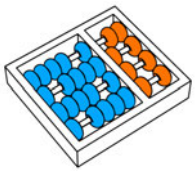
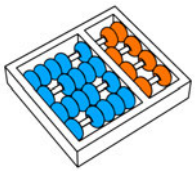


Tiago José de Carvalho

"Illumination Inconsistency Sleuthing for Exposing Fauxtography and Uncovering Composition Telltales in Digital Images"

"*Investigando Inconsistências de Iluminação para Detectar Fotos Fraudulentas e Descobrir Traços de Composições em Imagens Digitais*"

CAMPINAS
2014



University of Campinas
Institute of Computing

Universidade Estadual de Campinas
Instituto de Computação

Tiago José de Carvalho

"Illumination Inconsistency Sleuthing for Exposing Fauxtography and Uncovering Composition Telltales in Digital Images"

Supervisor: / *Orientador:* **Prof. Dr. Anderson de Rezende Rocha**

Co-Supervisor: / *Co-orientador:* **Prof. Dr. Hélio Pedrini**

"Investigando Inconsistências de Iluminação para Detectar Fotos Fraudulentas e Descobrir Traços de Composições em Imagens Digitais"

PhD Thesis presented to the Post Graduate Program of the Institute of Computing of the University of Campinas to obtain a PhD degree in Computer Science.

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Ciência da Computação do Instituto de Computação da Universidade Estadual de Campinas para obtenção do título de Doutor em Ciência da Computação.

This volume corresponds to the final version of the Thesis defended by Tiago José de Carvalho, under the supervision of Prof. Dr. Anderson de Rezende Rocha.

Este exemplar corresponde à versão final da Tese defendida por Tiago José de Carvalho, sob orientação de Prof. Dr. Anderson de Rezende Rocha.

Supervisor's signature / *Assinatura do Orientador*

CAMPINAS
2014

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Matemática, Estatística e Computação Científica
Maria Fabiana Bezerra Muller - CRB 8/6162

C253i

Carvalho, Tiago José de, 1985-
Illumination inconsistency sleuthing for exposing fauxtography and uncovering composition telltales in digital images / Tiago José de Carvalho. – Campinas, SP : [s.n.], 2014.

Orientador: Anderson de Rezende Rocha.
Coorientador: Hélio Pedrini.
Tese (doutorado) – Universidade Estadual de Campinas, Instituto de Computação.

1. Análise forense de imagem. 2. Computação forense. 3. Visão por computador. 4. Aprendizado de máquina. I. Rocha, Anderson de Rezende,1980-. II. Pedrini, Hélio,1963-. III. Universidade Estadual de Campinas. Instituto de Computação. IV. Título.

Informações para Biblioteca Digital

Título em outro idioma: Investigando inconsistências de iluminação para detectar fotos fraudulentas e descobrir traços de composições em imagens digitais
Palavras-chave em inglês:
Forensic image analysis
Digital forensics
Computer vision
Machine learning
Área de concentração: Ciência da Computação
Titulação: Doutor em Ciência da Computação
Banca examinadora:
Anderson de Rezende Rocha [Orientador]
Siome Klein Goldenstein
José Mario de Martino
Willian Robson Schwartz
Paulo André Vechiatto Miranda
Data de defesa: 21-03-2014
Programa de Pós-Graduação: Ciência da Computação

TERMO DE APROVAÇÃO

Defesa de Tese de Doutorado em Ciência da Computação, apresentada pelo(a) Doutorando(a) **Tiago José de Carvalho**, aprovado(a) em 21 de março de 2014, perante a Banca Examinadora composta pelos professores doutores:

Prof. Dr. William Robson Schwartz
Titular

Prof. Dr. Paulo André Vechiatto de Miranda
Titular

Prof. Dr. Siome Klein Goldenstein
Titular

Prof. Dr. Jose Mario De Martino
Titular

Prof. Dr. Anderson de Rezende Rocha
Presidente

Institute of Computing /*Instituto de Computação*
University of Campinas /*Universidade Estadual de Campinas*

Illumination Inconsistency Sleuthing for Exposing Fauxtography and Uncovering Composition Telltales in Digital Images

Tiago José de Carvalho[1]

March 21, 2014

Examiner Board/*Banca Examinadora*:

- Prof. Dr. Anderson de Rezende Rocha (Supervisor/*Orientador*)
- Prof. Dr. Siome Klein Goldenstein (Internal Member)
  IC - UNICAMP
- Prof. Dr. José Mario de Martino (Internal Member)
  FEEC - UNICAMP
- Prof. Dr. William Robson Schwartz (External Member)
  DCC - UFMG
- Prof. Dr. Paulo André Vechiatto Miranda (External Member)
  IME - USP
- Prof. Dr. Neucimar Jerônimo Leite (Substitute/*Suplente*)
  IC - UNICAMP
- Prof. Dr. Alexandre Xavier Falcão (Substitute/*Suplente*)
  IC - UNICAMP
- Prof. Dr. João Paulo Papa (External Substitute/*Suplente*)
  Unesp - Bauru

[1]Financial support: CNPq scholarship (Grant #40916/2012-1) 2012–2014

Abstract

Once taken for granted as genuine, photographs are no longer considered as a *piece of truth*. With the advance of digital image processing and computer graphics techniques, it has been easier than ever to manipulate images and forge new realities within minutes. Unfortunately, most of the times, these modifications seek to deceive viewers, change opinions or even affect how people perceive reality. Therefore, it is paramount to devise and deploy efficient and effective detection techniques. From all types of image forgeries, composition images are specially interesting. This type of forgery uses parts of two or more images to construct a *new reality* from scenes that never happened. Among all different telltales investigated for detecting image compositions, image-illumination inconsistencies are considered the most promising since a perfect light matching in a forged image is still difficult to achieve. This thesis builds upon the hypothesis that *image illumination inconsistencies are strong and powerful evidence of image composition* and presents four original and effective approaches to detect image forgeries. The first method explores eye specular highlight telltales to estimate the light source and viewer positions in an image. The second and third approaches explore *metamerism*, when the colors of two objects may appear to match under one light source but appear completely different under another one. Finally, the last approach relies on user's interaction to specify 3-D normals of suspect objects in an image from which the 3-D light source position can be estimated. Together, these approaches bring to the forensic community important contributions which certainly will be a strong tool against image forgeries.

Resumo

Antes tomadas como naturalmente genuínas, fotografias não mais podem ser consideradas como sinônimo de verdade. Com os avanços nas técnicas de processamento de imagens e computação gráfica, manipular imagens tornou-se mais fácil do que nunca, permitindo que pessoas sejam capazes de criar novas realidades em minutos. Infelizmente, tais modificações, na maioria das vezes, têm como objetivo enganar os observadores, mudar opiniões ou ainda, afetar como as pessoas enxergam a realidade. Assim, torna-se imprescindível o desenvolvimento de técnicas de detecção de falsificações eficientes e eficazes. De todos os tipos de falsificações de imagens, composições são de especial interesse. Esse tipo de falsificação usa partes de duas ou mais imagens para construir *uma nova realidade* exibindo para o observador situações que nunca aconteceram. Entre todos os diferentes tipos de pistas investigadas para detecção de composições, as abordagens baseadas em inconsistências de iluminação são consideradas as mais promissoras uma vez que um ajuste perfeito de iluminação em uma imagem falsificada é extremamente difícil de ser alcançado. Neste contexto, esta tese, a qual é fundamentada na hipótese de que *inconsistências de iluminação encontradas em uma imagem são fortes evidências de que a mesma é produto de uma composição*, apresenta abordagens originais e eficazes para detecção de imagens falsificadas. O primeiro método apresentado explora o reflexo da luz nos olhos para estimar as posições da fonte de luz e do observador da cena. A segunda e a terceira abordagens apresentadas exploram um fenômeno, que ocorre com as cores, denominado *metamerismo*, o qual descreve o fato de que duas cores podem aparentar similaridade quando iluminadas por uma fonte de luz mas podem parecer totalmente diferentes quando iluminadas por outra fonte de luz. Por fim, nossa última abordagem baseia-se na interação com o usuário que deve inserir normais 3-D em objetos suspeitos da imagem de modo a permitir um cálculo mais preciso da posicão 3-D da fonte de luz na imagem. Juntas, essas quatro abordagens trazem importantes contribuições para a comunidade forense e certamente serão uma poderosa ferramenta contra falsificações de imagens.

Acknowledgements

It is impressive how fast time goes by and how unpredictable things suddenly happen in our lives. Six years ago, I used to live in a small town with my parents in a really predictable life. Then, looking for something new, I have decided to change my life, restart in a new city and pursue a dream. But the path until this dream come true would not be easy. Nights without sleep, thousands of working hours facing stressful and challenging situations looking for solutions of new problems every day. Today, everything seems worth it and a dream becomes reality in a different city, with a different way of life, and always surrounded by people whom I love. People as my wife, Fernanda, one of the most important people in my life. A person that is with me in all moments, positives and negatives. A person who always supports me in my crazy dreams, giving me love, affection and friendship. And how not remembering my parents, Licinha and Norival?! They always helped me in the most difficult situations, standing by my side all the time, even living 700km away. My sister, Maria, a person that I have seen growing up, that I took care and that today certainly is so happy as I am. And there are so many other people really important to me that I would like to honor and thank. My friends whose names are impossible to enumerate, (if I start, I would need an extra page just for this) but who are the family that I chose and who are not family by blood, but family by love. I also thank my blood family which is much more than just relatives. They represent the real meaning of the word *family*. Also, I thank my advisors Anderson and Hélio whom taught me lessons day after day being responsible for the biggest part of this conquer. My father and mother in law, whom are like parents for me. The institutions which funded my scholarship and research (IF Sudeste de Minas, CNPq, Capes, Faepex). Unicamp by the opportunity to be here today. However, above all, I would like to thank God. Twice in my life, I have faced big fights for my life, and I believe that the fact of winning these battles and to be here today it is because of His help. Finally, I would like to thank everyone who believed and supported me in this four years of Ph.D. (plus two years of masters).

From my wholeheart, thank you!

"What is really good is to fight with determination, embrace life and live it with passion. Lose your battles with class and dare to win because the world belongs to those who dare to live. Life is worth too much to be insignificant."

Charlie Chaplin

List of Symbols

$\vec{L}$	Light Source Direction	10
$\mathbf{B}$	Irradiance	10
$\mathbf{R}$	Reflectance	10
Ω	Surface of the Sphere	10
$d\Omega$	Area Differential on the Sphere	10
$W(\vec{L})$	Lighting Environment	10
$\vec{N}$	Surface Normal Direction	10
I	Image	12
I_m	Face Rendered Model	12
E	Error Function	12
$\mathcal{R}$	Rotation Matrix	12
$\vec{t}$	Translation Vector	12
f	Focal Length	12
ρ	Principal Point	12
$\mathbf{e}$	Color of the Illuminant	15
λ	Scale Factor	15
$\mathbf{n}$	Order of Derivative	15
$\mathbf{p}$	Minkowski norm	15
σ	Gaussian Smoothing	15
$f_s(x, y)$	Shadowed Surface	16
$f_n(x, y)$	No-Shadowed Surface	16
$\mathbf{C}$	Shadow Matte Value	16
$\mathbf{D}$	Inconsistency Vector	16
φ	Density Function	16
$\vec{V}$	Viewer Direction	19
$\mathcal{H}$	Projective Transformation Matrix	20
$\mathbf{X}$	World Points	20
$\mathbf{x}$	Image Points	20
C	Circle Center in World Coordinates	20

r	Circle Radius	20
$\mathbf{P}$	Parametrized Circle	20
$\hat{\mathbf{X}}$	Model Points	20
$\mathcal{K}$	Intrinsic Camera Matrix	21
θ_x	Rotation Angle Around X axis	21
θ_y	Rotation Angle Around Y axis	21
θ_z	Rotation Angle Around Z axis	21
$\hat{\mathcal{H}}$	Transformation Matrix Between Camera and World Coordinates	21
$\vec{v}$	Viewer Direction on Camera Coordinates	22
$\vec{S}$	Specular Highlight in World Coordinates	22
$\mathbf{X}_s$	Specular Highlight Position in World Coordinates	22
$\mathbf{x}_s$	Specular Highlight Position in Image Coordinates	22
$\vec{l}$	Light Source Direction in Camera Coordinates	23
$\dot{\mathbf{x}}$	Estimated Light Source Position in Image Coordinates	23
$\vec{n}$	Surface Normal Direction on Camera Coordinates	23
Θ	Angular Error	23
$\ddot{\mathbf{x}}$	Estimated Viewer Position in Image Coordinates	24
$\mathbf{f}(\mathbf{x})$	Observed RGB Color from a Pixel at Location $\mathbf{x}$	38
ω	Spectrum of Visible Light	38
β	Wavelength of the Light	38
$\mathbf{e}(\beta, \mathbf{x})$	Spectrum of the Illuminant	38
$\mathbf{s}(\beta, \mathbf{x})$	Surface Reflectance of an Object	38
$\mathbf{c}(\beta)$	Color Sensitivities of the Camera	38
∂	Differential Operator	39
$\Gamma(\mathbf{x})$	Intensity in the Pixel at the Position $\mathbf{x}$	39
$\chi_c(\mathbf{x})$	Chromaticity in the Pixel at the Position $\mathbf{x}$	39
γ	Chromaticity of the Illuminant	39
c	Color Channel	40
$\vec{g}$	Eigenvector	41
$\mathbf{g}$	Eigenvalue	41
$\mathcal{D}$	Triplet (CCM, Color Space, Description Technique)	65
$\mathcal{P}$	Pair of $\mathcal{D}$	66
$\mathcal{C}$	Set of Classifiers	67
$\mathcal{C}^*$	Sub-Set of $\mathcal{C}$	68
$\mathbf{c}_i$	i^{th} Classifier in a Set of Classifiers	67
T	Training Set	67
V	Validation Set	67
$\mathcal{S}$	Set of $\mathcal{P}$ that Describes an Image I	69

$\mathcal{T}$ Threshold 76
ϕ Face 78
$\mathbf{A}$ Ambient Lighting 88
ϑ Slant 90
ϱ Tilt 90
$\mathbf{b}$ Lighting Parameters 91
Φ Azimuth 91
Υ Elevation 91

Contents

Abstract **ix**

Resumo **xi**

Acknowledgements **xiii**

Epigraph **xv**

1 Introduction **1**
1.1 Image Composition: a Special Type of Forgeries 2
1.2 Inconsistencies in the Illumination: a Hypothesis 3
1.3 Scientific Contributions 4
1.4 Thesis Structure 6

2 Related Work **7**
2.1 Methods Based on Inconsistencies in the Light Setting 9
2.2 Methods Based on Inconsistencies in Light Color 14
2.3 Methods Based on Inconsistencies in Shadows 15

3 Eye Specular Highlight Telltales for Digital Forensics **19**
3.1 Background 19
3.2 Proposed Approach 23
3.3 Experiments and Results 25
3.4 Final Remarks 28

4 Exposing Digital Image Forgeries by Illumination Color Classification **31**
4.1 Background 31
4.1.1 Related Concepts 31
4.1.2 Related Work 32
4.2 Proposed Approach 34

4.2.1 Challenges in Exploiting Illuminant Maps . . . 34
4.2.2 Methodology Overview . . . 37
4.2.3 Dense Local Illuminant Estimation . . . 37
4.2.4 Face Extraction . . . 41
4.2.5 Texture Description: SASI Algorithm . . . 41
4.2.6 Interpretation of Illuminant Edges: HOGedge Algorithm . . . 44
4.2.7 Face Pair . . . 47
4.2.8 Classification . . . 48
4.3 Experiments and Results . . . 50
4.3.1 Evaluation Data . . . 50
4.3.2 Human Performance in Spliced Image Detection . . . 50
4.3.3 Performance of Forgery Detection using Semi-Automatic Face Annotation in DSO-1 . . . 52
4.3.4 Fully Automated versus Semi-Automatic Face Detection . . . 54
4.3.5 Comparison with State-of-the-Art Methods . . . 55
4.3.6 Detection after Additional Image Processing . . . 57
4.3.7 Performance of Forgery Detection using a Cross-Database Approach 58
4.4 Final Remarks . . . 59

5 Splicing Detection via Illuminant Maps: More than Meets the Eye 61
5.1 Background . . . 61
5.2 Proposed Approach . . . 61
5.2.1 Forgery Detection . . . 62
5.2.2 Description . . . 63
5.2.3 Face Pair Classification . . . 67
5.2.4 Forgery Classification . . . 69
5.2.5 Forgery Detection . . . 69
5.3 Experiments and Results . . . 70
5.3.1 Datasets and Experimental Setup . . . 71
5.3.2 Round #1: Finding the best kNN classifier . . . 72
5.3.3 Round #2: Performance on DSO-1 dataset . . . 73
5.3.4 Round #3: Behavior of the method by increasing the number of IMs 76
5.3.5 Round #4: Forgery detection on DSO-1 dataset . . . 77
5.3.6 Round #5: Performance on DSI-1 dataset . . . 79
5.3.7 Round #6: Qualitative Analysis of Famous Cases involving Questioned Images . . . 80
5.4 Final Remarks . . . 84

6 Exposing Photo Manipulation From User-Guided 3-D Lighting Analysis 87
6.1 Background 87
6.2 Proposed Approach 88
6.2.1 User-Assisted 3-D Shape Estimation 88
6.2.2 3-D Light Estimation 90
6.2.3 Modeling Uncertainty 92
6.2.4 Forgery Detection Process 93
6.3 Experiments and Results 95
6.3.1 Round #1 95
6.3.2 Round #2 96
6.3.3 Round #3 97
6.4 Final Remarks 97

7 Conclusions and Research Directions 101

Bibliography 105

List of Tables

2.1 Literature methods based on illumination inconsistencies. 8

3.1 Equal Error Rate – Four proposed approaches and the original work method by Johnson and Farid [48]. 28

5.1 Different descriptors used in this work. Each table row represents an image descriptor and it is composed of the combination (triplet) of an illuminant map, a color space (onto which IMs have been converted) and description technique used to extract the desired property. 66

5.2 Accuracy computed for kNN technique using different k values and types of image descriptors. Performed experiments using validation set and 5-fold cross-validation protocol have been applied. All results are in %. 73

5.3 Classification results obtained from the methodology described in Section 5.2 with a 5-fold cross-validation protocol for different number of classifiers ($|\mathcal{C}^*|$). All results are in %. 75

5.4 Classification results for the methodology described in Section 5.2 with a 5-fold cross-validation protocol for different number of classifiers ($|\mathcal{C}^*|$) exploring the addition of new illuminant maps to the pipeline. All results are in %. 78

5.5 Accuracy for each color descriptor on fake face detection approach. All results are in %. 79

5.6 Accuracy computed through approach described in Section 5.2 for 5-fold cross-validation protocol in different number of classifiers ($|\mathcal{C}^*|$). All results are in %. 79

7.1 Proposed methods and their respective application scenarios. 103

List of Figures

1.1 *The two ways of life* is a photograph produced by Oscar G. Rejland in 1857 using more than 30 analog photographs. 2
1.2 An example of an image composition creation process. 3
1.3 Doctored and Original images involving former Egyptian president, Hosni Mubarak. Pictures published on BBC (http://www.bbc.co.uk/news/world-middle-east-11313738) and GettyImages (http://www.gettyimages.com). . 4

2.1 Images obtained from [45] depicting the estimated light source direction for each person in the image. 10
2.2 Image composition and their spherical harmonics. Original images obtained from [47]. 11
2.3 Image depicting results from using Kee and Farid's [50] method. Original images obtained from [50]. 13
2.4 Illustration of Kee and Farid's proposed approach [51]. The red regions represent correct constraints. The blue region exposes a forgery since its constraint point is in a region totally different form the other ones. Original images obtained from [51]. 17

3.1 The three stages of Johnson and Farid's approach based on eye specular highlights [48]. 20
3.2 Proposed extension of Johnson and Farid's approach. Light green boxes indicate the introduced extensions. 25
3.3 Examples the images used in the experiments of our first approach. 26
3.4 Comparison of classification results for Johnson and Farid's [48] approach against our approach. 27

4.1 From left to right: an image, its illuminant map and the distance map generated using Riess and Angelopoulou's [73] method. Original images obtained from [73]. 33
4.2 Example of illuminant map that directly shows an inconsistency. 35

4.3 Example of illuminant maps for an original image (a - b) and a spliced image (c - d). The illuminant maps are created with the IIC-based illuminant estimator (see Section 4.2.3). 36
4.4 Overview of the proposed method. 38
4.5 Illustration of the inverse intensity-chromaticity space (blue color channel). (a) depicts synthetic image (violet and green balls) while (b) depicts that specular pixels from (a) converge towards the blue portion of the illuminant color (recovered at the y-axis intercept). Highly specular pixels are shown in red. 40
4.6 An original image and its gray world map. Highlighted regions in the gray world map show a similar appearance. 42
4.7 An example of how different illuminant maps are (in texture aspects) under different light sources. (a) and (d) are two people's faces extracted from the same image. (b) and (e) display their illuminant maps, respectively, and (c) and (f) depicts illuminant maps in grayscale. Regions with same color (red, yellow and green) depict some similarity. On the other hand, (f) depicts the same person (a) in a similar position but extracted from a different image (consequently, illuminated by a different light source). The grayscale illuminant map (h) is quite different from (c) in highlighted regions. 43
4.8 An example of discontinuities generated by different illuminants. The illuminant map (b) has been calculated from splicing image depicted in (a). The person on the left does not show discontinuities in the highlighted regions (green and yellow). On the other hand, the alien part (person on the right) presents discontinuities in the same regions highlighted in the person on the left. 44
4.9 Overview of the proposed HOGedge algorithm. 45
4.10 (a) The gray world IM for the left face in Figure 4.6(b). (b) The result of the Canny edge detector when applied on this IM. (c) The final edge points after filtering using a square region. 46
4.11 Average signatures from original and spliced images. The horizontal axis corresponds to different feature dimensions, while the vertical axis represents the average feature value for different combinations of descriptors and illuminant maps. 49
4.12 Original (left) and spliced images (right) from both databases. 51
4.13 Comparison of different variants of the algorithm using semi-automatic (corner clicking) annotated faces. 53
4.14 Experiments showing the differences for automatic and semi-automatic face detection. 55

4.15 Different types of face location. Automatic and semi-automatic locations select a considerable part of the background, whereas manual location is restricted to face regions. 56
4.16 Comparative results between our method and state-of-the-art approaches performed using DSO-1. 58
4.17 ROC curve provided by cross-database experiment. 59

5.1 Overview of the proposed image forgery classification and detection methodology. 62
5.2 Image description pipeline. Steps *Choice of Color Spaces* and *Features From IMs* can use many different variants which allow us to characterize IMs
ering a wide range of cues and telltales. 67
5.3 Proposed framework for detecting image splicing. 68
5.4 Differences in ICC and GGE illuminant maps. The highlighted regions exemplify how the difference between ICC and GGE is increased on fake images. On the forehead of the person highlighted as pristine (a person that originally was in the picture), the difference between the colors of IIC and GGE, in similar regions, is very small. On the other hand, on the forehead of the person highlighted as fake (an alien introduced into the image), the difference between the colors of IIC and GGE is large (from green to purple). The same thing happens in the cheeks. 70
5.5 Images (a) and (b) depict, respectively, examples of pristine and fake images from DSO-1 dataset, whereas images (c) and (d) depict, respectively, examples of pristine and fake images from DSI-1 dataset. 72
5.6 Comparison between results reported by the approach proposed in this chapter and the approach proposed in Chapter 4 over DSO-1 dataset. Note the proposed method is superior in true positives and true negatives rates, producing an expressive lower rate of false positives and false negatives. . . 74
5.7 Classification histograms created during training of the selection process described in Section 5.2.3 for DSO-1 dataset. 75
5.8 Classification accuracies of all non-complex classifiers (kNN-5) used in our experiments. The blue line shows the actual threshold $\mathcal{T}$ described in Section 5.2 used for selecting the most appropriate classification techniques during training. In green, we highlight the 20 classifiers selected for performing the fusion and creating the final classification engine. 76
5.9 (a) IMs estimated from RWGGE; (b) IMs estimated from White Patch. . . 77

5.10 Comparison between current chapter approach and the one proposed in Chapter 4 over DSI-1 dataset. Current approach is superior in true positives and true negatives rates, producing an expressive lower rate of false positives and false negatives. 80

5.11 Questioned images involving Brazil's former president. (a) depicts the original image, which has been taken by photographer Ricardo Stucker, and (b) depicts the fake one, whereby Rosemary Novoa de Noronha's face (left side) is composed with the image. 81

5.12 The situation room images. (a) depicts the original image released by American government; (b) depicts one among many fake images broadcasted in the Internet. 82

5.13 IMs extracted from Figure 5.12(b). Successive JPEG compressions applied on the image make it almost impossible to detect a forgery by a visual analysis of IMs. 82

5.14 Dimitri de Angelis used Adobe Photoshop to falsify images side by side with celebrities. 83

5.15 IMs extracted from Figure 5.14(b). Successive JPEG compressions applied on the image, allied with a very low resolution, formed large blocks of same illuminant, leading our method to misclassify the image. 84

6.1 A rendered 3-D object with user-specified probes that capture the local 3-D structure. A magnified view of two probes is shown on the top right. . . . 89

6.2 Surface normal obtained using a small circular red probe in a shaded sphere in the image plane. We define a local coordinate system by $\mathbf{b}_1$, $\mathbf{b}_2$, and $\mathbf{b}_3$. The axis $\mathbf{b}_1$ is defined as the ray that connects the base of the probe and the center of projection (CoP). The slant of the 3-D normal $\vec{N}$ is specified by a rotation ϑ around $\mathbf{b}_3$, while the normal's tilt ϱ is implicitly defined by the axes $\mathbf{b}_2$ and $\mathbf{b}_3$, Equation (6.3). 91

6.3 Visualization of slant model for correction of errors constructed from data collected in a psychophysical study provided by Cole et al. [18]. 93

6.4 Visualization of tilt model for correction of errors constructed from data collected in a psychophysical study provided by Cole et al. [18]. 94

6.5 Car Model 95

6.6 Guittar Model 96

6.7 Bourbon Model 96

6.8 Bunny Model. 96

6.9 From left to right and top to bottom, the confidence intervals for the lighting estimate from one through five objects in the same scene, rendered under the same lighting. As expected and desired, this interval becomes smaller as more objects are detected, making it more easier to detect a forgery. Confidence intervals are shown at 60%, 90% (bold), 95% and 99% (bold). The location of the actual light source is noted by a black dot. . . . 98
6.10 Different objects and their respectively light source probability region extracted from a fake image. The light source probability region estimated for the fake object (j) is totally different from the light source probability region provided by the other objects. 99
6.11 (a) result for probability regions intersection from pristine objects and (b) absence of intersection between region from pristine objects and fake object. . . . 100

Chapter 1

Introduction

In a world where technology is improved daily at a remarkable speed, it is easy to face situations previously seen just in science fiction. One example is the use of advanced computational methods to solve crimes, an ordinary situation which usually occurs in TV shows such as the famous *Crime Scene Investigation (CSI)*[1], a crime drama television series. However, technology improvements are, at the same time, a boon and a bane. Although it empowers people to improve their quality of life, it also brings huge drawbacks such as increasing the number of crimes involving digital documents (e.g., images). Such cases have two main support factors: the low cost and easy accessibility of acquisition devices, increasing the number of digital images produced everyday, and the rapid evolution of image manipulation software packages that allow ordinary people to quickly grasp sophisticated concepts and produce excellent masterpieces of falsification.

Image manipulation ranges from simple color adjustment tweaks, which is considered an innocent operation, to the creation of synthetic images to deceive viewers. Images manipulated with the purpose of manipulating and misleading user opinions are present in almost all communication channels including newspapers, magazines, outdoors, TV shows, Internet, and even scientific papers [76]. However, image manipulations are not a product of the digital age. Figure 1.1 depicts a photograph known as *The Two Ways of Life* produced in 1857 using more than 30 analogical photographs[2].

Facts as the aforementioned ones harm our trust in the content of images. Hany Farid[3] defines the impact of image manipulations over people's trust as:

> In a scenario who's became ductile day after day, any manipulation produce uncertainty, no matter how tiny it is, so that confidence is eroded [29].

[1] `http://en.wikipedia.org/wiki/CSI:_Crime_Scene_Investigation`

[2] This and other historic cases of image manipulation are discussed in detail in [13].

[3] `http://www.cs.dartmouth.edu/farid/Hany_Farid/Home.html`

Figure 1.1: *The two ways of life* is a photograph produced by Oscar G. Rejland in 1857 using more than 30 analog photographs.

Trying to rescue this confidence, several researchers have been developing a new research area named *Digital Forensics*. According to Edward Delp[4], Digital Forensics is defined as

> (...) the collection of scientific techniques for the preservation, collection, validation, identification, analysis, interpretation, documentation, and presentation of digital evidence derived from digital sources for the purpose of facilitating or furthering the reconstruction of events, usually of a criminal nature [22].

Digital Forensics mainly targets three kinds of problems: source attribution, synthetic image detection and image composition detection [13, 76].

1.1 Image Composition: a Special Type of Forgeries

Our work focuses on one of the most common types of image manipulations: *splicing* or *composition*. Image splicing consists of using parts of two or more images to compose a new image that never took place in space and time. This composition process includes all the necessary operations (such as brightness and contrast adjustment, affine transformations,

[4] `https://engineering.purdue.edu/~ace/`

color changes, etc.) to construct realist images able to deceive viewer. In this process, normally, we refer to the parts coming from other images as aliens and the image receiving the other parts as host. Figure 1.2 depicts an example of some operations applied to construct a realist composition.

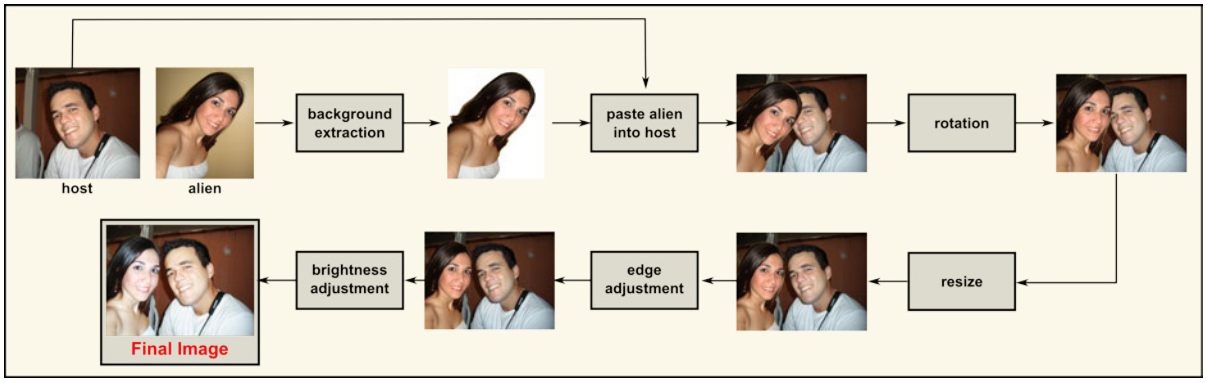


Figure 1.2: An example of an image composition creation process.

Image composition involving people are very popular and are employed with very different objectives. In one of the most recent cases of splicing involving famous people, the conman Dimitri de Angelis photoshoped himself side by side to famous people (e.g., former US president Bill Clinton and Russian president Mikhail Gorbachev). De Angelis has used these pictures to influence and dupe investors, garbing their trust. However, in March 2013 he was sentenced to twelve years in prison because of these frauds.

Another famous composition example dates from 2010 when *Al-Ahram*, a famous Egyptian newspaper, altered a photograph to make its own President Hosni Mubarak look like the host of White House talks over the Israeli-Palestinian conflict as Figure 1.3(a) depicts. However, in the original image, the actual leader of the meeting was US President Barack Obama.

Cases such as this one show how present image composition is in our daily lives. Unfortunately, it also decreases our trust on images and highlights the need for developing methods for recovering back such confidence.

1.2 Inconsistencies in the Illumination: a Hypothesis

Methods for detecting image composition are no longer just in the realm of science fiction. They have become actual and powerful tools in the forensic analysis process. Different types of methods have been proposed for detecting image composition. Methods based on inconsistencies in compatibility metrics [25], JPEG compression features [42] and perspective constraints [94] are just a few examples of inconsistencies explored to detect

(a) Doctored Image (b) Original Image

Figure 1.3: Doctored and Original images involving former Egyptian president, Hosni Mubarak. Pictures published on BBC (http://www.bbc.co.uk/news/world-middle-east-11313738) and GettyImages (http://www.gettyimages.com).

forgeries.

After studying and analyzing the advantages and drawbacks of different types of methods for detecting image composition, our work herein relies on the following research hypothesis

Image illumination inconsistencies are strong and powerful evidence of image composition.

This hypothesis has already been used by some researchers in the literature whose work will be detailed in the next chapter, and it is specially useful for detecting image composition because, even for expert counterfeiters, a perfect illumination match is extremely hard to achieve. Also, there are some experiments that show how difficult is for humans perceive image illumination inconsistencies [68]. Due to this difficulty, all methods proposed herein explore some kind of image illumination inconsistency.

1.3 Scientific Contributions

In a real forensic scenario, there is no silver bullet able to solve all problems once and for all. Experts apply different approaches together to increase confidence on the analysis and avoid missing any trace of tampering. Each one of the proposed methods herein brings with it many scientific contributions from which we highlight:

- Eye Specular Highlight Telltales for Digital Forensics: a Machine Learning Approach [79]:

1. proposition of new features not explored before;
2. use of machine learning approaches (single and multiple classifier combination) for the decision-making process instead of relying on simple and limited hypothesis testings;
3. reduction in the classification error in more than 20% when compared to the prior work.

- Exposing Digital Image Forgeries by Illumination Color Classification [14]:

1. interpretation of the illumination distribution in an image as object texture for feature computation;
2. proposition of a novel edge-based characterization method for illuminant maps which explores edge attributes related to the illumination process;
3. the creation of a benchmark dataset comprising 100 skillfully created forgeries and 100 original photographs;
4. quantitative and qualitative evaluations with users using the Mechanical Turk giving us important insights about the difficulty in detecting forgeries in digital images.

- Splicing Detection through Illuminant Maps: More than Meets the Eye [5]:

1. the exploration of other color spaces for digital forensics not addressed in Chapter 4 and the assessment of their pros and cons;
2. the incorporation of color descriptors, which showed to be very effective when characterizing illuminant maps;
3. a full study of the effectiveness and complementarity of many different image descriptors applied on illuminant maps to detect image illumination inconsistencies;
4. fitting of a machine learning framework for our approach, which automatically selects the best combination of all the factors of interest (e.g., color constancy maps, color space, descriptor, classifier);
5. the introduction of a new approach to detecting the most likely doctored part in fake images;

[5] T. Carvalho, F. Faria, R. Torres, H. Pedrini, and A. Rocha. Splicing detection through color constancy maps: More than meets the eye. Submitted to Elsevier Forensics Science International (FSI), 2014.

6. an evaluation on the impact of the number of color constancy maps and their importance to characterize an image in the composition detection task;
7. an improvement of 15 percentage points in classification accuracy when compared to the state-of-the-art results reported in Chapter 4.

- Exposing Photo Manipulation From User-Guided 3-D Lighting Analysis [6]:

1. the possibility of estimating 3-D lighting properties of a scene from a single 2-D image without knowledge of the 3-D structure of the scene;
2. a study of user's skills on 3-D probes insertion for 3-D estimation of lighting properties in a forensic scenario.

1.4 Thesis Structure

This work is structured so that reader can easily understand each one of our contributions, why they are important for the forensic community, how each piece connects to each other and what are possible drawbacks of each proposed technique.

First and foremost, we organized this thesis as a collection of articles. Chapter 2 describes the main methods grounded on illumination inconsistencies for detecting image composition. Chapter 3 describes our first actual contribution for detecting image composition, which is based on eye specular highlights [79]. Chapter 4 describes our second contribution, result of a fruitful collaboration with researchers of the University of Erlangen-Nuremberg. The work is based on illuminant color characterization [14]. Chapter 4 describes our third contribution, result of a collaboration with researchers of Unicamp and it is an improvement upon our work proposed in Chapter 5. Chapter 6 presents our last contribution, result of a collaboration with researchers of Dartmouth College. This work uses the knowledge of users to estimate full 3-D light source position in images in order to point out possible forgery artifacts. Finally, Chapter 7 concludes our work putting our research in perspective and discussing new research opportunities.

[6] T. Carvalho, H. Farid, and E. Kee. Exposing Photo Manipulation From User-Guided 3-D Lighting Analysis. Submitted to IEEE International Conference on Image Processing (ICIP), 2014.

