



Pedro Filipe Marques de Almeida

Licenciado em Engenharia Informática

Análise e Identificação de Obras de Arte

Dissertação para obtenção do Grau de Mestre em
Engenharia Informática

Orientadores : Nuno Correia, Prof. Catedrático, Universidade
Nova de Lisboa
Rui Jesus, Prof. Adjunto, Instituto Superior de En-
genharia de Lisboa

Júri:

Presidente: Prof. Doutor Pedro Manuel Corrêa Calvente Barahona

Arguente: Prof. Doutor Pedro Miguel Torres Mendes Jorge

Vogal: Prof. Doutor Nuno Manuel Robalo Correia



FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Novembro, 2012

Análise e Identificação de Obras de Arte

Copyright © Pedro Filipe Marques de Almeida, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa

A Faculdade de Ciências e Tecnologia e a Universidade Nova de Lisboa têm o direito, perpétuo e sem limites geográficos, de arquivar e publicar esta dissertação através de exemplares impressos reproduzidos em papel ou de forma digital, ou por qualquer outro meio conhecido ou que venha a ser inventado, e de a divulgar através de repositórios científicos e de admitir a sua cópia e distribuição com objectivos educacionais ou de investigação, não comerciais, desde que seja dado crédito ao autor e editor.

À minha avó Eugénia

Agradecimentos

Durante o período em que a dissertação foi realizada vários elementos, quer institucionais quer individuais, contribuíram para o seu sucesso, tendo representado uma ajuda preciosa. Na vida deparamo-nos com inúmeras dificuldades, contudo com muito esforço e dedicação consegue-se ultrapassar essas dificuldades. Felizmente ao longo da dissertação surgiram ajudas muito úteis e as quais ficarão para sempre. Gostaria assim de expressar os meus sinceros agradecimentos:

- À Fundação para a Ciência e a Tecnologia (FCT) pelo apoio financeiro prestado no âmbito do projecto "História, Materiais e Técnicas de pintores portugueses, 1850-1918: Romantismo, Naturalismo e Modernismo" PTDC/ EAT-EAT/113612/ 2009.
- Ao meu orientador, Nuno Correia, pela confiança, disponibilidade, liberdade e suporte sempre que necessário.
- Ao meu orientador, Rui Jesus, pelo tempo disponibilizado, pelo seu conhecimento e pelo seu gosto em ajudar.
- À professora Maria João Melo, pela simpatia e motivação para chegar mais além;
- À dedicação excepcional da aluna de Doutoramento em Conservação e Restauro, Cristina Montagner, mostrando-se sempre disponível para ajudar, ouvir, discutir novas ideias e partir à descoberta de novas soluções.
- À Exma. Isabel Carlos, directora do Centro de Arte Moderna, e à Dra. Ana Vasconcelos, curadora do Centro de Arte Moderna – da Fundação Calouste Gulbenkian.
- Ao professor Sérgio Nascimento, Departamento de Física da Universidade do Minho, e à sua equipa: João Manuel Linhares, Hélder Tiago Correia e Osamu Masuda pela aquisição de imagens hiperespectrais das pinturas.
- Ao professor Marcello Picollo, Instituto de Física Aplicada "Nello Carrara", IFAC-CNR (Florença) pela análise utilizando a Fiber Optics Reflectance Spectroscopy (FORS).

- Ao professor João Pedro Martins de Almeida Lopes, Investigador Auxiliar na Faculdade de Farmácia, Universidade do Porto, pelo auxílio ao nível da análise de imagens hiperespectrais.
- Aos meus colegas de curso, Nuno Grade, Ana Fernandes, Sérgio Silva, André Simões pelo suporte dado ao longo da elaboração da dissertação.
- À minha namorada Nicole Silva, por todo o apoio, força, carinho, dedicação e paciência reveladas ao longo de todo o desenvolvimento.
- À minha família, em especial à minha mãe Emilia Almeida e minha avó Eugénia Caiado, e aos restantes amigos por todo apoio, em especial ao Ricardo Pedrosa, à Ana Cordeiro e à Sofia Pedrosa.

Resumo

Tradicionalmente, a análise de obras de arte era efectuada apenas por historiadores e outros peritos, sendo este um processo longo e complicado, exigindo a análise de várias evidências. Entre essas evidências destacam-se a análise da pincelada característica do pintor, a análise da assinatura do pintor caso exista e a análise dos materiais utilizados. Com o desenvolvimento das técnicas de processamento e análise de sinal e de imagem verificou-se que estas poderiam dar suporte aos historiadores na análise de obras de arte, ajudando a desvendar grande parte das incógnitas que ainda existem sobre alguns pintores. Assim, alguns museus decidiram criar repositórios digitais das suas colecções de modo a que vários investigadores possam trabalhar nesses documentos, evitando assim o contacto físico com a obra de arte.

