disparity images were evaluated by the KITTI online evaluation engine. The feedback MDE is shown in Table 5.2 in where the out-Noc is the percentage of erroneous pixels in non-occluded areas, the Out-All is the percentage of erroneous pixels in total, the Avg-Noc is computed by: [Average disparity / end-point error in non-occluded areas], and the Avg-All by [Average disparity / end-point error in total]. As can be seen, the mean DE of non-occlusion with 3-pixels threshold is 9.39% which is quite acceptable. It should be noted that this post-aggregation method can improve 5.52% as for SAD and 4.16% as for census cost function. Even for stable cost functions like DiffCensus, our method can still make an improvement. This proves that for color images, the proposed algorithm improves the quality of the disparity map.

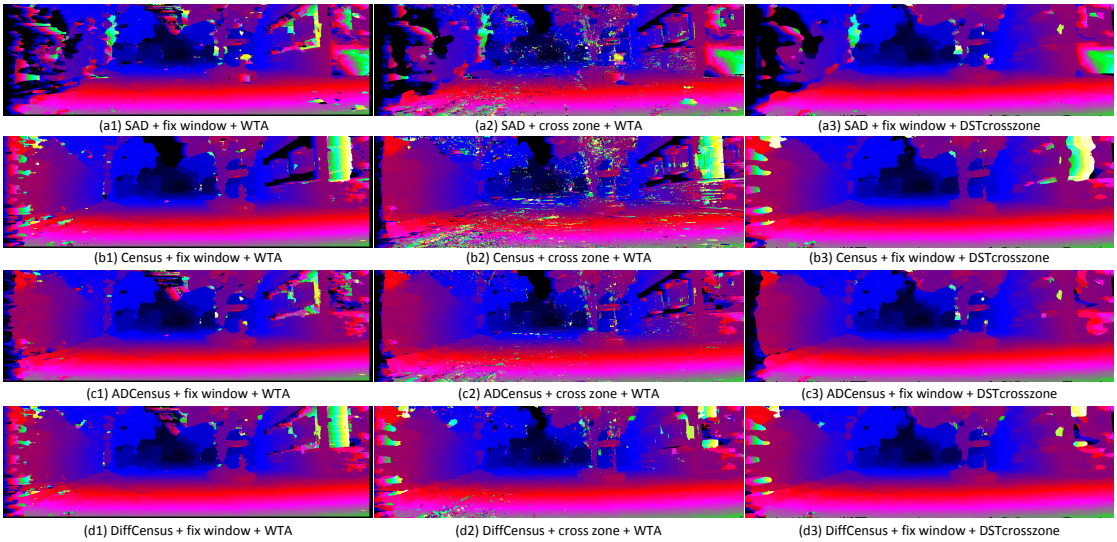
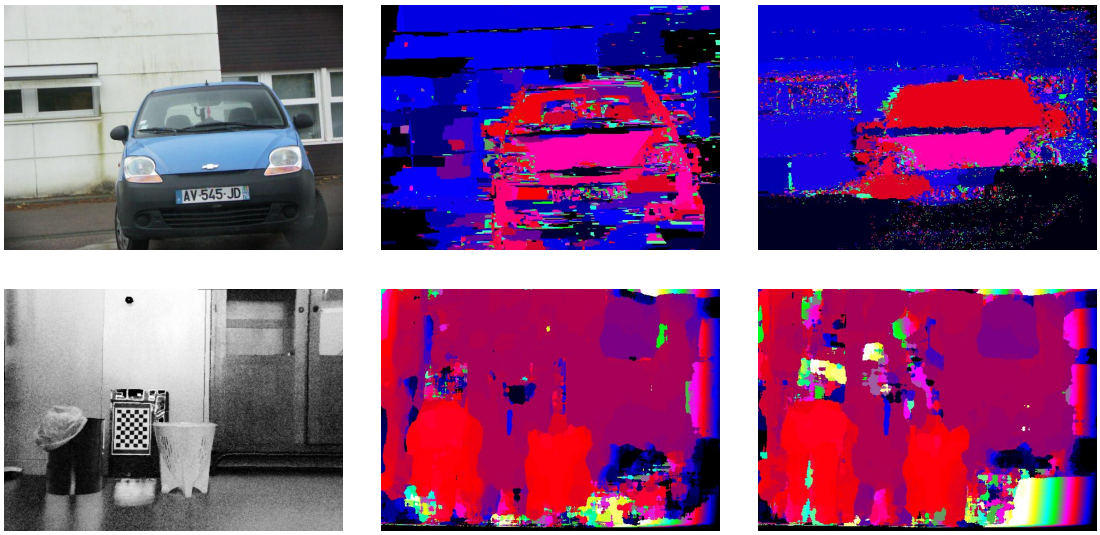


Figure 5.4 – The resulting disparity maps for different test cases: first column with the fix window and WTA, second column with the cross zone and WTA, last column with our method.