Chapter 2

Related Work

In Chapter 1, we defined image composition and discussed the importance of devising and developing methods able to detect this kind of forgery. Such methods are based on several kinds of different telltales left in the image during the process of composition and include compatibility metrics [25], JPEG compression features [42] and perspective constraints [94]. However, we are specially interested in methods that explore illumination inconsistencies to detect composition images.

We can divide methods that explore illumination inconsistencies into three main groups of methods:

- **methods based on inconsistencies in the light setting:** this group of methods encloses approaches that look for inconsistencies in the light position and in models that aim at reconstructing the scene illumination conditions. As examples of these methods, it is worth mentioning [45], [46], [48], [47], [64], [50], [16], [77], and [27];

- **methods based on inconsistencies in light color:** this group of methods encloses approaches that look for inconsistencies in the color of illuminants present in the scene. As examples of these methods, it is worth mentioning [33], [93], and [73];

- **methods based on inconsistencies in the shadows:** this group of methods encloses approaches that look for inconsistencies in the scene illumination using telltales derived from shadows. As examples of these methods, it is worth mentioning [95], [61], and [51].

Table 2.1 summarizes the most important literature methods and what they are based upon. The details of these methods will be explained along this chapter.

Table 2.1: Literature methods based on illumination inconsistencies.

Group	**Method**	**Characteristics**
1	Johnson and Farid [45]	Detect inconsistencies in 2-D light source direction estimated from objects occluding contours
1	Johnson and Farid [48]	Detect inconsistencies in 2-D light source direction estimated from eye specular highlights
1	Johnson and Farid [47]	Detect inconsistencies in 3-D light environment estimated using five first spherical harmonics
1	Yingda et al. [64]	Detect inconsistencies in 2-D light source direction which is estimated using surface normals calculated from each pixel in the image
1	Haipeng et al. [16]	Detect inconsistencies in 2-D light source direction using Hestenes-Powell multiplier method for calculating the light source direction
1	Kee and Farid [50]	Detect inconsistencies in 3-D light environment estimated from faces using nine spherical harmonics
1	Roy et al. [77]	Detect the difference in 2-D light source incident angle
1	Fan et al. [27]	Detect inconsistencies in 3-D light environment using a shaping from shading approach
2	Gholap and Bora [33]	Investigate illuminant colors estimating dichromatic planes from each specular highlight region of an image to detect inconsistencies and image forgeries
2	Riess and Angelopoulou [73]	Estimate illuminants locally from different parts of an image using an extension of the Inverse-Intensity Chromaticity Space to detect forgeries
2	Wu and Fang [93]	Estimate illuminant colors from overlapping blocks using reference blocks to detect forgeries
3	Zhang and Wang [95]	Uses planar homology to model the relationship of shadows in an image and discovering forgeries
3	Qiguang et al. [61]	Explores shadow photometric consistencies to detect image forgeries
3	Kee and Farid [51]	Constructs geometric constraints from shadows to detect forgeries

2.1 Methods Based on Inconsistencies in the Light Setting

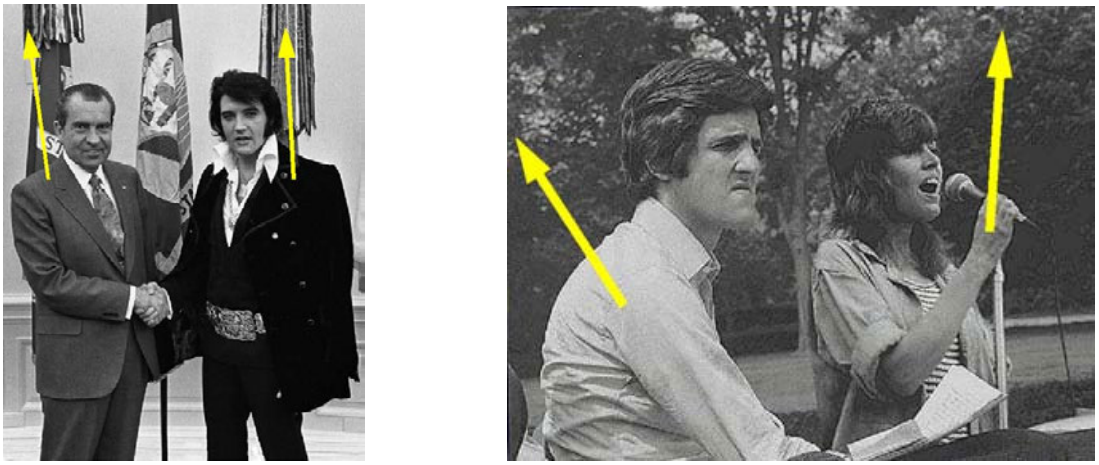
Johnson and Farid [45] proposed an approach based on illumination inconsistencies. They analyzed the light source direction from different objects in the same image trying to detect traces of tampering. The authors start by imposing different constraints for the problem:

1. all the analyzed objects have Lambertian surface;
2. surface reflectance is constant;
3. the object surface is illuminated by an infinitely distant light source.

Even using such restrictions, to estimate the light source position from any object in the image, it is necessary to have 3-D normals from, at least, four distinct points in the object. From one image and with objects of arbitrary geometry, this is a very hard task. To circumvent this geometry problem, the authors use a specific solution proposed by Nillius e Eklundh [66], which allows the estimation of two components of normals in the object occluding contour. Then, the authors can estimate the 2-D light source position for different objects in the same image and compare them. If the difference in the estimated light source position of different objects is larger than a threshold, the investigated image shows traces of tampering.

Figure 2.1 depicts an example of Johnson and Farid's method [45]. In spite of being a promising advance for detecting image tampering, this method still presents some problems, such as the inherently ambiguity of just estimating 2-D light source positions, fact that can confuse even an expert analyst. Another drawback relies on the limited applicability of the techniques, given that it targets only at outdoor images.

In another work, Johnson and Farid [46] explore chromaticity aberrations as an indicative of image forgery. Chromaticity deviation is the name of a process whereby a polychromatic ray of light splits into different light rays (according to their wavelength) when reaching the camera lenses. Using RGB images, the authors assume that the chromaticity deviation is constant (and dependent on each channel wavelength) for all color channels and create a model, based on image statistical properties, of how the ray light should split for each color channel. Given this premise and using the green channel as reference, the authors estimate deviations between the red and green channels and between the blue and green channels for selected parts (patches) of the image. Inconsistencies on this split pattern are used as telltales to detect forgeries. A drawback of this method is that chromaticity deviation depends on the camera lens used to take the picture. Therefore, image compositions created using images from the same camera can not depict necessary inconsistencies to detect forgeries with this approach.



(a) Original Image (b) Tampered Image

Figure 2.1: Images obtained from [45] depicting the estimated light source direction for each person in the image.

Johnson and Farid [48] also explored eye specular highlights for estimating light source position and detecting forgeries. This work is the foundation upon which we build our first proposed method and it will better explained in Chapter 3.

Johnson and Farid [47] detect traces of image tampering on complex light environments. For that, the authors assumed an infinity light source and Lambertian and convex surfaces. The authors modeled the problem assuming that object reflectance is constant and that the camera response function is linear. All these constraints allow the authors to represent irradiance $\mathbf{B}(\vec{N})$ parameterized by a surface normal vector (with unit length) $\vec{N}$ as a convolution between the reflectance surface function $\mathbf{R}(\vec{L},\vec{N})$ with lighting environment $W(\vec{L})$

$$\begin{aligned} \mathbf{B}(\vec{N}) &= \textstyle\int_{\Omega} \mathbf{W}(\vec{L})\mathbf{R}(\vec{L},\vec{N})d\Omega, \\ \mathbf{R}(\vec{L},\vec{N}) &= \max(\vec{L}\cdot\vec{N},0) \end{aligned} \tag{2.1}$$

where $\mathbf{W}(\vec{L})$ refers to the light intensity at light source incident direction $\vec{L}$, Ω is the surface of the sphere and $d\Omega$ is an area differential on the sphere.

Spherical harmonics define an orthonormal basis system over any sphere, similar to the Fourier Transform over a 1-D circle [82]. Therefore, Equation 2.1 can be rewritten[1] in function of the three first-order spherical harmonics (nine first terms)

$$\mathbf{B}(\vec{N}) \approx \sum_{n=0}^{2}\sum_{m=-n}^{n} \hat{r}_n l_{n,m} Y_{n,m}(\vec{N}) \tag{2.2}$$

[1]We refer the reader to the original paper for a more complete explanation [47].

where $\hat{r}_n$ are constants of Lambertian function in points of the analyzed surface, $Y_{n,m}(\cdot)$ is the mth spherical harmonic with order n, and $l_{n,m}$ is the ambient light coefficient from the mth spherical harmonic with order n. Given the difficulty of estimating 3-D normals from 2-D images, the authors assumed an image under orthographic projection, allowing estimations of normals along occluding contour (as done in [45], in an occluding contour, the z component from the normal surface is equal to zero). This problem simplification allows to represent Equation 2.2 using just five coefficients (spherical harmonics), which is enough in a forensic analysis. These coefficients compose an illumination vector and, given illumination vectors from two different objects in the same scene, they can be analyzed using correlation metrics. Figures 2.2(a-b) illustrate, respectively, an image generated by a process of composition and its spherical harmonics (obtained from three different objects)

(a) Composition image (b) Spherical harmonics from objects

Figure 2.2: Image composition and their spherical harmonics. Original images obtained from [47].

As drawbacks, these methods do not deal with images with extensive shadow regions and also only use simple correlation to compare different illumination representations.

Also exploring inconsistencies in light source position, Yingda et al. [64] proposed an approach similar to the one previously proposed by Johnson and Farid [45]. However, instead of using just surface normals on occluding contours, they proposed a simple method for estimating the surface normal for each pixel. Using a neighborhood of eight pixels around interest pixels, the pixel with highest intensity at this neighborhood is considered as the direction of the 2-D surface normal. Then, the authors divide the image into k blocks (assuming a diffuse reflectivity of unit value for each block) and model an error function and minimize such function via least square to estimate the light source position.

Different from Johnson and Farid [45], which just estimate light source for an infinity far away light source, this approach also deals with local light source positions. As Johnson

and Farid's [45] work, this approach also has, as main drawback, the ambiguity introduced by the estimation of 2-D light source position, which can lead to wrong conclusions about the analyzed image. Furthermore, the simple use of more intensity pixel in a neighborhood to determine the normal direction is a rough approach to a scenario where small details have important meaning.

Extending upon the work of Yingda et al. [64], Haipeng et al. [16] proposed a small modification of original method. The authors proposed to replace the least square minimization method by the Hestenes-Powell multiplier method for calculating the light source direction in the infinity far away light source scenario. This allowed the authors to estimate the light source direction of objects in the scene and their background. Finally, instead of comparing light source direction estimated from two or more objects in the same scene, the method detects inconsistencies comparing light source direction of object against light source direction of the object background. Since this method is essentially the same method presented by Yingda et al. [64] (with just a slight difference on the minimization method), it has the same previously mentioned drawbacksYingda et al. [64].

In a new approach, also using light direction estimation, Kee and Farid [50] specialized the approach proposed in [47] to deal with images containing people. Using a 3-D morphable model proposed in [9] to synthesize human faces, the authors generated 3-D faces using a linear combination of basic image faces. With this 3-D model, the authors circumvent the difficulties presented in [47], where just five spherical harmonics have been estimated. Once a 3-D model is created, it is registered with the image under investigation maximizing an objective function composed by intrinsic and extrinsic camera parameters. It maximizes the correlation between the image $I(\cdot)$ and the rendered model $I_m(\cdot)$.

$$E(\mathcal{R}, \vec{t}, f, c_x, c_y) = I(x, y) * I_m(x, y) \tag{2.3}$$

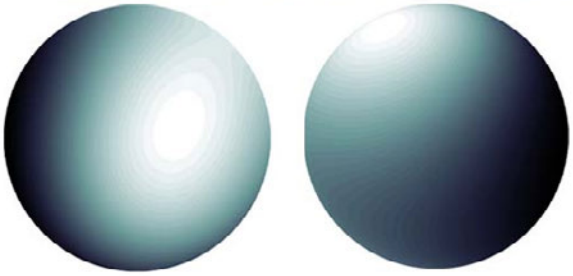
where $\mathbf{R}$ is the rotation matrix, $\vec{t}$ is a translation vector, f is the focal length and $\rho = (c_x, c_y)$ are the coordinates of the principal point ρ.

Figures 2.3(a-c) depict, respectively, the face 3-D model created using two images, the analyzed image and the resulting harmonics obtained from Figure 2.3(b). The major drawback of this method is the strong user dependence, fact that sometimes introduces failures in the analysis.

Roy et al. [77] proposed to identify image forgeries detecting the difference in light source incident angle. For that, the authors smooth the image noise using a max filter before applying a decorrelation stretch algorithm. To extract the shading (intensity) profile, the authors use the R channel from the resulting RGB improved image. Once the shading profile is estimated, the authors estimate the structural profile information using localized histogram equalization [39].

(a) 3-D face models generated from two images of the same person

(b) Composition Image



(c) Resulting Harmonics

Figure 2.3: Image depicting results from using Kee and Farid's [50] method. Original images obtained from [50].

From the localized histogram image, the authors select small blocks, which need to contain transitions from illumination to shadow, from interest objects. For each one of these blocks, the authors determine an interest point using a set of three points around it to estimate its normal. Finally, a joint intensity profile information (point intensity) and shading profile information (surface normal of the interest point) the authors are able to estimate the light source direction, which is used to detect forgeries – comparing directions provided by two different blocks it is possible to detect inconsistencies. The major problem of this method is its strongly dependence on image processing operations (as noise reduction and decorrelation stretching) since simple operations as a JPEG compression can destruct existent relations among pixel values.

Fan et al. [27] introduced two counter-forensic methods for showing how vulnerable lighting forensic methods based on 2-D information can be. More specifically, the authors presented two counter-forensic methods against Johnson and Farid's [47] method. Nevertheless, they proposed to explore the shape from shading algorithm to detect forgeries in

3-D complex light environment. The first counter-forensic method relies on the fact that methods for 2-D lighting estimation forgeries rely upon just on occluding contour regions. So, if a fake image is created and the pixel values along occluding contours of the fake part are modified so that to keep the same order from original part, methods relying on the 2-D information on occluding contours can be deceived.

The second counter-forensic method explores the weakness of spherical harmonics relationship. According to the authors, the method proposed by Johnson and Farid [47] also fails when a composition is created using parts of images with similar spherical harmonics. The reason is that the method is just able to detect five spherical harmonics and there are images where the detected and kept harmonics are similar, but the discarded ones are different. Both counter-forensic methods have been tested and their effectiveness have been proved. Finally, as a third contribution, the authors proposed to use a shape from shading approach, as proposed by Huang and Smith [43], to estimate 3-D surface normals (with unit length). Once that 3-D normals are available, the users can now estimate the nine spherical harmonics without a scenario restriction. Despite being a promising approach, the method presents some drawbacks. First, the applicability scenario is constrained to just outdoor images with an infinity light source; second, the normals are estimated by a minimization process which can introduce serious errors in the light source estimation; finally, the the method was tested only on simple objects.

2.2 Methods Based on Inconsistencies in Light Color

Continuing to investigate illumination inconsistencies, but now using different clues, Gholap and Bora [33] pioneered the use illuminant colors to investigate the presence, or not, of composition operations in digital images. For that, the authors used a dichromatic reflection model proposed by Tominaga and Wandell [89], which assumes a single light source to estimate illuminant colors from images. Dichromatic planes are estimated using principal component analysis (PCA) from each specular highlight region of an image. By applying a Singular Value Decomposition (SVD) on the RGB matrix extracted from highlighted regions, the authors extract the eigenvectors associated with the two most significant eigenvalues to construct the dichromatic plane. This plane is then mapped onto a straight line, named *dichromatic line*, in normalized r-g-chromaticity space. For distinct objects illuminated by the same light source, the intersection point produced by their dichromatic line intersection represents the illuminant color. If the image has more than one illuminant, it will present more than one intersection point, which is not expected to happen in pristine (non-forged images). This method represented the first important step toward forgery detection using illuminant colors, but has some limitations such as the need of well defined specular highlight regions for estimating the illuminants.

Following Gholap and Bora's work [33], Riess and Angelopoulou [73] used an extension of the Inverse-Intensity Chromaticity Space, originally proposed by Tan et al. [86], to estimate illuminants locally from different parts of an image for detecting forgeries. This work is the foundation upon which we build our second proposed method and it will better explained in Chapter 4.

Wu and Fang [93] proposed a new way to detect forgeries using illuminant colors. Their method divides a color image into overlapping blocks estimating the illuminant color for each block. To estimate the illuminant color, the authors proposed to use the algorithms Gray-World, Gray-Shadow and Gray-Edge [91], which are based on low-level image features and can be modeled as

$$\mathbf{e}(\mathbf{n}, \mathbf{p}, \sigma) = \frac{1}{\lambda} \left(\int \int \left| \nabla^{\mathbf{n}} f_\sigma (x, y) \right|^{\mathbf{p}} dxdy \right)^{\frac{1}{\mathbf{p}}} \tag{2.4}$$

where λ is a scale factor, $\mathbf{e}$ is the color of the illuminant, $\mathbf{n}$ is the order of derivative, $\mathbf{p}$ is Minkowski norm, σ is the scale parameter of a Gaussian filter. To estimate illuminants using the algorithms Gray-Shadow, first-order Gray-Edge and second-order Gray-Edge, the authors just use $\mathbf{e}(0,\mathbf{p},0)$, $\mathbf{e}(1,\mathbf{p},\sigma)$ and $\mathbf{e}(2,\mathbf{p},\sigma)$ respectively. Then, the authors use a maximum likelihood classifier proposed by Gijsenij and Gevers [34] to select the most appropriate method to represent each block. To detect forgeries, the authors choose some blocks as reference and estimate their illuminants. Afterwards, the angular error between reference blocks and a suspicious block is calculated. If this distance is greater than a threshold, this block is labeled as manipulated. This method is also strongly dependent on user's inputs. In addition, if the reference blocks are incorrectly chosen, for example, the performance of the method is strongly compromised.

2.3 Methods Based on Inconsistencies in Shadows

We have so far seen how light source position and light source color can be used for detecting image forgeries. We now turn our attention to the last group of methods based on illumination inconsistencies. Precisely, this section presents methods relying on shadow inconsistencies for detecting image forgeries.

Zhang and Wang [95] proposed an approach that utilizes the planar homology [83], which models the relationship of shadows in an image for discovering forgeries. Based on this model, the authors proposed to construct two geometric constraints: the first one is based on the relationship of connecting lines. A connecting line is a line that connects some object point with its shadow. According to planar homology, all of these connecting lines intersect in a vanishing point. The second constraint is based on the ratio of these connecting lines. In addition, the authors also proposed to explore the changing ratio

along the normal direction of the shadow boundaries (extracted from shading images [4]). Geometric and shadow photometric constraints together are used to detect image compositions. However, in spite of being a good initial step in forensic shadow analysis, the major drawback of the method is that it only works with images containing casting shadows, a very restricted scenario.

Qiguang et al. [61] also explored shadow photometric consistencies to detect image forgeries. The authors proposed to estimate the shadow matte value along shadow boundaries and use this value to detect forgeries. However, different from Zhang and Wang [95], to estimate the shadow matte value they analyze shadowed and non-shadowed regions, adapting a thin plate model [23] for their problem. Estimating two intensity surfaces, the shadowed surface ($f_s(x, y)$), which reflects intensity of shadow pixels, and non-shadowed surface ($f_n(x, y)$), which reflects the intensity of pixels without shadows, the authors define the shadow matte value as as

$$\mathbf{C} = mean\{f_n(x, y) - f_s(x, y)\} \tag{2.5}$$

Once $\mathbf{C}$ is defined, the authors estimate, for each color channel, an inconsistency vector $\mathbf{D}$ as

$$\mathbf{D} = \exp(\lambda) \cdot (\exp(-\mathbf{C}(x)) - \exp(-\mathbf{C}(y))) \tag{2.6}$$

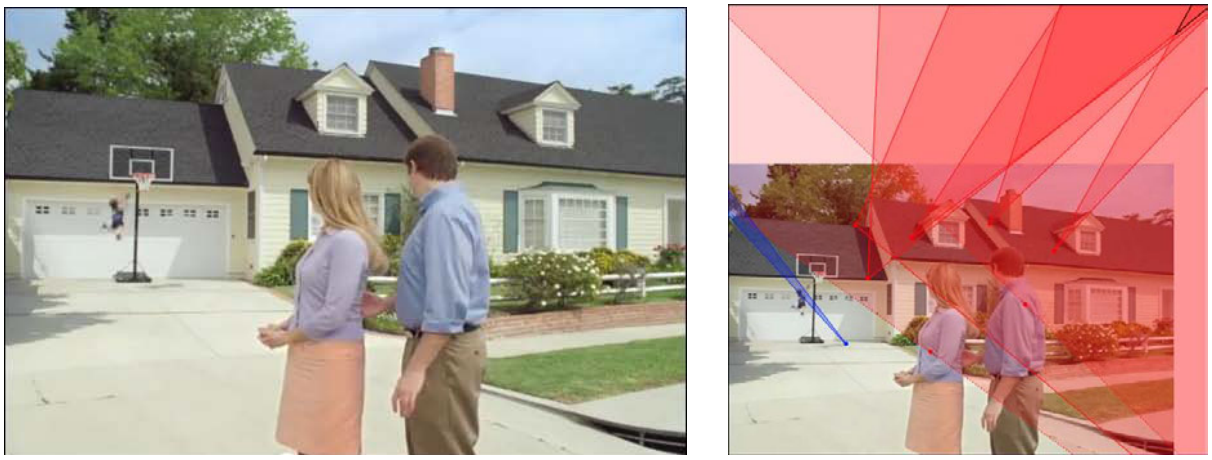
where λ is a scale factor. Finally, inconsistencies are identified measuring the error to satisfy a Gaussian distribution with the density function

$$\varphi(D) = \frac{1}{\sqrt{2\pi}} e^{\frac{-D^2}{2}} \tag{2.7}$$

In spite of its simplicity, this method represents a step forward in image forensic analysis. However, a counter forensic technique targeting this method could use an improved shadow boundary adjustment to compromise its accuracy.

In one of most recent approaches using inconsistencies in shadows to detecting image forgeries, Kee and Farid [51] used shadows constraints to detect forgeries. The authors used cast and attached shadows to estimate the light source position. According to the authors, a constraint provided by cast shadows is constructed connecting a point in shadow to points on an object that may have cast shadows. On the other hand, attached shadows refer to the shadow region generated when objects occlude the light from themselves. In this case, constraints are specified by half planes. Once both kinds of shadows are selected by a user, the algorithm estimates, for each selected shadow, a possible region for light source. Intersecting constraints from different selected shadows help constraining the problem and solving the ambiguity of the 2-D light position estimation. However, if some image part violates these constraints, the image is considered as a composition. Figure 2.4 illustrates an example of this algorithm result. Unfortunately, the process of

including shadow constraints needs high expertise and may often lead the user to a wrong analysis. Also, since the light source position is estimated just on the 2-D plane, as in other methods, this one also can still present some ambiguous results.



(a) Image (b) Shadow Constraints

Figure 2.4: Illustration of Kee and Farid's proposed approach [51]. The red regions represent correct constraints. The blue region exposes a forgery since its constraint point is in a region totally different form the other ones. Original images obtained from [51].

Along this chapter, we presented different methods for detecting image composition. All of them are based on different cues of illumination inconsistencies. However, there is no perfect method or silver bullet to solve all the problems and the forensic community is always looking for new approaches able to circumvent drawbacks and limitations of previous methods. Thinking about this, in the next chapter, we introduce our first, out of four, approach to detecting image composition.

Chapter 3

Eye Specular Highlight Telltales for Digital Forensics

As we presented in Chapter 2, several approaches explore illumination inconsistencies as telltales to detect image composition. As such, research on new telltales has received special attention from the forensic community, making the forgery process more difficult for the counterfeiters. In this chapter, we introduce a new method for pinpointing image telltales in eye specular highlights to detect forgeries. Parts of the contents and findings in this chapter were published in [79].

3.1 Background

The method proposed by Johnson and Farid [48] is based on the fact that the position of a specular highlight is determined by the relative positions of the light source, the reflective surface of the eye, and the viewer (i.e., the camera). Roughly speaking, the method can be divided into three stages, as Figure 3.1 depicts.

The first stage consists of estimating the direction of the light source for each eye in the picture. The authors assume that the eyes are perfect reflectors and use the law of reflection:

$$\vec{L} = 2(\vec{V}^T \vec{N})\vec{N} - \vec{V}, \tag{3.1}$$

where the 3-D vectors $\vec{L}$, $\vec{N}$ and $\vec{V}$ correspond to light source direction, the surface normal at the highlight, and the direction in which the highlight will be seen. Therefore, the light source direction $\vec{L}$ can be estimated from the surface normal $\vec{N}$ and viewer direction $\vec{V}$ at a specular highlight. However, it is difficult to estimate these two vectors in the 3-D space from a single 2-D image.

In order to circumvent this difficulty, it is possible to estimate a transformation ma-

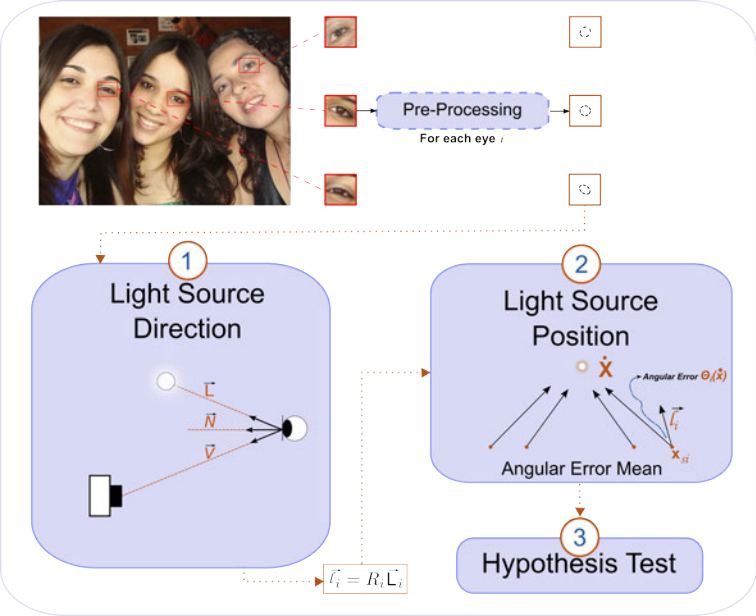


Figure 3.1: The three stages of Johnson and Farid's approach based on eye specular highlights [48].

trix $\mathcal{H}$ that maps 3-D world coordinates onto 2-D image coordinates by making some simplifying assumptions such as:

1. the limbus (the boundary between the sclera and iris) is modelled as a circle in the 3-D world system and as an ellipse in the 2-D image system;
2. the distortion of the ellipse with respect to the circle is related to the pose and position of the eye relative to the camera;
3. and points on a limbus are coplanar.

With these assumptions, $\mathcal{H}$ becomes a 3×3 planar projective transform, in which the world points $\mathbf{X}$ and image points $\mathbf{x}$ are represented by 2-D homogeneous vectors, $\mathbf{x} = \mathcal{H}\mathbf{X}$.

Then, to estimate the matrix $\mathcal{H}$ as well as the circle center point $C = (C_1, C_2, 1)^T$, and radius r (recall that C and r represent the limbus in world coordinates) first the authors define the error function:

$$E(\boldsymbol{P}; \mathcal{H}) = \sum_{i=1}^{m} \min_{\hat{\mathbf{X}}} \|\mathbf{x}_i - \mathcal{H}\hat{\mathbf{X}}_i\|^2, \tag{3.2}$$

where $\hat{\mathbf{X}}$ is on the circle parameterized by $\mathbf{P} = (C_1, C_2, r)^T$, and m is the total number of data points in the image system.

Once defined the error function, it encloses the sum of the squared errors between each data point, $\mathbf{x}$, and the closest point on the 3D model, $\hat{\mathbf{X}}$. So, they use an iterative and

non-linear least squares function, such as Levenberg-Marquardt iteration method [78], to solve it.

In the case when the focal length f is known, they decompose $\mathcal{H}$ in function of intrinsic and extrinsic camera parameters [40] as

$$\mathcal{H} = \lambda \mathcal{K} \begin{pmatrix} \vec{r_1} & \vec{r_2} & \vec{t} \end{pmatrix} \tag{3.3}$$

where λ defines a scale factor, $\vec{r_1}$ is a column vector that represents the first column of the rotation matrix $\mathcal{R}$, $\vec{r_2}$ is a column vector that represents the second column of rotation matrix $\mathcal{R}$, $\vec{t}$ is a column vector which represents translation and the intrinsic matrix K is

$$\mathcal{K} = \begin{pmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{3.4}$$

The next step estimates the matrix $\hat{\mathcal{H}}$ and $\mathcal{R}$, representing, respectively, the transformation of the world system in the camera system and the rotation between them, decomposing $\mathcal{H}$. $\hat{\mathcal{H}}$ is directly estimated from Equation 3.3, choosing λ such that $\mathbf{r_1}$ and $\mathbf{r_2}$ are unit vectors

$$\mathcal{H} = \lambda \mathcal{K} \begin{pmatrix} \vec{r_1} & \vec{r_2} & \vec{t} \end{pmatrix}, \tag{3.5}$$

$$\frac{1}{\lambda} \mathcal{K}^{-1} \mathcal{H} = \begin{pmatrix} \vec{r_1} & \vec{r_2} & \vec{t} \end{pmatrix}, \tag{3.6}$$

$$\hat{\mathcal{H}} = \begin{pmatrix} \vec{r_1} & \vec{r_2} & \vec{t} \end{pmatrix} \tag{3.7}$$

$\mathcal{R}$ can also be easily estimated from $\mathcal{H}$ as

$$\mathcal{R} = (\vec{r_1} \quad \vec{r_2} \quad \vec{r_1} \times \vec{r_2}) \tag{3.8}$$

where $\mathbf{r_1} \times \mathbf{r_2}$ is the cross product between $\mathbf{r_1}$ and $\mathbf{r_2}$.

However, in a real forensic scenario, many times the image focal length cannot be available (making impossible to estimate $\mathcal{K}$ and consequently $\hat{\mathcal{H}}$). To overpass this possible problem, the authors rely on the fact that the transformation matrix $\mathcal{H}$ is composed of eight unknowns: λ, f, the rotation angles to compose $\mathcal{R}$ matrix (θ_x, θ_y, θ_z), and the translation vector $\vec{t}$ (which has three components t_x, t_y, t_z). Using these unknowns, Equation 3.3 can be rewritten as

$$\mathcal{H} = \lambda \begin{pmatrix} f\cos(\theta_y)\cos(\theta_z) & f\cos(\theta_y)\sin(\theta_z) & ft_x \\ f(\sin(\theta_x)\sin(\theta_y)\cos(\theta_z) - \cos(\theta_x)\sin(\theta_z)) & f(\sin(\theta_x)\sin(\theta_y)\sin(\theta_z) + \cos(\theta_x)\cos(\theta_z)) & ft_y \\ \cos(\theta_x)\sin(\theta_y)\cos(\theta_z) + \sin(\theta_x)\sin(\theta_z) & \cos(\theta_x)\sin(\theta_y)\sin(\theta_z) - \sin(\theta_x)\cos(\theta_z) & t_z \end{pmatrix} \tag{3.9}$$

Then, taking the left upper corner matrix (2×2 matrix) from Equation 3.9 and using a non-linear least-squares approach, the following equation is minimized

$$\begin{aligned} E(\theta_x, \theta_y, \theta_z, \hat{f}) \quad = \quad & (\hat{f}\cos(\theta_y)\cos(\theta_z) - h_1)^2 + (\hat{f}\cos(\theta_y)\sin(\theta_z) - h_2)^2 \\ & + (\hat{f}(\sin(\theta_x)\sin(\theta_y)\cos(\theta_z) - \cos(\theta_x)\sin(\theta_z)) - h_4)^2 \\ & + (\hat{f}(\sin(\theta_x)\sin(\theta_y)\sin(\theta_z) + \cos(\theta_x)\cos(\theta_z)) - h_5)^2 \end{aligned} \tag{3.10}$$

where h_i is the ith entry of $\mathcal{H}$ in the Equation 3.9 and $\hat{f} = \lambda f$. Focal length is then estimated as

$$f = \frac{h_7^2 f_1 + h_8^2 f_2}{h_7^2 + h_8^2} \tag{3.11}$$

where

$$f_1 = \frac{\hat{f}(\cos(\theta_x)\sin(\theta_y)\cos(\theta_z) + \sin(\theta_x)\sin(\theta_z))}{h_7} \qquad f_2 = \frac{\hat{f}(\cos(\theta_x)\sin(\theta_y)\sin(\theta_z) - \sin(\theta_x)\cos(\theta_z))}{h_8} \tag{3.12}$$

Now, the camera direction $\vec{V}$ and the surface normal $\vec{N}$ can be calculated in the world coordinate system. $\vec{V}$ is $\mathcal{R}^{-1}\vec{v}$, where $\vec{v}$ represents the direction of the camera in the camera system, and it can be calculated by

$$\vec{v} = -\frac{\mathbf{x}_c}{\|\mathbf{x}_c\|} \tag{3.13}$$

where $\mathbf{x}_c = \hat{\mathcal{H}}\mathbf{X}_C$ and $\mathbf{X}_C = \{C_1 \quad C_2 \quad 1\}$.

To estimate $\vec{N}$, first we need to define $\vec{S} = \{S_x, S_y\}$, which represents the specular highlight in the world coordinates, measured with respect to the center of the limbus in human eye model. $\vec{S}$ is estimated as

$$\vec{S} = \frac{p}{r}(\mathbf{X}_s - \mathbf{P}), \tag{3.14}$$

where p is a constant obtained from 3-D model of human eye, r is the previously defined radius of the limbus in world coordinates, $\mathbf{P}$ is the previously defined parametrized circle in world coordinates which matches with limbus) and $\mathbf{X}_s$ is

$$\mathbf{X}_s = \mathcal{H}^{-1}\mathbf{x}_s, \tag{3.15}$$

with $\mathbf{x}_s$ representing 2-D position of specular highlights in image coordinates.

Then, the surface normal $\vec{N}$ at a specular highlight is computed in world coordinates as

$$\vec{N} = \begin{pmatrix} S_x & + & kV_x \\ S_y & + & kV_z \\ q & + & kV_z \end{pmatrix} \tag{3.16}$$

where $\vec{V} = \{V_x, V_y, V_z\}$, q is a constant obtained from 3-D model of human eye and k is obtained by solving the following quadratic system

$$k^2 + 2(S_x V_x + S_y V_y + q V_z) + (S_x^2 + S_y^2 + q^2) = 0, \tag{3.17}$$

The same surface normal in camera coordinates is $\vec{n} = \mathcal{R}\vec{N}$.

Finally, the first stage of the method in [48] is completed by calculating the light source direction $\vec{L}$ by replacing $\vec{V}$ and $\vec{N}$ in Equation 3.1. In order to compare light source estimates in the image system, the light source estimate is converted to camera coordinates: $\vec{l} = \mathcal{R}\vec{L}$.

The second stage is based on the assumption that all estimated directions $\vec{l_i}$ converge toward the position of the actual light source in the scene, where $i = 1, ..., n$ and n is the number of specular highlights in the picture. This position can be estimated by minimizing the error function

$$E(\dot{\mathbf{x}}) = \sum_{i=1}^{n} \Theta_i(\dot{\mathbf{x}}), \tag{3.18}$$

where $\Theta_i(\dot{\mathbf{x}})$ represents the angle between the position of actual light source in the scene ($\dot{\mathbf{x}}$) and the estimated light source direction $\vec{l_i}$, at the ith specular highlight ($\mathbf{x}_{si}$). Additionally, $\Theta_i(\dot{\mathbf{x}})$ is given by

$$\Theta_i(\dot{\mathbf{x}}) = \arccos\left(\vec{l_i}^T \frac{\dot{\mathbf{x}} - \mathbf{x}_{si}}{||\dot{\mathbf{x}} - \mathbf{x}_{si}||}\right). \tag{3.19}$$

In the third and last stage of Johnson and Farid's [48] method, the authors verify image consistency in the forensic scenario. For an image that has undergone composition, it is expected that the angular errors between eye specular highlights are higher than in pristine images. Based on this statement, the authors use a classical hypothesis test with 1% significance level over the average angular error to identify whether or not the image under investigation is the result of a composition.

The authors tested their technique for estimating the 3-D light direction on synthetic images of eyes that were rendered using the `PBRT` environment [71] and with a few real images. To come out with a decision for a given image, the authors determine whether the specular highlights in an image are inconsistent considering only the average angular error and a classical hypothesis test which might be rather limiting in a real forensic scenario.

3.2 Proposed Approach

In this section, we propose some extensions to Johnson and Farid [48]'s approach by using more discriminative features in the problem characterization stage and a supervised machine learning classifier in the decision stage.