Entre os interessados neste tipo de ferramentas encontram-se os museus e os departamentos policiais. Os museus pretendem autenticar e restaurar todas as suas obras. Os departamentos policiais que combatem o contrabando e falsificação de obras de arte estão também interessados pois isto permitiria validar a autenticidade de obras apreendidas com maior rapidez.

Neste trabalho foram consideradas duas fases principais, incidindo o estudo sobre as obras do pintor Amadeo de Souza-Cardoso. Numa primeira fase efectuou-se a análise das assinaturas através da aplicação de técnicas de processamento de imagem e reconhecimento de padrões, com vista a verificação de autenticidade e atribuição de autor assim como também a agrupamento de assinaturas semelhantes. A segunda fase direccionou-se para a análise dos materiais utilizados nas pinturas a partir das imagens hiperespectrais dos quadros do pintor em análise.

Palavras-chave: Processamento de imagem, Autenticidade, Atribuição, Assinaturas, Análise de Materiais, Imagens Hiperespectrais, Classificação

Abstract

Traditionally, the analysis of works of art was made only by historians and other experts, which is a long and complicated process, requiring the analysis of several evidences. Among these evidences are the analysis of the painter's brushstroke, the analysis of painter's signature and the analysis of the materials used. With the development of techniques for image and signal processing and analysis it was found that they could support the historians in the analysis of works of art, helping to solve many of the unknowns of which still exist about some painters. Some museums have decided to create digital repositories from their collection so that researchers can work in these documents, thus avoiding physical contact with the work of art.

Among those interested in this kind of tools are the museums and police departments. The museums want to authenticate and restore all their works of art. Police departments that fight smuggling and forgery of works of art are also interested because it would allow to validate the authenticity of confiscated works more quickly.

This work considered two main phases, focusing on the works of painter Amadeo de Souza-Cardoso. In a first phase the signatures were analysed, by the application of image processing and pattern recognition techniques, in order to verify the authenticity and attribution as well as to group similar signatures. The second phase was directed to the analysis of materials used in paintings starting from hyperspectral images of the paintings.

Keywords: Image processing, Authenticity, Attribution, Signatures, Material Analysis, Hyperspectral Images, Classification

Conteúdo

1	Introdução	1
1.1	Motivação	1
1.2	Descrição e Objectivos	2
1.3	Solução Apresentada	3
1.4	Principais Contribuições	4
1.5	Estrutura do Documento	5
2	Trabalho Relacionado	7
2.1	Introdução	7
2.2	Extracção de Características	8
2.2.1	SIFT – Scale Invariant Feature Transform	8
2.2.2	Gabor Filter	9
2.2.3	Bag-of-Features	10
2.3	Análise Computacional de Obras de Arte	11
2.3.1	Técnica de Extracção da Orientação Dominante	11
2.3.2	Classificação de Pinturas utilizando Características Locais e Globais	12
2.3.3	Autenticidade de Pinturas Baseada na Utilização de Wavelets	14
2.4	Análise de Assinaturas Digitais	19
2.4.1	Autenticação Baseada nas Características Geométricas das Assina- turas	19
2.4.2	Autenticação de Texto Utilizando Características de Textura	21
2.4.3	Identificação de Assinaturas Utilizando Pontos de Interesse	22
2.4.4	Classificação de Assinaturas através de SVMs	23
2.5	Aquisição e Análise de Imagens Hiperespectrais	25
2.5.1	Aquisição de Imagens Hiperespectrais	25
2.5.2	Objectivos para Visualização de Imagens Hiperespectrais	26
2.5.3	Visualização de Imagens Hiperespectrais Através do Mapeamento de Três Comprimentos de Onda	27