5.6.2 Results of the fusion with polarization

Since there are no public available polarimetry data set, we test the proposed method with the fusion rule on our self-collected stereo images to give preliminary results, where the resulting disparity maps and original images (being I_0) are shown in Figure 5.5.

The threshold in the fusion rule is set to $th = 0.13$ (After several th values tested, we the best one according the best visual result). The DOP image is used as polarization image for fusion, since it produces better results than the AOP image. The reason is that, the AOP of a pixel is more related to the incident angle. When the view changes, the incident angle also changes accordingly, and that the AOP consequently changes a lot between left and right images. The DOP, however, relies more to the surface material, which does not strongly change from the left to the right views. In Figure 5.5, the black parts in the polarization results correspond to the pixels on which no appropriate disparity can be found with the constraint given by the DOP. The disparity map are quiet noisy compared to the result from the KITTI data set, since our acquired images are much more noisy that those from KITTI. It is observed from the disparity map that, as an advantage, on the part of the windshield of the car the polarization result performs better. This fact is encouraging, since for traditional



Original image (left) | Disparity map (color) | Disparity map (polarization)

Figure 5.5 – The results of fusion with DOP.

color-based methods, the transparent object is always a problem even for the Lidar detector. However, on the image with some indoor objects, the polarization result on the reflective floor produce even worse result than the color one. On the metal cabinet behind the chessboard, the polarization also produces much more error than the color based one. These errors are reduced when increasing the fusion threshold, however, still few improvement can be observed compared to the color-based result.

From the results shown above, we can conclude that the polarization image, when used as a constraint, may have some improvement for transparent objects. However, more effort should be made to handle the error that it brought, e.g. the erroneous results on high reflective areas.

5.7 Discussion

In this chapter, we firstly proposed a DST-based post-aggregation method for stereo matching, which is proved to improve the disparity generated by the color images. A fusion rule is then proposed that uses polarization images as a constraint. The DOP image is found to produce better fusion results than the AOP image. It is straightforward, since the AOP changes largely according to different views, while the DOP does not.

The obtained results affirm that the fusion results are found to be limited: some improvement are observed on the transparent objects, however, the error that it brought on other areas, e.g., high reflective areas, make these results less acceptable. The reason might be that polarization attributes that we have found, are not constant from different views. The change of these attributes from the left to the right views may be irregularly, e.g. the DOP of a pixel p get larger in right image as $DOP_L(p) < DOP_R(p)$, and its neighbor pixel q may get smaller, as $DOP_L(q) > DOP_R(q)$. This is totally contradictory to the basic of stereo matching.

For the future work, to applied polarization onto stereo matching, one should find an attribute that is more or less constant from left to right images: it does not need to be a new attribute, it can be a new measurement of distance by the polarization state. An example of such distance can be found in [24]. This distance is modeled for insects detection, it does not suit for outdoor scenarios. However, by carefully modeling a new distance for outdoor scenes, the result should be improved. An other problem is that, the noise should also be carefully treated: the polarization state of the neighboring pixels are naturally different, since they describe the reflected light from different positions, different incident lights etc. Yet in the point of view of image processing, the basic assumption is that the neighboring pixels should have homogeneous attributes. With properly denoising the polarization images, the result my also be improved.

6 Conclusion and perspectives

In this thesis we have focused to improve the performance of computer vision tasks using polarization images.

Specularity removal Since polarization images provide complementary color constraints for specular components, we proposed a polarization-based specularity removal method using global energy minimization. The global energy function is constructed by the independency assumption and the constraint given by a first approximate solution.

As a future work, we should further adapt this algorithm to outdoor scenes to corporate with ADAS (e.g. through region segmentation), as well as the fast implementation to produce real-time results. We also suggest to apply inpainting method for the area where the light intensity surpasses the camera sensor range, which causes the over saturation on the image. The reason is that the over saturation has caused the lost of information, which cannot be recovered through specularity removal methods. Inpainting techniques may simulate the original information through assumption of homogeneity on texture or other features.

Polarization as features Light polarization patterns, including specular highlights, also carry different surface information: surface material, roughness, geometric structure etc. For this reason, beyond removing the specularity, polarization images can also be used as image features for tasks such as image semantic segmentation and car detection.