We make the important observation that in the forensic scenario, beyond the angular error average, there are other important characteristics that could also be taken into account in the decision-making stage in order to improve the quality of any eye-highlight-based detector.

Therefore, we first decide to take into account the standard deviation of angular errors ($\Theta_i(\dot{\mathbf{x}})$), given that even in the original images there is a non-null standard deviation. This is due to the successive minimization of functions and simplification of the problem, adopted in the previous steps.

Another key feature is related to the position of the viewer (the device that captured the image). In a pristine image (one that is not a result of a composition) the camera position must be the same for all persons in the photograph, i.e., the estimated directions $\vec{v}$ must converge to a single camera position.

To find the camera position and take it into account, we minimize the following function

$$E(\ddot{\mathbf{x}}) = \sum_{i=1}^{n} \Theta_i(\ddot{\mathbf{x}}), \tag{3.20}$$

where $\Theta_i(\ddot{\mathbf{x}})$ represents the angle between the estimated direction of the viewer $\boldsymbol{v}_i$, and the direction of the actual viewer in the scene, at the ith specular highlight $\mathbf{x}_{si}$

$$\Theta_i(\ddot{\mathbf{x}}) = \arccos\left(\vec{v}_i^T \frac{\ddot{\mathbf{x}} - \mathbf{x}_{si}}{||\ddot{\mathbf{x}} - \mathbf{x}_{si}||}\right). \tag{3.21}$$

Considering $\ddot{\mathbf{x}}$ to be the actual viewer position obtained with Eq. 3.20, the angular error at the ith specular highlight is $\Theta_i(\ddot{\mathbf{x}})$. In order to use this information in the decision-making stage, we can average all the available angular errors. In this case, it is also important to analyze the standard deviation of angular errors $\Theta_i(\ddot{\mathbf{x}})$.

Our extended approach now comprises four characteristics of the image instead of just one as the prior work we rely upon:

1. **LME:** mean of the angular errors $\Theta_i(\dot{\mathbf{x}})$, related to the light source $\vec{l}$;
2. **LSE:** standard deviation of the angular errors $\Theta_i(\dot{\mathbf{x}})$, related to the light source $\vec{l}$;
3. **VME:** mean of the angular errors $\Theta_i(\ddot{\mathbf{x}})$, related to the viewer $\vec{v}$;
4. **VSE:** standard desviation of the angular errors $\Theta_i(\ddot{\mathbf{x}})$, related to the viewer $\vec{v}$.

In order to set forth the standards for a more general and easy to extend smart detector, instead of using a simple hypothesis testing in the decision stage, we turn to a supervised machine learning scenario in which we feed a Support Vector Machine classifier (SVM) or a combination of those with the calculated features. Figure 3.2 depicts an overview of our proposed extensions.

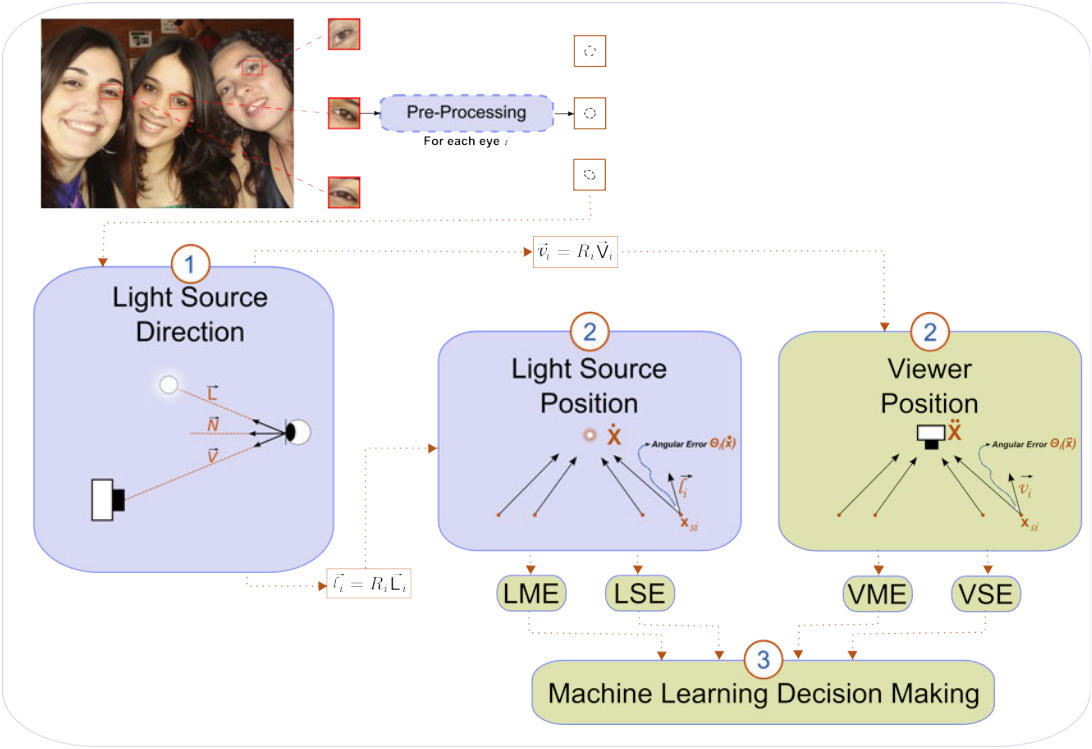


Figure 3.2: Proposed extension of Johnson and Farid's approach. Light green boxes indicate the introduced extensions.

3.3 Experiments and Results

Although the method proposed by Johnson and Farid in [48] has a great potential, the authors have validated their approach by using mainly PBRT synthetic images which is rather limiting. In contrast, in our approach we perform our experiments using a data set comprising everyday pictures typically with more than three megapixels in resolution.

We acquired 120 images from daily scenes. From that, 60 images are normal (without any tampering or processing) and the other 60 images contain different manipulations. To create the manipulated images, we chose a host image and, from another image (alien), we selected an arbitrary face, pasting the alien part in the host. Since the method just analyzes the eyes, no additional fine adjustments were performed. Also, all of the images have more than three mega-pixels given that we need a clear view of the eye. This process guarantees that all the images depict two or more people with visible eyes, as Figure 3.3 illustrates.

The experiment pipeline starts with the limbus point extraction for every person in every image. The limbus point extraction can be performed using a manual marker

(a) Pristine (No manipulation)

(b) Fake

Figure 3.3: Examples the images used in the experiments of our first approach.

around the iris, or with an automatic method such as [41]. Since this is not the primary focus in our approach, we used manual markers. Afterwards, we characterize each image considering the features described in Section 3.2 obtaining a feature vector for each one. We then feed two-class classifiers with these features in order to achieve a final outcome. For this task, we used Support Vector Machine (SVM) with a standard RBF kernel. For a fair comparison, we perform five-fold cross-validation in all the experiments.

As some features in our proposed method rely upon non-linear minimization methods, which are initialized with random seeds, we can extract features using different seeds with no additional mathematical effort. Therefore, the proposed approach extracts the four features proposed in Section 3.2 five times for each image using different seeds each time, producing five different sets of features for each image. By doing so, we also present results for a pool of five classifiers, each classifier fed with a different set of features, analyzing an image in conjunction with a classifier-fusion fashion approach.

Finally, we assess the proposed features under the light of four classifier approaches:

1. **Single Classifier (SC):** a single classifier fed with the proposed features to predict the class (pristine or fake).

2. **Classifier Fusion with Majority Voting (MV):** a new sample is classified by a pool of five classifiers. Each classifier casts for a class vote in a winner-takes-all approach.

3. **Classifier Fusion with OR Rule (One Pristine):** similar to MV except that the decision rule states as non-fake if at least one classifier casts a vote in this direction.

4. **Classifier Fusion with OR Rule (One Fake):** similar to MV except that the decision rule states as fake if at least one classifier casts a vote in this direction.

To show the behavior of each round compared with Johnson and Farid's approach we used a ROC curve, in which the y-axis (Sensitivity) represents the fake images correctly classified as fakes and the x-axis (1 - Specificity) represents pristine images incorrectly classified. Figure 3.4 shows the results for our proposed approach (with four different classifier decision rules) compared to the results of Johnson and Farid's approach.

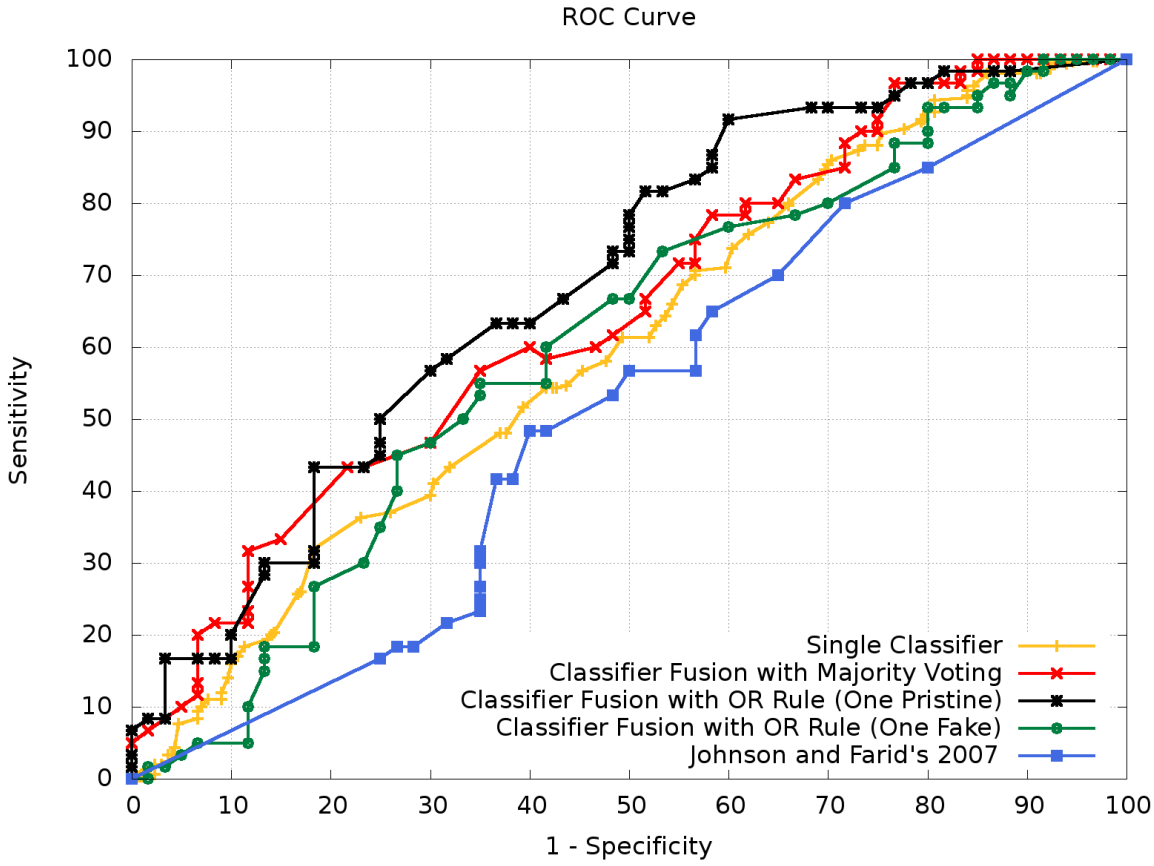


Figure 3.4: Comparison of classification results for Johnson and Farid's [48] approach against our approach.

All the proposed classification decision rules perform better than the prior work we rely upon in the approach proposed in this chapter, which is highlighted by their superior position in the graph of Figure 3.4. This allows us to come up with two conclusions: first, the new proposed features indeed make difference and contribute to the final classification decision; and second, different classifiers can take advantage of different seeds used in the

calculation of the features. Note that with 40% specificity, we detect 92% of fakes correctly while the Johnson and Farid's prior work achieves $\cong 64\%$.

Another way to compare our approach to Johnson and Farid's one is to assess the classification behavior on the *Equal Error Rate* (EER) point. Table 3.1 shows this comparison.

The best proposed method – *Classifier Fusion with OR Rule (One Pristine)* – decreases the classification error in 21% when compared to Johnson and Farid's approach at the EER point. Even if we consider just a single classifier (no fusion at all), the proposed extension performs 7% better than the prior work considering the EER point.

Table 3.1: Equal Error Rate – Four proposed approaches and the original work method by Johnson and Farid [48].

Method	EER (%)	Accuracy (%)	Improv. over prior work (%)
Single Classifier	44	56	07
Fusion MV	40	60	15
Fusion One Pristine	**37**	**63**	**21**
Fusion One Fake	41	59	13
Johnson and Farid [48]	48	52	–

3.4 Final Remarks

In this chapter, we presented our first approach to detect composition images using illumination inconsistencies. We extended Johnson and Farid's [48] prior work in such a way we now derive more discriminative features for detecting traces of tampering in composites of people and use the calculated features with decision-making classifiers based on simple, yet powerful, combinations of the well-known Support Vector Machines.

The new features (such as the viewer/camera estimated position) and the new decision-making process reduced the classification error in more than 20% when compared to the prior work. To validate our ideas, we have used a data set of real composites and images typically with more than three mega-pixels in resolution[1].

It is worth noting, however, the classification results are still affected by some drawbacks. First of all, the accuracy of the light direction estimation relies heavily on the camera calibration step. If the eyes are occluded by eyelids or are too small, the limbus selection becomes too difficult to accomplish, demanding an experienced user. Second, the focal length estimation method, which is common in a forensic scenario, is often affected

[1] http://ic.unicamp.br/∼rocha/pub/downloads/2014-tiago-carvalho-thesis/icip-eyes-database.zip

by numerical instabilities due to the starting conditions of the minimization function suggested in [48].

The aforementioned problems inspired us to develop a new method able to detect splicing images containing people that is not strongly dependent on their eyes. In the next chapter, we present such method that relies upon the analysis of the illuminant color in the image.

Chapter 4

Exposing Digital Image Forgeries by Illumination Color Classification

Different from our approach presented in Chapter 3, which is based on inconsistencies in the light setting, the approach proposed in this chapter relies on inconsistencies in light color. We extend upon the work of Riess and Angelopoulou [73] and analyze illuminant color estimates from local image regions. The resulting method is an important step toward minimizing user interaction for an illuminant-based tampering detection. Parts of the contents and findings in this chapter were published in [14].

4.1 Background

This section comprises two essential parts: related concepts and related work.

4.1.1 Related Concepts

According to colorimetry, light is composed of electromagnetic waves perceptible to human vision. These electromagnetic waves can be classified by their wavelength which varies from a very narrow, a short-wavelength edge between 360 and 400 nm, to a long-wavelength edge, between 760 and 830 nm [67].

There are two kinds of lights: monochromatic light, which is light that cannot be separated into components, and polychromatic lights (i.e., the white light provided by sunlight)i, which are composed by a mixture of monochromatic lights with different wavelengths. A spectrum is a band of color observed when a beam of white light is separated into components of light that are arranged in the order of their wavelengths. When the amount of one or more spectrum/bands is decreased in intensity, the light resulting from combination of these bands is a colored light, different from the original white light [67].

This new light is characterized by a specific spectral power distributions (SPDs), which represent the intensity of each band present in resulting light.

There is a large amount of different SPDs and the CIE has standardized a few of them, which are called *illuminants* [80]. In a rough way, an illuminant color (sometimes called a light-source color) can also be understood as the color of a light that appears to be emitted from a light source [67].

It is important to note here two facts: the first of them refers to the difference between illuminants and light sources. A light source is a natural or artificial light emitter and an illuminant is a specific SPD. Second, it is important to bear in mind that even the same light source can generate different illuminants. The illuminant formed by the sun, for example, varies in its appearance during the day and time of year as well as with the weather. We only capture the same illuminant, measuring the sunlight at the same place at the same time.

Complementing the definition of illuminant, comes the one related to *metamerism*. Its formal definition is: *Two specimens having identical tristimulus values for a given reference illuminant and reference observer are metameric if their spectral radiance distributions differ within the visible spectrum.* [80]. In other words, sometimes two objects composed by different materials (which provide different color stimuli) can cause sensation of identical appearance depending of the observer or scene illuminants. Illuminant metamerism results from scene illuminant changes (keeping the same observer) while observer metamerism results from observer changes (keeping the same illuminant) [1].

In this sense, keeping illuminant metamerism in mind (just under a very specific illuminant, two objects with different materials will depict very similar appearance), we can explore illuminants and metamerism in forensics to check the consistency of similar objects in a scene. If two objects with very similar color stimuli (e.g., human skin) depict inconsistent appearance (different illuminants), it means they might have undergone different illumination conditions hinting at a possible image composition. On the other hand, if we have a photograph with two people and the color appearance on the faces of such people are consistent, it is likely they have undergone similar lighting conditions (except in a very specific condition of metamerism).

4.1.2 Related Work

Riess and Angelopoulou [73] proposed to use a color-based method that investigates illuminant colors to detect forgeries in forensic scenario. Their method comprises four main steps:

[1] Datacolor Metamerism.`http://industrial.datacolor.com/support/wp-content/uploads/2013/01/Metamerism.pdf`. Accessed: 2013-12-23.

1. segmentation of the image in many small segments, grouping regions of approximately same color. These segments are named *superpixels*. Each one of these superpixels has its illuminant color estimated locally using an extension of the physic-model proposed by Tan et al. [86].

2. selection of superpixels to be further investigated by the user;

3. estimation of the illuminant color, which is performed twice, one for every superpixel and another one for the selected superpixels;

4. distance calculation from the selected superpixels to the other ones generating a distance map, which is the base for an expert analysis regarding forgery detection.

Figure 4.1 depicts an example of the generated illuminant and distance maps using Riess and Angelopoulou's [73] approach.

(a) Image (b) Illuminant Map from (a) (c) Distance map from (b)

Figure 4.1: From left to right: an image, its illuminant map and the distance map generated using Riess and Angelopoulou's [73] method. Original images obtained from [73].

The authors do not provide a numerical decision criterion for tampering detection. Thus, an expert is left with the difficult task of visually examining an illuminant map for evidence of tampering. The involved challenges are further discussed in Section 4.2.1.

On the other hand, in the field of color constancy, descriptors for the illuminant color have been extensively studied. Most research in color constancy focus on uniformly illuminated scenes containing a single dominant illuminant. Bianco and Schettini [7], for example, proposed a machine-learning based illuminant estimator specific for faces. However, their method has two main drawbacks that prevent us from implementing it in local illuminant estimation: (1) it is focused on a single illuminant estimation; (2) the illuminant estimation depends on a big cluster of similar color pixels which, many times, is not achieved by local illuminant estimation. This is just one of many examples of single illuminant estimation algorithms[2]. In order to use the color of the incident illumination

[2]See [2, 3, 35] for a complete overview of illuminants estimation algorithms for single illuminants

as a sign of image tampering, we require multiple, spatially-bound illuminant estimates. So far, limited research has been done in this direction. The work by Bleier et al. [10] indicates that many off-the-shelf single-illuminant algorithms do not scale well on smaller image regions. Thus, problem-specific illuminant estimators are required.

Besides the work of [73], Ebner [26] presented an early approach to multi-illuminant estimation. Assuming smoothly blending illuminants, the author proposes a diffusion process to recover the illumination distribution. In practice, this approach oversmooths the illuminant boundaries. Gijsenij et al. [37] proposed a pixelwise illuminant estimator. It allows to segment an image into regions illuminated by distinct illuminants. Differently illuminated regions can have crisp transitions, for instance between sunlit and shadow areas. While this is an interesting approach, a single illuminant estimator can always fail. Thus, for forensic purposes, we prefer a scheme that combines the results of multiple illuminant estimators. Earlier, Kawakami et al. [49] proposed a physics-based approach that is custom-tailored for discriminating shadow/sunlit regions. However, for our work, we consider the restriction to outdoor images overly limiting.

In this chapter, we build upon the ideas of [73] and [93] and use the relatively rich illumination information provided by both physics-based and statistics-based color constancy methods [73, 91] to detect image composites. Decisions with respect to the illuminant color estimators are completely taken away from the user, which differentiates this work from prior solutions.

4.2 Proposed Approach

Before effectively describing the approach proposed in this chapter, we first highlight the main challenges when using illuminant maps to detect image composition.

4.2.1 Challenges in Exploiting Illuminant Maps

To illustrate the challenges of directly exploiting illuminant estimates, we briefly examine the illuminant maps generated by the method of Riess and Angelopoulou [73]. In this approach, an image is subdivided into regions of similar color (superpixels). An illuminant color is locally estimated using the pixels within each superpixel (for details, see [73] and Section 4.2.3). Recoloring each superpixel with its local illuminant color estimate yields a so-called *illuminant map*. A human expert can then investigate the input image and the illuminant map to detect inconsistencies.

Figure 4.2 shows an example image and its illuminant map, in which an inconsistency can be directly shown: the inserted mandarin orange in the top right exhibits multiple green spots in the illuminant map. All other fruits in the scene show a gradual transition

from red to blue. The inserted mandarin orange is the only onethat deviates from this pattern. In practice, however, such analysis is often challenging, as shown in Figure 4.3.

(a) Fake Image (b) Illuminant Map from (a)

Figure 4.2: Example of illuminant map that directly shows an inconsistency.

The top left image is original, while the bottom image is a composite with the rightmost girl inserted. Several illuminant estimates are clear outliers, such as the hair of the girl on the left in the bottom image, which is estimated as strongly red illuminated. Thus, from an expert's viewpoint, it is reasonable to discard such regions and to focus on more reliable regions, e. g., the faces. In Figure 4.3, however, it is difficult to justify a tampering decision by comparing the color distributions in the facial regions. It is also challenging to argue, based on these illuminant maps, that the rightmost girl in the bottom image has been inserted, while, e. g., the rightmost boy in the top image is original.

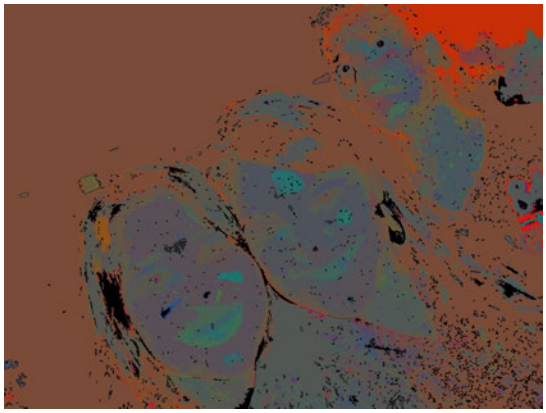
Although other methods operate differently, the involved challenges are similar. For instance, the approach by Gholap and Bora [33] is severely affected by clipping and camera white-balancing, which is almost always applied on images from off-the-shelf cameras. Wu and Fang [93] implicitly create illuminant maps and require comparison to a reference region. However, different choices of reference regions lead to different results, and this makes this method error-prone.

Thus, while illuminant maps are an important intermediate representation, we emphasize that further, automated processing is required to avoid biased or debatable human decisions. Hence, we propose a pattern recognition scheme operating on illuminant maps. The features are designed to capture the shape of the superpixels in conjunction with the

color distribution. In this spirit, our goal is to replace the expert-in-the-loop, by only requiring annotations of faces in the image.

Note that, the estimation of the illuminant color is error-prone and affected by the materials in the scene. However, (cf. also Figure 4.2), estimates on objects of similar material exhibit a lower relative error. Thus, we limit our detector to skin, and in particular to faces. Pigmentation is the most obvious difference in skin characteristics between different ethnicities. This pigmentation difference depends on many factors as quantity of melanin, amount of UV exposure, genetics, melanosome content and type of pigments found in the skin [44]. However, this intra-material variation is typically smaller than that of other materials possibly occurring in a scene.

(a) Original Image



(b) Illuminant Map from (a)

(c) Fake Image

(d) Illuminant Map from (c)

Figure 4.3: Example of illuminant maps for an original image (a - b) and a spliced image (c - d). The illuminant maps are created with the IIC-based illuminant estimator (see Section 4.2.3).

4.2.2 Methodology Overview

We classify the illumination for each pair of faces in the image as either consistent or inconsistent. Throughout the Chapter we abbreviate illuminant estimation as IE, and illuminant maps as IM. The proposed method consists of five main components:

1. *Dense Local Illuminant Estimation (IE):* The input image is segmented into homogeneous regions. Per illuminant estimator, a new image is created where each region is colored with the extracted illuminant color. This resulting intermediate representation is called illuminant map (IM).

2. *Face Extraction:* This is the only step that may require human interaction. An operator sets a bounding box around each face (e. g., by clicking on two corners of the bounding box) in the image that should be investigated. Alternatively, an automated face detector can be employed. We then crop every bounding box out of each illuminant map, so that only the illuminant estimates of the face regions remain.

3. *Computation of Illuminant Features:* for all face regions, texture-based and gradient-based features are computed on the IM values. Each one of them encodes complementary information for classification.

4. *Paired Face Features:* Our goal is to assess whether a pair of faces in an image is consistently illuminated. For an image with n_f faces, we construct $\binom{n_f}{2}$ joint feature vectors, consisting of all possible pairs of faces.

5. *Classification:* We use a machine learning approach to automatically classify the feature vectors. We consider an image as a forgery if at least one pair of faces in the image is classified as inconsistently illuminated.

Figure 4.4 summarizes these steps. In the remainder of this section, we present the details of these components.

4.2.3 Dense Local Illuminant Estimation

To compute a dense set of localized illuminant color estimates, the input image is segmented into superpixels, i. e., regions of approximately constant chromaticity, using the algorithm by Felzenszwalb and Huttenlocher [30]. Per superpixel, the color of the illuminant is estimated. We use two separate illuminant color estimators: the statistical generalized gray world estimates and the physics-based inverse-intensity chromaticity space,

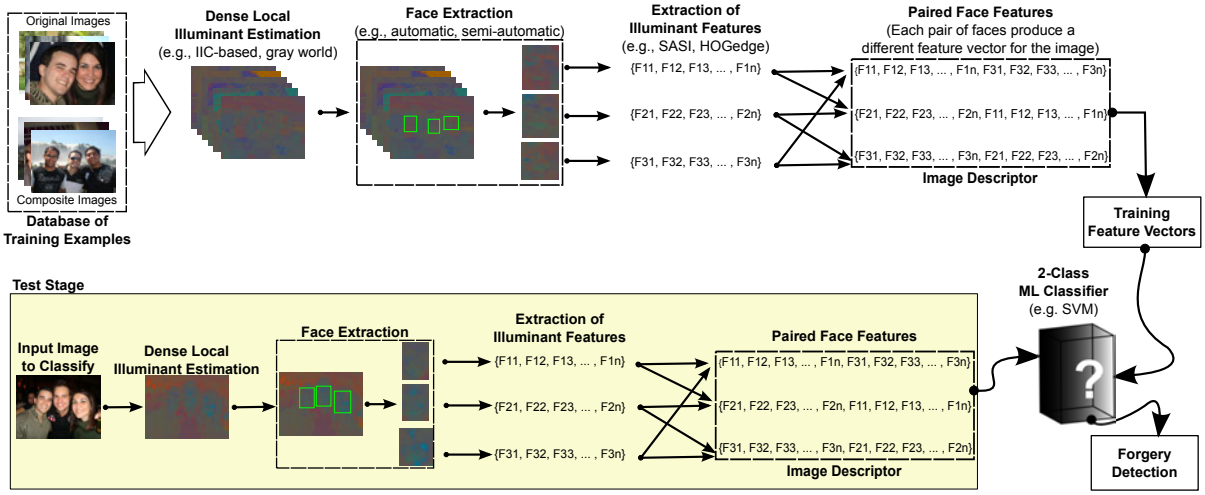


Figure 4.4: Overview of the proposed method.

as we explain below. We obtain, in total, two illuminant maps by recoloring each superpixel with the estimated illuminant chromaticities of each one of the estimators. Both illuminant maps are independently analyzed in the subsequent steps.

Generalized Gray World Estimates

The classical gray world assumption by Buchsbaum [11] states that the average color of a scene is gray. Thus, a deviation of the average of the image intensities from the expected gray color is due to the illuminant. Although this assumption is nowadays considered to be overly simplified [3], it has inspired the further design of statistical descriptors for color constancy. We follow an extension of this idea, the generalized gray world approach by van de Weijer et al. [91].

Let $\mathbf{f}(\mathbf{x}) = (\Gamma_R(\mathbf{x}), \Gamma_G(\mathbf{x}), \Gamma_B(\mathbf{x}))^{\mathrm{T}}$ denote the observed RGB color of a pixel at location $\mathbf{x}$ and $\Gamma_i(\mathbf{x})$ denote the intensity of the pixel in channel i at position $\mathbf{x}$. Van de Weijer et al. [91] assume purely diffuse reflection and linear camera response. Then, $\mathbf{f}(\mathbf{x})$ is formed by

$$\mathbf{f}(\mathbf{x}) = \int_{\omega} \mathbf{e}(\beta, \mathbf{x})\mathbf{s}(\beta, \mathbf{x})\mathbf{c}(\beta)d\beta \ , \tag{4.1}$$

where ω denotes the spectrum of visible light, β denotes the wavelength of the light, $\mathbf{e}(\beta, \mathbf{x})$ denotes the spectrum of the illuminant, $\mathbf{s}(\beta, \mathbf{x})$ the surface reflectance of an object, and $\mathbf{c}(\beta)$ the color sensitivities of the camera (i. e., one function per color channel). Van de Weijer et al. [91] extended the original grayworld hypothesis through the incorporation of three parameters:

- Derivative order $\mathbf{n}$: the assumption that the average of the illuminants is achromatic can be extended to the absolute value of the sum of the derivatives of the image.

- Minkowski norm $\mathbf{p}$: instead of simply adding intensities or derivatives, respectively, greater robustness can be achieved by computing the $\mathbf{p}$-th Minkowski norm of these values.

- Gaussian smoothing σ: to reduce image noise, one can smooth the image prior to processing with a Gaussian kernel of standard deviation σ.

Putting these three aspects together, van de Weijer et al. proposed to estimate the color of the illuminant $\mathbf{e}$ as

$$\lambda \mathbf{e}^{\mathbf{n},\mathbf{p},\sigma} = \left(\int \left| \frac{\partial^{\mathbf{n}} \mathbf{\Gamma}^{\sigma}(\mathbf{x})}{\partial \mathbf{x}^{\mathbf{n}}} \right|^{\mathbf{p}} d\mathbf{x} \right)^{1/\mathbf{p}} \quad . \tag{4.2}$$

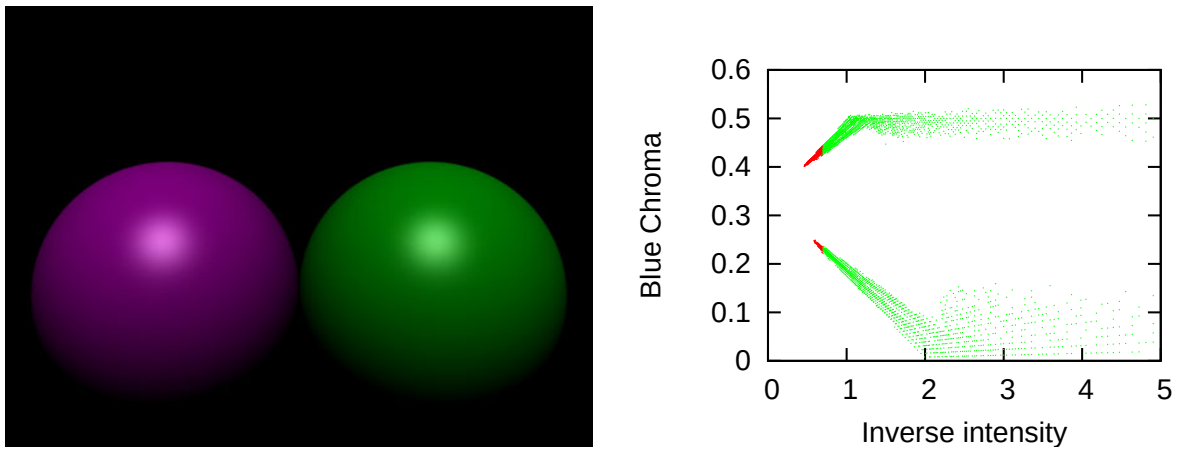
Here, the integral is computed over all pixels in the image, where $\mathbf{x}$ denotes a particular image position (pixel coordinate). Furthermore, λ denotes a scaling factor, $|\cdot|$ the absolute value, ∂ the differential operator, and $\Gamma^{\sigma}(\mathbf{x})$ the observed intensities at position $\mathbf{x}$, smoothed with a Gaussian kernel σ. Note that $\mathbf{e}$ can be computed separately for each color channel. Compared to the original gray world algorithm, the derivative operator increases the robustness against homogeneously colored regions of varying sizes. Additionally, the Minkowski norm emphasizes strong derivatives over weaker derivatives, so that specular edges are better exploited [36].

Inverse Intensity-Chromaticity Estimates

The second illuminant estimator we consider in this paper is the so-called inverse intensity-chromaticity (IIC) space. It was originally proposed by Tan et al. [86]. In contrast to the previous approach, the observed image intensities are assumed to exhibit a mixture of diffuse and specular reflectance. Pure specularities are assumed to consist of only the color of the illuminant. Let (as above) $\mathbf{f}(\mathbf{x}) = (\Gamma_R(\mathbf{x}), \Gamma_G(\mathbf{x}), \Gamma_B(\mathbf{x}))^{\mathrm{T}}$ be a column vector of the observed RGB colors of a pixel. Then, using the same notation as for the generalized gray world model, $\mathbf{f}(\mathbf{x})$ is modelled as

$$\mathbf{f}(\mathbf{x}) = \int_{\omega} (\mathbf{e}(\beta, \mathbf{x}) \mathbf{s}(\beta, \mathbf{x}) + \mathbf{e}(\beta, \mathbf{x})) \mathbf{c}(\beta) d\beta \quad . \tag{4.3}$$

Let $\Gamma_c(\mathbf{x})$ be the intensity and $\chi_c(\mathbf{x})$ be the chromaticity (i. e., normalized RGB-value) of a color channel $c \in \{R, G, B\}$ at position $\mathbf{x}$, respectively. In addition, let γ_c be the



(a) Synthetic Image (b) Inverse intensity Chromaticity Space

Figure 4.5: Illustration of the inverse intensity-chromaticity space (blue color channel). (a) depicts synthetic image (violet and green balls) while (b) depicts that specular pixels from (a) converge towards the blue portion of the illuminant color (recovered at the y-axis intercept). Highly specular pixels are shown in red.

chromaticity of the illuminant in channel c. Then, after a somewhat laborious calculation, Tan et al. [86] derived a linear relationship between $\mathbf{f}(\mathbf{x})$, $\chi_c(\mathbf{x})$ and γ_c by showing that

$$\chi_c(\mathbf{x}) = m(\mathbf{x}) \frac{1}{\sum_{i \in \{R,G,B\}} \Gamma_i(\mathbf{x})} + \gamma_c \ . \tag{4.4}$$

Here, $m(\mathbf{x})$ mainly captures geometric influences, i. e., light position, surface orientation and camera position. Although $m(\mathbf{x})$ can not be analytically computed, an approximate solution is feasible. More importantly, the only aspect of interest in illuminant color estimation is the y-intercept γ_c. This can be directly estimated by analyzing the distribution of pixels in IIC space. The IIC space is a per-channel 2D space, where the horizontal axis is the inverse of the sum of the chromaticities per pixel, $1/\sum_i \Gamma_i(\mathbf{x})$, and the vertical axis is the pixel chromaticity for that particular channel. Per color channel c, the pixels within a superpixel are projected onto inverse intensity-chromaticity (IIC) space.

Figure 4.5 depicts an exemplary IIC diagram for the blue channel. A synthetic image is rendered (a) and projected onto IIC space (b). Pixels from the green and purple balls form two clusters. The clusters have spikes that point towards the same location at the y-axis. Considering only such spikes from each cluster, the illuminant chromaticity is estimated from the joint y-axis intercept of all spikes in IIC space [86].

In natural images, noise dominates the IIC diagrams. Riess and Angelopoulou [73] proposed to compute these estimates over a large number of small image patches. The

final illuminant estimate is computed by a majority vote of these estimates. Prior to the voting, two constraints are imposed on a patch to improve noise resilience. If a patch does not satisfy these constraints, it is excluded from voting.