2.5.4	Visualização de Imagens Hiperespectrais <i>Principal-Components-Based</i>	28
2.5.5	Visualização de Imagens Hiperespectrais Através de Classificação Supervisionada	30
2.5.6	Visualização de Imagens Hiperespectrais Através de Classificação Não-Supervisionada	32
2.6	Análise de Imagens Hiperespectrais em Obras de Arte	32
2.6.1	Identificação de Materiais em Obras de Arte	32
2.6.2	Mapeamento e Identificação de Materiais Utilizados por Picasso	35
2.7	Discussão	36
3	Análise de Obras de Arte	39
3.1	Introdução	39
3.2	Análise de Assinaturas	40
3.2.1	Pré-Processamento dos Dados	40
3.2.2	Extracção e Organização de Características	43
3.2.3	Identificação de Períodos Artísticos	45
3.2.4	Classificação de Pinturas Baseada em Características das Assinaturas	47
3.3	Classificação de Pinturas Baseada em Características de Pincelada	49
3.4	Mapeamento de Materiais	50
3.4.1	Organização e Pré-Processamento dos Dados	50
3.4.2	Mapeamento de Materiais Através de Classificação Supervisionada	51
3.4.3	Mapeamento de Materiais Através de Classificação Não Supervisionada	55
3.4.4	Mapeamento de Materiais Através de Classificação com Base de Dados Externa	56
3.5	Classificação de Pinturas Baseada em Diferentes Tipos de Análise	68
3.6	Ambiente de Desenvolvimento	70
3.7	Discussão	70
4	Resultados	71
4.1	Resultados da Análise de Assinaturas	71
4.1.1	Identificação de Períodos Artísticos	71
4.1.2	Classificação de Pinturas Baseada em Características das Assinaturas	73
4.2	Classificação de Pinturas Baseada em Características de Pincelada	78
4.3	Resultados do Mapeamento de Materiais	79
4.3.1	Mapeamento de Materiais através de Classificação Supervisionada	80
4.3.2	Mapeamento de Materiais através de Classificação Não-Supervisionada	83
4.3.3	Resultados do Mapeamento de Materiais através de Classificação com Base de Dados Externa	84
4.4	Classificação de Pinturas Baseada em Diferentes Tipos de Análise	91
4.4.1	Atribuição de Pinturas	91

<i>CONTEÚDO</i>	xv
4.4.2 Atribuição e Autenticação de Pinturas	92
4.5 Discussão	94
5 Conclusões e Trabalho Futuro	99
5.1 Conclusões	99
5.2 Trabalho Futuro	101
6 Apêndice	111
6.1 Lista de Elementos Químicos Materiais Puros	111

Lista de Figuras

2.1	SIFT Keypoints [Low04].	8
2.2	Visualização de características extraídas utilizando o Gabor Filter [PD05] (a) imagem digital, (b) Aplicação de um filtro de Gabor.	9
2.3	Exemplo de construção do histograma que representa uma imagem [FFLT].	11
2.4	Fase de filtragem (a, b). Fase de extracção da orientação (c, d, e) [BPvdH09].	12
2.5	Arquitectura da infraestrutura construída [She09].	13
2.6	<i>Wavelet Templates</i> utilizados por cada grupo de investigação [JHB ⁺ 08]. . .	15
2.7	Extracção das características geométricas da pincelada [JHB ⁺ 08].	16
2.8	Extracção dos valores de energia [JHB ⁺ 08].	18
2.9	Extracção de Características Espaciais [jai02]	19
2.10	(a) Vertical Splitting, (b) Horizontal Splitting [MRB06].	20
2.11	Diagrama de blocos da técnica apresentada em [SBT98].	21
2.12	Pontos de interesse extraídos nas assinaturas [rui08].	22
2.13	(a)-(d) <i>Matches</i> entre pares de assinaturas [rui08].	22
2.14	Extracção de características: (a) pré-processamento (b) altura (c) histo- grama vertical máximo (d) histograma horizontal máximo (e) centro ho- rizontal (f) centro vertical (g) número local horizontal máximo (h) número local vertical máximo (i) pontos de contorno (j) características de grelha [OSK05].	24
2.15	Estrutura da SVM construída [OSK05].	25
2.16	Resultados de aplicação do método [OSK05].	25
2.17	Esquema do sistema de captação de dados. (c1) controlo digital do filtro, (c2) controlo analógico do filtro (c3) controlo do <i>shutter</i> e tempo de exposi- ção (c4) sinal digital da imagem [dAeTdC04].	26
2.18	Mapeamento de dois conjuntos diferentes de comprimentos de onda nos canais RGB [TKDO03].	27
2.19	Aplicação de diferentes algoritmos a imagens hiperespectrais (a) PCA, CLS and MCR-ALS [JAMB08].	28