We proposed to integrate the polarization setup into image semantic segmentation,

where HOG, LBP and Lab features are extracted from both polarization and color images to concatenate a feature vector. The joint boosting classifier is used to select a feature at each iteration, which serves as model learning and feature fusion at the same time. The experimental results shows that this proposed method improves especially the image labeling accuracy for the dust set, which implies that the polarization imaging would be a potential solution for poor illumination cases. In the future work, ones should apply this method in a much larger data base with fast implementation to give more general and realistic results.

The polarization is also used as features to detect cars in road scenes. Among all the polarization features we acquired, a feature selection is firstly performed to select the most informative features as well as remove redundancy. The DPM detector is trained separately on the selected polarization images and color images. A fusion rule is proposed to combine the two trained detector and produce the final detection results. Experimental results shows that the proposed method can largely reduces the false detection.

As a limitation, in this proposed method, the polarization-based feature is not formed to generate new true positive detection boxes. The reason why we did not take new detection from the polar-based model in the fusion rule, is that this model is not robust enough, and that the new detection boxes it provides are usually incorrect. To deal with this limitation, as a future work, the polarization features should be properly fused with the color one inside the model and inside the training loop. A stable model which properly integrates the color-based feature and the polar-based feature might able to both reduce the false detection and produce new true positive bounding boxes as well.

Stereo matching As the special polarization patterns shown on the high reflective and transparent objects, we intend to apply polarization into stereo matching to improve the disparity map in these regions. We firstly proposed a post-aggregation stereo matching method for color-based images, which is tested on a well-known stereo matching benchmark to show the feasibility of this method. A fusion rule is then proposed to constrain the disparity map using polarization images. However, the given disparity map is not satisfying. It showed some improvement on transparent objects, which over performs other sensors such as Lidar. Yet on high reflective surfaces, the disparity map is not better than the one use color only images. The reason is that the specular reflection changes position with the change of views, which

is contradictory to the general assumption in stereo matching. To further improve this result, one may consider to perform an object detection using polarization, and then use the detection result as prior knowledge to improve the accuracy of the disparity map.

In this thesis, we showed that polarization images can be used to improve computer vision applications, through remove the image highlight or analyze the image with the presence of highlights. This showed the potential of polarization to be used to improve various ADAS-related applications and to go beyond color images limitations.

Publications

Fan Wang, Alina Miron, Samia Ainouz, Abdelaziz Bensrhair. POST-AGGREGATION STEREO MATCHING METHOD USING DEMPSTER-SHAFER THEORY. IEEE International Conference on Image Processing, 2014.

Fan Wang, Samia Ainouz, Caroline Petitjean, Abdelaziz Bensrhair. POLARIZATION-BASED SPECULARITY REMOVAL METHOD WITH GLOBAL ENERGY MINIMIZATION. IEEE International Conference on Image Processing, 2016.

Fan Wang, Samia Ainouz, Abdelaziz Bensrhair. ROBUST VEHICLE DETECTION VIA THE FUSION OF POLARIZATION AND COLOR IMAGES. British machine vision conference, 2016. *submitted*.

Fan Wang, Samia Ainouz, Caroline Petitjean, Abdelaziz Bensrhair. LOCAL-GLOBAL COMBINED DIFFUSE AND SPECULAR REFLECTION SEPARATION METHOD BASED ON POLARIZATION. Computer Vision and Image Understanding. *submitted*.

A Appendix A

In section 4.3–4.5, we have introduced the DPM model and the LSVM algorithm proposed by [16]. Here, more detail of the model matching procedure, the stochastic gradient descent algorithm and the theory of data-mining hard examples are described, as well as the post-processing method used in [16].

A.1 Model matching

To detect objects in an image, the overall score for each root location according to the best possible placement of the parts is computed,

$$score(z_0) = \max_{z_1,\cdots,z_n} score(z_0,\cdots z_n). \tag{A.1}$$

A detection is performed by highscore root locations, well corresponding part filter locations describe a full object hypothesis. Note that an object may yield several instances with different configuration in locations.

The dynamic programming and the generalized distance transforms (min-convolutions) is used to compute the best locations for the parts. More in detail, let $R_{i,l} = F_i' \cdot \phi(H,z) = F_i' \cdot \phi(H,(x,y,l))$ be an array storing the response of the $i-$th model filter in the $l-$th level of the feature pyramid. The matching algorithm starts by computing these responses. Note that $R_{i,l}$ is a cross-correlation between F_i' and level l of the feature pyramid.