In practice, these constraints are straightforward to compute. The pixel colors of a patch are projected onto IIC space. Principal component analysis on the distribution of the patch-pixels in IIC space yields two eigenvalues $\mathbf{g}_1$, $\mathbf{g}_2$ and their associated eigenvectors $\vec{g}_1$ and $\vec{g}_2$. Let $\mathbf{g}_1$ be the larger eigenvalue. Then, $\vec{g}_1$ is the principal axis of the pixel distribution in IIC space. In the two-dimensional IIC-space, the principal axis can be interpreted as a line whose slope can be directly computed from $\vec{g}_1$. Additionally, $\mathbf{g}_1$ and $\mathbf{g}_2$ can be used to compute the eccentricity $\sqrt{1-\sqrt{\mathbf{g}_2}/\sqrt{\mathbf{g}_1}}$ as a metric for the shape of the distribution. Both constraints are associated with this eigenanalysis[3]. The first constraint is that the slope must exceed a minimum of 0.003. The second constraint is that the eccentricity has to exceed a minimum of 0.2.

4.2.4 Face Extraction

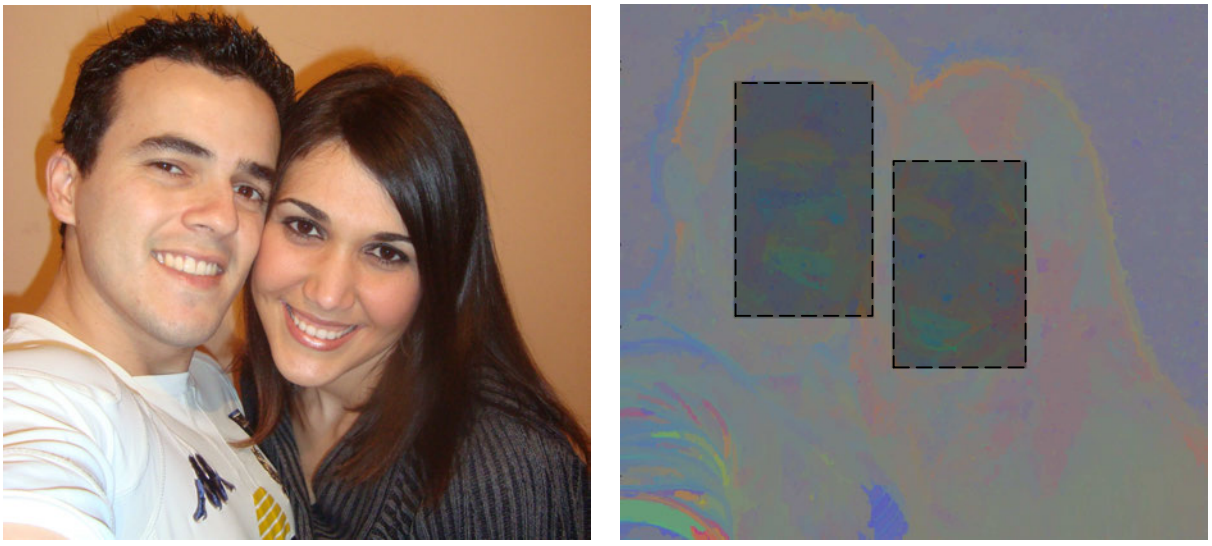
We require bounding boxes around all faces in an image that should be part of the investigation. For obtaining the bounding boxes, we could in principle use an automated algorithm, e. g., the one by Schwartz et al. [81]. However, we prefer a human operator for this task for two main reasons: a) this minimizes false detections or missed faces; b) scene context is important when judging the lighting situation. For instance, consider an image where all persons of interest are illuminated by flashlight. The illuminants are expected to agree with one another. Conversely, assume that a person in the foreground is illuminated by flashlight, and a person in the background is illuminated by ambient light. Then, a difference in the color of the illuminants is expected. Such differences are hard to distinguish in a fully-automated manner, but can be easily excluded in manual annotation.

We illustrate this setup in Figure 4.6. The faces in Figure 4.6(a) can be assumed to be exposed to the same illuminant. As Figure 4.6(b) shows, the corresponding gray world illuminant map for these two faces also has similar values.

4.2.5 Texture Description: SASI Algorithm

When analyzing an illuminant map, we figured out that two or more people illuminated by similar light source tend to present illuminant maps in their faces with similar texture, while people under different light source tend to present different texture in their illuminant maps. Even when we observe the same person in the same position but under

[3]The parameter values were previously investigated by Riess and Angelopoulou [73, 74]. In this paper, we rely on their findings.



(a) Original Image (b) IM with highlighted similar parts

Figure 4.6: An original image and its gray world map. Highlighted regions in the gray world map show a similar appearance.

different illumination, they present illuminant maps with different texture. Figure 4.7 depicts an example showing similarity and difference in illuminant maps when considering texture appearance.

We use the Statistical Analysis of Structural Information (SASI) descriptor by Çarkacıoğlu and Yarman-Vural [15] to extract texture information from illuminant maps. Recently, Penatti et al. [70] pointed out that SASI performs remarkably well. For our application, the most important advantage of SASI is its capability of capturing small granularities and discontinuities in texture patterns. Distinct illuminant colors interact differently with the underlying surfaces, thus generating distinct illumination *texture*. This can be a very fine texture, whose subtleties are best captured by SASI.

SASI is a generic descriptor that measures the structural properties of textures. It is based on the autocorrelation of horizontal, vertical and diagonal pixel lines over an image at different scales. Instead of computing the autocorrelation for every possible shift, only a small number of shifts is considered. One autocorrelation is computed using a specific fixed orientation, scale, and shift. Computing the mean and standard deviation of all such pixel values yields two feature dimensions. Repeating this computation for varying orientations, scales and shifts yields a 128-dimensional feature vector. As a final step, this vector is normalized by subtracting its mean value, and dividing it by its standard deviation. For details, please refer to [15].

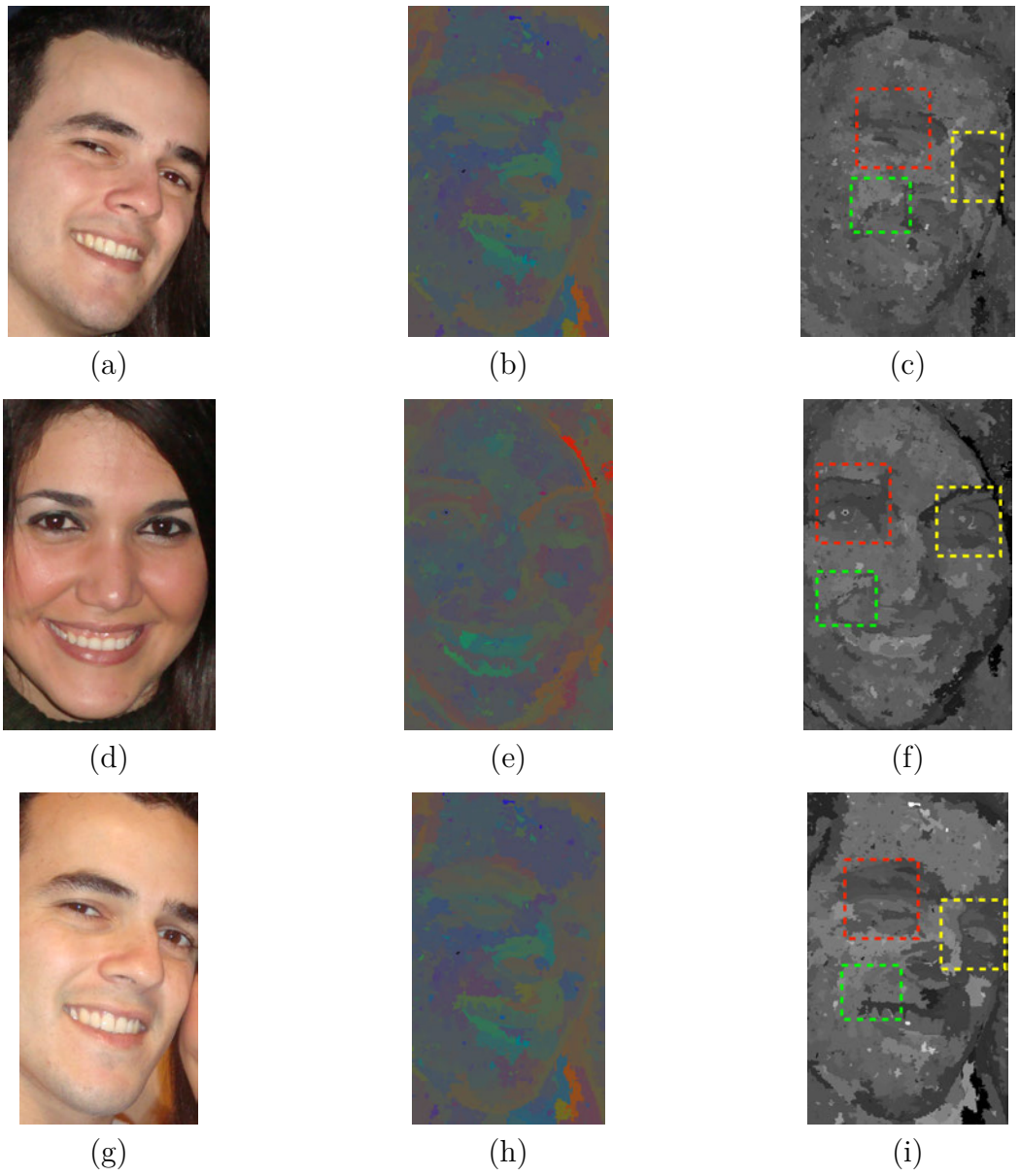
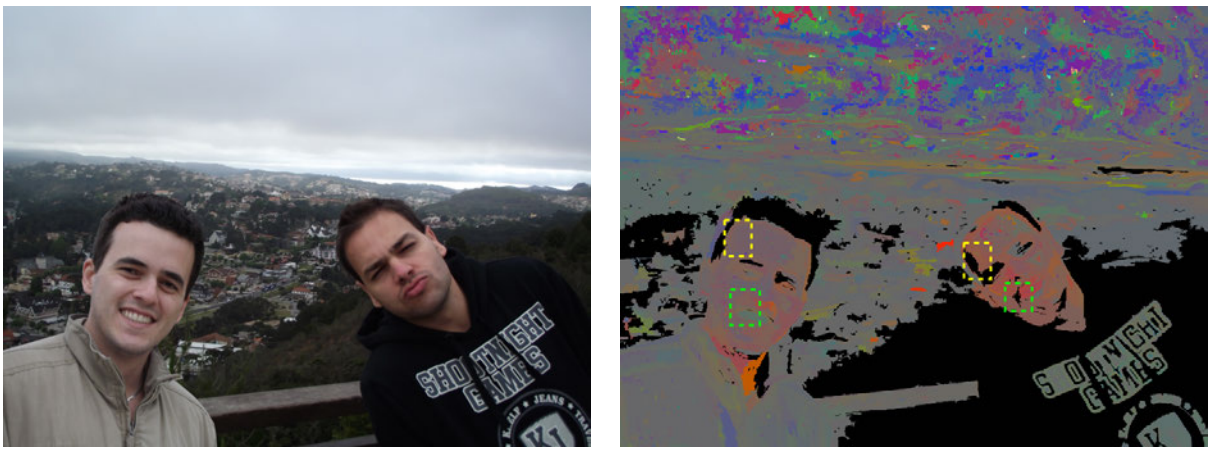


Figure 4.7: An example of how different illuminant maps are (in texture aspects) under different light sources. (a) and (d) are two people's faces extracted from the same image. (b) and (e) display their illuminant maps, respectively, and (c) and (f) depicts illuminant maps in grayscale. Regions with same color (red, yellow and green) depict some similarity. On the other hand, (f) depicts the same person (a) in a similar position but extracted from a different image (consequently, illuminated by a different light source). The grayscale illuminant map (h) is quite different from (c) in highlighted regions.

4.2.6 Interpretation of Illuminant Edges: HOGedge Algorithm

Differing illuminant estimates in neighboring segments can lead to discontinuities in the illuminant map. Dissimilar illuminant estimates can occur for a number of reasons: changing geometry, changing material, noise, retouching or changes in the incident light. Figure 4.8 depicts an example of such discontinuities.



(a) (b)

Figure 4.8: An example of discontinuities generated by different illuminants. The illuminant map (b) has been calculated from splicing image depicted in (a). The person on the left does not show discontinuities in the highlighted regions (green and yellow). On the other hand, the alien part (person on the right) presents discontinuities in the same regions highlighted in the person on the left.

Thus, one can interpret an illuminant estimate as a low-level descriptor of the underlying image statistics. We observed that the edges, e. g., computed by a Canny edge detector, detect in several cases a combination of the segment borders and isophotes (i. e., areas of similar incident light in the image). When an image is spliced, the statistics of these edges is likely to differ from original images. To characterize such edge discontinuities, we propose a new algorithm called *HOGedge*. It is based on the well-known HOG-descriptor, and computes visual dictionaries of gradient intensities in edge points. The full algorithm is described below. Figure 4.9 shows an algorithmic overview of the method. We first extract approximately equally distributed candidate points on the edges of illuminant maps. At these points, HOG descriptors are computed. These descriptors are summarized in a visual-word dictionary. Each of these steps is presented in greater detail next.

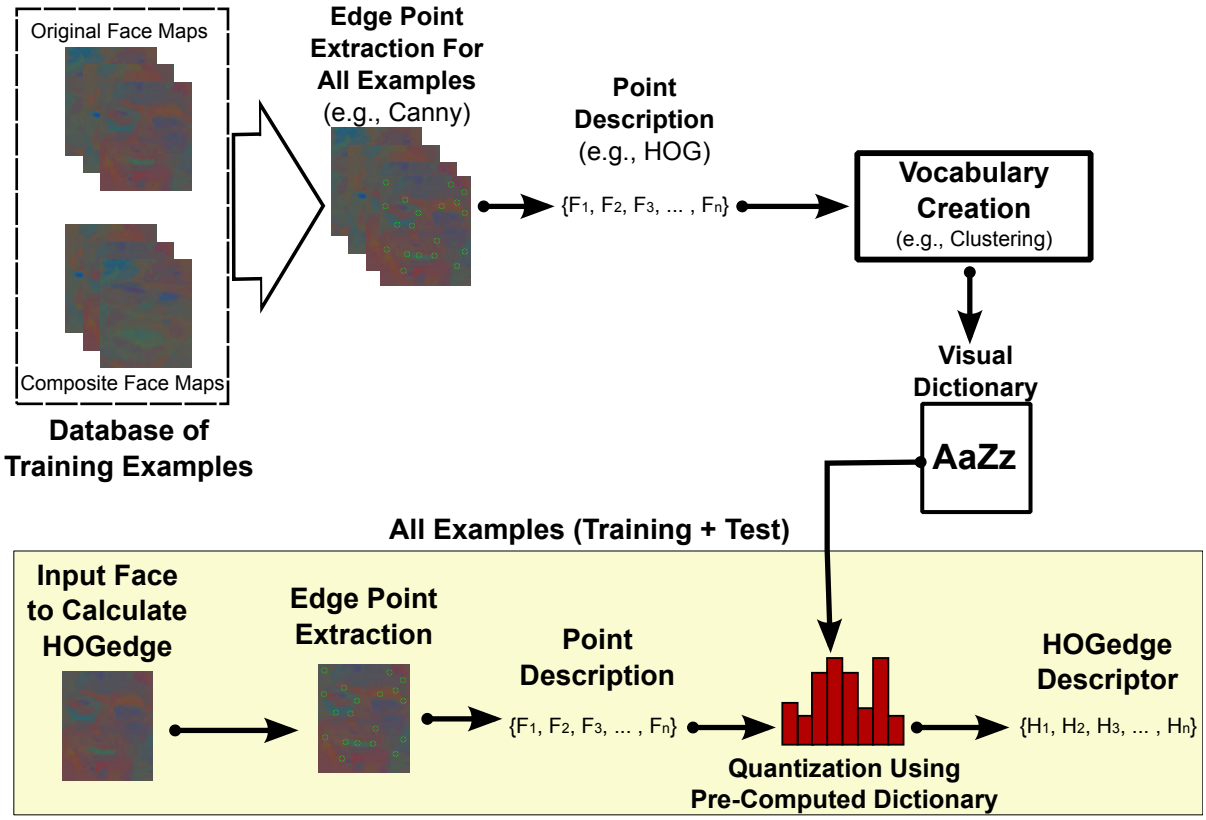


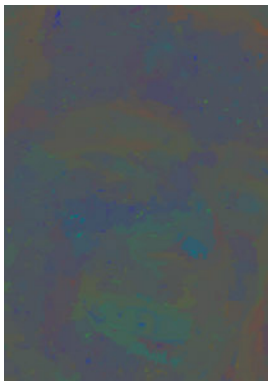
Figure 4.9: Overview of the proposed HOGedge algorithm.

Extraction of Edge Points

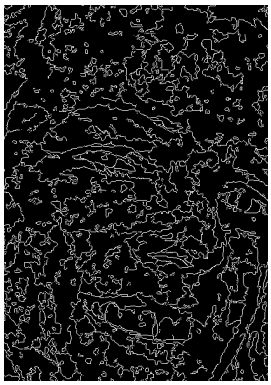
Given a face region from an illuminant map, we first extract edge points using the Canny edge detector [12]. This yields a large number of spatially close edge points. To reduce the number of points, we filter the Canny output using the following rule: starting from a seed point, we eliminate all other edge pixels in a region of interest (ROI) centered around the seed point. The edge points that are closest to the ROI (but outside of it) are chosen as seed points for the next iteration. By iterating this process over the entire image, we reduce the number of points but still ensure that every face has a comparable density of points. Figure 4.10 depicts an example of the resulting points.

Point Description

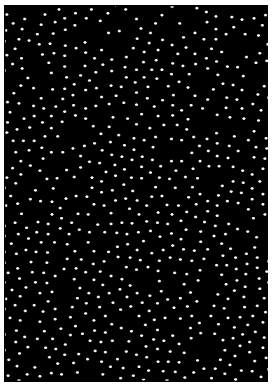
We compute Histograms of Oriented Gradients (HOG) [21] to describe the distribution of the selected edge points. HOG is based on normalized local histograms of image gradient orientations in a dense grid. The HOG descriptor is constructed around each of the edge



(a) IM



(b) Extracted Edge Points



(c) Filtered Edge Points

Figure 4.10: (a) The gray world IM for the left face in Figure 4.6(b). (b) The result of the Canny edge detector when applied on this IM. (c) The final edge points after filtering using a square region.

points. The neighborhood of such an edge point is called a cell. Each cell provides a local 1-D histogram of quantized gradient directions using all cell pixels. To construct the feature vector, the histograms of all cells within a spatially larger region are combined and contrast-normalized. We use the HOG output as a feature vector for the subsequent steps.

Visual Vocabulary

The number of extracted HOG vectors varies depending on the size and structure of the face under examination. We use visual dictionaries [20] to obtain feature vectors of fixed length. Visual dictionaries constitute a robust representation, where each face is treated as a set of region descriptors. The spatial location of each region is discarded [92].

To construct our visual dictionary, we subdivide the training data into feature vectors from original and doctored images. Each group is clustered in n clusters using the k-means algorithm [8]. Then, a visual dictionary with $2n$ visual words is constructed, where each word is a cluster center. Thus, the visual dictionary summarizes the most representative feature vectors of the training set. Algorithm 1 shows the pseudocode for the dictionary creation.

Quantization Using the Pre-Computed Visual Dictionary

For evaluation, the HOG feature vectors are mapped to the visual dictionary. Each feature vector in an image is represented by the closest word in the dictionary (with respect to the Euclidean distance). A histogram of word counts represents the distribution of HOG

Algorithm 1 HOGedge – Visual dictionary creation

Input: V_{TR} (training database examples)
n (the number of visual words per class)
Output: V_D (visual dictionary containing $2n$ visual words)

```
V_D ← ∅;
V_NF ← ∅;
V_DF ← ∅;
for each face IM i ∈ V_TR do
  V_EP ← edge points extracted from i;
  for each point j ∈ V_EP do
    FV ← apply HOG in image i at position j;
    if i is a doctored face then
      V_DF ← {V_DF ∪ FV};
    else
      V_NF ← {V_NF ∪ FV};
    end if
  end for
end for
Cluster V_DF using n centers;
Cluster V_NF using n centers;
V_D ← {centers of V_DF ∪ centers of V_NF};
return V_D;
```

feature vectors in a face. Algorithm 2 shows the pseudocode for the application of the visual dictionary on IMs.

4.2.7 Face Pair

To compare two faces, we combine the same descriptors for each of the two faces. For instance, we can concatenate the SASI-descriptors that were computed on gray world. The idea is that a feature concatenation from two faces is different when one of the faces is an original and one is spliced. For an image containing n_f faces ($n_f \geq 2$), the number of face pairs is $(n_f(n_f - 1))/2$.

The SASI descriptor and HOGedge algorithm capture two different properties of the face regions. From a signal processing point of view, both them are *signatures* with different behavior. Figure 4.11 shows a very high-level visualization of the distinct information that is captured by these two descriptors. For one of the folds from our experiments (see Section 4.3.3), we computed the mean value and standard deviation per feature dimension. For a less cluttered plot, we only visualize the feature dimensions with the largest difference in the mean values for this fold. This experiment empirically demonstrates two

Algorithm 2 HOGedge – Face characterization

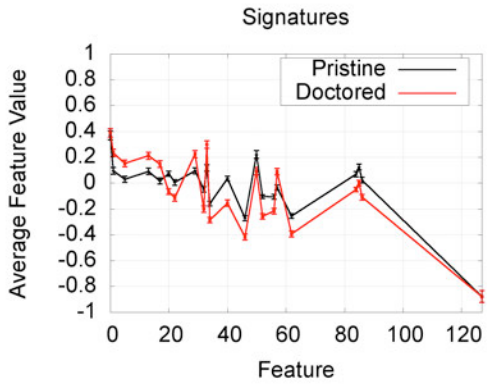
Input: V_D (visual dictionary pre-computed with $2n$ visual words)
IM (illuminant map from a face)
Output: HFV (HOGedge feature vector)

```
HFV ← 2n-dimensional vector, initialized to all zeros;
V_FV ← ∅;
V_EP ← edge points extracted from IM;
for each point i ∈ V_EP do
  FV ← apply HOG in image IM at position j;
  V_FV ← {V_FV ∪ FV};
end for
for each feature vector i ∈ V_FV do
  lower_distance ← +∞;
  position ← −1;
  for each visual word j ∈ V_D do
    distance ← Euclidean distance between i and j;
    if distance < lower_distance then
      lower_distance ← distance;
      position ← position of j in V_D;
    end if
  end for
  HFV [position] ← HFV [position] + 1;
end for
return HFV;
```

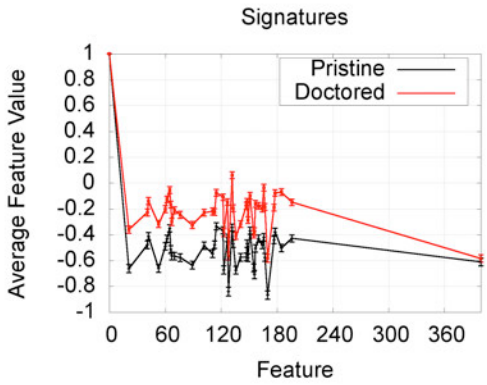
points. Firstly, SASI and HOGedge, in combination with the IIC-based and gray world illuminant maps create features that discriminate well between original and tampered images, in at least some dimensions. Secondly, the dimensions, where these features have distinct value, vary between the four combinations of the feature vectors. We exploit this property during classification by fusing the output of the classification on both feature sets, as described in the next section.

4.2.8 Classification

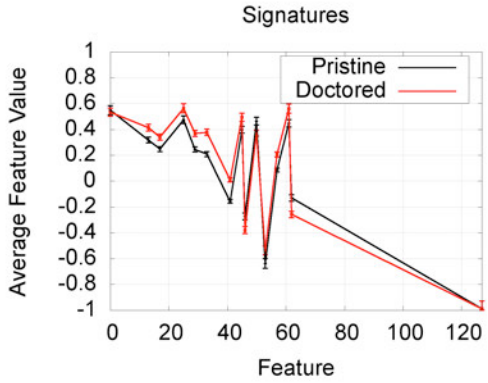
We classify the illumination for each pair of faces in an image as either consistent or inconsistent. Assuming all selected faces are illuminated by the same light source, we tag an image as manipulated if one pair is classified as inconsistent. Individual feature vectors, i. e., SASI or HOGedge features on either gray world or IIC-based illuminant maps, are classified using a support vector machine (SVM) classifier with a radial basis function (RBF) kernel.



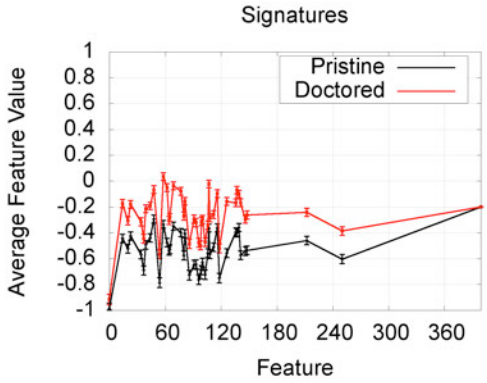
(a) SASI extracted from IIC



(b) HOGedge extracted from IIC



(c) SASI extracted from Gray-World



(d) HOGedge extracted from Gray-World

Figure 4.11: Average signatures from original and spliced images. The horizontal axis corresponds to different feature dimensions, while the vertical axis represents the average feature value for different combinations of descriptors and illuminant maps.

The information provided by the SASI features is complementary to the information from the HOGedge features. Thus, we use a machine learning-based fusion technique for improving the detection performance. Inspired by the work of Ludwig et al. [62], we use a late fusion technique named SVM-Meta Fusion. We classify each combination of illuminant map and feature type independently (i. e., SASI-Gray-World, SASI-IIC, HOGedge-Gray-World and HOGedge-IIC) using a two-class SVM classifier to obtain the distance between the image's feature vectors and the classifier decision boundary. SVM-Meta Fusion then merges the marginal distances provided by all m individual classifiers to build a new feature vector. Another SVM classifier (i. e., on meta level) classifies the combined feature vector.

4.3 Experiments and Results

To validate our approach, we performed six rounds of experiments using two different databases of images involving people. We show results using classical ROC curves where *sensitivity* represents the number of composite images correctly classified and *specificity* represents the number of original images (non-manipulated) correctly classified.

4.3.1 Evaluation Data

To quantitatively evaluate the proposed algorithm, and to compare it to related work, we considered two datasets. One consists of images that we captured ourselves, while the second one contains images collected from the internet. Additionally, we validated the quality of the forgeries using a human study on the first dataset. Human performance can be seen as a baseline for our experiments.

DSO-1

This is our first dataset and it was created by ourselves. It is composed of 200 indoor and outdoor images with image resolution of $2,048 \times 1,536$ pixels. Out of this set of images, 100 are original, i. e., have no adjustments whatsoever, and 100 are forged. The forgeries were created by adding one or more individuals in a source image that already contained one or more people. When necessary, we complemented an image splicing operation with post-processing operations (such as color and brightness adjustments) in order to increase photorealism.

DSI-1

This is our second dataset and it is composed of 50 images (25 original and 25 doctored) downloaded from different websites in the Internet with different resolutions[4].

Figure 4.12 depicts some example images from our databases.

4.3.2 Human Performance in Spliced Image Detection

To demonstrate the quality of DSO-1 and the difficulty in discriminating original and tampered images, we performed an experiment where we asked humans to mark images as tampered or original. To accomplish this task, we have used Amazon Mechanical Turk[5].

[4]Original images were downloaded from Flickr (http://www.flickr.com) and doctored images were collected from different websites such as Worth 1000 (http://www.worth1000.com/), Benetton Group 2011 (http://press.benettongroup.com/), Planet Hiltron (http://www.facebook.com/pages/Planet-Hiltron/150175044998030), etc.

[5]https://www.mturk.com/mturk/welcome

(a) DSO-1 Original image

(b) DSO-1 Spliced image

(c) DSI-1 Original image

(d) DSI-1 Spliced image

Figure 4.12: Original (left) and spliced images (right) from both databases.

Note that on Mechanical Turk categorization experiments, each batch is evaluated only by experienced users which generally leads to a higher confidence in the outcome of the task. In our experiment, we setup five identical categorization experiments, where each one of them is called a batch. Within a batch, all DSO-1 images have been evaluated. For each image, two users were asked to tag the image as original or manipulated. Each image was assessed by ten different users, each user expended on average 47 seconds to tag an image. The final accuracy, averaged over all experiments, was 64.6%. However, for spliced images, the users achieved only an average accuracy of 38.3%, while human accuracy on the original images was 90.9%. The kappa-value, which measures the degree of agreement between an arbitrary number of raters in deciding the class of a sample, based on the Fleiss [31] model, is 0.11. Despite being subjective, this kappa-value, according to the Landis and Koch [59] scale, suggests a slight degree of agreement between users, which further supports our conjecture about the difficulty of forgery detection in DSO-1 images.

4.3.3 Performance of Forgery Detection using Semi-Automatic Face Annotation in DSO-1

We compare five variants of the method proposed in this paper. Throughout this section, we manually annotated the faces using corner clicking (see Section 4.3.4). In the classification stage, we use a five-fold cross validation protocol, an SVM classifier with an RBF kernel, and classical grid search for adjusting parameters in training samples [8]. Due to the different number of faces per image, the number of feature vectors for the original and the spliced images is not exactly equal. To address this issue during training, we weighted feature vectors from original and composite images. Let w_o and w_c denote the number of feature vectors from original and composite images, respectively. To obtain a proportional class weighting, we set the weight of features from original images to $w_c/(w_o + w_c)$ and the weight of features from composite images to $w_o/(w_o + w_c)$.

We compared the five variants SASI-IIC, SASI-Gray-World, HOGedge-IIC, HOGedge-Gray-World and Metafusion. Compound names, such as SASI-IIC, indicate the data source (in this case: IIC-based illuminant maps) and the subsequent feature extraction method (in this case: SASI). The single components are configured as follows:

- **IIC:** IIC-based illuminant maps are computed as described in [73].
- **Gray-World:** Gray world illuminant maps are computed by setting $\mathbf{n} = 1$, $\mathbf{p} = 1$, and $\sigma = 3$ in Equation 4.2.
- **SASI:** The SASI descriptor is calculated over the Y channel from the YC_bC_r color space. All remaining parameters are chosen as presented in [70][6].
- **HOGedge:** Edge detection is performed on the Y channel of the YC_bC_r color space, with a Canny low threshold of 0 and a high threshold of 10. The square region for edge point filtering was set to 32×32 pixels. Furthermore, we used 8-pixel cells without normalization in HOG. If applied on IIC-based illuminant maps, we computed 100 visual words for both the original and the tampered images (i. e., the dictionary consisted of 200 visual words). On gray world illuminant maps, the size of the visual word dictionary was set to 75 for each class, leading to a dictionary of 150 visual words.
- **Metafusion:** We implemented a late fusion as explained in Section 4.2.8. As input, it uses SASI-IIC, SASI-Gray-World, and HOGedge-IIC. We excluded HOGedge-Gray-World from the input methods, as its weaker performance leads to a slightly worse combined classification rate (see below).

[6] We gratefully thank the authors for the source code.

Figure 4.13 depicts a ROC curve of the performance of each method using the corner clicking face localization. The area under the curve (AUC) is computed to obtain a single numerical measure for each result.

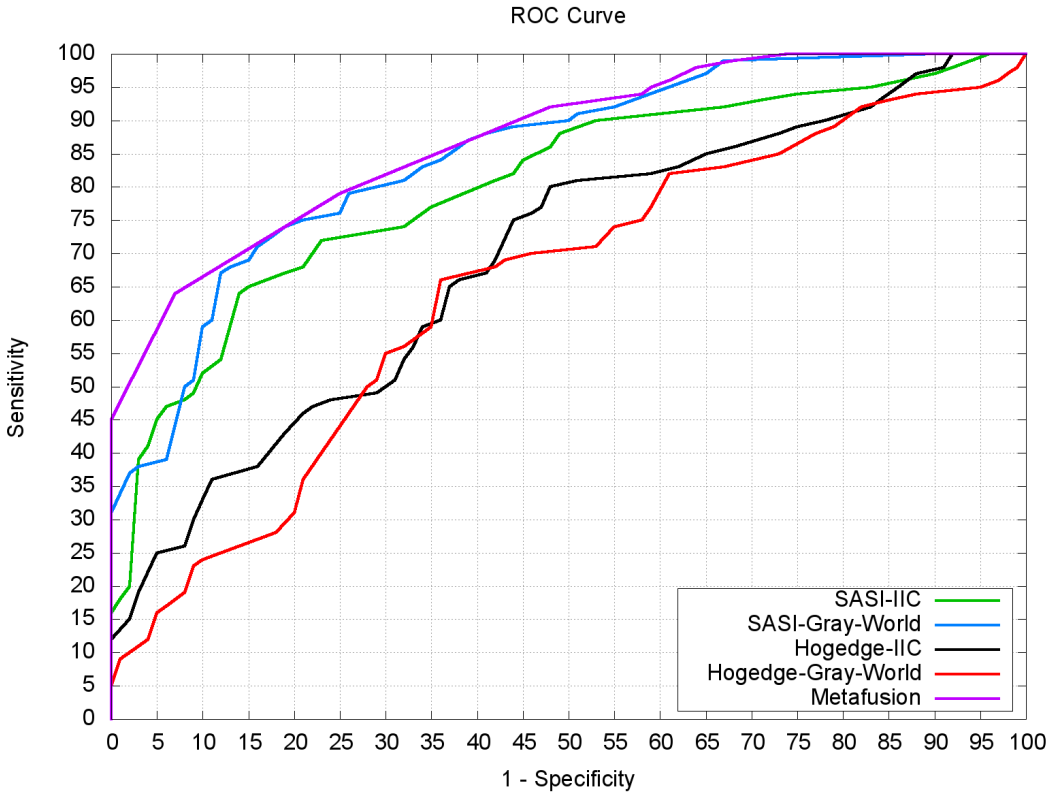


Figure 4.13: Comparison of different variants of the algorithm using semi-automatic (corner clicking) annotated faces.

From the evaluated variants, Metafusion performs best, resulting in an AUC of 86.3%. In particular for high specificity (i. e., few false alarms), the method has a much higher sensitivity compared to the other variants. Thus, when the detection threshold is set to a high specificity, and a photograph is classified as composite, Metafusion provides to an expert high confidence that the image is indeed manipulated.

Note also that Metafusion clearly outperforms human assessment in the baseline Mechanical Turk experiment (see Section 4.3.2). Part of this improvement comes from the fact that Metafusion achieves, on spliced images alone, an average accuracy of 67%, while human performance was only 38.3%.

The second best variant is SASI-Gray-World, with an AUC of 84.0%. In particular for a specificity below 80.0%, the sensitivity is comparable to Metafusion. SASI-IIC achieved

an AUC of 79.4%, followed by HOGedge-IIC with an AUC of 69.9% and HOGedge-Gray-World with an AUC of 64.7%. The weak performance of HOGedge-Gray-World comes from the fact that illuminant color estimates from the gray world algorithm vary more smoothly than IIC-based estimates. Thus, the differences in the illuminant map gradients (as extracted by the HOGedge algorithm) are generally smaller.

4.3.4 Fully Automated versus Semi-Automatic Face Detection

In order to test the impact of automated face detection, we re-evaluated the best performing variant, Metafusion, on three versions of automation in face detection and annotation.

- **Automatic Detection:** we used the PLS-based face detector [81] to detect faces in the images. In our experiments, the PLS detector successfully located all present faces in only 65% of the images. We then performed a 3-fold cross validation on this 65% of the images. For training the classifier, we used the manually annotated bounding boxes. In the test images, we used the bounding boxes found by the automated detector.

- **Semi-Automatic Detection 1 (Eye Clicking):** an expert does not necessarily have to mark a bounding box. In this variant, the expert clicks on the eye positions. The Euclidean distance between the eyes is the used to construct a bounding box for the face area. For classifier training and testing we use the same setup and images as in the automatic detection.

- **Semi-Automatic Detection 2 (Corner Clicking):** in this variant, we applied the same marking procedure as in the previous experiment and the same classifier training/testing procedure as in automatic detection.

Figure 4.14 depicts the results of this experiment. The semi-automatic detection using corner clicking resulted in an AUC of 78.0%, while the semi-automatic using eye clicking and the fully-automatic approaches yielded an AUC of 63.5% and AUC of 63.0%, respectively. Thus, as it can also be seen in Figures 4.15(a), (b) and (c), proper face location is important for improved performance.