2.20	Imagens pseudo-cor obtidas através do mapeamento directo dos <i>scores</i> nos canais (R, G, B). (a) (P1, P2, P3) (b) (P2, P2, P3)[TKDO03].	29
2.21	Classificação de uma imagem hiperspectral com a respectiva legenda de cores [GBHD04].	31
2.22	(a) A vermelho os <i>pixels</i> classificados como <i>lapis lazuli</i> (configuração VIS + UV). (b) A vermelho os <i>pixels</i> classificados como <i>lead-tin yellow</i> e <i>lead white</i> . (c) Exemplo de espectros adquiridos com cada um dos instrumentos [PDMDRP08].	34
2.23	(Esquerda) Imagem da pintura <i>Harlequin Musician</i> . (Centro-Esquerda) Cubo hiperspectral e <i>clustering</i> de <i>pixels</i> . (Centro-Direita) Espectro dos materiais que constituem a base de dados. (Direita) Mapeamento dos materiais. [DZT ⁺ 10]	35
2.24	(A) Imagem de luminescência infravermelhos. (B) Mapeamento dos espectros de emissão primários. (C) Imagem da pintura <i>Harlequin Musician</i> . (D) Mapeamento dos espectros de referência que são luminescentes no cubo hiperspectral. Em baixo, os espectros obtidos utilizando Luminescence Imaging Spectroscopy e Hyperspectral Camera [DZT ⁺ 10].	36
3.1	Diagrama do sistema desenvolvido para análise de obras de arte.	40
3.2	Extracção dos pontos de interesse utilizando a SIFT. (a) Aplicadas à imagem original. (b) Aplicadas à imagem após aplicação da extracção do <i>background</i>	41
3.3	Exemplos de imagens binárias obtidas após o processo de segmentação.	42
3.4	Extracção dos <i>background</i> da imagem. (a) Resultado da extracção automática. (b) Resultado da extracção manual.	42
3.5	Extracção de características. (a) Pontos de interesse SIFT. (b) Aplicação dos filtros de Gabor.	43
3.6	Esquema de construção do vocabulário de "palavras" visuais [FFLT].	44
3.7	Diagrama de blocos do método de agrupamento de assinaturas.	46
3.8	Diagrama de blocos do método de classificação de assinaturas.	48
3.9	Pontos de interesse extraídos utilizando a técnica SIFT numa pintura de Amadeo.	49
3.10	Organização da imagem hiperspectral.	51
3.11	Pré-Processamento das imagens hiperspectrais. (a) Imagem original. (b) Imagem recortada.	51
3.12	Processo de construção da base de dados de referência	52
3.13	Exemplo de variabilidade de um material ao longo de uma mesma superfície. A preto o espectro médio calculado.	52
3.14	Espectros com forma semelhante mas pertencentes a materiais diferentes (a) Base de suporte de uma pintura (b) Pigmento cinzento.	53
3.15	Paleta de tintas do pintor Amadeo de Souza-Cardoso.	58

3.16 Fluxograma de decisão construído para classificação.	59
3.17 Análise pontual EDXRF efectuada na pintura Mucha	60
3.18 Segmentação através de intervalos de tolerância de cor. (a) Ponto na pintura onde foi efectuada a amostra. (b) A branco todos os <i>pixels</i> identificados como pertencentes à amostra.	61
3.19 Exemplos de classificação. (a) Identificação do pigmento branco (A branco). (b) Identificação do pigmento preto (A preto). (c) Identificação das zonas sem análise EDXRF (A roxo). (d) Identificação das zonas com pigmentos puros. (e) Identificação das zonas com misturas de pigmentos (A preto).	63
3.20 (a) Espectro original da imagem hiperespectral. (b) Espectro após aplicação de filtro Savitzky–Golay de ordem binomial e com tamanho do filtro 9.	64
3.21 Exemplo de soma linear de dois espectros. (a) Espectro Red Vermillion. (b) Espectro Yellow Ochre. (c) Espectro soma 60% Red Vermillion e 40% Yellow Ochre.	67
3.22 Diagrama de blocos do método de classificação combinada de pinturas conjugando vários tipos de informação.	68
4.1 Imagens presentes no <i>cluster</i> 1	72
4.2 Imagens presentes no <i>cluster</i> 2	73
4.3 Exemplos de imagens de entrada no classificador (a) Classe "Amadeo". (b) Classe "Outros Artistas".	75
4.4 Exemplos de imagens de entrada no classificador (a) Classe "Amadeo". (b) Classe "Falsificações".	76
4.5 Exemplos de imagens de entrada no classificador (a) Classe "Amadeo". (b) Classe "Não Amadeo".	77
4.6 Exemplos de imagens de entrada no classificador (a) Classe "Amadeo", imagens cedidas pelo Centro de Arte Moderna. (b) Classe "Não Amadeo", imagens retiradas de http://www.guggenheim.org	78
4.7 Construção da base de dados a partir de amostras retiradas da imagem hiperespectral e respectiva paleta de cores atribuída a cada um dos pigmentos.	80
4.8 Resultado visual da classificação. (a) Todos os pigmentos mapeados. (b) Apenas mapeado o pigmento 8. (c) Apenas mapeado o pigmento 3.	81
4.9 Amostragem das regiões de interesse para validação do classificador	82
4.10 Resultado visual da classificação. (a) Todos os <i>clusters</i> mapeados. (b) Apenas mapeado o <i>cluster</i> 6. (c) Apenas mapeado o <i>cluster</i> 1.	84
4.11 Pontos da pintura onde foi efectuada a análise FORS	85
4.12 Resultado visual da classificação. (a) Todos os pigmentos de referência mapeados. (b) Apenas mapeado o pigmento 5. (c) Apenas mapeado o pigmento 2.	85