After computing filter responses, it is transformed to generate the spatial uncertainty

as:

$$D_{i,l}(x,y) = \max_{dx,dy}(R_{i,l}(x+dx, y+dy) - d_i \cdot \phi_d(dx,dy)), \tag{A.2}$$

where the high filter scores are spread to nearby locations with the deformation costs. For example, the value $D_{i,l}(x,y)$ is the maximum contribution of the $i-$th part to the score of a root location that places the anchor of this part at position (x,y) in level l. The transformed array, $D_{i,l}$, can be computed in linear time from the response array, $R_{i,l}$, using the generalized distance transform algorithm from [**?**].

The overall root scores at each level can be expressed by the sum of the root filter response at that level, plus shifted versions of transformed and subsampled part responses,

$$score(x_0, y_0, l_0) = R_{0,l_0}(x_0, y_0) + \sum_{i=1}^{n} D_{i,l_0-\lambda}(2(x_0, y_0) + v_i) + b. \tag{A.3}$$

Recall that λ defines the level difference between part and root filters (root filter is applied in a more 'rough' level than the part filters).

Note that in Equation (A.3), we just independently pick the best location for each part filter, since there are interactions between part filters. More specifically, the contribution of each part filter to a defined root location is computed and recorded in the transformed arrays $D_{i,l}$, so that we efficiently obtain the total score of a root position at level l by adding up the root filter response and the contributions from each part.

In addition, the computation of $D_{i,l}$ simultaneously compute optimal displacements for a part as a function of its anchor position,

$$P_{i,l}(x,y) = arg \max_{dx,dy}(R_{i,l}(x+dx, y+dy) - d_i \cdot \phi_d(dx,dy)). \tag{A.4}$$

After finding a root location x_0, y_0, l_0 with high score we can find the corresponding part locations by looking up the optimal displacements in $P_{i,l_0-\lambda}(2(x_0, y_0) + v_i)$.

A.2 Stochastic gradient descent

In this section, we describe the stochastic gradient descent [16] used in 'step 2) optimize filters' in detail.

For an arbitrary training set D, let $z_i(\beta) = arg\max_{z \in Z(xi)} \beta \cdot \Phi(x_i, z)$, then $f_\beta(x) = \beta \cdot \Phi(x_i, z_i(\beta))$.

The sub-gradient of the LSVM object function

$$\nabla L_D(\beta) = \beta + C \sum_{i=1}^{n} h(\beta, x_i, y_i) \tag{A.5}$$

is computed as follows,

$$h(\beta, x_i, y_i) = \begin{cases} 0 & if\ y_i f_\beta(x_i) \geq 1 \\ -y_i \Phi(x_i, z_i(\beta)) & otherwise \end{cases} . \tag{A.6}$$

In the stochastic gradient descent, the ∇L_D is approximated using a subset of the examples and the step is taken in its negative direction. Using a single example, $< x_i, y_i >$, the $\sum_{i-1}^{n} h(\beta, x_i, y_i)$ is approximated to $nh(\beta, x_i, y_i)$. This algorithm repeatedly updates β as follows:

1) Let α_t be the learning rate for iteration t.

2) Let i be a random example.

3) Let $z_i = arg\max_{z \in Z(x_i)} \beta \cdot \Phi(x_i, z)$.

4) If $y_i f_\beta(x_i) = y_i(\beta \cdot \Phi(x_i, z_i)) \geq 1$, set $\beta := \beta - \alpha_t \beta$.

5) Else set $\beta := \beta - \alpha_t(\beta - Cny_i \Phi(x_i, z_i))$.

The time of convergence for this algorithm depends on the number of training examples. In particular, for the object detection cases, if there are many 'easy' examples, step 2 will often pick one of these and we do not make much progress.

A.3 Data-mining hard examples

Let $C_1 \subseteq D$ be an initial cache of examples. The model is trained repeatly with the updates of the cache as:

1) Let $\beta_t := \beta^*(C_t)$ (train the model using C_t).

2) If $H(\beta_t, D) \subseteq C_t$ stop and return β_t

3) Let $C'_t := C_t \setminus X$ for any X such that $X \subseteq E(\beta_t, C_t)$ (shrink the cache).

4) Let $C_{t+1} := C'_t \cup X$ for any X such that $X \subseteq D$ and $X \cap H(\beta_t, D) \setminus C_t \neq \emptyset$(grow the cache).