Although automatic face detection algorithms have improved over the years, we find user-selected faces more reliable for a forensic setup mainly because automatic face detection algorithms are not accurate in bounding box detection (location and size). In our experiments, automatic and eye clicking detection have generated an average bounding box size which was 38.4% and 24.7% larger than corner clicking detection, respectively. Thus, such bounding boxes include part of the background in a region that should contain just face information. The precision of bounding box location in automatic detection and

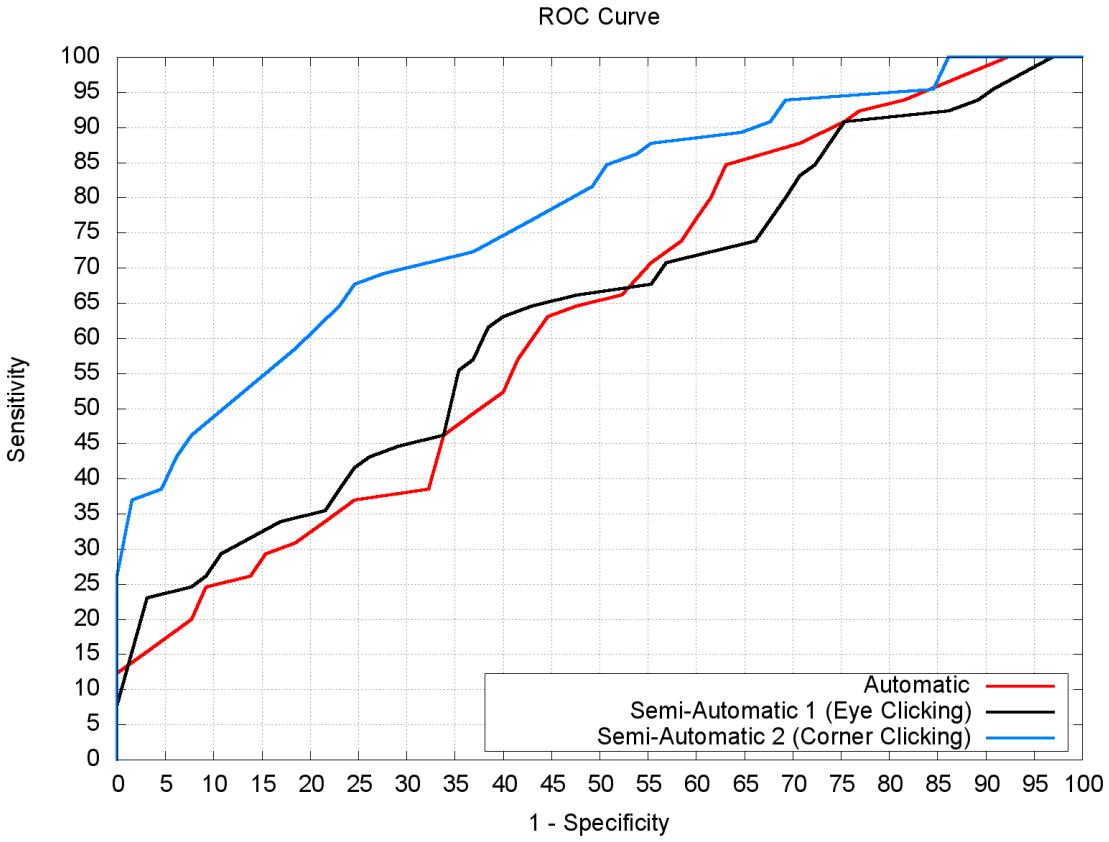


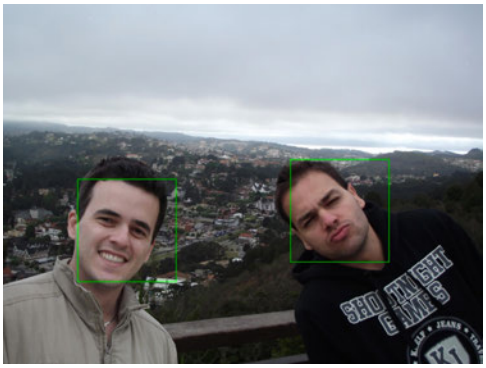
Figure 4.14: Experiments showing the differences for automatic and semi-automatic face detection.

eye clicking has also been worse than semi-automatic using corner clicking. Note, however, that the selection of faces under similar illumination conditions is a minor interaction that requires no particular knowledge in image processing or image forensics.

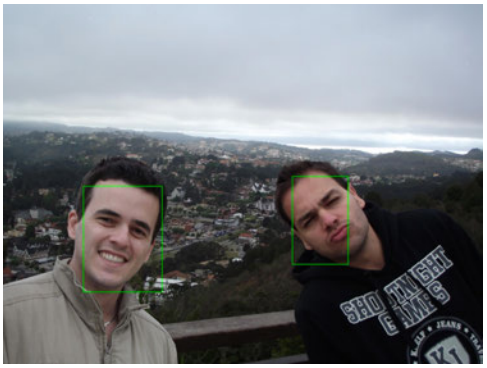
4.3.5 Comparison with State-of-the-Art Methods

For experimental comparison, we implemented the methods by Gholap and Bora [33] and Wu and Fang [93]. Note that neither of these works includes a quantitative performance analysis. Thus, to our knowledge, this is the first direct comparison of illuminant color-based forensic algorithms.

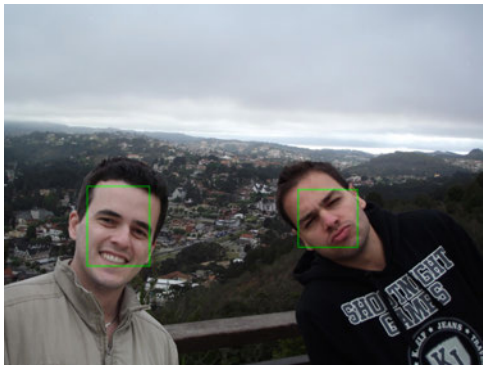
For the algorithm by Gholap and Bora [33], three partially specular regions per image were manually annotated. For manipulated images, it is guaranteed that at least one of the regions belongs to the tampered part of the image, and one region to the original part. Fully saturated pixels were excluded from the computation, as they have presumably



(a) Automatic



(b) Semi-automatic (Eye Clicking)



(c) Semi-automatic (Corner Clicking)

Figure 4.15: Different types of face location. Automatic and semi-automatic locations select a considerable part of the background, whereas manual location is restricted to face regions.

been clipped by the camera. Camera gamma was approximately inverted by assuming a value of 2.2. The maximum distance of the dichromatic lines per image were computed. The threshold for discriminating original and tampered images was set via five-fold cross-validation, yielding a detection rate of 55.5% on DSO-1.

In the implementation of the method by Wu and Fang, the Weibull distribution is computed in order to perform image classification prior to illuminant estimation. The training of the image classifier was performed on the ground truth dataset by Ciurea and Funt [17] as proposed in the work [93]. As the resolution of this dataset is relatively low, we performed the training on a central part of the images containing 180×240 pixels (excluding the ground-truth area). To provide images of the same resolution for illuminant classification, we manually annotated the face regions in DSO-1 with bounding boxes of fixed size ratio. Setting this ratio to 3:4, each face was then rescaled to a size of 180×240 pixels. As the selection of suitable reference regions is not well-defined (and also highly image-dependent), we directly compare the illuminant estimates of the faces in the scene.

Here, the best result was obtained with three-fold cross-validation, yielding a detection rate of 57%. We performed five-fold cross-validation, as in the previous experiments. The results drop to 53% detection rate, which suggeststhat this algorithm is not very stable with respect to the selection of the data.

To reduce any bias that could be introduced from training on the dataset by Ciurea and Funt, we repeated the image classifier training on the reprocessed ground truth dataset by Gehler [7]. During training, care was taken to exclude the ground truth information from the data. Repeating the remaining classification yielded a best result of 54.5% on two-fold cross-validation, or 53.5% for five-fold cross-validation.

Figure 4.16 shows the ROC curves for both methods. The results of our method clearly outperform the state-of-the-art. However, these results also underline the challenge in exploiting illuminant color as a forensic cue on real-world images. Thus, we hope our database will have a significant impact in the development of new illuminant-based forgery detection algorithms.

4.3.6 Detection after Additional Image Processing

We also evaluated the robustness of our method against different processing operations. The results are computed on DSO-1. Apart from the additional preprocessing steps, the evaluation protocol was identical to the one described in Section 4.3.3. In a first experiment, we examined the impact of JPEG compression. Using `libJPEG`, the images were recompressed at the JPEG quality levels 70, 80 and 90. The detection rates were 63.5%, 64% and 69%, respectively. Using `imagemagick`, we conducted a second experiment adding per image a random amount of Gaussian noise, with an attenuated value varying between 1% and 5%.On average, we obtained an accuracy of 59%. Finally, again using `imagemagick`, we randomly varied the brightness and/or contrast of the image by either +5% or −5%. These brightness/contrast manipulations resulted in an accuracy of 61.5%.

These results are expected. For instance, the performance deterioration after strong JPEG compression introduces blocking artifacts in the segmentation of the illuminant maps. One could consider compensating for the JPEG artifacts with a deblocking algorithm. Still, JPEG compression is known to be a challenging scenario in several classes of forensic algorithms [72, 53, 63]

One could also consider optimizing the machine-learning part of the algorithm. However, here, we did not fine-tune the algorithm for such operations, as postprocessing can be addressed by specialized detectors, such as the work by Bayram et al. for brightness and

[7] L. Shi and B. Funt. Re-processed Version of the Gehler Color Constancy Dataset of 568 Images. `http://www.cs.sfu.ca/~colour/data/shi_gehler/`, January 2011.

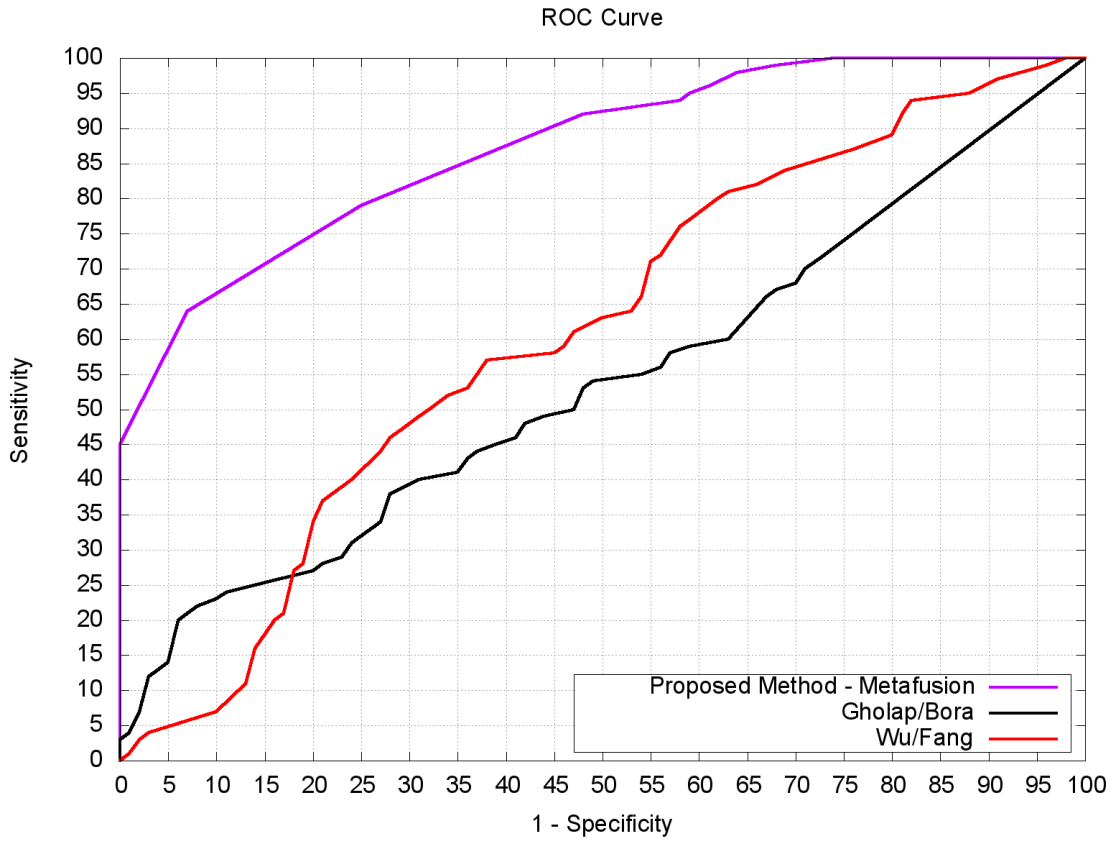


Figure 4.16: Comparative results between our method and state-of-the-art approaches performed using DSO-1.

contrast changes [5], combined with one of the recent JPEG-specific algorithms (e. g., [6]).

4.3.7 Performance of Forgery Detection using a Cross-Database Approach

To evaluate the generalization of the algorithm with respect to the training data, we followed an experimental design similar to the one proposed by Rocha et al. [75]. We performed a cross-database experiment, using DSO-1 as training set and the 50 images of DSI-1 (internet images) as test set. We used the pre-trained Metafusion classifier from the best performing fold in Section 4.3.3 without further modification. Figure 4.17 shows the ROC curve for this experiment. The results of this experiment are similar to the best ROC curve in Section 4.3.3, with an AUC of 82.6%. This indicates that the proposed method offers a degree of generalization to images from different sources and to faces of varying sizes.

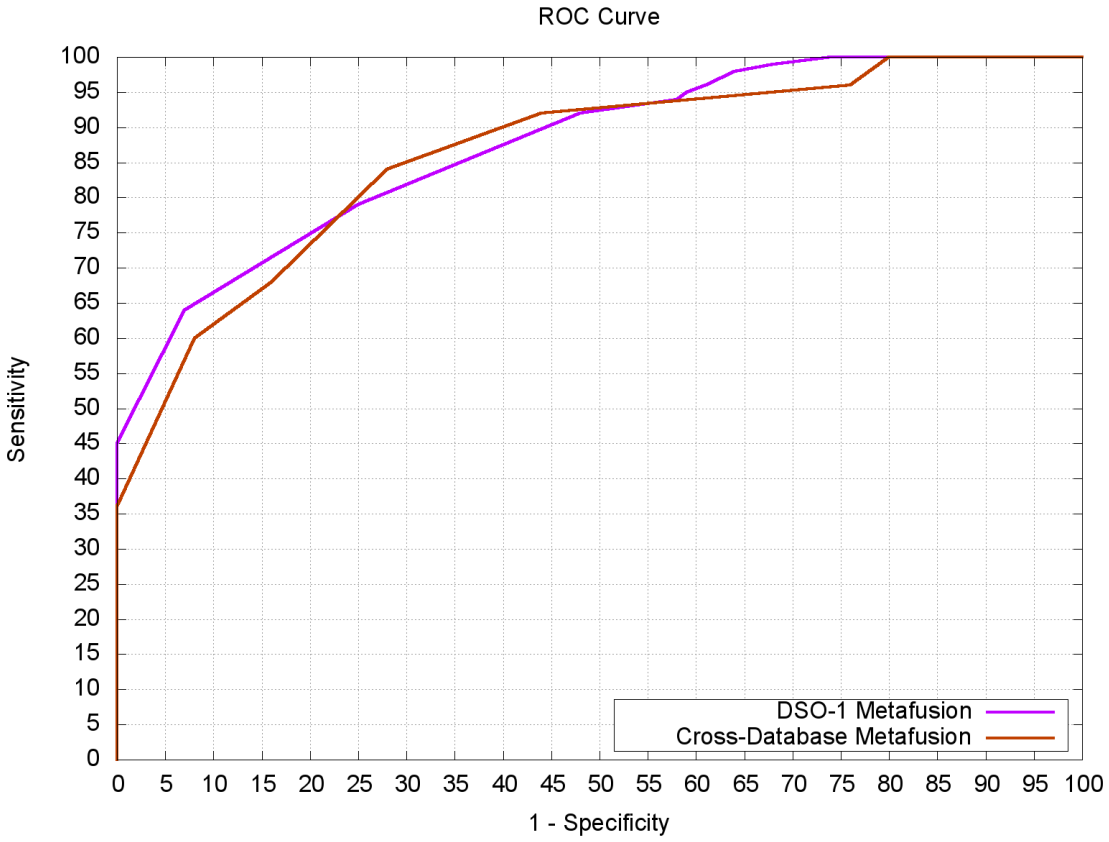


Figure 4.17: ROC curve provided by cross-database experiment.

4.4 Final Remarks

In this work, we presented a new method for detecting forged images of people using the illuminant color. We estimate the illuminant color using a statistical gray edge method and a physics-based method which exploits the inverse intensity-chromaticity color space. We treat these illuminant maps as texture maps. We also extract information on the distribution of edges on these maps. In order to describe the edge information, we propose a new algorithm based on edge-points and the HOG descriptor, called HOGedge. We combine these complementary cues (texture- and edge-baed) using machine learning late fusion. Our results are encouraging, yielding an AUC of over 86% correct classification. Good results are also achieved over internet images and under cross-database training/testing.

Although the proposed method is custom-tailored to detect splicing on images containing faces, there is no principal hindrance in applying it to other, problem-specific

materials in the scene.

The proposed method requires only a minimum amount of human interaction and provides a crisp statement on the authenticity of the image. Additionally, it is a significant advancement in the exploitation of illuminant color as a forensic cue. Prior color-based work either assumes complex user interaction or imposes very limiting assumptions.

Although promising as forensic evidence, methods that operate on illuminant color are inherently prone to estimation errors. Thus, we expect that further improvements can be achieved when more advanced illuminant color estimators become available.

Reasonably effective skin detection methods have been presented in the computer vision literature in the past years. Incorporating such techniques can further expand the applicability of our method. Such an improvement could be employed, for instance, in detecting pornography compositions which, according to forensic practitioners, have become increasingly common nowadays.

Chapter 5

Splicing Detection via Illuminant Maps: More than Meets the Eye

In the previous chapter, we have introduced a new method based on illuminant color analysis for detecting forgeries on image compositions containing people. However, its effectiveness still needed to be improved for real forensic applications. Furthermore, some important telltales, such as illuminants colors, have not been statistically analyzed in the method. In this chapter, we introduce a new method for analyzing illuminant maps, which uses more discriminative features and a robust machine learning framework able to determine the most complementary set of features to be applied in illuminant map analyses. Parts of the contents and findings in this chapter were submitted to a forensic journal [1].

5.1 Background

The method proposed in Chapter 4 is currently the state-of-the-art of methods based on inconsistencies in light color. Therefore, the background for this chapter is actually the whole Chapter 4. We refer the reader to that chapter for more details.

5.2 Proposed Approach

The approach proposed in this chapter have been developed to correct some drawbacks and mainly to achieve an improved accuracy over the approach presented in Chapter 4. This section describes in details each step of the improved image forgery detection approach.

[1] T. Carvalho, F. Faria, R. Torres, H. Pedrini, and A. Rocha. Splicing detection through color constancy maps: More than meets the eye. Submitted to Elsevier Forensics Science International (FSI), 2014.

5.2.1 Forgery Detection

Most of the times, the splicing detection process relies on the expert's experience and background knowledge. This process usually is time consuming and error prone once that image splicing is more and more sophisticated and an aural (e.g., visual) analysis may not be enough to detect forgeries.

Our approach to detecting image splicing, which is specific for detecting composites of people, is developed aiming at minimizing the user interaction. The splicing detection task performed by our approach consists in labelling a new image among two pre-defined classes (real and fake) and later pointing the face with higher probability to be the fake face. In this process, a classification model is created to indicate the class to which a new image belongs.

The detection methodology comprises four main steps:

1. **Description**: relies on algorithms to estimate IMs and extract image visual cues (e.g., color, texture, and shape), encoding the extracted information into feature vectors;

2. **Face Pair Classification**: relies on algorithms that use image feature vectors to learn intra- and inter-class patterns of the images to classify each new image feature vector;

3. **Forgery Classification**: consists in labelling a new image into one of existing known classes (real and fake) based on the previously learned classification model and description techniques;

4. **Forgery Detection**: once knowing that an image is fake, this task aims at identifying which face(s) are more likely to be fake in the image.

Figure 5.1 depicts a coarse view of our method which shall be refined later on.

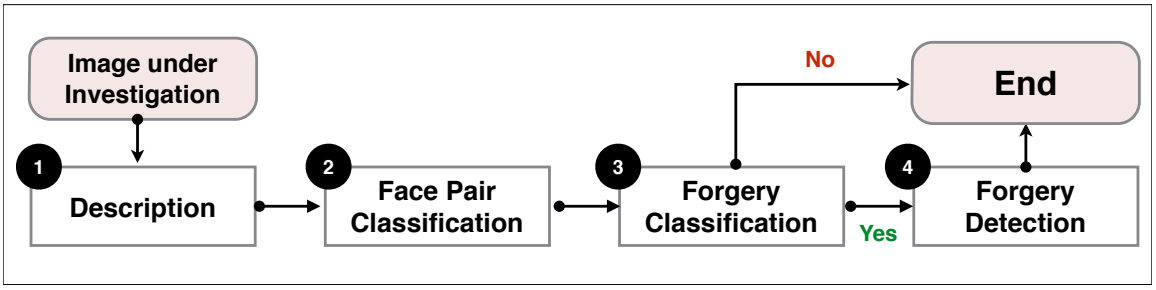


Figure 5.1: Overview of the proposed image forgery classification and detection methodology.

5.2.2 Description

Image descriptors have been used in many different problems in the literature, such as content-based image retrieval [52], medical image analysis [75], and geographical information systems analysis [24] to name just a few.

The method proposed in Chapter 4 represents an important step toward a better analysis of IMs given that most of the times, analyzing IMs directly to detect forgeries is not an easy task. Although effective, in Chapter 4, we just explored a limited range of image descriptors to develop an automatic forgery detector. Also, we did not explore many complementary properties in the analysis, restricting the investigation to only four different ways of IMs characterization.

Bearing in mind that in a real forensic scenario, an improved accuracy in fake detection is much more important than a real-time application, in this chapter, we propose to augment the description complexity of images in order to achieve an improved classification accuracy.

Our method employs a combination of different IMs, color spaces, and image descriptors to explore different and complementary properties to characterize images in the process of detecting fake images. This description process comprises a pipeline of five steps, which we describe next.

IM Estimation

In general, the literature describes two types of algorithms for estimating IMs: statistics-based and physics-based. They capture different information from image illumination and, here, these different types of information have been used to produce complementary features in the fake detection process.

For capturing statistical-based information, we use the generalized grayworld estimates algorithm (GGE) proposed by van de Weijer et al. [91]. This algorithm, estimates the illuminant $\mathbf{e}$ from pixels as

$$\lambda \mathbf{e}^{\mathbf{n},\mathbf{p},\sigma} = \left(\int \left| \frac{\partial^{\mathbf{n}} \Gamma^{\sigma}(\mathbf{x})}{\partial \mathbf{x}^{\mathbf{n}}} \right|^{\mathbf{p}} d\mathbf{x} \right)^{1/\mathbf{p}} \quad . \tag{5.1}$$

where $\mathbf{x}$ denotes a pixel coordinate, λ is a scale factor, $|\cdot|$ is the absolute value, ∂ the differential operator, $\Gamma^{\sigma}(\mathbf{x})$ is the observed intensities at position $\mathbf{x}$, smoothed with a Gaussian kernel σ, $\mathbf{p}$ is the Minkowski norm, and $\mathbf{n}$ is derivative order.

On the other hand, for capturing physics-based information, we use the inverse-intensity chromaticity space (IIC), an extension from the method proposed by Tan et al. [86], where the intensity $\Gamma_c(\mathbf{x})$ and the chromaticity $\chi_c(\mathbf{x})$ (i.e.,normalized

RGB-value) of a color channel $c \in \{R, G, B\}$ at position $\mathbf{x}$ is represented by

$$\chi_c(\mathbf{x}) = m(\mathbf{x}) \frac{1}{\sum_{i \in \{R,G,B\}} \Gamma_i(\mathbf{x})} + \gamma_c \tag{5.2}$$

In this equation, γ_c represents the chromaticity of the illuminant in channel c and $m(\mathbf{x})$ mainly captures geometric influences, i. e., light position, surface orientation, and camera position and is feasible approximate, as described in [86].

Choice of Color Space and Face Extraction

IMs are usually represented in RGB space, however, when characterizing such maps there is no hard constraint regarding its color space. Therefore, it might be the case that some properties present in the maps are better highlighted in alternative color spaces. Therefore, given that some description techniques are more suitable for specific color spaces, this step converts illuminant maps into different color spaces.

In Chapter 4, we have used IMs in YCbCr space only. In this chapter, we propose to augment the number of color spaces available in order to capture the smallest nuances present in such maps not visible in the original color spaces. For that, we consider additionally the Lab, HSV and RGB color spaces [1]. We have chosen such color spaces because Lab and HSV, as well as YCbCr, are color spaces that allow us to separate the illuminance channel from other chromaticity channels, which is useful when applying texture and shape descriptors. In addition, we have chosen RGB because it is the most used color space when using color descriptors and is a natural choice since most cameras capture images originally in such space.

Once we define a color space, we extract all faces present in the investigated image using a manual bounding box defined by the user.

Feature Extraction from IMs

From each extracted face in the previous step, we now need to find telltales that allow us to correctly identify splicing images. Such information is present in different visual properties (e.g., texture, shape and color, among others) on the illuminant maps. For that, we take advantage of image description techniques.

Texture, for instance, allows us to characterize faces whereby illuminants are disposed similarly when comparing two faces. The SASI [15] technique, that was used in Chapter 4, presented a good performance, therefore, we keep it in our current analysis. Furthermore, guided by the excellent results reported in a recent study by Penatti et al. [70], we also included LAS [87] technique. Complementarily, we also incorporated the Unser [90] descriptor, which presents a lower complexity and generates compact feature vectors when compared to SASI and LAS.

Unlike texture properties, shape properties present in fake faces have different pixel intensities when compared to shapes present in faces that originally belong to the analyzed image in an IM. In this sense, in Chapter 4, we proposed the HOGedge algorithm, which led to a classification AUC close to 70%. Due to its performance, here, we replace it by two other shape techniques, EOAC [65] and SPYTEC [60]. EOAC is based on shape orientations and correlation between neighboring shapes. These are properties that are potentially useful for forgery detection using IMs given that neighboring shape in regions of composed faces tend not to be correlated. We selected SPYTEC since it uses the wavelet transform, which captures multi-scale information normally not directly visible in the image.

According to Riess and Angelopoulou [73], when IMs are analyzed by an expert for detecting forgeries, the main observed feature is color. Thus, in this chapter, we decided to add color description techniques as an important visual cue to be encoded into the process of image description. The considered color description techniques are ACC [42], BIC [84], CCV [69] and LCH [85].

ACC is a technique based on color correlograms and encodes image spatial information. This is very important on IM analysis given that similar spatial regions (e.g., cheeks and lips) from two different faces should present similar colors in the map. BIC technique presents a simple and effective description algorithm, which reportedly presented good performance in the study carried out in Penatti et al. [70]. It captures border and interior properties in an image and encodes them in a quantized histogram. CCV is a segmentation-based color technique and we selected it because it is a well-known color technique in the literature and usually is used as a baseline in several analysis. Complementarily, LCH is a simple local color description technique which encodes color distributions of fixed-size regions of the image. This might be useful when comparing illuminants from similar regions in two different faces.

Face Characterization and Paired Face Features

Given that in this chapter we consider more than one variant of IMs, color spaces and description techniques, let $\mathcal{D}$ be an image descriptor composed of the triplet (IMs, color space, and description technique). Assuming all possible combinations of such triplets according to the IMs, color spaces and description techniques we consider herein, we have 54 different image descriptors. Table 5.1 shows all image descriptors used in this work.

Finally, to detect a forgery, we need to analyze whether a suspicious part of the image is consistent or not with other parts from the same image. Specifically, when we try to detect forgeries involving composites of people faces, we need to compare if a suspicious face is consistent with other faces in the image. In the worst case, all faces are suspicious and need to be compared to the others. Thus, instead of analyzing each

Table 5.1: Different descriptors used in this work. Each table row represents an image descriptor and it is composed of the combination (triplet) of an illuminant map, a color space (onto which IMs have been converted) and description technique used to extract the desired property.

IM	Color Space	Description Technique	Kind
GGE	Lab	ACC	Color
GGE	RGB	ACC	Color
GGE	YCbCr	ACC	Color
GGE	Lab	BIC	Color
GGE	RGB	BIC	Color
GGE	YCbCr	BIC	Color
GGE	Lab	CCV	Color
GGE	RGB	CCV	Color
GGE	YCbCr	CCV	Color
GGE	HSV	EOAC	Shape
GGE	Lab	EOAC	Shape
GGE	YCbCr	EOAC	Shape
GGE	HSV	LAS	Texture
GGE	Lab	LAS	Texture
GGE	YCbCr	LAS	Texture
GGE	Lab	LCH	Color
GGE	RGB	LCH	Color
GGE	YCbCr	LCH	Color
GGE	HSV	SASI	Texture
GGE	Lab	SASI	Texture
GGE	YCbCr	SASI	Texture
GGE	HSV	SPYTEC	Shape
GGE	Lab	SPYTEC	Shape
GGE	YCbCr	SPYTEC	Shape
GGE	HSV	UNSER	Texture
GGE	Lab	UNSER	Texture
GGE	YCbCr	UNSER	Texture

IM	Color Space	Description Technique	Kind
IIC	Lab	ACC	Color
IIC	RGB	ACC	Color
IIC	YCbCr	ACC	Color
IIC	Lab	BIC	Color
IIC	RGB	BIC	Color
IIC	YCbCr	BIC	Color
IIC	Lab	CCV	Color
IIC	RGB	CCV	Color
IIC	YCbCr	CCV	Color
IIC	HSV	EOAC	Shape
IIC	Lab	EOAC	Shape
IIC	YCbCr	EOAC	Shape
IIC	HSV	LAS	Texture
IIC	Lab	LAS	Texture
IIC	YCbCr	LAS	Texture
IIC	Lab	LCH	Color
IIC	RGB	LCH	Color
IIC	YCbCr	LCH	Color
IIC	HSV	SASI	Texture
IIC	Lab	SASI	Texture
IIC	YCbCr	SASI	Texture
IIC	HSV	SPYTEC	Shape
IIC	Lab	SPYTEC	Shape
IIC	YCbCr	SPYTEC	Shape
IIC	HSV	UNSER	Texture
IIC	Lab	UNSER	Texture
IIC	YCbCr	UNSER	Texture

image face separately, after building $\mathcal{D}$ for each face in the image, it encodes the feature vectors of each pair of faces under analysis into one feature vector. Given an image under investigation, it is characterized by the different feature vectors, and paired vectors $\mathcal{P}$ are created through direct concatenation between two feature vectors $\mathcal{D}$ of the same type for each face. Figure 5.2 depicts the full description pipeline.

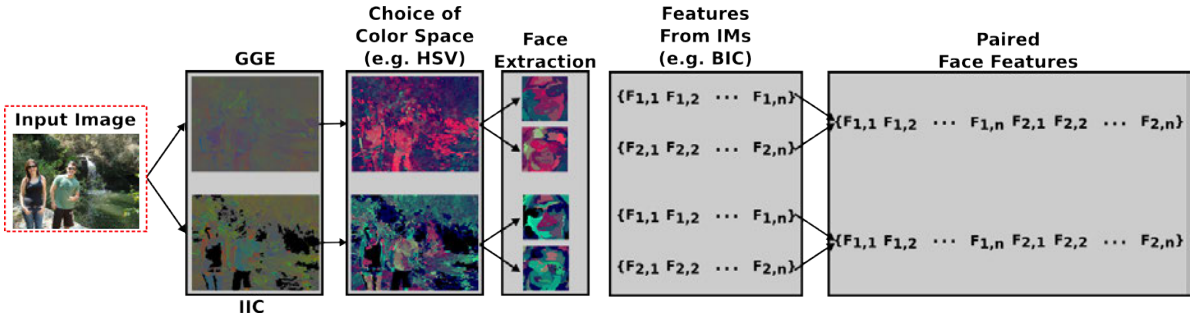


Figure 5.2: Image description pipeline. Steps *Choice of Color Spaces* and *Features From IMs* can use many different variants which allow us to characterize IMs gathering a wide range of cues and telltales.

5.2.3 Face Pair Classification

In this section, we show details about the classification step. When using different IMs, color spaces, and description techniques, the obvious question is how to automatically select the most important ones to keep and combine them toward an improved classification performance. For this purpose, we take advantage of the classifier selection and fusion introduced in Faria et al. [28].

Classifier Selection and Fusion

Let $\mathcal{C}$ be a set of classifiers in which each classifier $\mathbf{c}_j \in \mathcal{C}$ $(1 < j \leq |\mathcal{C}|)$ is composed of a tuple comprising a learning method (e.g., Naïve Bayes, k-Nearest Neighbors, and Support Vector Machines) and a single image descriptor $\mathcal{P}$.

Initially, all classifiers $\mathbf{c}_j \in \mathcal{C}$ are trained on the elements of a training set T. Next, the outcome of each classifier on the validation set V, different from T, is computed and stored into a matrix M_V, where $|M_V| = |V| \times |\mathcal{C}|$ and $|V|$ is the number of images in a validation set V. The actual training and validation data points are known *a priori*.

In the following, M_V is used as input to select a set $\mathcal{C}^* \subset \mathcal{C}$ of classifiers that are good candidates to be combined. To perform it, for each par of classifier $(\mathbf{c}_i, \mathbf{c}_j)$ we calculate five diversity measures

$$\mathrm{COR}(\mathbf{c}_i, \mathbf{c}_j) = \frac{\mathbf{ad} - \mathbf{bc}}{\sqrt{(\mathbf{a}+\mathbf{b})(\mathbf{c}+\mathbf{d})(\mathbf{a}+\mathbf{c})(\mathbf{b}+\mathbf{d})}}, \tag{5.3}$$

$$\mathrm{DFM}(\mathbf{c}_i, \mathbf{c}_j) = \mathbf{d}, \tag{5.4}$$

$$\mathrm{DM}(\mathbf{c}_i, \mathbf{c}_j) = \frac{\mathbf{b}+\mathbf{c}}{\mathbf{a}+\mathbf{b}+\mathbf{c}+\mathbf{d}}, \tag{5.5}$$

$$\mathrm{IA}(\mathbf{c}_i, \mathbf{c}_j) = \frac{\mathbf{2(ac} - \mathbf{bd})}{(\mathbf{a}+\mathbf{b})(\mathbf{c}+\mathbf{d}) + (\mathbf{a}+\mathbf{c})(\mathbf{b}+\mathbf{d})}, \tag{5.6}$$

$$\text{QSTAT}(\mathbf{c}_i, \mathbf{c}_j) = \frac{\mathbf{ad} - \mathbf{bc}}{\mathbf{ad} + \mathbf{bc}}, \tag{5.7}$$

where COR is Correlation Coefficient p, DFM is Double-Fault Measure, DM is Disagreement Measure, IA is Interrater Agreement k and QSTAT is Q-Statistic [57]. Furthermore, $\mathbf{a}$ is the amount of correctly classified images in the validation set by both classifiers. The vectors $\mathbf{b}$ and $\mathbf{c}$ are, respectively, the amount of images correctly classified by $\mathbf{c}_j$ but missed by $\mathbf{c}_i$ and amount of images correctly classified by $\mathbf{c}_i$ but missed by $\mathbf{c}_j$. Lastly, $\mathbf{d}$ is the amount of images misclassified by both classifiers.

These diversity measures now represent scores for pairs of classifiers. A ranked list, which is sorted based on pairs of classifiers scores, is created. As last step of the selection process, a subset of classifiers is chosen using this ranked list with the mean threshold of the pair. In other words, diversity measures are computed to achieve the degree of agreement/disagreement between all available classifiers in $\mathcal{C}$ set. Finally, $\mathcal{C}^*$, containing the most frequent and promising classifiers, are selected [28].

Given a set of paired feature vectors $\mathcal{P}$ of two faces extracted from a new image I, we use each classifier $c_b \in \mathcal{C}^*$ $(1 < b \leq |\mathcal{C}^*|)$ to determine the label (forgery or real) of these feature vectors, producing b outcomes. The b outcomes are used as input of a fusion technique (in this case, majority voting) that takes the final decision regarding the definition of each paired feature vector $\mathcal{P}$ extracted from I.

Figure 5.3 depicts a fine-grained view of the forgery detection framework. Figure 5.3(b) shows the entire classifier selection and fusion process.