4.13	Amostragem das regiões de interesse para validação do classificador . . .	86
4.14	(a) Imagem RGB da pintura Mucha. (b) Resultado geral da classificação para teste. (c) Visualização individual de uma mistura de Red Vermillion e Yellow Cr, dois pigmentos puros	87
4.15	(a) Imagem RGB da pintura Doutor. (b) Resultado geral da classificação para teste. (c) Visualização individual de uma mistura de Red Vermillion e Yellow Ochre, dois pigmentos puros	88
4.16	(a) Imagem RGB da pintura Geometrico. (b) Resultado geral da classificação para teste. (c) Visualização individual de um pigmento puro, o Red Vermillion	88

Lista de Tabelas

3.1	Exemplo da metodologia de validação da classificação dos materiais puros.	65
4.1	Distribuição dos dados para atribuição de assinaturas.	75
4.2	Resultados obtidos na classificação de atribuição de assinaturas.	75
4.3	Distribuição dos dados para autenticidade de assinaturas.	76
4.4	Resultados obtidos na verificação de autenticidade de assinaturas.	77
4.5	Distribuição dos dados para atribuição e autenticidade de assinaturas.	78
4.6	Resultados obtidos na classificação de atribuição e autenticidade de assinaturas.	78
4.7	Distribuição dos dados para atribuição e autenticidade de pinturas com base nas características de pincelada.	79
4.8	Resultados obtidos na classificação de pinturas com base nas características de pincelada.	79
4.9	Resultados obtidos na classificação supervisionada com base de dados interna de acordo com as áreas definidas na figura 4.9. PC - Número de <i>pixels</i> classificados correctamente. T - Número total de <i>pixels</i> . A - <i>Accuracy</i> da classificação.	83
4.10	Resultados obtidos na classificação com recurso a base de dados externa. PC - Número de <i>pixels</i> classificados correctamente. T - Número total de <i>pixels</i> . A - <i>Accuracy</i> da classificação.	86
4.11	Resultados da classificação aplicada aos vários casos de estudo.	89
4.12	Distribuição dos dados para atribuição de pinturas com base no classificador conjunto de pinturas.	91
4.13	Resultados obtidos na classificação de atribuição pinturas com base no classificador combinado utilizando a SIFT.	92
4.14	Resultados obtidos na classificação de atribuição pinturas com base no classificador combinado utilizando os filtros de Gabor.	92

4.15	Distribuição dos dados para atribuição e autenticação de pinturas com base no classificador conjunto de pinturas.	93
4.16	Resultados obtidos na classificação de atribuição e autenticidade pinturas com base no classificador combinado utilizando a SIFT.	93
4.17	Resultados obtidos na classificação de atribuição e autenticidade pinturas com base no classificador combinado utilizando os filtros de Gabor.	93
4.18	Resultados obtidos para os três casos de estudo submetidos ao classificador combinado de pinturas.	94
6.1	Elementos químicos presentes em cada um dos materiais puros de Amadeo de Souza Cardoso.	111



Introdução

A vida de cada pintor continua a ser uma incógnita para cada historiador de arte, havendo um grande interesse em descobrir alguns dos segredos mais bem guardados. As ferramentas de processamento de imagem podem ser uma solução para muitos dos problemas dos historiadores.

1.1 Motivação

Na primeira década do século XXI surgiram inúmeras técnicas de análise e processamento de imagem e também novas formas de analisar sinais complexos. O incremento do poder computacional fez com que surgissem novas técnicas, mais complexas e precisas, de analisar imagens digitais. Inúmeras áreas, desde a medicina à história de arte viram neste facto uma oportunidade para aplicarem algum desse conhecimento nos seus métodos. Ao nível da história de arte, técnicas como o reconhecimento automático de padrões em imagens digitais, desenvolvidas nos últimos anos, podem ser aplicadas a imagens de obras de arte com o objectivo de validar a sua autenticidade e auxiliar na sua conservação. O desenvolvimento tecnológico também contribuiu para o surgimento de câmaras hiperespectrais, capazes de captar dados como a luz reflectida num objecto numa determinada gama de frequências, podendo constituir uma importante fonte de informação. Esta tecnologia tem sido utilizada ao longo dos anos para captação e posterior análise de imagens de terreno através de imagens de satélite. Contudo, começam a surgir aplicações deste tipo de técnicas na área da história de arte, nomeadamente ao nível da análise de materiais utilizados na pintura. As imagens de satélite permitem identificar sobretudo zonas com determinados materiais ou com a presença de determinadas espécies de árvores e minerais. No caso de uma imagem hiperespectral de uma pintura,