In this algorithm, the cache is shrinked by removing easy examples from C_t(examples that are outside the margin defined by β_t) in step 3). In step 4), the cache is grown by adding examples from D (at least one new example that is inside the margin defined by β_t). If there does not exist sum example, the algorithm would have returned in step 2.

The following theorem shows that when we stop we have found $\beta^*(D)$.

Theorem 1: Let $C \subseteq D$ and $\beta = \beta^*(C)$. If $H(\beta, D) \subseteq C$ then $\beta = \beta^*(D)$ [16].

Proof. $C \subseteq D$ implies $L_D(\beta^*(D)) > L_C(\beta^*(C)) = L_C(\beta)$. SInce $H(\beta, D) \subseteq C$ all examples in $D \setminus C$ have zero loss on β. This implies $L_C(\beta) = L_D(\beta)$. We conclude $L_D(\beta^*(D)) \geq L_D(\beta)$, and because L_D has a unique minimum $\beta = \beta^*(D)$. ■

The next result shows the algorithm will stop after a finite number of iterations. Intuitively this follows from the fact that $L_{C_t}(\beta^*(C_t)$ grows in each iteration, but it is bounded by $L_D(\beta^*(D))$.

Theorem 2: The data-mining algorithm terminates [16].

Proof. When we shrink the cache C'_t contains all examples from C_t with non-zero loss in a ball around β_t. This implies $L_{C'_t}$ is identical to L_{C_t} in a ball around β_t, and since β_t is a minimum of L_{C_t} is also must be a minimum of $L_{C'_t}$. Thus $L_{C'_t}(\beta^*(C'_t)) = L_{C_t}(\beta^*(C_t))$.

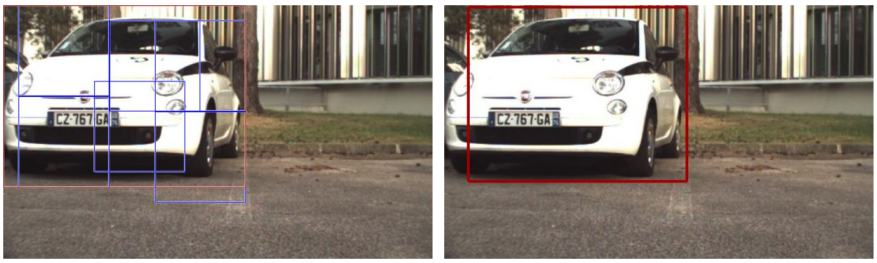


Figure A.1 – A car detection and the bounding box predicted from the object location configuration.

When we grow the cache $C_{t+1 \setminus C'_t}$ contains at least one example $< x, y >$ with non-zero loss at β_t. Since $C'_t \subseteq C_{t+1}$ we have $L_{C_{t+1}}(\beta) \geq L_{C'_t}(\beta)$ for all β. If $\beta^*(C_{t+1}) \neq \beta^*(C'_t)$ then $L_{C_{t+1}}(\beta^*(C_{t+1})) \geq L_{C'_t}(\beta^*(C'_t))$ because $L_{C'_t}$ has a unique minimum. If $\beta^*(C_{t+1}) = \beta^*(C'_t)$ then $L_{C_{t+1}}(\beta^*(C_{t+1})) \geq L_{C'_t}(\beta^*(C'_t))$ due to $< x, y >$.

We conclude $L_{C_{t+1}}(\beta^*(C_{t+1})) \geq L_{C_t}(\beta^*(C_t))$. Since there are finitely many caches the loss in the cache can only grow a finite number of times. ■

A.4 Post-processing

A.4.1 Bounding box prediction

By have the complete configuration of an object hypothesis, z, The bounding box is predicted using its root and relative part filter locations. It is implemented by mapping a feature vector $g(z)$, to the upper-left, (x_1, y_1), and lower-right, (x_2, y_2), corners of the bounding box. As the model with n parts, $g(z)$ is a $2n+3$ dimensional vector containing the width of the root filter in image pixels (this provides scale information) and the location of the upper-left corner of each filter in the image.

After the training process, the output of the detector on each instance is used to learn linear functions for predicting x_1, y_1, x_2 and y_2 from $g(z)$ (via linear least-square regression). Figure A.1 illustrates an example of the predicted bounding box and the configuration of the detected car.

A.4.2 Non-maximum suppression

In the matching process, as for each instance, we usually have multiple overlapping detections, the non-maximum suppression is used to eliminating these repeated detections. More specifically, in an image, we have a set of detections D, where each detection is defined by a bounding box and a score. We sort the detection in D by the corresponding score, and greedily select the highest scoring bounding boxes while skipping the bounding boxes that have at leat 50% overlaping with the selected ones.