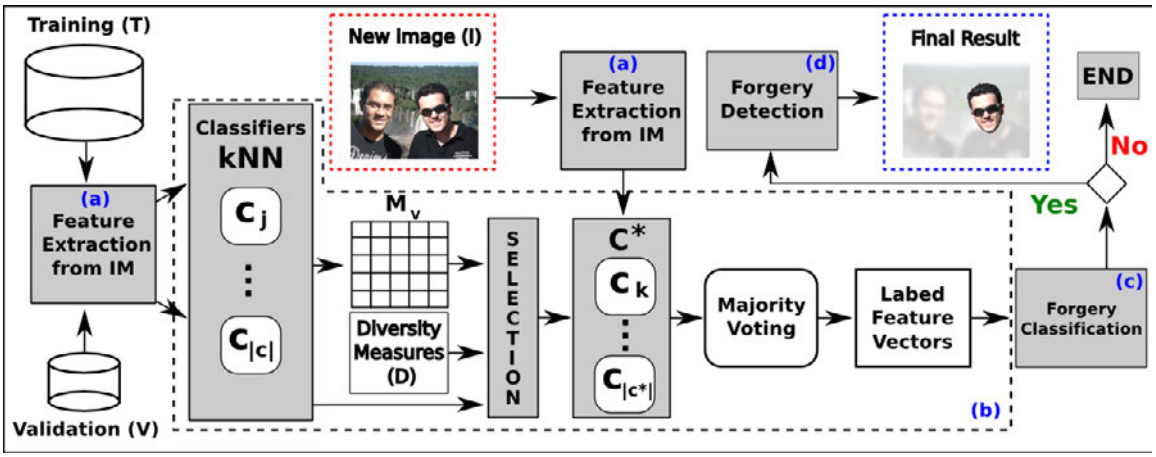


Figure 5.3: Proposed framework for detecting image splicing.

We should point out that the fusion technique used in the original framework [28] has been exchanged from support vector machines to majority voting. It's because when the original framework has been used, the support vector machines technique created a model very specialized for detecting original images, which increased the number of false

negatives. However, in a real forensic scenario we look for decreasing the false negative rate and to achieve it, we adopted a majority voting technique as an alternative.

5.2.4 Forgery Classification

It is important to notice that, sometimes, one image I is described by more than one paired vector $\mathcal{P}$ given that it might have more than two people present. Given an image I that contains q people, it is characterized by a set $\mathcal{S} = \{\mathcal{P}_1, \mathcal{P}_2, \cdots, \mathcal{P}_m\}$ being $m = \frac{q \times (q-1)}{2}$ and $q \geq 2$. In cases of $m \geq 2$, we adopt a strategy that prioritizes forgery detection. Hence, if any paired feature vector $\mathcal{P} \in \mathcal{S}$ is classified as fake, we classify the image I as a fake image. Otherwise, we classify it as pristine or non-fake.

5.2.5 Forgery Detection

Given an image I, already classified as fake in the first part of the method, it is important to refine the analysis and point out which part of the image is actually the result of a composition. This step was overlooked in the approach presented in Chapter 4. To perform such task, we can not use the same face pair feature vectors used in the last step (*Forgery Classification*), since we would find the pair with highest probability instead of the face with highest probability to be fake.

When analyzing the IMs, we realized that, in an image containing just pristine faces, the difference between colors depicted by GGE and IIC at the same position from same face is small. Notwithstanding, when an image contains a fake face, the difference between colors depicted by GGE and IIC, at the same position, is increased for this particular face. Figure 5.4 depicts an example of this fact.

In addition, we observed that, unlike colors, the superpixels disposal in both maps are very similar, for pristine and fake faces, resulting in faces with very similar texture and shapes in both GGE and IIC maps. This similarity fact makes the difference between GGE and IIC for texture and shape almost inconspicuous.

Despite not being sufficient for classifying an image as fake or not, since such variation may be very soft sometimes, this singular color changing characteristic helped us to develop a method for detecting the face with highest probability to be fake.

Given an image already classified as fake (see Section 5.2.4), we propose to extract, for each face into the image, its GGE and IIC maps, convert them into the desired color space, and use a single image color descriptor to extract feature vectors from GGE and from IIC. Then, we calculate the Manhattan distance between these two feature vectors which will result in a special feature vector that roughly measures how GGE and IIC from the same face are different, in terms of illuminant colors, considering the chosen color feature vector. Then, we train a Support Vector Machine (SVM) [8] with a radial

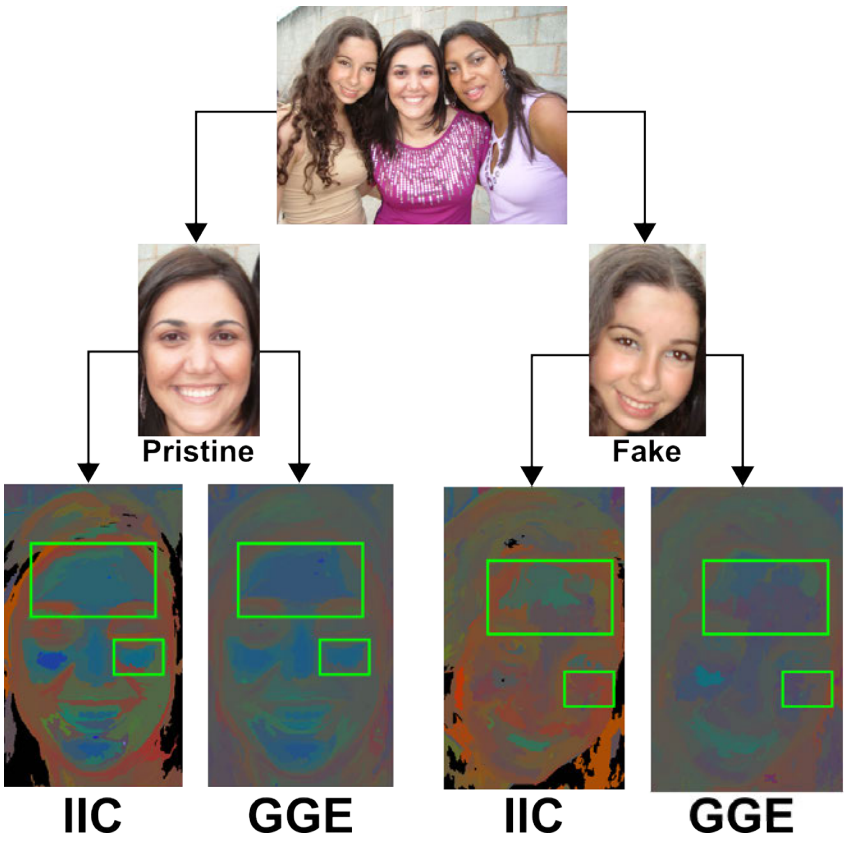


Figure 5.4: Differences in ICC and GGE illuminant maps. The highlighted regions exemplify how the difference between ICC and GGE is increased on fake images. On the forehead of the person highlighted as pristine (a person that originally was in the picture), the difference between the colors of IIC and GGE, in similar regions, is very small. On the other hand, on the forehead of the person highlighted as fake (an alien introduced into the image), the difference between the colors of IIC and GGE is large (from green to purple). The same thing happens in the cheeks.

basis function (RBF) kernel to give us probabilities of being fake for each analyzed face. The face with the highest probability of being fake is pointed out as the fake face from the image.

It is important to note here this trained classifier is specially trained to favor the fake class, therefore, it must be used after the forgery classification step described earlier.

5.3 Experiments and Results

This section describes the experiments we performed to show the effectiveness of the proposed method as well as to compare it with results presented on Chapter 4 counterparts.

We performed six rounds of experiments.

Round 1 is intended to show the best k-nearest neighbor (kNN) classifier to be used in additional rounds of tests. Instead of focusing on a more complex and complicated classifier, we select the simplest one possible for the individual learners in order to show the power of the features we employ as well as the utility of our proposed method for selecting the most appropriate combinations of features, color spaces, and IMs. Rounds 2 and 3 of experiments aim at exposing the proposed method behavior under different conditions. In these two rounds of experiments, we employed a 5-fold cross validation protocol in which we hold four folds for training and one for testing cycling the testing sets five times to evaluate the classifiers variability under different training sets. Round 4 explores the ability of the proposed method to find the actual forged face in an image, whereas Round 5 shows specific tests with original and montage photos from the Internet. Finally, Round 6 shows a qualitative analysis of famous cases involving questioned images.

5.3.1 Datasets and Experimental Setup

To provide a fair comparison with experiments performed on Chapter 4, we have used the same datasets DSO-1 and DSI-1[2].

DSO-1 dataset is composed of 200 indoor and outdoor images, comprising 100 original and 100 fake images, with an image resolution of $2,048 \times 1,536$ pixels. DSI-1 dataset is composed of 50 images (25 original and 25 doctored) downloaded from the Internet with different resolutions. In addition, we have used the same users' marks of faces as Chapter 4. Figure 5.5 (a) and (b) depict examples of DSO-1 dataset, whereas Figure 5.5 (c) and (d) depict examples of DSI-1 dataset.

We have used the 5-fold cross-validation protocol, which allowed us to report results that are directly and easily comparable in the testing scenarios.

Another important point of this chapter is the form we present the obtained results. We use the average accuracy across the five 5-fold cross-validation protocol and its standard deviation. However, to be comparable with the results reported in Chapter 4, we also present Sensitivity (which represents the number of true positives or the number of fake images correctly classified) and Specificity (which represents the number of true negatives or the number of pristine images correctly classified).

For all image descriptors herein, we have used the standard configuration proposed by Penatti et al. [70].

[2]http://ic.unicamp.br/ ∼rocha/pub/downloads/2014-tiago-carvalho-thesis/fsi-database.zip

(a) Pristine

(b) Fake

(c) Pristine

(d) Fake

Figure 5.5: Images (a) and (b) depict, respectively, examples of pristine and fake images from DSO-1 dataset, whereas images (c) and (d) depict, respectively, examples of pristine and fake images from DSI-1 dataset.

5.3.2 Round #1: Finding the best kNN classifier

After characterizing an image with a specific image descriptor, we need to choose the appropriate learning method to perform the classification. The method proposed here focuses on using complementary information to describe the IMs. Therefore, instead of using a powerful machine learning classifier such as Support Vector Machines, we use a simple learning method, the k-Nearest Neighbor (kNN) [8]. Another advantage is that, in a dense space as ours, which comprises many different characterization techniques, kNN classifier tends to present an improved behavior achieving efficient and effective results. However, even with a simple learning method such as as kNN, we still need to determine the most appropriate value for parameter k. This round of experiments aims at exploring

best k which will be used in the remaining set of experiments.

For this experiment, to describe each paired vector of the face $\mathcal{P}$, we have extracted all image descriptors from IIC in color space YCbCr. This configuration has been chosen because it was one of the combinations proposed in Chapter 4 and because the IM produced by IIC was used twice in the metafusion explained in Section 4.2.8. We have used DSO-1 with a 5-fold cross-validation protocol from which three folds are used for training, one for validation and one for testing.

Table 5.2 shows the results for the entire set of image descriptors we consider herein. kNN-5 and kNN-15 yielded the best classification accuracies in three of the image descriptors. As we mentioned before, this chapter focuses on looking for best complementary ways to describe IMs. Hence, we decided to choose kNN-5 that is simpler and faster than the alternatives. From now on, all the experiments reported in this work considers the kNN-5 classifier.

Table 5.2: Accuracy computed for kNN technique using different k values and types of image descriptors. Performed experiments using validation set and 5-fold cross-validation protocol have been applied. All results are in %.

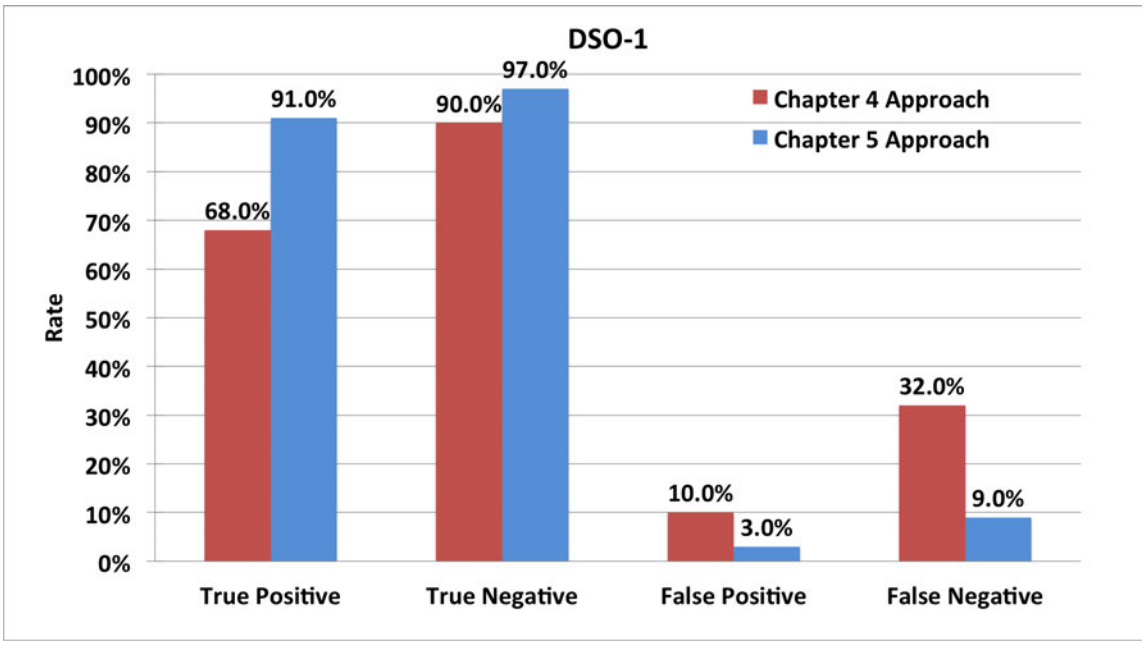
Descriptors	**kNN-**1	**kNN-**3	kNN-5	**kNN-**7	**kNN-**9	**kNN-**11	**kNN-**13	**kNN-**15
ACC	72.0	72.8	73.0	72.5	**73.8**	72.6	73.3	73.5
BIC	70.7	71.5	72.7	76.4	77.2	76.4	76.2	**77.3**
CCV	70.9	70.7	**74.0**	75.0	72.5	72.2	71.5	71.8
EOAC	64.8	65.4	**65.5**	65.2	63.9	61.7	61.9	60.7
LAS	67.3	69.1	71.0	**72.3**	72.2	71.5	71.2	70.3
LCH	61.9	**64.0**	62.2	62.1	63.7	62.2	63.7	63.3
SASI	67.9	70.3	**71.6**	69.9	70.1	70.3	69.9	69.4
SPYTEC	63.0	62.4	62.7	64.5	64.5	64.5	65.4	**66.5**
UNSER	65.0	66.9	67.0	67.8	67.1	67.9	68.5	**69.7**

5.3.3 Round #2: Performance on DSO-1 dataset

We now apply the proposed method for classifying an image as fake or real (the actual detection/localization of the forgery shall be explored in Section 5.3.5). For this experiment, we consider the DSO-1 dataset.

We have used all 54 image descriptors with kNN-5 learning technique resulting in 54 different classifiers. Recall that a classifier is composed of one descriptor and one learning technique. By using the modified combination technique we propose, we select the best combination $|\mathcal{C}^*|$ of different classifiers. Having tested different numbers of combinations, using $|\mathcal{C}^*| = \{5, 10, 15, \ldots, 54\}$, we achieve an average accuracy of 94.0% (with a Sensitivity of 91.0% and Specificity of 97.0%) with a standard deviation of 4.5% using all

54 classifiers $\mathcal{C}$. This result is 15 percentage points better than the result reported in Chapter 4 (despite reported result being an AUC of 86.0%, in the best operational point, with 68.0% of Sensitivity and 90.0% of Specificity, the accuracy is 79.0%). For better visualization, Figure 5.6 depicts a direct comparison between the accuracy of both results as a bar graph.



DSO-1 dataset

Figure 5.6: Comparison between results reported by the approach proposed in this chapter and the approach proposed in Chapter 4 over DSO-1 dataset. Note the proposed method is superior in true positives and true negatives rates, producing an expressive lower rate of false positives and false negatives.

Table 5.3 shows the results of all tested combinations of $|\mathcal{C}^*|$ on each testing fold and their average and standard deviation.

Given that the forensic scenario is more interested in a high classification accuracy than a real-time application (our method takes around three minutes to extract all features from an investigated image), the use of all 54 classifiers is not a major problem. However, the result using only the best subset of them ($|\mathcal{C}^*| = 20$ classifiers) achieves an average accuracy of 90.5% (with a Sensitivity of 84.0% and a Specificity of 97.0%) with a standard deviation of 2.1%, which is a remarkable result compared to the results reported on Chapter 4.

The selection process is performed as described in Section 5.2.3 and is based on the histogram depicted in Figure 5.7. The classifier selection approach takes into account

Table 5.3: Classification results obtained from the methodology described in Section 5.2 with a 5-fold cross-validation protocol for different number of classifiers ($|\mathcal{C}^*|$). All results are in %.

DSO-1 dataset											
Run	**Number of Classifiers $\|\mathcal{C}^*\|$**										
	5	**10**	**15**	**20**	**25**	**30**	**35**	**40**	**45**	**50**	**54 (ALL)**
1	90.0	85.0	92.5	**90.0**	90.0	95.0	90.0	87.5	87.5	90.0	92.5
2	90.0	87.5	87.5	**90.0**	90.0	90.0	87.5	90.0	90.0	90.0	90.0
3	95.0	92.5	92.5	**92.5**	95.0	95.0	95.0	95.0	95.0	95.0	97.5
4	67.5	82.5	95.0	**92.5**	92.5	95.0	97.5	97.5	95.0	100.0	100.0
5	82.5	80.0	80.0	**87.5**	85.0	90.0	90.0	90.0	87.5	87.5	90.0
Final(Avg)	85.0	85.5	89.5	**90.5**	90.5	92.0	92.0	91.0	91.0	92.5	94.0
Std. Dev.	10.7	4.8	6.0	**2.1**	3.7	2.7	4.1	4.1	3.8	5.0	4.5

both the accuracy performance of classifiers and their correlation.

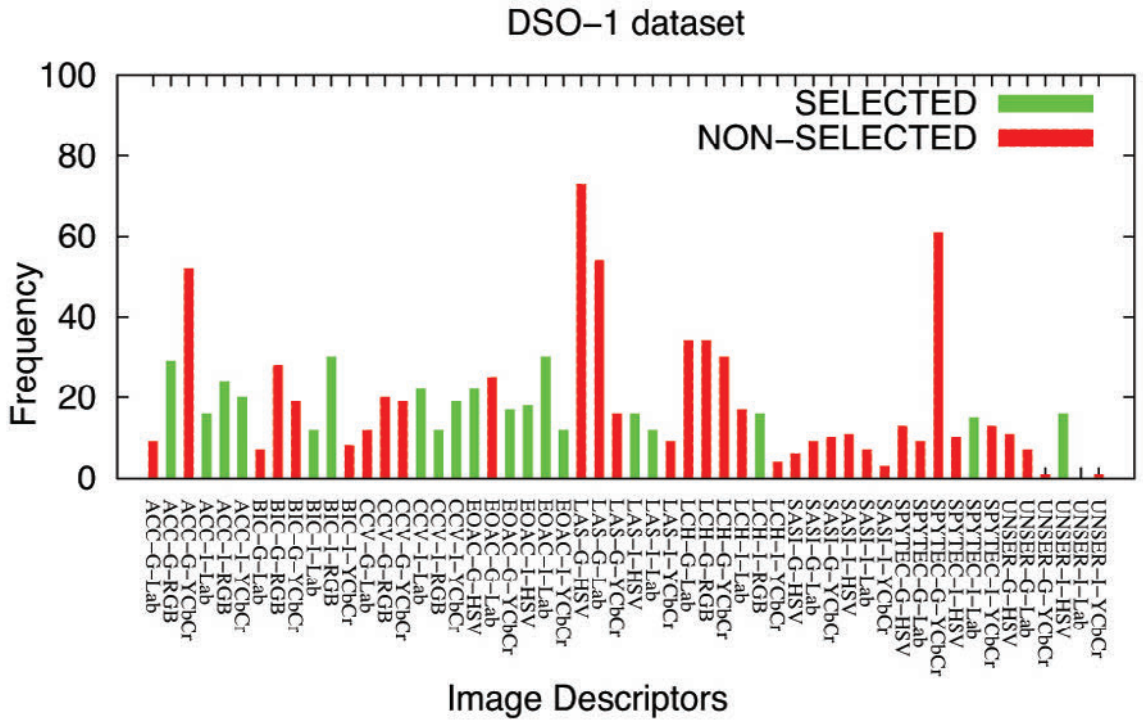


Figure 5.7: Classification histograms created during training of the selection process described in Section 5.2.3 for DSO-1 dataset.

Figure 5.8 depicts, in green, the $|\mathcal{C}^*|$ classifiers selected. It is important to highlight that all three kinds of descriptors (texture-, color- and shape-based ones) contribute for the best setup, reinforcing two of ours most important contributions in this chapter: the

importance of complementary information to describe the IMs and the value of color descriptors in IMs description process.

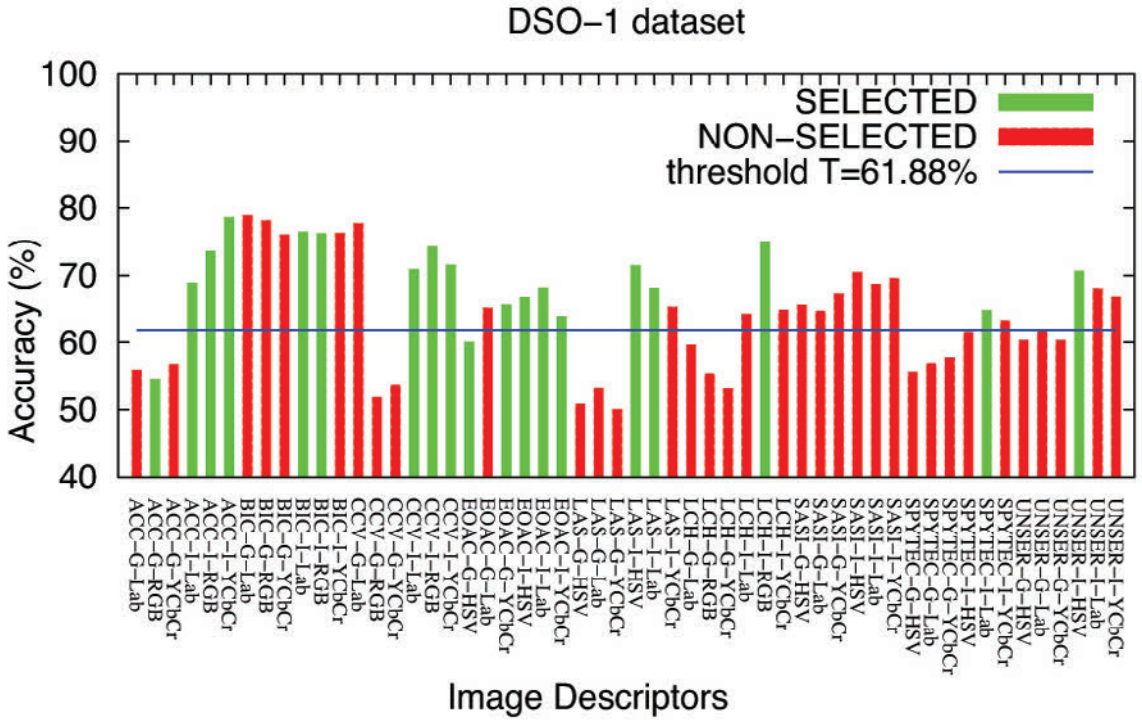


Figure 5.8: Classification accuracies of all non-complex classifiers (kNN-5) used in our experiments. The blue line shows the actual threshold $\mathcal{T}$ described in Section 5.2 used for selecting the most appropriate classification techniques during training. In green, we highlight the 20 classifiers selected for performing the fusion and creating the final classification engine.

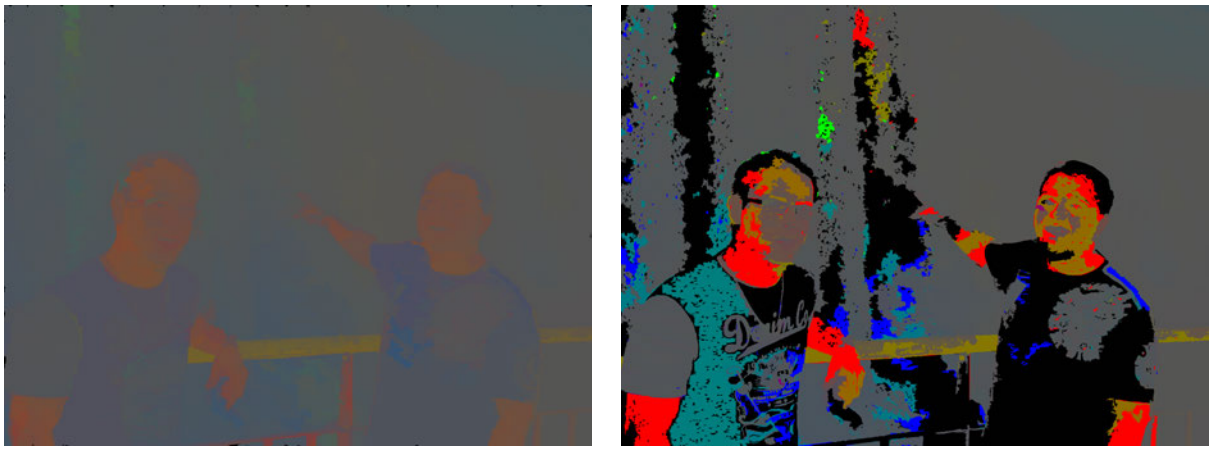
5.3.4 Round #3: Behavior of the method by increasing the number of IMs

Our method explores two different and complementary types of IMs: statistical-based and physics-based. However, these two large classes of IMs encompass many different methods than just IIC (physics) and GGE (statistics). However, many of them, such as [32], [38] and [34], are strongly dependent on a training stage. This kind of dependence in IMs estimation could restrict the applicability of the method, so we avoid using such methods in our IM estimation.

On other hand, it is possible to observe that, when we change the parameters $\mathbf{n}$, $\mathbf{p}$, and σ in Equation 5.1, different types of IMs are created. Our GGE is generated using $\mathbf{n} = 1$,

$\mathbf{p} = 1$, and $\sigma = 3$, whose parameters have been determined in Chapter 4 experiments. However, according to Gijsenij and Gevers [34], the best parameters to estimate GGE for real world images are $\mathbf{n} = 0$, $\mathbf{p} = 13$, and $\sigma = 2$.

Therefore, in this round of experiments, we introduce two new IMs in our method: a GGE-estimated map using $\mathbf{n} = 0$, $\mathbf{p} = 13$, and $\sigma = 2$ (as discussed in Gijsenij and Gevers [34]), which we named RWGGE; and the White Patch algorithm proposed by [58], which is estimated through Equation 5.1 with $\mathbf{n} = 0$, $\mathbf{p} = \infty$, and $\sigma = 0$. Figures 5.9 (a) and (b) depict, respectively, examples of RWGGE and White Patch IMs.



(a) RWGGE IM (b) White Patch IM

Figure 5.9: (a) IMs estimated from RWGGE; (b) IMs estimated from White Patch.

After introducing these two new IMs in our pipeline, we end up with $\mathcal{C} = 108$ different classifiers instead of $\mathcal{C} = 54$ in the standard configuration. Hence, we have used the two best configurations found in Round #1 and #2 (20 and all $\mathcal{C}$ classifiers) to check if considering more IMs is effective to improve the classification accuracy. Table 5.4 shows the results for this experiment.

The results show that the inclusion of a larger number of IMs does not necessarily increase the classification accuracy of the method. White Patch map, for instance, introduces too much noise since the IM estimation is now saturated in many parts of face. RWGGE, on other hand, produces a homogeneous IM in the entire face, which decreases the representation of the texture descriptors, leading to a lower final classification accuracy.

5.3.5 Round #4: Forgery detection on DSO-1 dataset

We now use the proposed methodology in Section 5.2.5 to actually detect the face with the highest probability of being the fake face in an image tagged as fake by the classifier

Table 5.4: Classification results for the methodology described in Section 5.2 with a 5-fold cross-validation protocol for different number of classifiers ($|\mathcal{C}^*|$) exploring the addition of new illuminant maps to the pipeline. All results are in %.

DSO-1 dataset		
Run	**Number of Classifiers $\|\mathcal{C}^*\|$**	
	20	**108**
1	90.0	82.5
2	87.5	90.0
3	92.5	90.0
4	95.0	87.5
5	85.0	82.5
Final(Avg)	**90.0**	86.5
Std. Dev.	3.9	3.8

previously proposed.

First we extract each face ϕ of an image I. For each ϕ, we estimate the illuminant maps IIC and GGE, keeping it on RGB color space and describing it by using a color descriptor (e.g., BIC). Once each face is described by two different feature vectors, one extracted from IIC and one extracted from GGE, we create the final feature vector that describes each ϕ as the difference, through Manhattan distance, between these two vectors.

Using the same 5-fold cross-validation protocol, we now train an SVM[3] classifier using an RBF kernel and with the option to return the probability of each class after classification. At this stage, our priority is to identify fake faces, so we increase the weight of the fake class to ensure such priority. We use a weight of 1.55 for fake class and 0.45 for pristine class (in LibSVM the sum of both weight classes needs to be 2). We use the standard grid search algorithm for determining the SVM parameters during training.

In this round of experiments, we assume that I has already been classified as fake by the classifier proposed in Section 5.2. Therefore, we just apply the fake face detector over images already classified as fake images. Once all the faces have been classified, we analyze the probability for fake class reported by the SVM classifier for each one of them. The face with the highest probability is pointed out as the most probable of being fake.

Table 5.5 reports the detection accuracy for each one of the color descriptors used at this round of experiments.

It is important to highlight here that sometimes an image has more than one fake face. However, the proposed method currently points out only the one with the highest probability to be fake. We are now investigating alternatives to extend this for additional

[3]We have used LibSVM implementation http://www.csie.ntu.edu.tw/∼cjlin/libsvm/ with its standard configuration (As of Jan. 2014).

Table 5.5: Accuracy for each color descriptor on fake face detection approach. All results are in %.

Descriptors	Accuracy (Avg.)	Std. Dev.
ACC	76.0	5.8
BIC	**85.0**	**6.3**
CCV	83.0	9.8
LCH	69.0	7.3

faces.

5.3.6 Round #5: Performance on DSI-1 dataset

In this round of experiments, we repeat the setup proposed in Chapter 4. By using DSO-1 as training samples, we classify DSI-1 samples. In other words, we perform a cross-dataset validation in which we train our method with images from DSO-1 and test it against images from the Internet (DSI-1).

As described in Round #2, we classified each one of the 54 $\mathcal{C}$ classifiers from one image through a kNN-5 and we selected the best combination of them using the modified combination approach. We achieved an average classification accuracy of 83.6% (with a Sensitivity of 75.2% and a Specificity of 92.0%) with a standard deviation of 5.0% using 20 classifiers. This result is around 8 percentage points better than the result reported in Chapter 4 (reported AUC is 83.0%, however, the best operational point is 64.0% of Sensitivity and 88.0% of Specificity with a classification accuracy of 76.0%).

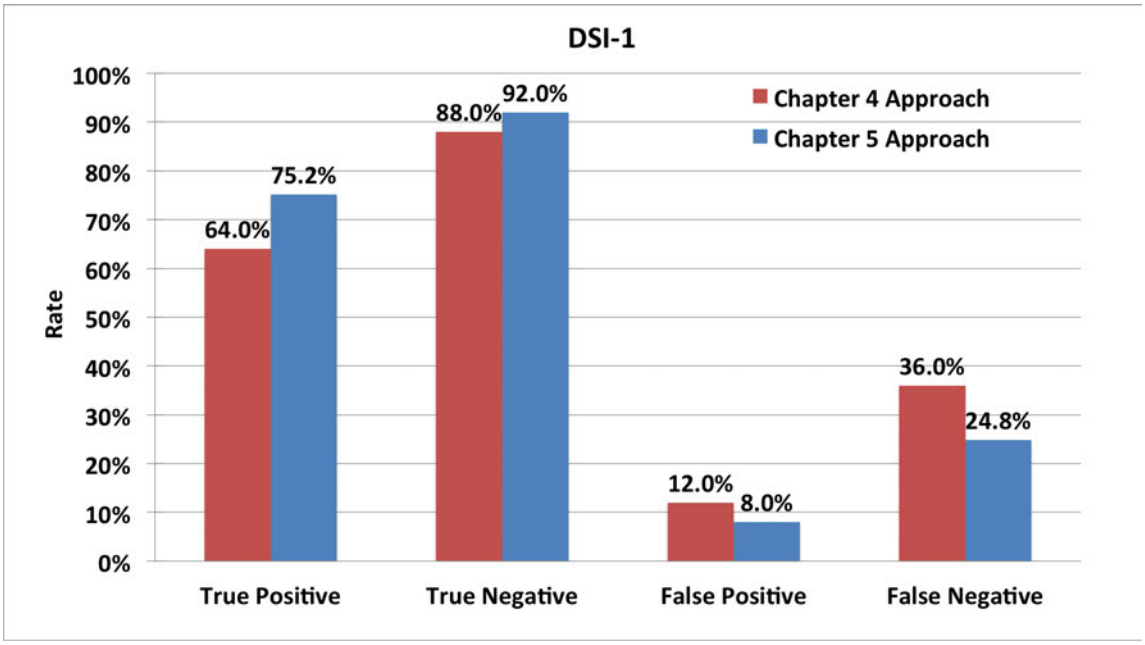
Table 5.6 shows the results of all tested combinations of $|\mathcal{C}^*|$ on each testing fold, as well as their average and standard deviation.

Table 5.6: Accuracy computed through approach described in Section 5.2 for 5-fold cross-validation protocol in different number of classifiers ($|\mathcal{C}^*|$). All results are in %.

DSI-1 dataset											
Run	**Number of Classifiers $\|\mathcal{C}^*\|$**										
	5	**10**	**15**	**20**	**25**	**30**	**35**	**40**	**45**	**50**	**54 (ALL)**
1	88.0	90.0	82.0	**92.0**	90.0	88.0	86.0	84.0	84.0	84.0	84.0
2	80.0	76.0	78.0	**80.0**	80.0	80.0	84.0	86.0	88.0	88.0	86.0
3	62.0	80.0	82.0	**82.0**	82.0	86.0	84.0	78.0	82.0	80.0	80.0
4	76.0	78.0	80.0	**80.0**	78.0	68.0	72.0	74.0	72.0	78.0	74.0
5	70.0	82.0	78.0	**84.0**	88.0	84.0	84.0	86.0	84.0	90.0	90.0
Final(Avg)	75.2	81.2	80.0	**83.6**	83.6	81.2	82.0	81.6	82.0	84.0	82.8
Std. Dev.	9.9	5.4	2.0	**5.0**	5.2	7.9	5.7	5.4	6.0	5.1	6.1

As introduced in Round # 2, we also show a comparison between our current results

and results obtained on Chapter 4 on DSI-1 dataset as a bar graph (Figure 5.10).



DSI-1 dataset

Figure 5.10: Comparison between current chapter approach and the one proposed in Chapter 4 over DSI-1 dataset. Current approach is superior in true positives and true negatives rates, producing an expressive lower rate of false positives and false negatives.

5.3.7 Round #6: Qualitative Analysis of Famous Cases involving Questioned Images

In this round of experiments, we perform a qualitative analysis of famous cases involving questioned images. To perform it, we use the previously trained classification models of Section 5.3.2. We classify the suspicious image using the model built for each training set and if any of them reports the image as a fake one, we classify it as ultimately fake.

Brazil's former president

On November 23, 2012 Brazilian Federal Police started an operation named *Safe Harbor operation*, which dismantled an undercover gang on federal agencies for fraudulent technical advices. One of the gang's leaders was Rosemary Novoa de Noronha [4].

[4]Veja Magazine. Operação Porto Seguro. `http://veja.abril.com.br/tema/operacao-porto-seguro`. Accessed: 2013-12-19.

Eagle to have their spot under the cameras and a 15-minute fame, at the same time, people started to broadcast on the Internet, images where Brazil's former president Luiz Inácio Lula da Silva appeared in personal life moments side by side with de Noronha. Shortly after, another image in exactly the same scenario started to be broadcasted, however, at this time, without de Noronha in the scene.

We analyzed both images, which are depicted in Figures 5.11 (a) and (b), using our proposed method. Figure 5.11 (a) has been classified as pristine on all five classification folds, while Figure 5.11 (b) has been classified as fake on all classification folds.

(a) Pristine

(b) Fake

Figure 5.11: Questioned images involving Brazil's former president. (a) depicts the original image, which has been taken by photographer Ricardo Stucker, and (b) depicts the fake one, whereby Rosemary Novoa de Noronha's face (left side) is composed with the image.

The situation room image

Another recent forgery that quickly spread out on the Internet was based on an image depicting the Situation Room [5] when the Operation Neptune's Spear, a mission against Osama bin Laden, was happening. The original image depicts U.S. President Barack Obama along with members of his national security team during the operation Neptune's Spear on May 1, 2011.