a sua análise pode permitir identificar os materiais utilizados pelo pintor. Assim, tanto o reconhecimento de padrões em imagens digitais como a análise de materiais poderão fornecer informações importantes acerca de um pintor e da época em que este viveu.

O processo de análise de obras de arte poderá obter benefícios com a inclusão de algumas técnicas de processamento de imagem como auxílio aos historiadores no processo de análise. Processos como a análise de autenticação de uma pintura, análise de atribuição de autor de uma pintura, análise dos materiais utilizados na mesma, análise de zonas da pintura com materiais semelhantes podem ser agilizados, recorrendo deste modo a um conjunto de ferramentas construídas para o efeito.

1.2 Descrição e Objectivos

Esta dissertação enquadra-se no âmbito do projecto “História, Materiais e Técnicas de pintores portugueses, 1850-1918: Romantismo, Naturalismo e Modernismo”, actualmente em execução na Universidade Nova de Lisboa, Faculdade de Ciências e Tecnologia. Este é um projecto multidisciplinar pois envolve uma colaboração entre o Departamento de Informática, o Departamento de Conservação e Restauro, vários historiadores, o Museu Nacional de Arte Contemporânea-Museu do Chiado, o Centro de Arte Moderna-Fundação Calouste Gulbenkian, a Casa Museu Anastácio Gonçalves e a Polícia Judiciária, tendo também o apoio da Fundação para a Ciência e Tecnologia. O projecto centra-se na análise e identificação de pinturas de Amadeo de Souza-Cardoso, pintor modernista português (1887-1918). O objectivo da dissertação é construir um sistema para classificação de pinturas em termos de autenticidade e atribuição de autor com base nas imagens digitais das pinturas, nas imagens digitais das assinaturas presentes nas pinturas e nas imagens hiperespectrais. Contudo ao longo do desenvolvimento foram surgindo objectivos secundários de acordo com as necessidades dos historiadores. Os principais objectivos da dissertação são:

1. Verificação da autenticidade e atribuição de autor de pinturas baseada nas características das assinaturas dos pintores presentes nas obras de arte.
2. Identificação de períodos na vida artística do pintor de acordo com as características visuais das assinaturas.
3. Reconstrução do classificador de autenticidade e atribuição de pinturas baseado em características de pincelada, anteriormente implementado no projecto.
4. Análise dos materiais utilizados nas pinturas através do estudo de imagens hiperespectrais captadas em várias pinturas do pintor em análise. Esta análise permitirá também efectuar a verificação de autenticidade e atribuição de pinturas.
5. Classificação de pinturas baseada em diversos tipos de informação, combinando a classificação de pinturas baseada na análise de pincelada, com a classificação de pinturas baseada nas características das assinaturas e com a análise dos materiais.

Face à existência de um conjunto de assinaturas autênticas e assinaturas falsificadas do pintor em análise, surgiu a necessidade de desenvolver uma técnica para verificação de pinturas em termos de autenticidade e atribuição de autor que pudesse ser integrada com a análise da textura já implementada de modo a construir um classificador mais robusto. Além disso, os historiadores envolvidos no projecto tinham ainda o interesse de identificar períodos de assinaturas de forma automática, podendo efectuar um estudo mais aprofundado sobre determinados períodos da vida artística do pintor.

O classificador de pinturas baseado em características de pincelada já havia sido implementado anteriormente no projecto, todavia de modo a atingir o objectivo principal desta dissertação foi necessário reconstruí-lo e adaptá-lo de modo a ser integrado com outros tipos de informação. Os resultados obtidos com este classificador podem ser consultados no trabalho descrito em [MJC⁺12].

A análise de imagens hiperespectrais permite uma análise mais aprofundada de uma pintura recorrendo a vários tipos de informação. Através da luz reflectida por cada um dos pigmentos utilizados pelo pintor, é possível identificar regiões pintadas com o mesmo material numa pintura. Normalmente este processo era efectuado de forma pontual, havendo assim um desaproveitamento da restante informação contida na imagem hiperespectral. Esta componente da dissertação surge deste modo com o objectivo de construir um método de análise mais automático, assim como um método de classificação de autenticidade e atribuição de pinturas baseado na análise dos materiais.