Bibliography

[1] Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, and Sabine Susstrunk. Slic superpixels compared to state-of-the-art superpixel methods. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 34(11):2274–2282, 2012.

[2] Samia Ainouz, Olivier Morel, David Fofi, Saleh Mosaddegh, and Abdelaziz Bensrhair. Adaptive processing of catadioptric images using polarization imaging: towards a pola-catadioptric model. *Optical Engineering*, 52(3):037001–037001, 2013.

[3] Gary Atkinson, Edwin R Hancock, et al. Shape estimation using polarization and shading from two views. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 29(11):2001–2017, 2007.

[4] Max Born and Emil Wolf. *Principles of optics: electromagnetic theory of propagation, interference and diffraction of light*. Cambridge university press, 1999.

[5] Yuri Boykov and Vladimir Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 26(9):1124–1137, 2004.

[6] Yuri Boykov, Olga Veksler, and Ramin Zabih. Fast approximate energy minimization via graph cuts. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(11):1222–1239, 2001.

[7] Arthur D Costea and Sergiu Nedevschi. Multi-class segmentation for traffic scenarios at over 50 fps. In *Intelligent Vehicles Symposium Proceedings, 2014 IEEE*, pages 1390–1395. IEEE, 2014.

[8] Franklin C Crow. Summed-area tables for texture mapping. *ACM SIGGRAPH computer graphics*, 18(3):207–212, 1984.

[9] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 886–893. IEEE, 2005.

[10] Johannes F De Boer, Thomas E Milner, Martin JC van Gemert, and J Stuart Nelson. Two-dimensional birefringence imaging in biological tissue by polarization-sensitive optical coherence tomography. *Optics letters*, 22(12):934–936, 1997.

[11] Arthur P Dempster. Upper and lower probabilities induced by a multivalued mapping. *The annals of mathematical statistics*, 38(2):325–339, 1967.

[12] Arthur P Dempster et al. Upper and lower probabilities generated by a random closed interval. *The Annals of Mathematical Statistics*, 39(3):957–966, 1968.

[13] Thierry Denoeux. A k-nearest neighbor classification rule based on dempster-shafer theory. *Systems, Man and Cybernetics, IEEE Transactions on*, 25(5):804–813, 1995.

[14] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The pascal visual object classes (voc) challenge. *International Journal of Computer Vision*, 88(2):303–338, June 2010.

[15] Clement Farabet, Camille Couprie, Laurent Najman, and Yann LeCun. Learning hierarchical features for scene labeling. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(8):1915–1929, 2013.

[16] Pedro F Felzenszwalb, Ross B Girshick, David McAllester, and Deva Ramanan. Object detection with discriminatively trained part-based models. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(9):1627–1645, 2010.

[17] Fluxdata. Imaging polarimeters fd-1655-p. http://www.fluxdata.com/imaging-polarimeters.

[18] Jannik Fritsch, Tobias Kuehnl, and Andreas Geiger. A new performance measure and evaluation benchmark for road detection algorithms. In *International Conference on Intelligent Transportation Systems (ITSC)*, 2013.

[19] Andreas Geiger, Philip Lenz, Christoph Stiller, and Raquel Urtasun. Vision meets robotics: The kitti dataset. *International Journal of Robotics Research (IJRR)*, 2013.

[20] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.

[21] P. Javier Herrera Gonzalo Pajares and Jes¨¨2s M. de la Cruz. Combining stereovision matching constraints for solving the correspondence problem. *Advances in Theory and Applications of Stereo Vision*.

[22] Saurabh Gupta, Pablo Arbeláez, Ross Girshick, and Jitendra Malik. Indoor scene understanding with rgb-d images: Bottom-up segmentation, object detection and semantic segmentation. *International Journal of Computer Vision*, pages 1–17, 2014.

[23] Roger T Hanlon, Michael R Maxwell, Nadav Shashar, Ellis R Loew, and Kim-Laura Boyle. An ethogram of body patterning behavior in the biomedically and commercially valuable squid loligo pealei off cape cod, massachusetts. *The Biological Bulletin*, 197(1):49–62, 1999.

[24] Martin J How and N Justin Marshall. Polarization distance: a framework for modelling object detection by polarization vision systems. *Proceedings of the Royal Society of London B: Biological Sciences*, 281(1776):20131632, 2014.