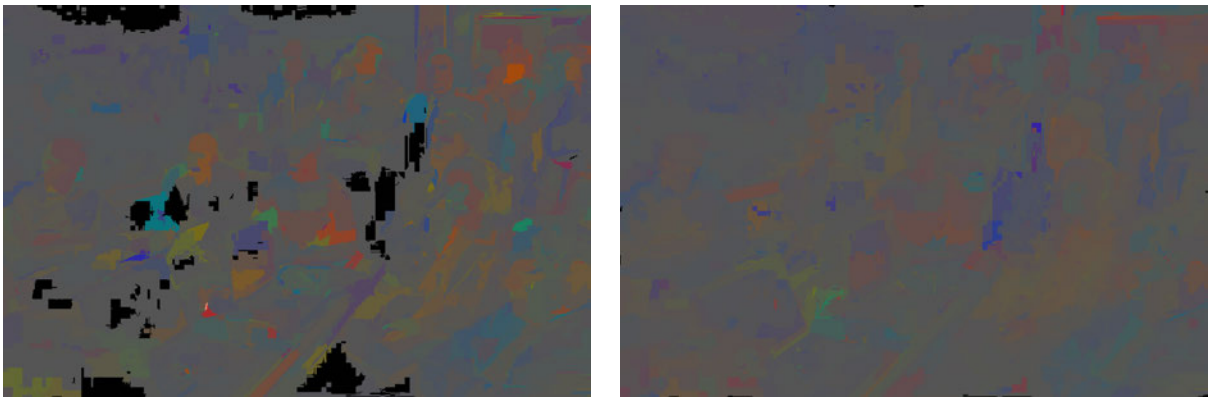
Shortly after the release of the original image, several fake images depicting the same scene had been broadcasted in the Internet. One of the most famous among them depicts Italian soccer player Mario Balotelli in the center of image.

[5]Original image from `http://upload.wikimedia.org/wikipedia/commons/a/ac/Obama_and_Biden_await_updates_on_bin_Laden.jpg` (As of Jan. 2014).

We analyzed both images, the original (officially broadcasted by the White House) and the fake one. Both images are depicted in Figures 5.12 (a) and (b).

(a) Pristine (b) Fake

Figure 5.12: The situation room images. (a) depicts the original image released by American government; (b) depicts one among many fake images broadcasted in the Internet.



(a) IIC (b) GGE

Figure 5.13: IMs extracted from Figure 5.12(b). Successive JPEG compressions applied on the image make it almost impossible to detect a forgery by a visual analysis of IMs.

Given that the image containing player Mario Balotelli has undergone several compressions (which slightly compromises IMs estimation), our method classifies this image as real in two out of the five trained classifiers. For the original one, all of the five classifiers pointed out the image as pristine. Since the majority of the classifiers pointed out the image with the Italian player as fake (3 out of 5), we decide the final class as fake which is the correct one.

Figures 5.13 (a) and (b) depict, respectively, IIC and GGE maps produced by the fake image containing Italian player Mario Balotelli. Just performing a visual analysis on these

maps is almost impossible to detect any pattern capable of indicating a forgery. However, once that our method explores complementary statistical information on texture, shape and color, it was able to detect the forgery.

Dimitri de Angeli's Case

In March 2013, Dimitri de Angelis was found guilty and sentenced to serve 12 years in jail for swindling investors in more than 8 million dollars. To garner the investor's confidence, de Angelis produced several images, side by side with celebrities, using Adobe Photoshop.

We analyzed two of these images: one whereby he is shaking hand of US former president Bill Clinton and other whereby he is side by side with former basketball player Dennis Rodman.

(a) Dennis Rodman.

(b) Bill Clinton

Figure 5.14: Dimitri de Angelis used Adobe Photoshop to falsify images side by side with celebrities.

Our method classified Figure 5.14(a) as a fake image with all five classifiers. Unfortunately, Figure 5.14(b) has been misclassified as pristine. This happened because this image has a very low resolution and has undergone strong JPEG compression harming the IM estimation. Then, instead of estimating many different local illuminants in many parts of the face, IMs estimate just a large illuminant comprising the entire face as depicted in Figures 5.15(a) and (b). This fact allied with a skilled composition which probably also performed light matching leading the faces to have compatible illumination (in Figure 5.14 (b) both faces have a frontal light) led our method to a misclassification.

(a) IIC (b) GGE

Figure 5.15: IMs extracted from Figure 5.14(b). Successive JPEG compressions applied on the image, allied with a very low resolution, formed large blocks of same illuminant, leading our method to misclassify the image.

5.4 Final Remarks

Image composition involving people is one of the most common tasks nowadays. The reasons vary from simple jokes with colleagues to harmful montages defaming or impersonating third parties. Independently on the reasons, it is paramount to design and deploy appropriate solutions to detect such activities.

It is not only the number of montages that is increasing. Their complexity is following the same path. A few years ago, a montage involving people normally depicted a person innocently put side by side with another one. Nowadays, complex scenes involving politicians, celebrities and child pornography are in place. Recently, we helped to solve a case involving a high profile politician purportedly having sex with two children according to eight digital photographs. A careful analysis of the case involving light inconsistencies checking as well as border telltales showed that all photographs were the result of image composition operations.

Unfortunately, although technology is capable of helping us solving such problems, most of the available solutions still rely on experts' knowledge and background to perform well. Taking a different path, in this paper we explored the color phenomenon of metamerism and how the appearance of a color in an image change under a specific type of lighting. More specifically, we investigated how the human material skin changes under different illumination conditions. We captured this behavior through image illuminants and creating what we call illuminant maps for each investigated image.

In the approaches proposed in Chapters 4 and 5, we analyzed illuminant maps entailing the interaction between the light source and the materials of interest in a scene. We expect that similar materials illuminated by a common light source have similar properties in such maps. To capture such properties, in this chapter we explored image descriptors that analyze color, texture and shape cues. The color descriptors identify if similar parts of the object are colored in the IM in a similar way since the illumination is common. The texture descriptors verify the distribution of colors through super pixels in a given region. Finally, shape descriptors encompass properties related to the object borders in such color maps. In Chapter 4, we investigated only two descriptors when analyzing an image. In this chapter, we presented a new approach to detecting composites of people that explore complementary information for characterizing images. However, instead of just stockpiling a huge number of image descriptors, we need to effectively find the most appropriate ones for the task. For that, we adapt an automatic way of selecting and combining the best image descriptors with their appropriated color spaces and illuminant maps. The final classifier is fast and effective for determining whether an image is real or fake.

We also proposed a method for effectively pointing out the region of an image that was forged. The automatic forgery classification, in addition to the actual forgery localization, represents an invaluable asset for forgery experts with a 94% classification rate, a remarkable 72% error reduction when compared to the method proposed in Chapter 4.

Future developments of this work may include the extension of the method for considering additional and different parts of body (e.g., all skin spots of the human body visible in an image). Given that our method compares skin material, it is feasible to use additional body parts, such as arms and legs, to increase the detection and confidence of the method.

Chapter 6

Exposing Photo Manipulation From User-Guided 3-D Lighting Analysis

The approaches presented in the previous chapters were specifically designed to detect forgeries involving people. However, sometimes an image composition can involve different elements. A car or a building can be introduced into the scene with specific purposes. In this chapter, we introduce our last contribution, which focuses on detecting 3-D light source inconsistencies in scenes with arbitrary objects using a user's guided approach. Parts of the contents and findings in this chapter will be submitted to an image processing conference [1].

6.1 Background

As previously described in Chapter 2, Johnson and Farid [45] have proposed an approach using 2-D light direction to detecting tampered images base based on some assumptions

1. all the analyzed objects have Lambertian surface;

2. surface reflectance is constant;

3. the object surface is illuminated by an infinitely distant light source.

The authors modeled the intensity of each pixel into the image as a relationship between the surface normals, light source position and and ambient light as

$$\Gamma(x, y) = \mathbf{R}(\vec{N}(x, y) \cdot \vec{L}) + A, \tag{6.1}$$

[1] T. Carvalho, H. Farid, and E. Kee. Exposing Photo Manipulation From User-Guided 3-D Lighting Analysis. Submitted to IEEE International Conference on Image Processing (ICIP), 2014.

where $\Gamma(x, y)$ is the intensity of the pixel at the position (x, y), $\mathbf{R}$ is the constant reflectance value, $\vec{N}(x, y)$ is the surface normal at the position (x, y), $\vec{L}$ is the light source direction and $\mathbf{A}$ is the ambient term.

Taking this model as starting point, the authors showed that light source position can be estimated by solving the following linear system

$$\begin{pmatrix} \vec{N}_x(x_1, y_1) & \vec{N}_y(x_1, y_1) & 1 \\ \vec{N}_x(x_2, y_2) & \vec{N}_y(x_2, y_2) & 1 \\ & \vdots & 1 \\ \vec{N}_x(x_p, y_p) & \vec{N}_y(x_p, y_p) & 1 \end{pmatrix} \begin{pmatrix} \vec{L}_x \\ \vec{L}_y \\ \mathbf{A} \end{pmatrix} = \begin{pmatrix} \Gamma(x_1, y_1) \\ \Gamma(x_1, y_1) \\ \vdots \\ \Gamma(x_1, y_1) \end{pmatrix} \tag{6.2}$$

where $\vec{N}_p = \{\vec{N}_x(x_p, y_p), \vec{N}_y(x_p, y_p)\}$ is the pth normal surface extracted along the occluding contour of some Lambertian object into the scene, $\mathbf{A}$ is the ambient term, $\Gamma(x_p, y_p)$ is the pixel intensity where normal surface has been extracted and $\vec{L} = \{\vec{L}_x, \vec{L}_y\}$ are the x and y components of light source direction.

However, since this solution just estimates 2-D light source direction, ambiguity in the answer can be embedded, many times compromising the effectiveness of performed analysis.

6.2 Proposed Approach

In the approach proposed in this chapter, we seek to estimate 3-D lighting from objects or people in a single image, relying on an analyst to specify the required 3-D shape from which lighting is estimated. To perform it, we describe a full work flow where first we use user-interface for obtaining these shape estimates. Secondly, we estimate 3-D lighting from these shapes estimates, performing a perturbation analysis that contends with any errors or biases in the user-specified 3-D shape and, finally, proposing a probabilistic technique for combining multiple lighting estimates to determine if they are physically consistent with a single light source.

6.2.1 User-Assisted 3-D Shape Estimation

The projection of a 3-D scene onto a 2-D image sensor results in a basic loss of information. Generally speaking, recovering 3-D shape from a single 2-D image is at best a difficult problem and at worst an under-constrained problem. There is, however, good evidence from the human perception literature that human observers are fairly good at estimating 3-D shape from a variety of cues including, foreshortening, shading, and familiarity [18, 54, 56, 88]. To this end, we ask an analyst to specify the local 3-D shape of surfaces. We have found that with minimal training, this task is relatively easy and accurate.

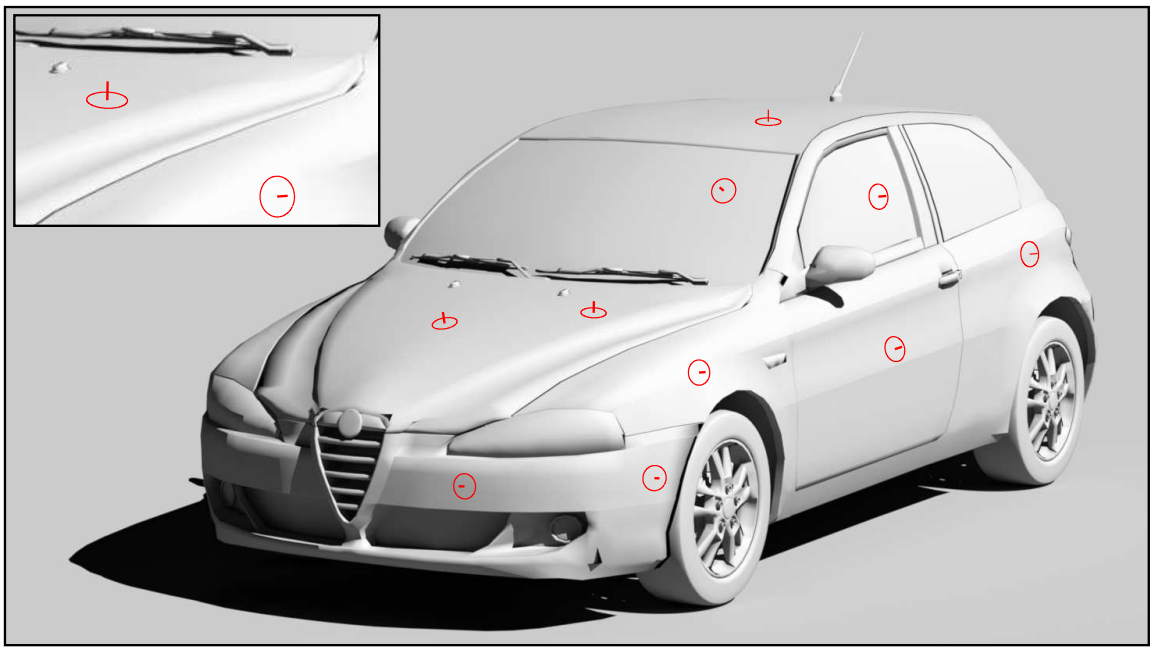


Figure 6.1: A rendered 3-D object with user-specified probes that capture the local 3-D structure. A magnified view of two probes is shown on the top right.

An analyst estimates the local 3-D shape at different locations on an object by adjusting the orientation of a small 3-D probe. The probe consists of a circular base and a small vector (the stem) orthogonal to the base. An analyst orients a virtual 3-D probe so that when the probe is projected onto the image, the stem appears to be orthogonal to the object surface. Figure 6.1 depicts an example of several such probes on a 3-D rendered model of a car.

With the click of a mouse, an analyst can place a probe at any point $\mathbf{x}$ in the image. This initial mouse click specifies the location of the probe base. As the analyst drags their mouse, he/she controls the orientation of the probe by way of the 2-D vector $\boldsymbol{v}$ from the probe base to the mouse location. This vector is restricted by the interface to have a maximum value of ϖ pixels, and is not displayed.

Probes are displayed to the analyst by constructing them in 3-D, and projecting them onto the image. The 3-D probe is constructed in a coordinate system that is local to the object, Figure 6.2, defined by three mutually orthogonal vectors

$$\mathbf{b}_1 = \begin{bmatrix} \mathbf{x} - \rho \\ f \end{bmatrix} \qquad \mathbf{b}_2 = \begin{bmatrix} \boldsymbol{v} \\ \frac{1}{f}\boldsymbol{v} \cdot (\mathbf{x} - \rho) \end{bmatrix} \qquad \mathbf{b}_3 = \mathbf{b}_1 \times \mathbf{b}_2 \quad , \tag{6.3}$$

where $\mathbf{x}$ is the location of the probe base in the image, and f and ρ are a focal length and principal point (discussed shortly). The 3-D probe is constructed by first initializing it into a default orientation in which its stem, a unit vector, is coincident with $\mathbf{b}_1$, and

the circular base lies in the plane spanned by $\mathbf{b}_2$ and $\mathbf{b}_3$, Figure 6.2. The 3-D probe is then adjusted to correspond with the analyst's desired orientation, which is uniquely defined by their 2-D mouse position $\boldsymbol{v}$. The 3-D probe is parameterized by a slant and tilt, Figure 6.2. The length of the vector $\boldsymbol{v}$ specifies a slant rotation, $\vartheta = \sin^{-1}(\|\boldsymbol{v}\|/\varpi)$, of the probe around $\mathbf{b}_3$. The tilt, $\varrho = \tan^{-1}(v_y/v_x)$, is embodied in the definition of the coordinate system (Equation 6.3).

The construction of the 3-D probe requires the specification of a focal length f and principal point ρ, Equation (6.3). There are, however, two imaging systems that need to be considered. The first is that of the observer relative to the display [19]. This imaging system dictates the appearance of the probe when it is projected into the image plane. In that case, we assume an orthographic projection with $\rho = \boldsymbol{0}$, as in [54, 18]. The second imaging system is that of the camera which recorded the image. This imaging system dictates how the surface normal $\vec{N}$ is constructed to estimate the lighting (Section 6.2.2). If the focal length and principal point are unknown then they can be set to a typical mid-range value, and $\rho = \boldsymbol{0}$.

The slant/tilt convention accounts for linear perspective, and for the analyst's interpretation of the photo [55, 19, 51]. A slant of 0 corresponds to a probe that is aligned with the 3-D camera ray, $\mathbf{b}_1$. In this case the probe stem projects to a single point within the circular base [55]. A slant of $\pi/2$ corresponds to a probe that lies on an occluding boundary in the photo. In this case, the probe projects to a T-shape with the stem coincident with the axis $\mathbf{b}_2$, and with the circular base laying in the plane spanned by axes $\mathbf{b}_1$ and $\mathbf{b}_3$. This 3-D geometry is consistent given the analyst's orthographic interpretation of a photo, as derived in [51].

With user-assisted 3-D surface normals in hand, we can now proceed with estimating the 3-D lighting properties of the scene.

6.2.2 3-D Light Estimation

We begin with the standard assumptions that a scene is illuminated by a single distant point light source (e.g., the sun) and that an illuminated surface is Lambertian and of constant reflectance. Under these assumptions, the intensity of a surface patch is given by

$$\Gamma \;=\; \vec{N} \cdot \vec{L} + \mathbf{A}, \tag{6.4}$$

where $\vec{N} = \{\vec{N}_x, \vec{N}_y, \vec{N}_z\}$ is the 3-D surface normal, $\vec{L} = \{L_x, L_y, L_z\}$ specifies the direction to the light source (the magnitude of $\vec{L}$ is proportional to the light brightness), and the constant ambient light term $\mathbf{A}$ approximates indirect illumination. Note that this expression assumes that the angle between the surface normal and light is less than 90°.

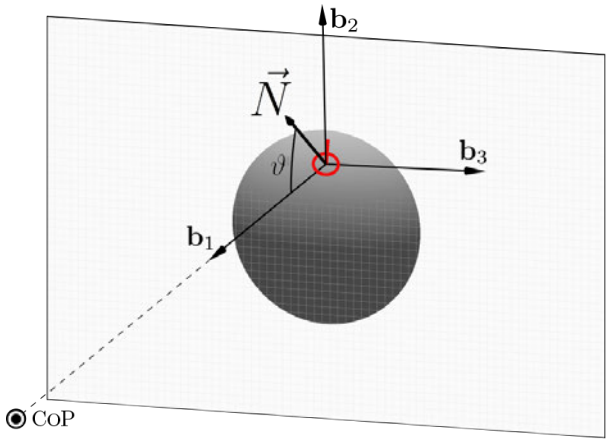


Figure 6.2: Surface normal obtained using a small circular red probe in a shaded sphere in the image plane. We define a local coordinate system by $\mathbf{b}_1$, $\mathbf{b}_2$, and $\mathbf{b}_3$. The axis $\mathbf{b}_1$ is defined as the ray that connects the base of the probe and the center of projection (CoP). The slant of the 3-D normal $\vec{N}$ is specified by a rotation ϑ around $\mathbf{b}_3$, while the normal's tilt ϱ is implicitly defined by the axes $\mathbf{b}_2$ and $\mathbf{b}_3$, Equation (6.3).

The four components of this lighting model (light direction and ambient term) can be estimated from $k \geq 4$ surface patches with known surface normals. The equation for each surface normal and corresponding intensity are packed into the following linear system

$$\begin{pmatrix} \vec{N}_1 & 1 \\ \vec{N}_2 & 1 \\ \vdots & \vdots \\ \vec{N}_k & 1 \end{pmatrix} \begin{pmatrix} \vec{L} \\ \mathbf{A} \end{pmatrix} = \mathbf{\Gamma} \tag{6.5}$$

$$\mathbf{N}\mathbf{b} = \mathbf{\Gamma} \ , \tag{6.6}$$

where $\mathbf{\Gamma}$ is a k-vector of observed intensity for each surface patch. The lighting parameters $\mathbf{b}$ can be estimated by using standard least squares

$$\mathbf{b} = (\mathbf{N}^T\mathbf{N})^{-1}\mathbf{N}^T\mathbf{\Gamma}, \tag{6.7}$$

where the first three components of $\mathbf{b}$ correspond to the estimated light direction. Because we assume a distant light source, this light estimate can be normalized to be unit sum and visualized in terms of azimuth $\Phi \in [-\pi, \pi]$ and elevation $\Upsilon \in [-\pi/2, \pi/2]$ given by

$$\Phi = \tan^{-1}(-L_x/L_z) \qquad \Upsilon = \sin^{-1}(L_y/\|\vec{L}\|). \tag{6.8}$$

In practice, there will be errors in the estimated light direction due to errors in the user-specified 3-D surface normals, deviations of the imaging model from our assumptions, signal-to-noise ratio in the image, etc. To contend with such errors, we perform a perturbation analysis yielding a probabilistic measure of the light direction.

6.2.3 Modeling Uncertainty

For simplicity, we assume that the dominant source of error is the analyst's estimate of the 3-D normals. A model for these errors is generated from large-scale psychophysical studies in which observers were presented with one of twelve different 3-D models, and asked to orient probes, such as those used here to specify the object shape [18]. The objects were shaded with a simple outdoor lighting environment. Using Amazon's Mechanical Turk, a total of $45,241$ probe settings from 560 observers were collected.

From this data, we construct a probability distribution for the actual light slant and tilt conditioned on the estimated slant and tilt. Specifically, for slant, our model considers the error between an input user slant value and its ground truth. For tilt, our model also considers the dependency between slant and tilt modelling tilt error relative to ground truth slant.

Figures 6.3 and 6.4 depict a view of our models as 2-D histograms. The color pallet in the right of each figure points out the probability of error for each bin. In tilt model, depicted by Figure 6.4, errors with higher probability are concentrated near 0 degrees horizontal axis (white line). In slant model, depicted by Figure 6.3, on the other hand, the errors are more spread out vertically, which point out that users are relatively better to estimate tilt but they are not so accurate on slant estimation.

We then model the uncertainty in the analyst's estimated light position using these error tables. This process can be described in three main steps:

1. randomly draw an error for slant (E_ϑ) and for tilt (E_ϱ) from previously constructed models;

2. for each one of these errors, weight it by some calculated weight. Specifically, this step have been inspired by the fact that user's behavior on slant and tilt perception is different. Also, empirically we have found that slant and tilt influence the estimation of light source position in different ways. While slant has a strong influence in light source position along elevation axis, tilt has a major influence along azimuth axis. The weights are calculated as
$$E_\vartheta = \frac{(\pi - \Phi_i)}{2\pi} \tag{6.9}$$
$$E_\varrho = \frac{(\pi - 2\Upsilon_i)}{2\pi} \tag{6.10}$$
where Φ_i and Υ_i represent, respectively, the azimuth and elevation position from estimated light source using probes provided by user (without any uncertainty correction).

3. incorporate these errors in original slant/tilt input values and re-estimating the light position $\vec{L}$

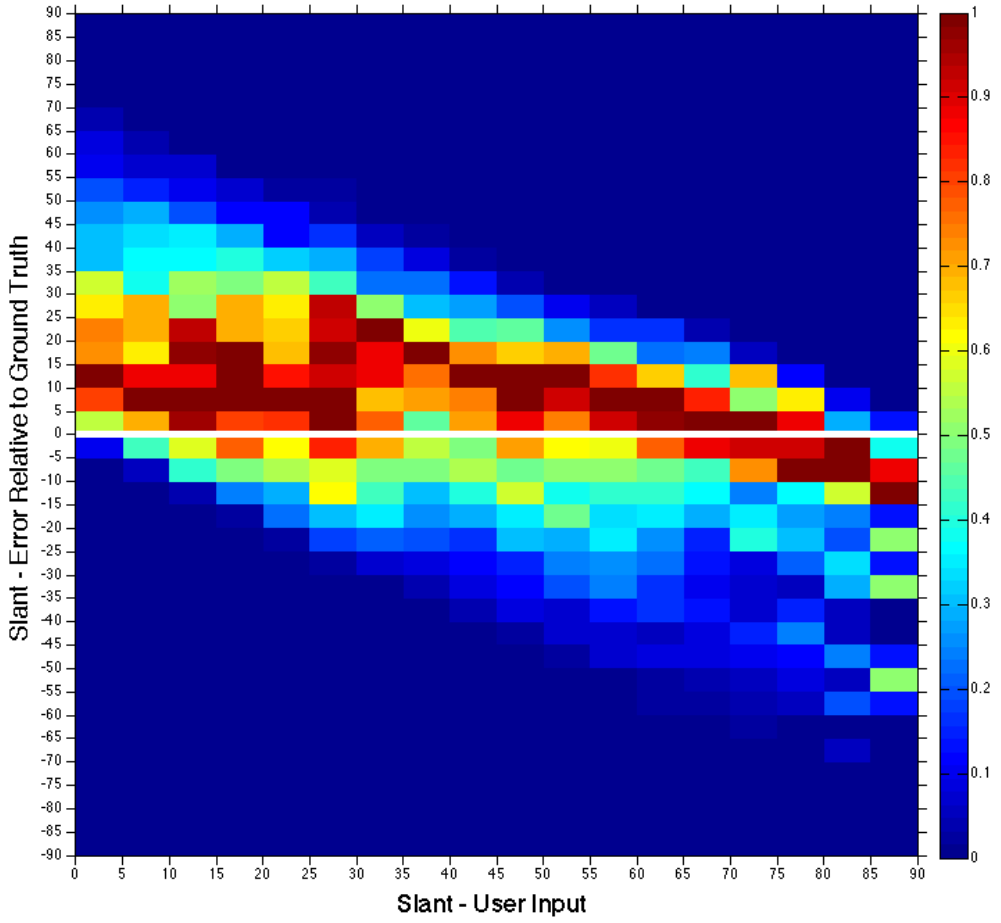


Figure 6.3: Visualization of slant model for correction of errors constructed from data collected in a psychophysical study provided by Cole et al. [18].

Each estimated light position contributes with a small Gaussian density in the estimated light azimuth/elevation space. These densities are accumulated across $20,000$ random draws, producing a kernel-density estimation of the uncertainty in the analyst's estimate of lighting.

6.2.4 Forgery Detection Process

Once we can produce a kernel-density estimation in azimuth/elevation space using probes from objects, we can now use it to detect forgeries. Suppose that we have a scene with k suspicious objects. To analyze the consistency of these k objects, first we ask an analyst to input as many as possible probes for each one of these objects. Thus, for each object, we use all its probes to estimate a kernel-density distribution. Then, a confidence region

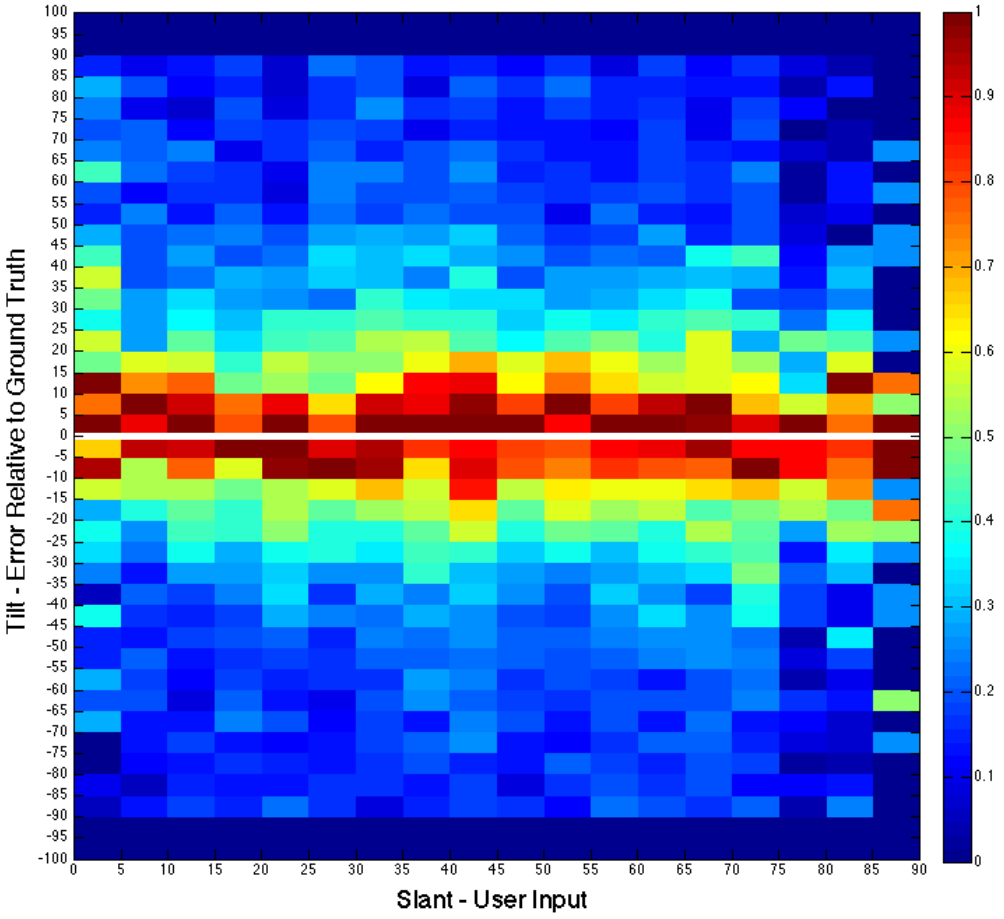


Figure 6.4: Visualization of tilt model for correction of errors constructed from data collected in a psychophysical study provided by Cole et al. [18].

(e.g., 99%) is computed for each distribution. We have now k confidence regions for the image. The physical consistency of this image is determined by intersecting these confidence regions.

For pristine images, this intersection[2] process will generate a feasible region in azimuth/elevation space. However, for a fake image, the alien object will produce a confidence region in azimuth/elevation space distant from all the other regions (produced by pristine objects). So, when intersecting the region produced by the fake object with the region produced by pristine objects, result region will be empty, characterizing a fake image.

[2]By intersecting confidence regions, rather than multiplying probabilities, every constraint must be satisfied for the lighting to be consistent.

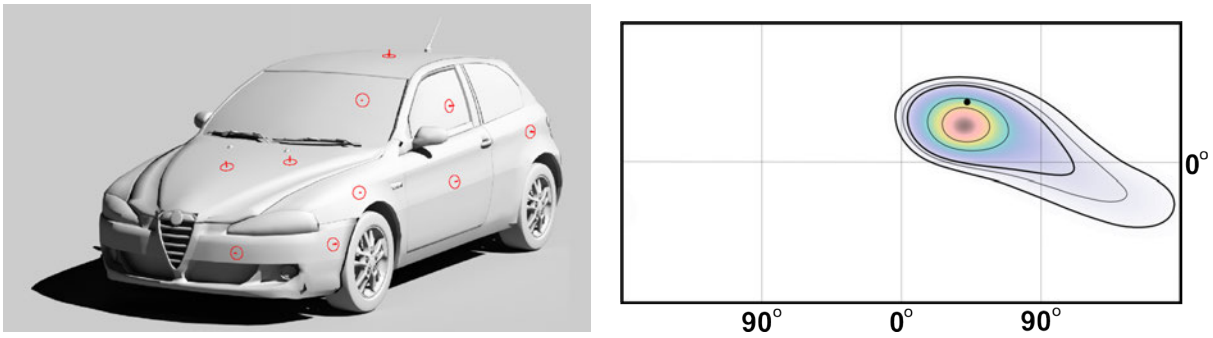
It is important to highlight here that, we just verify consistency among objects where the analyst have input probes. If an image depicts k objects, for example, but the analyst input probes just on two of them, our method will verify if these two objects match the 3-D light source position between them. In this example case, nothing can be ensured about the other objects in the image.

6.3 Experiments and Results

In this section, we performed three rounds of experiments to present the behavior of our method. In the first one, we investigate the behavior of the proposed method in controlled scenario. In the second one, we present results that show how confidence intervals intersection reduce feasible region of light source. Finally, in the last one we apply our method in one forgery constructed from a real world image.

6.3.1 Round #1

We rendered ten objects under six different lighting conditions. Sixteen previously untrained users were each instructed to place probes on ten objects. Shown in the left column of Figures 6.5, 6.6, 6.7 and 6.8 are four representative objects with the user-selected probes and in the right of each figure is the resulting estimate of light source specified as confidence intervals in an azimuth/elevation space. The small black dot in each figure corresponds to the actual light position. The contours correspond to, from outside to inside, probabilities ranging of 99%, 95%, 90%, 60% and 30%. On average, users were able to estimate the azimuth and elevation with an average accuracy of 11.1 and 20.6 degrees with a standard deviation of 9.4 and 13.3 degrees, respectively. On average, a user placed 12 probes on an object in 2.4 minutes.



(a) Model with Probes (b) Light Source Position

Figure 6.5: Car Model

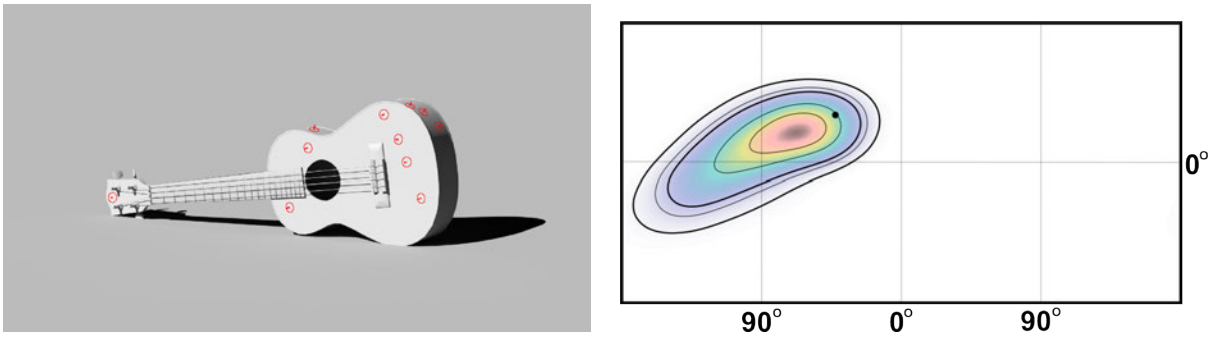


Figure 6.6: Guittar Model

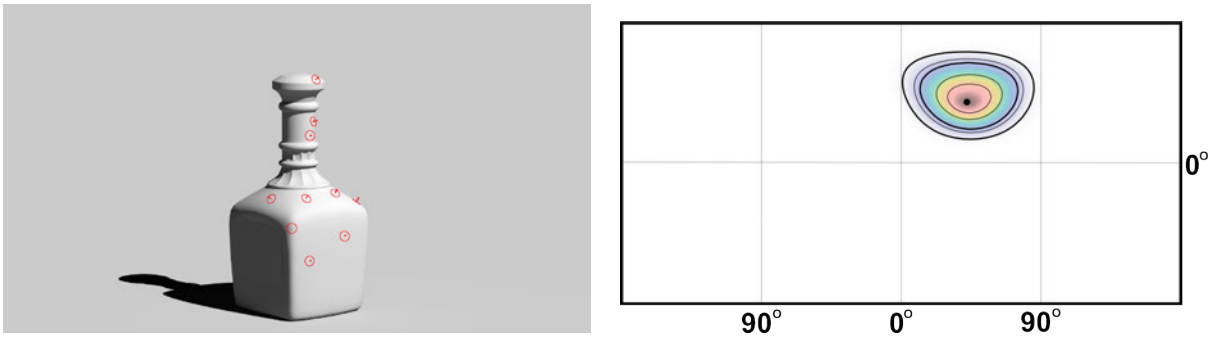


Figure 6.7: Bourbon Model

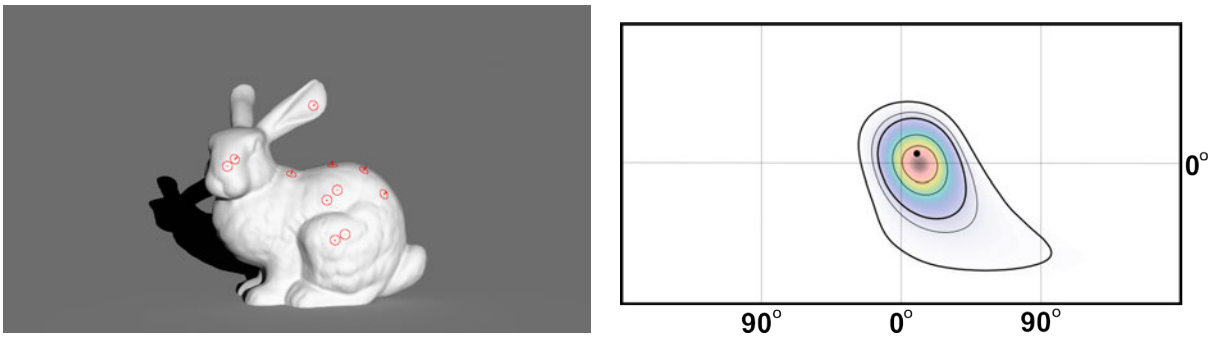


Figure 6.8: Bunny Model.