A construção de um classificador combinado de pinturas que possa ser utilizado pelos museus para validação de autenticidade e atribuição do autor de pinturas consiste num dos objectivos principais da dissertação, assim como se trata também do objectivo final do projecto em que a dissertação se insere. Cada um dos objectivos anteriores contribui directamente para a realização deste objectivo, tendo surgido vários objectivos secundários ao longo do desenvolvimento da dissertação.

1.3 Solução Apresentada

No âmbito da dissertação foram desenvolvidas um conjunto de técnicas com vista a auxiliar os historiadores e conservadores no processo de análise de obras de arte. Essas técnicas foram desenvolvidas de forma isolada, contudo o seu conjunto constitui um sistema para análise de obras de arte com vários módulos. No futuro, espera-se ser possível desenvolver um sistema de verificação de pinturas em termos de autenticidade e atribuição de autor que possa ser utilizado por cada um dos museus nas suas colecções e também pela polícia judiciária em pinturas apreendidas. O desenvolvimento dividiu-se em quatro fases, cada uma delas com um âmbito distinto, mas com o objectivo comum de ajudarem o historiador de arte no processo de análise de um determinado pintor.

A primeira fase englobou a análise das assinaturas em duas vertentes, verificação de autenticidade e atribuição de autor à obra e numa segunda vertente o agrupamento de assinaturas com características visuais semelhantes. O pintor em análise, Amadeo

de Souza-Cardoso, utilizou um conjunto de assinaturas para autenticar as suas obras. O primeiro objectivo foi a construção de um classificador que permitisse distinguir entre assinaturas autênticas do pintor e assinaturas falsificadas ou pertencentes a outros pintores. Outro dos objectivos foi a construção de um algoritmo de agrupamento de assinaturas visualmente semelhantes, de modo a identificar possíveis períodos na vida artística do pintor.

Numa segunda fase efectuou-se a reconstrução do classificador de autenticidade e atribuição de pinturas baseado nas características da pincelada. Como já foi afirmado, este já tinha sido construído anteriormente no âmbito do projecto mas havia a necessidade de o adaptar aos objectivos pretendidos.

A terceira fase foi direccionada para a análise das imagens hiperespectrais. Tal como foi relatado anteriormente, havia a necessidade de tornar a análise de uma imagem hiperespectral um processo mais automático. Tendo como base as imagens hiperespectrais captadas em quadros do pintor em análise, procedeu-se ao desenvolvimento de um conjunto de técnicas para auxiliar os historiadores no processo de identificação de zonas da pintura pintadas com o mesmo material. A identificação de zonas com materiais semelhantes ao longo de uma ou várias pinturas foi a principal prioridade, contudo era desejo dos historiadores poder perceber se existiam zonas da pintura que não tinham sido feitas pela mão do artista ou se constituíam misturas de pigmentos puros.

Por fim, a última fase consistiu no desenvolvimento de um sistema integrado de classificação de pinturas baseado em várias evidências, quer seja através das características de pincelada presentes na pintura, através das características das assinaturas ou da análise dos materiais utilizados pelo artista. Este sistema é um dos principais desafios que se pretende resolver com o desenvolvimento da dissertação, englobando assim a verificação de autenticidade e a atribuição de autor de pinturas.

1.4 Principais Contribuições

Este trabalho não se restringiu ao desenvolvimento de uma aplicação ou plataforma mas sim ao desenvolvimento de um conjunto de técnicas que podem ser aplicadas a vários conteúdos e adaptadas de acordo com o caso de estudo que se pretende analisar. As principais contribuições da dissertação são:

1. Um sistema para verificação de autenticidade e atribuição de autor de pinturas baseado nas características das assinaturas de pintores. Este engloba o desenvolvimento de um algoritmo de classificação de imagens baseado em reconhecimento de padrões, cujo resultado permite afirmar com um determinado grau de certeza se uma assinatura foi feita pela mão do pintor ou pertence a um outro pintor, podendo também ser uma falsificação.
2. Uma ferramenta para agrupamento de assinaturas de acordo com características visuais semelhantes.