[25] Aapo Hyvärinen, Juha Karhunen, and Erkki Oja. *Independent component analysis*, volume 46. John Wiley & Sons, 2004.

[26] Takeo Kanade and Masatoshi Okutomi. A stereo matching algorithm with an adaptive window: Theory and experiment. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 16(9):920–932, 1994.

[27] Dae Woong Kim, Stephen Lin, Ki-Sang Hong, and Heung-Yeung Shum. Variational specular separation using color and polarization. In *MVA*, pages 176–179, 2002.

[28] Hyeongwoo Kim, Hailin Jin, Sunil Hadap, and Inso Kweon. Specular reflection separation using dark channel prior. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1460–1467, 2013.

[29] Gudrun J Klinker, Steven A Shafer, and Takeo Kanade. The measurement of highlights in color images. *International Journal of Computer Vision*, 2(1):7–32, 1988.

[30] Vladimir Kolmogorov and Ramin Zabin. What energy functions can be minimized via graph cuts? *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 26(2):147–159, 2004.

[31] Lubor Ladicky, Christopher Russell, Pushmeet Kohli, and Philip HS Torr. Associative hierarchical crfs for object class image segmentation. In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 739–746. IEEE, 2009.

[32] Alex Levinshtein, Adrian Stere, Kiriakos N Kutulakos, David J Fleet, Sven J Dickinson, and Kaleem Siddiqi. Turbopixels: Fast superpixels using geometric flows. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(12):2290–2297, 2009.

[33] Shih-Schön Lin, Konstantin M Yemelyanov, Edward N Pugh Jr, and Nader Engheta. Polarization enhanced visual surveillance techniques. In *Networking, Sensing and Control, 2004 IEEE International Conference on*, volume 1, pages 216–221. IEEE, 2004.

[34] David G Lowe. Object recognition from local scale-invariant features. In *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, volume 2, pages 1150–1157. Ieee, 1999.

[35] Alina Miron, Samia Ainouz, Alexandrina Rogozan, and Abdelaziz Bensrhair. A robust cost function for stereo matching of road scenes. *Pattern Recognition Letters*, 38:70–77, 2014.

[36] Francesco Mitis, Dinesh Sethi, World Health Organization, et al. *European facts and global status report on road safety 2013*. World Health Organization, Regional Office for Europe, 2013.

[37] Daisuke Miyazaki, Masataka Kagesawa, and Katsushi Ikeuchi. Transparent surface modeling from a pair of polarization images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 26(1):73–82, 2004.

[38] Olivier Morel, Fabrice Meriaudeau, Christophe Stolz, and Patrick Gorria. Polarization imaging applied to 3d reconstruction of specular metallic surfaces. In

Electronic Imaging 2005, pages 178–186. International Society for Optics and Photonics, 2005.

[39] Olivier Morel, Christophe Stolz, Fabrice Meriaudeau, and Patrick Gorria. Active lighting applied to three-dimensional reconstruction of specular metallic surfaces by polarization imaging. *Applied optics*, 45(17):4062–4068, 2006.

[40] Shree K Nayar, Xi-Sheng Fang, and Terrance Boult. Separation of reflection components using color and polarization. *International Journal of Computer Vision*, 21(3):163–186, 1997.

[41] T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(7):971–987, Jul 2002.

[42] World Health Organization. *Global status report on road safety: time for action*. World Health Organization, 2009.

[43] Stefan Rahmann. Inferring 3d scene structure from a single polarization image. In *Industrial Lasers and Inspection (EUROPTO Series)*, pages 22–33. International Society for Optics and Photonics, 1999.

[44] Stefan Rahmann. Reconstruction of quadrics from two polarization views. In *Pattern Recognition and Image Analysis*, pages 810–820. Springer, 2003.

[45] Bryan C Russell, Antonio Torralba, Kevin P Murphy, and William T Freeman. Labelme: a database and web-based tool for image annotation. *International journal of computer vision*, 77(1-3):157–173, 2008.

[46] Megumi Saito, Yoichi Sato, Katsushi Ikeuchi, and Hiroshi Kashiwagi. Measurement of surface orientations of transparent objects using polarization in highlight. *Systems and Computers in Japan*, 32(5):64–71, 2001.

[47] Yoav Y Schechner, Srinivasa G Narasimhan, and Shree K Nayar. Instant dehazing of images using polarization. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, volume 1, pages I–325. IEEE, 2001.

[48] Yoav Y Schechner, Srinivasa G Narasimhan, and Shree K Nayar. Polarization-based vision through haze. *Applied Optics*, 42(3):511–525, 2003.