6.3.2 Round #2

In a real-world analysis, an analyst will be able to specify the 3-D shape of multiple objects which can then be combined to yield an increasingly more specific estimate of

lighting position. Figure 6.9 depicts, for example, the results of sequentially intersecting the estimated light source position from five objects in the same scene. From left to right and top to bottom, the resulting light source probability region get smaller every time than a new probability region is included. Of course, the smaller this confidence region, the more effective this technique will be in detecting inconsistent lighting.

6.3.3 Round #3

As we present in Section 6.2.4, we can now use light source region detection to expose forgeries. Figure 6.10 (a) depicts a forgery image (we have added a trash bin in the bottom left corner of the image). The first step to detect a forgery is chose which objects we want to investigate at this image, inputing probes at these objects as we can see in Figure 6.10(b), (d), (f), (h) and (j). So, for each object we calculate the probability region as depicted n Figure 6.10(c), (e), (g), (i) and (k).

We have now five light source probability regions, one for each object, in azimuth/elevation space. The probability region provided by pristine objects, which have originally been in the same image, when intersected produce a not empty region, as depicted in Figure 6.11(a). However, if we try to intersect the probability region provided by the trash can, it will produce an empty azimuth/elevation map. Figure 6.11(b) depicts in the same azimuth/elevation map, the intersection region depicted in Figure 6.11(a) and the probability region provided by the trash can. Clearly, there is no intersection between these two regions, which means that the trash can is an alien object relative to other analyzed objects.

6.4 Final Remarks

In this chapter, we have presented a new approach to detecting image compositions from inconsistencies in light source. Given a set of user-marked 3-D normals, we are able to estimate 3-D light source position from arbitrary objects in a scene without any additional information. To account for the error embedded in light source position estimation process, we have constructed an uncertainty model using data from an extensive psychophysical study which have measured users skills to perceive normals direction. Then, we have estimated the light source position many times, constructing confidence regions of possible light source positions. In a forensic scenario, when the intersection of suspicious parts produce an empty confidence region, there is an indication of image tampering.

The approach presented herein represents an important step forward in the forensic scenario mainly because it is able to detect the 3-D light source position from a single image without any a priori knowledge. Such fact makes the task of creating a composition

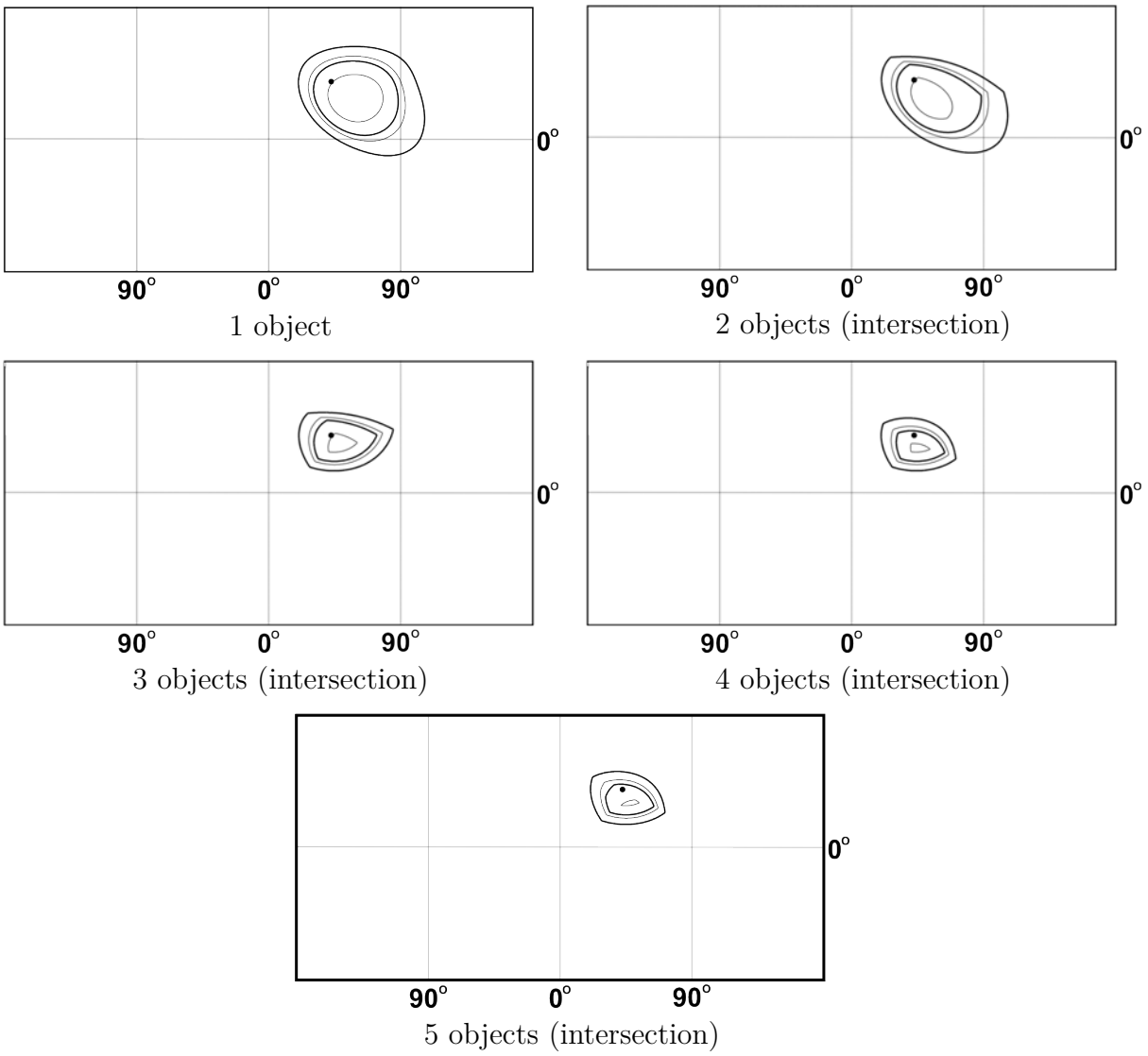
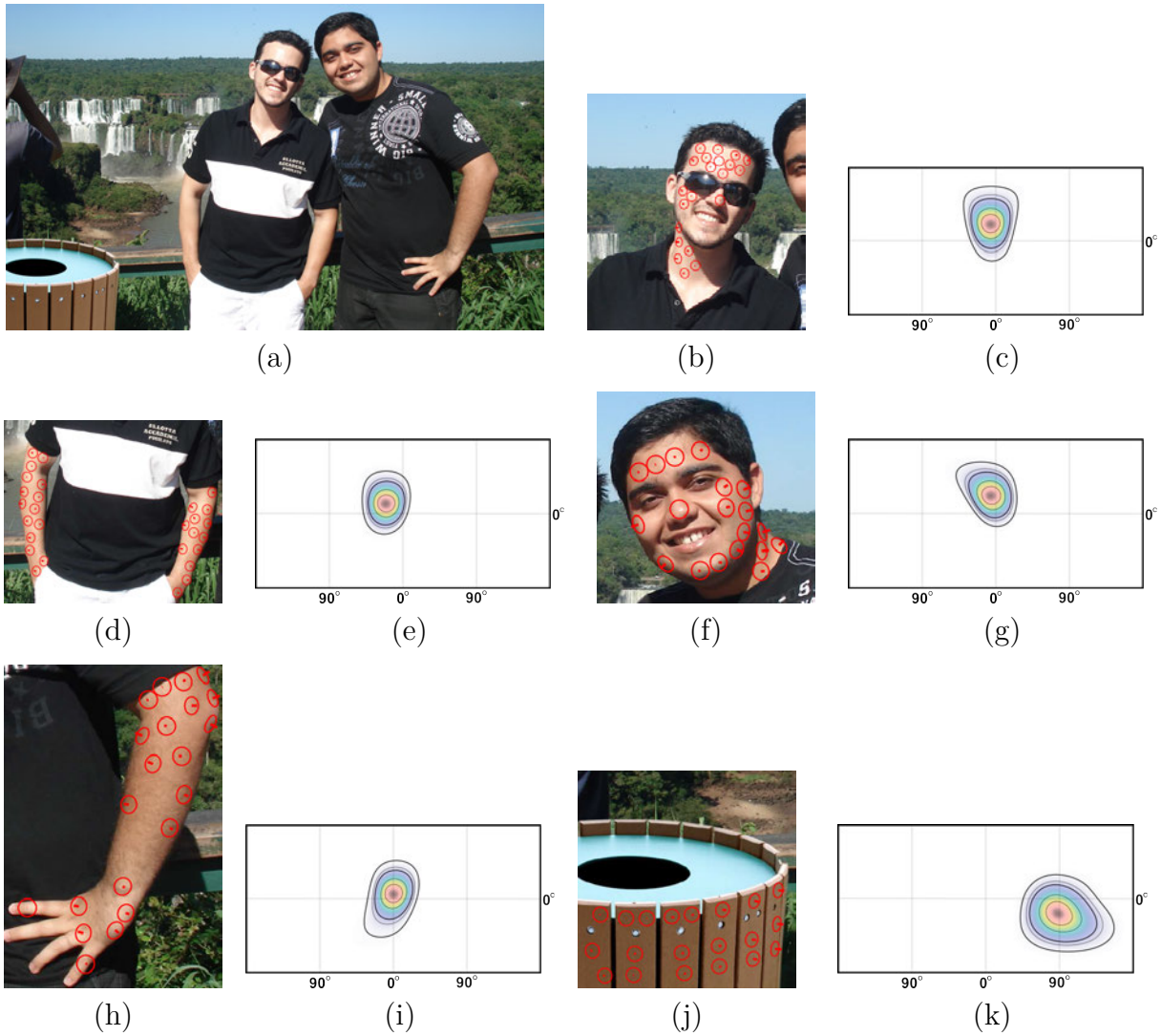


Figure 6.9: From left to right and top to bottom, the confidence intervals for the lighting estimate from one through five objects in the same scene, rendered under the same lighting. As expected and desired, this interval becomes smaller as more objects are detected, making it more easier to detect a forgery. Confidence intervals are shown at 60%, 90% (bold), 95% and 99% (bold). The location of the actual light source is noted by a black dot.

image harder since counterfeiters need now to consider 3-D light information in the scene.

As proposals for future work, we intend to investigate better ways to compensate user's errors in normal estimates, which consequently will generate smaller confidence regions in azimuth/elevation. A small confidence region allows us to estimate light source position



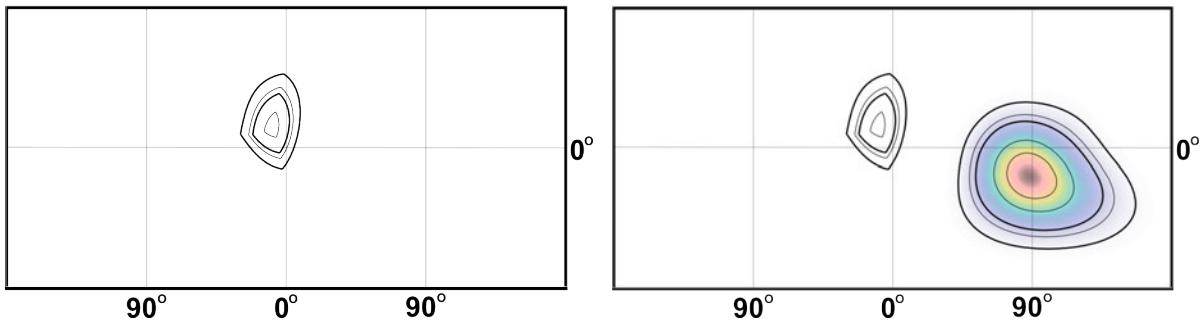
(a) (b) (c)

(d) (e) (f) (g)

(h) (i) (j) (k)

Figure 6.10: Different objects and their respectively light source probability region extracted from a fake image. The light source probability region estimated for the fake object (j) is totally different from the light source probability region provided by the other objects.

with a higher precision, improving the confidence of the method.



(a) (b)

Figure 6.11: (a) result for probability regions intersection from pristine objects and (b) absence of intersection between region from pristine objects and fake object.

Chapter 7

Conclusions and Research Directions

Technology improvements are responsible for uncountable society advances. However, they are not always used in favor of constructive reasons. Many times, malicious people use such resources to take advantage from other people. In computer science, it could not be different. Specifically, when it comes to digital documents, often malevolent people use manipulation tools for creating documents, in special fake images, for self benefit. Image composition is among the most common types of image manipulation and consists of using parts of two or more images to create a new fake one. In this context, this work has presented four methods that rely on illumination inconstancies for detecting this image compositions.

Our first approach uses eye specular highlights to detect image composition containing two or more people. By estimating light source and viewer position, we are able to construct discriminative features for the image which associate with machine learning methods allow for an improvement of more than 20% error reduction when compared to the state-of-the-art method. Since it is based on eye specular highlights, our proposed approach has as main advantage the fact that such specific part of the image is difficult to manipulate with precision without leaving additional telltales. On the other hand, as drawback, the method is specific for scenes where eyes are visible and in adequate resolution, since it depends on iris contour marks. Also, the manual iris marking step can sometimes introduce human errors to the process, which can compromise the method's accuracy. To overcome this limitation, in our second and third approaches, we explore a different type of light property. We decide to explore *metamerism*, a color phenomenon whereby two colors may appear to match under one light source, but appear completely different under another one.

In our second approach, we build texture and shape representations from local illuminant maps extracted from the images. Using such texture and edge descriptors, we extract complementary features which have been combined through a strong machine

learning method (SVM) to achieve an AUC of 86% (with an accuracy rate of 79%) in classification of image composition containing people. Another important contribution to the forensic community introduced by this part of our work is the creation of DSO-1 database, a realistic and high-resolution image set comprising 200 images (100 normal and 100 doctored). Compared to other state-of-the-art methods based on illuminant colors, besides its higher accuracy, our method is also less user dependent, and its decision step is totally performed by machine learning algorithms. Unfortunately, this approach has two main drawbacks that restrict its applicability: the first one is the fact that an accuracy of 79% is not sufficient for a strong inference on image classification in the forensic scenario; the second one is that the approach discards an important information, which is color, for the analysis of illuminants. Both problems inspired us to construct our third approach.

Our third approach builds upon our second one by analyzing more discriminative features and using a robust machine learning framework to combine them. Instead of using just four different ways to describe illuminant maps, we took advantage of a wide set of combinations involving different types of illuminant maps, color space and image features. Features based on color of illuminant maps, not addressed before, are now used complementarily with texture and shape information to describe illuminant maps. Furthermore, from this complete set of different features extracted from illuminant maps, we are now able to detect their best combination, which allows us to achieve a remarkable classification accuracy of 94%. This is a significant step forward for the forensic community given that now a fast and effective analysis for composite images involving people can be performed in short time. However, although image composition involving people is one of the most usual ways of modifying images, other elements can also be inserted into them. To address this issue, we proposed our last approach.

In our fourth approach, we insert the user back in the loop to solve a more complex problem: the one of detecting image splicing regardless of their type. For that, we consider user knowledge to estimate 3-D shapes from images. Using a simple interface, we show that an expert is able to specify 3-D surface normals in a single image from which the 3-D light source position is estimated. The light source estimation process is repeated several times, always trying to correct embedded user errors. Such correction is performed by using a statistical model, which is generated from large-scale psychophysical studies with users. As a result, we estimate not just a position for light source, but a region in 3-D space containing light source. Regions from different objects in the same figure can be intersected and, when a forgery is present, its intersection with other image objects produces an empty region, pointing out a forgery. This method corrects the limitation of detecting forgeries only for images containing people. However, the downside is that we again have a strong dependence on the user.

The main conclusion of this work is that forensic methods are in constant development.

Table 7.1: Proposed methods and their respective application scenarios.

Method	**Possible Application Scenarios**
Method based on eye specular highlights (Chapter 3)	Indoor and outdoor images containing two or more people and where the eyes are well visible
Methods based on illuminant colors analysis (Chapters 4 and 5)	Indoor and outdoor images containing two or more people and where the faces are visible
Method based on 3-D light source analysis (Chapter 6)	Outdoor images containing arbitrary objects

They have their pros and cons and there is no *silver bullet* able to detect all types of image composition and at high accuracy. The method described in Chapter 3 could not be applied to an image depicting a beach scenario, for instance, with people using sunglasses. However, this kind of image could be analyzed with the method proposed in Chapters 4, 5 and 6. Similar analyses can be drawn for several other situations. Indoor images most of the times present many different local light sources. This scenario prevents us to use the approach proposed in Chapter 6 since it has been developed for outdoor scenes with an infinite light source. However, if the scene contains people, we can perform an analysis using our approaches proposed in Chapter 3, 4 and 5. Outdoor images, where the suspicious object is not a person, is an additional example of how our techniques work complementary. Despite our approaches from Chapters 3, 4 and 5 can just be applied to detect image compositions involving people, by using our last approach proposed in Chapter 6, we can analyze any Lambertian object in this type of scenario. Table 7.1 summarizes the main application scenarios where the proposed methods can be applied.

All these examples make clear that using methods for capturing different types of telltales, as we have proposed along this work, allows for a more complete investigation of suspicious images increasing the confidence of the process. Also, proposing methods which are grounded on different telltales contribute with the forensic community so it can investigate images provided by a large number of different scenarios.

As for research directions and future work, we suggest different contributions for each one of our proposed approaches. For the eye specular highlight approach, two interesting extensions would be adapting an automatic iris detection method to replace user manual marks and exploring different non-linear optimization algorithms in the light source and viewer estimation. For the approaches that explore metamerism and illuminant color, an interesting future work would be improving the location of the actual forgery face (treating more than one forgery face) and proposing forms to compare illuminants provided by different body parts from the same person. This last one, would remove the necessity of having two or more people in the image to detect forgeries and would be very useful for pornography image composition detection. The influence of ethnicity in forgery detection using illuminant color can also be investigated as an interesting extension. As for our last

approach which estimates 3-D light source positions, we envision at least two essential extensions. The first one refers to the fact that a better error correction function needs to be performed, giving us more precise confidence regions while the second one refers to the fact that this work should incorporate other forensic methods, as the one proposed by Kee and Farid [51], to increase its confidence on the light source position estimation.

Bibliography

[1] M.H. Asmare, V.S. Asirvadam, and L. Iznita. Color Space Selection for Color Image Enhancement Applications. In *Intl. Conference on Signal Acquisition and Processing*, pages 208–212, 2009.

[2] K. Barnard, V. Cardei, and B. Funt. A Comparison of Computational Color Constancy Algorithms – Part I: Methodology and Experiments With Synthesized Data. *IEEE Transactions on Image Processing (T.IP)*, 11(9):972–983, Sep 2002.

[3] K. Barnard, L. Martin, A. Coath, and B. Funt. A Comparison of Computational Color Constancy Algorithms – Part II: Experiments With Image Data. *IEEE Transactions on Image Processing (T.IP)*, 11(9):985–996, Sep 2002.

[4] H. G. Barrow and J. M. Tenenbaum. *Recovering Intrinsic Scene Characteristics from Images*. Academic Press, 1978.

[5] S. Bayram, I. Avcibaş, B. Sankur, and N. Memon. Image Manipulation Detection with Binary Similarity Measures. In *European Signal Processing Conference (EUSIPCO)*, volume I, pages 752–755, 2005.

[6] T. Bianchi and A. Piva. Detection of Non-Aligned Double JPEG Compression Based on Integer Periodicity Maps. *IEEE Transactions on Information Forensics and Security (T.IFS)*, 7(2):842–848, April 2012.

[7] S. Bianco and R. Schettini. Color Constancy using Faces. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Providence, RI, USA, June 2012.

[8] C. M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

[9] V. Blanz and T. Vetter. A Morphable Model for the Synthesis of 3-D Faces. In *ACM Annual Conference on Computer Graphics and Interactive Technique (SIGGRAPH)*, pages 187–194, 1999.

[10] M. Bleier, C. Riess, S. Beigpour, E. Eibenberger, E. Angelopoulou, T. Tröger, and A. Kaup. Color Constancy and Non-Uniform Illumination: Can Existing Algorithms Work? In *IEEE Color and Photometry in Computer Vision Workshop*, pages 774–781, 2011.

[11] G. Buchsbaum. A Spatial Processor Model for Color Perception. *Journal of the Franklin Institute*, 310(1):1–26, July 1980.

[12] J. Canny. A Computational Approach to Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence (T.PAMI)*, 8(6):679–698, 1986.

[13] T. Carvalho, A. Pinto, E. Silva, F. da Costa, G. Pinheiro, and A. Rocha. *Escola Regional de Informática de Minas Gerais*, chapter Crime Scene Investigation (CSI): da Ficção à Realidade, pages 1–23. UFJF, 2012.

[14] T. Carvalho, C. Riess, E. Angelopoulou, H. Pedrini, and A. Rocha. Exposing Digital Image Forgeries by Illumination Color Classification. *IEEE Transactions on Information Forensics and Security (T.IFS)*, 8(7):1182–1194, 2013.

[15] A. Çarkacıoğlu and F. T. Yarman-Vural. SASI: A Generic Texture Descriptor for Image Retrieval. *Pattern Recognition*, 36(11):2615–2633, 2003.

[16] H. Chen, X. Shen, and Y. Lv. Blind Identification Method for Authenticity of Infinite Light Source Images. In *IEEE Intl. Conference on Frontier of Computer Science and Technology (FCST)*, pages 131–135, 2010.

[17] F. Ciurea and B. Funt. A Large Image Database for Color Constancy Research. In *Color Imaging Conference: Color Science and Engineering Systems, Technologies, Applications (CIC)*, pages 160–164, Scottsdale, AZ, USA, November 2003.

[18] F. Cole, K. Sanik, D. DeCarlo, A. Finkelstein, T. Funkhouser, S. Rusinkiewicz, and M. Singh. How Well Do Line Drawings Depict Shape? *ACM Transactions on Graphics (ToG)*, 28(3), August 2009.

[19] E. A. Cooper, E. A. Piazza, and M. S. Banks. The Perceptual Basis of Common Photographic Practice. *Journal of Vision*, 12(5), 2012.

[20] G. Csurka, C. R. Dance, L. Fan, J. Willamowski, and C. Bray. Visual Categorization With Bags of Keypoints. In *Workshop on Statistical Learning in Computer Vision*, pages 1–8, 2004.

[21] N. Dalal and B. Triggs. Histograms of Oriented Gradients for Human Detection. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 886–893, 2005.

[22] E. J. Delp, N. Memon, and M. Wu. Digital Forensics. *IEEE Signal Processing Magazine*, 26(3):14–15, March 2009.

[23] P. Destuynder and M. Salaun. *Mathematical Analysis of Thin Plate Models*. Springer, 1996.

[24] J. A. dos Santos, P. H. Gosselin, S. Philipp-Foliguet, R. S. Torres, and A. X. Falcao. Interactive Multiscale Classification of High-Resolution Remote Sensing Images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 6(4):2020–2034, 2013.

[25] M. Doyoddorj and K. Rhee. A Blind Forgery Detection Scheme Using Image Compatibility Metrics. In *IEEE Intl. Symposium on Industrial Electronics (ISIE)*, pages 1–6, 2013.

[26] M. Ebner. Color Constancy Using Local Color Shifts. In *European Conference in Computer Vision (ECCV)*, pages 276–287, 2004.

[27] W. Fan, K. Wang, F. Cayre, and Z. Xiong. 3D Lighting-Based Image Forgery Detection Using Shape-from-Shading. In *European Signal Processing Conference*, pages 1777–1781, aug. 2012.

[28] Fabio A. Faria, Jefersson A. dos Santos, Anderson Rocha, and Ricardo da S. Torres. A Framework for Selection and Fusion of Pattern Classifiers in Multimedia Recognition. *Pattern Recognition Letters*, 39(0):52–64, 2013.

[29] H. Farid. *Deception: Methods, Motives, Contexts and Consequences*, chapter Digital Doctoring: Can We Trust Photographs?, pages 95–108. Stanford University Press, 2009.

[30] P. F. Felzenszwalb and D. P. Huttenlocher. Efficient Graph-Based Image Segmentation. *Springer Intl. Journal of Computer Vision (IJCV)*, 59(2):167–181, 2004.

[31] J. L. Fleiss. Measuring Nominal Scale Agreement Among Many Raters. *Psychological Bulletin*, 76(5):378–382, 1971.

[32] P. V. Gehler, C. Rother, A. Blake, T. Minka, and T. Sharp. Bayesian Color Constancy Revisited. In *Intl. Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8, 06 2008.

[33] S. Gholap and P. K. Bora. Illuminant Colour Based Image Forensics. In *IEEE Region 10 Conference*, pages 1–5, 2008.

[34] A. Gijsenij and T. Gevers. Color Constancy Using Natural Image Statistics and Scene Semantics. *IEEE Transactions on Pattern Analysis and Machine Intelligence (T.PAMI)*, 33(4):687–698, 2011.

[35] A. Gijsenij, T. Gevers, and J. van de Weijer. Computational Color Constancy: Survey and Experiments. *IEEE Transactions on Image Processing (T.IP)*, 20(9):2475–2489, September 2011.

[36] A. Gijsenij, T. Gevers, and J. van de Weijer. Improving Color Constancy by Photometric Edge Weighting. *IEEE Pattern Analysis and Machine Intelligence (PAMI)*, 34(5):918–929, May 2012.

[37] A. Gijsenij, R. Lu, and T. Gevers. Color Constancy for Multiple Light Sources. *IEEE Transactions on Image Processing (T.IP)*, 21(2):697–707, 2012.

[38] Arjan Gijsenij, Theo Gevers, and Joost Weijer. Generalized gamut mapping using image derivative structures for color constancy. *Int. Journal of Computer Vision*, 86(2-3):127–139, January 2010.

[39] R. C. Gonzalez and R. E. Woods. *Digital Image Processing*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 2nd edition, 2001.

[40] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, New York, NY, USA, 2 edition, 2003.

[41] Z. He, T. Tan, Z. Sun, and X. Qiu. Toward accurate and fast iris segmentation for iris biometrics. *IEEE Transactions on Pattern Analysis and Machine Intelligence (T.PAMI)*, 31(9):1670–1684, 2009.

[42] J. Huang, R. Kumar, M. Mitra, W. Zhu, and R. Zabih. Image Indexing Using Color Correlograms. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 762–768, 1997.

[43] R. Huang and W. A. P. Smith. Shape-from-Shading Under Complex Natural Illumination. In *IEEE Intl. Conference on Image Processing (ICIP)*, pages 13–16, 2011.

[44] T. Igarashi, K. Nishino, and S. K. Nayar. The Appearance of Human Skin: A Survey. *Foundations and Trends in Computer Graphics and Vision*, 3(1):1–95, 2007.

[45] M. K. Johnson and H. Farid. Exposing Digital Forgeries by Detecting Inconsistencies in Lighting. In *ACM Workshop on Multimedia and Security (MM&Sec)*, pages 1–10, New York, NY, USA, 2005. ACM.

[46] M. K. Johnson and H. Farid. Exposing Digital Forgeries Through Chromatic Aberration. In *ACM Workshop on Multimedia and Security (MM&Sec)*, pages 48–55. ACM, 2006.

[47] M. K. Johnson and H. Farid. Exposing Digital Forgeries in Complex Lighting Environments. *IEEE Transactions on Information Forensics and Security (T.IFS)*, 2(3):450–461, 2007.

[48] M. K. Johnson and H. Farid. Exposing Digital Forgeries Through Specular Highlights on the Eye. In Teddy Furon, François Cayre, Gwenaël J. Doërr, and Patrick Bas, editors, *ACM Information Hiding Workshop (IHW)*, volume 4567 of *Lecture Notes in Computer Science*, pages 311–325, 2008.

[49] R. Kawakami, K. Ikeuchi, and R. T. Tan. Consistent Surface Color for Texturing Large Objects in Outdoor Scenes. In *IEEE Intl. Conference on Computer Vision (ICCV)*, pages 1200–1207, 2005.

[50] E. Kee and H. Farid. Exposing Digital Forgeries from 3-D Lighting Environments. In *IEEE Intl. Workshop on Information Forensics and Security (WIFS)*, pages 1–6, dec. 2010.

[51] E. Kee, J. O'brien, and H. Farid. Exposing Photo Manipulation with Inconsistent Shadows. *ACM Transactions on Graphics (ToG)*, 32(3):1–12, July 2013.

[52] Petrina A. S. Kimura, João M. B. Cavalcanti, Patricia Correia Saraiva, Ricardo da Silva Torres, and Marcos André Gonçalves. Evaluating Retrieval Effectiveness of Descriptors for Searching in Large Image Databases. *Journal of Information and Data Management*, 2(3):305–320, 2011.

[53] M. Kirchner. Linear Row and Column Predictors for the Analysis of Resized Images. In *ACM Workshop on Multimedia and Security (MM&Sec)*, pages 13–18, September 2010.

[54] J. J. Koenderink, A. Van Doorn, and A. Kappers. Surface Perception in Pictures. *Percept Psychophys*, 52(5):487–496, 1992.

[55] J. J. Koenderink, A. J. van Doorn, H. de Ridder, and S. Oomes. Visual Rays are Parallel. *Perception*, 39(9):1163–1171, 2010.

[56] J. J. Koenderink, A. J. van Doorn, A. M. L. Kappers, and J. T. Todd. Ambiguity and the Mental Eye in Pictorial Relief. *Perception*, 30(4):431–448, 2001.

[57] L. I. Kuncheva and C. J. Whitaker. Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy. *Machine Learning*, 51(2):181–207, 2003.

[58] Edwin H. Land. The Retinex Theory of Color Vision. *Scientific American*, 237(6):108–128, December 1977.

[59] J. Richard Landis and Gary G. Koch. The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1):159–174, 1977.

[60] Dong-Ho Lee and Hyoung-Joo Kim. A Fast Content-Based Indexing and Retrieval Technique by the Shape Information in Large Image Database. *Journal of Systems and Software*, 56(2):165–182, March 2001.

[61] Q. Liu, X. Cao, C. Deng, and X. Guo. Identifying image composites through shadow matte consistency. *IEEE Transactions on Information Forensics and Security (T.IFS)*, 6(3):1111–1122, 2011.

[62] O. Ludwig, D. Delgado, V. Goncalves, and U. Nunes. Trainable Classifier-Fusion Schemes: An Application to Pedestrian Detection. In *IEEE Intl. Conference on Intelligent Transportation Systems*, pages 1–6, 2009.

[63] J. Lukáš, J. Fridrich, and M. Goljan. Digital Camera Identification From Sensor Pattern Noise. *IEEE Transactions on Information Forensics and Security (T.IFS)*, 1(2):205–214, June 2006.

[64] Y. Lv, X. Shen, and H. Chen. Identifying Image Authenticity by Detecting Inconsistency in Light Source Direction. In *Intl. Conference on Information Engineering and Computer Science (ICIECS)*, pages 1–5, 2009.

[65] Fariborz Mahmoudi, Jamshid Shanbehzadeh, Amir-Masoud Eftekhari-Moghadam, and Hamid Soltanian-Zadeh. Image Retrieval Based on Shape Similarity by Edge Orientation Autocorrelogram. *Pattern Recognition*, 36(8):1725–1736, 2003.

[66] P. Nillius and J.O. Eklundh. Automatic Estimation of the Projected Light Source Direction. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1076–1083, 2001.

[67] N. Ohta, A. R. Robertson, and A. A. Robertson. *Colorimetry: Fundamentals and Applications*, volume 2. J. Wiley, 2005.

[68] Y. Ostrovsky, P. Cavanagh, and P. Sinha. Perceiving illumination inconsistencies in scenes. *Perception*, 34(11):1301–1314, 2005.

[69] G. Pass, R. Zabih, and J. Miller. Comparing Images Using Color Coherence Vectors. In *ACM Intl. Conference on Multimedia*, pages 65–73, 1996.

[70] Otavio Penatti, Eduardo Valle, and Ricardo da S. Torres. Comparative Study of Global Color and Texture Descriptors for Web Image Retrieval. *Journal of Visual Communication and Image Representation (JVCI)*, 23(2):359–380, 2012.

[71] M. Pharr and G. Humphreys. *Physically Based Rendering: From Theory To Implementation*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2nd edition, 2010.

[72] A. C. Popescu and H. Farid. Statistical Tools for Digital Forensics. In *Information Hiding Conference (IHW)*, pages 395–407, June 2005.

[73] C. Riess and E. Angelopoulou. Scene Illumination as an Indicator of Image Manipulation. In *ACM Information Hiding Workshop (IHW)*, volume 6387, pages 66–80, 2010.

[74] C. Riess, E. Eibenberger, and E. Angelopoulou. Illuminant Color Estimation for Real-World Mixed-Illuminant Scenes. In *IEEE Color and Photometry in Computer Vision Workshop*, November 2011.

[75] A. Rocha, T. Carvalho, H. F. Jelinek, S. Goldenstein, and J. Wainer. Points of Interest and Visual Dictionaries for Automatic Retinal Lesion Detection. *IEEE Transactions on Biomedical Engineering (T.BME)*, 59(8):2244–2253, 2012.

[76] A. Rocha, W. Scheirer, T. E. Boult, and S. Goldenstein. Vision of the Unseen: Current Trends and Challenges in Digital Image and Video Forensics. *ACM Computer Survey*, 43(4):1–42, 2011.

[77] A. K. Roy, S. K. Mitra, and R. Agrawal. A novel method for detecting light source for digital images forensic. *Opto-Electronics Review*, 19(2):211–218, 2011.

[78] A. Ruszczyǹski. *Nonlinear Optimization*. Princeton University Press, 2006.

[79] P. Saboia, T. Carvalho, and A. Rocha. Eye Specular Highlights Telltales for Digital Forensics: A Machine Learning Approach. In *IEEE Intl. Conference on Image Processing (ICIP)*, pages 1937–1940, 2011.

[80] J. Schanda. *Colorimetry: Understanding the CIE System*. Wiley, 2007.

[81] W. R. Schwartz, A. Kembhavi, D. Harwood, and L. S. Davis. Human Detection Using Partial Least Squares Analysis. In *IEEE Intl. Conference on Computer Vision (ICCV)*, pages 24–31, 2009.

[82] P. Sloan, J. Kautz, and J. Snyder. Precomputed Radiance Transfer for Real-Time Rendering in Dynamic, Low-Frequency Lighting Environments. *ACM Transactions on Graphics (ToG)*, 21(3):527–536, 2002.

[83] C. E. Springer. *Geometry and Analysis of Projective Spaces*. Freeman, 1964.

[84] R. Stehling, M. Nascimento, and A. Falcao. A Compact and Efficient Image Retrieval Approach Based on Border/Interior Pixel Classification. In *ACM Conference on Information and Knowledge Management*, pages 102–109, 2002.

[85] M.J. Swain and D.H. Ballard. Color Indexing. *Intl. Journal of Computer Vision*, 7(1):11–32, 1991.

[86] R. T. Tan, K. Nishino, and K. Ikeuchi. Color Constancy Through Inverse-Intensity Chromaticity Space. *Journal of the Optical Society of America A*, 21:321–334, 2004.

[87] B. Tao and B. Dickinson. Texture Recognition and Image Retrieval Using Gradient Indexing. *Elsevier Journal of Visual Communication and Image Representation (JVCI)*, 11(3):327–342, 2000.

[88] J. Todd, J. J. Koenderink, A. J. van Doorn, and A. M. L. Kappers. Effects of Changing Viewing Conditions on the Perceived Structure of Smoothly Curved Surfaces. *Journal of Experimental Psychology: Human Perception and Performance*, 22(3):695–706, 1996.

[89] S. Tominaga and B. A. Wandell. Standard Surface-Reflectance Model and Illuminant Estimation. *Journal of the Optical Society of America A*, 6(4):576–584, Apr 1989.

[90] M. Unser. Sum and Difference Histograms for Texture Classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence (T.PAMI)*, 8(1):118–125, 1986.

[91] J. van de Weijer, T. Gevers, and A. Gijsenij. Edge-Based Color Constancy. *IEEE Transactions on Image Processing (T.IP)*, 16(9):2207–2214, 2007.

[92] J. Winn, A. Criminisi, and T. Minka. Object Categorization by Learned Universal Visual Dictionary. In *IEEE Intl. Conference on Computer Vision (ICCV)*, pages 1800–1807, 2005.

[93] X. Wu and Z. Fang. Image Splicing Detection Using Illuminant Color Inconsistency. In *IEEE Intl. Conference on Multimedia Information Networking and Security (MINES)*, pages 600–603, 2011.

[94] H. Yao, S. Wang, Y. Zhao, and X. Zhang. Detecting Image Forgery Using Perspective Constraints. *IEEE Signal Processing Letters (SPL)*, 19(3):123–126, 2012.

[95] W. Zhang, X. Cao, J. Zhang, J. Zhu, and P. Wang. Detecting Photographic Composites Using Shadows. In *IEEE Intl. Conference on Multimedia and Expo (ICME)*, pages 1042–1045, 2009.