[49] Kari Sentz and Scott Ferson. *Combination of evidence in Dempster-Shafer theory*, volume 4015. Citeseer, 2002.

[50] Steven A Shafer. Using color to separate reflection components. *Color Research & Application*, 10(4):210–218, 1985.

[51] Hong-Bin Shen and Kuo-Chen Chou. Predicting protein subnuclear location with optimized evidence-theoretic k-nearest classifier and pseudo amino acid composition. *Biochemical and Biophysical Research Communications*, 337(3):752–756, 2005.

[52] Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(8):888–905, 2000.

[53] Jamie Shotton, Matthew Johnson, and Roberto Cipolla. Semantic texton forests for image categorization and segmentation. In *Computer vision and pattern recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE, 2008.

[54] William A. Shurcliff. *Polarization Light: Production and Use*. Havard University Press.

[55] Philippe Smets. The combination of evidence in the transferable belief model. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 12(5):447–458, 1990.

[56] Jian Sun, Nan-Ning Zheng, and Heung-Yeung Shum. Stereo matching using belief propagation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25(7):787–800, 2003.

[57] Robby T Tan and Katsushi Ikeuchi. Separating reflection components of textured surfaces using a single image. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(2):178–193, 2005.

[58] R.T. Tan, K. Nishino, and K. Ikeuchi. Separating reflection components based on chromaticity and noise analysis. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 26(10):1373–1379, Oct 2004.

[59] Joseph Tighe and Svetlana Lazebnik. Superparsing. *International Journal of Computer Vision*, 101(2):329–349, 2013.

[60] Antonio Torralba, Kevin P Murphy, and William T Freeman. Sharing features: efficient boosting procedures for multiclass object detection. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 2, pages II–762. IEEE, 2004.

[61] Shinji Umeyama and Guy Godin. Separation of diffuse and specular components of surface reflection by use of polarization and statistical analysis of images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 26(5):639–647, 2004.

[62] Fan Wang, Alina Miron, Samia Ainouz, and Abdelaziz Bensrhair. Post-aggregation stereo matching method using dempster-shafer theory. In *Proceedings of the IEEE International conference of image processing*, 2014.

[63] Fan Wang, Caroline Petitjean, Samia Ainouz, and Abdelaziz Bensrhair. Polarization-based specularity removal method with global energy mnimization. In *Proceedings of the IEEE International conference of image processing*, 2016.

[64] Lawrence B Wolff. Polarization vision: a new sensory approach to image understanding. *Image and Vision computing*, 15(2):81–93, 1997.

[65] Lawrence B Wolff and Andreas G Andreou. Polarization camera sensors. *Image and Vision Computing*, 13(6):497–510, 1995.

[66] Lawrence B. Wolff and Terrance E. Boult. Constraining object features using a polarization reflectance model. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (7):635–657, 1991.

[67] Tianfu Wu, Bo Li, and Song-Chun Zhu. Learning and-or models to represent context and occlusion for car detection and viewpoint estimation. 2015.

[68] Jianwei Yang, Zhaowei Cai, Longyin Wen, Zhen Lei, Guodong Guo, and Stan Z. Li. A new projection space for separation of specular-diffuse reflection components in color images. In *Proceedings of the 11th Asian Conference on Computer Vision - Volume Part IV*, ACCV'12, pages 418–429, Berlin, Heidelberg, 2013. Springer-Verlag.

[69] Jianwei Yang, Lixing Liu, and S.Z. Li. Separating specular and diffuse reflection components in the hsi color space. In *Computer Vision Workshops (ICCVW), 2013 IEEE International Conference on*, pages 891–898, Dec 2013.

[70] Kuk-Jin Yoon and In So Kweon. Adaptive support-weight approach for correspondence search. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(4):650–656, 2006.

[71] Ramin Zabih and John Woodfill. Non-parametric local transforms for computing visual correspondence. In *Computer Vision¡ªECCV'94*, pages 151–158. Springer, 1994.

[72] Ke Zhang, Jiangbo Lu, and Gauthier Lafruit. Cross-based local stereo matching using orthogonal integral images. *Circuits and Systems for Video Technology, IEEE Transactions on*, 19(7):1073–1079, 2009.

[73] Hongyuan Zhu, Fanman Meng, Jianfei Cai, and Shijian Lu. Beyond pixels: A comprehensive survey from bottom-up to semantic image segmentation and cosegmentation. *arXiv preprint arXiv:1502.00717*, 2015.