3. Reconstrução do classificador de pinturas com base nas características de pincelada de modo a ser integrado no classificador combinado de pinturas.
4. Um método para agrupamento de *pixels* semelhantes numa imagem hiperespectral através de semelhanças entre os espectros. Este método pretende auxiliar o historiador na identificação visual de zonas da pintura com pigmentos semelhantes. A partir da identificação de zonas semelhantes será também possível encontrar zonas da pintura que não tenham sido feitas pelo pintor em estudo ou zonas com misturas de pigmentos.
5. Um sistema para validação de autenticidade e atribuição de autor de pinturas baseado em características da pincelada do artista, em características das assinaturas do artista e com base na análise dos materiais presentes numa pintura. Assim, cada pintura é classificada com um maior grau de certeza devido à utilização de vários tipos de evidências.

1.5 Estrutura do Documento

A estrutura do documento é a seguinte:

1. **Introdução:** Neste capítulo apresenta-se uma visão geral da dissertação, realçando a motivação, o contexto em que se insere, a solução proposta e as principais contribuições previstas.
2. **Trabalho Relacionado:** Apresenta uma visão geral e resumida de alguns trabalhos já realizados e que se relacionam com a solução apresentada.
3. **Análise de Obras de Arte:** Este capítulo relata o desenvolvimento das técnicas de reconhecimento de padrões relatadas anteriormente de acordo com o conjunto de objectivos definidos.
4. **Resultados:** Este capítulo descreve a forma como as diferentes técnicas foram avaliadas e os respectivos resultados da avaliação de cada uma delas de modo a mostrar a sua robustez e eficácia.
5. **Conclusões e Trabalho Futuro:** Por fim, apresentam-se as conclusões resultantes de todo o trabalho desenvolvido e um conjunto de indicações para o trabalho futuro no seguimento do respectivo desenvolvimento.



Trabalho Relacionado

Este capítulo descreve alguns trabalhos de investigação que se relacionam directamente com o âmbito da dissertação, sendo apenas focados os pormenores mais relevantes.

2.1 Introdução

Este capítulo apresenta trabalhos realizados na análise de obras de arte que utilizam técnicas de reconhecimento de padrões. O desenvolvimento da dissertação incidiu numa primeira fase sobre reconhecimento de padrões em imagens digitais, processo que é efectuado em duas fases principais. Primeiro é feita a extracção de informação a partir da imagem digital RGB. Essa informação pode ser baseada na cor ou na textura da imagem. A secção "Extracção de Características" descreve algumas das técnicas computacionais que permitem tanto a extracção de características em imagens como a organização das mesmas para futura classificação baseada em padrões semelhantes. Em segundo lugar, surge a classificação com vista ao reconhecimento de padrões semelhantes ao longo das imagens. Assim, a cada imagem de entrada do algoritmo é atribuída uma etiqueta de acordo com a construção do classificador. A secção "Análise Computacional de Obras de Arte" descreve alguns dos métodos de classificação existentes aplicados na área da análise de obras de arte. A secção "Análise de Assinaturas Digitais" foca algumas das características específicas extraídas para a análise de assinaturas incidindo principalmente nos métodos de classificação de assinaturas.

Além da análise de imagens digitais RGB, o desenvolvimento da dissertação incidiu também sobre a análise de imagens hiperespectrais. A quarta secção designada "Aquisição e Análise de Imagens Hiperespectrais" aborda a análise de imagens hiperespectrais, descrevendo como estas são adquiridas. A secção "Análise de Imagens Hiperespectrais

em Obras de Arte" apresenta alguns dos trabalhos de investigação que efectuaram análise de materiais em obras de arte a partir de imagens hiperespectrais, tentando utilizar vários tipos de informação.

Por fim, a última secção apresenta uma discussão sobre os trabalhos analisados, fazendo a respectiva ligação com os problemas que se pretendem resolver com o desenvolvimento da dissertação.

2.2 Extracção de Características

Esta secção apresenta de forma resumida algumas das técnicas para extracção e organização de características de textura em imagens digitais.

2.2.1 SIFT – Scale Invariant Feature Transform

Este método foi proposto por David Lowe¹ [Low04], sendo um dos mais utilizados actualmente por vários sistemas de reconhecimento de objectos a partir de um conjunto de características (*features*). O método detecta um conjunto de pontos de interesse numa imagem, sendo esses pontos de interesse organizados em vectores de características.

As localizações dos pontos de interesse candidatos são obtidas como máximos e mínimos do resultado de aplicação de funções Gaussian em várias escalas. Após obtidas as localizações, é aplicado um modelo a cada ponto de interesse de modo a seleccionar os pontos de interesse de acordo com a sua estabilidade, por exemplo, pontos com pouco contraste devem ser rejeitados (são mais sensíveis ao ruído). Baseado nos gradientes locais atribui-se uma ou mais orientações a cada ponto de interesse (ver figura 2.1). Estes pontos de interesse são representados sob a forma de um descritor e agrupados no conjunto de descritores de uma imagem.

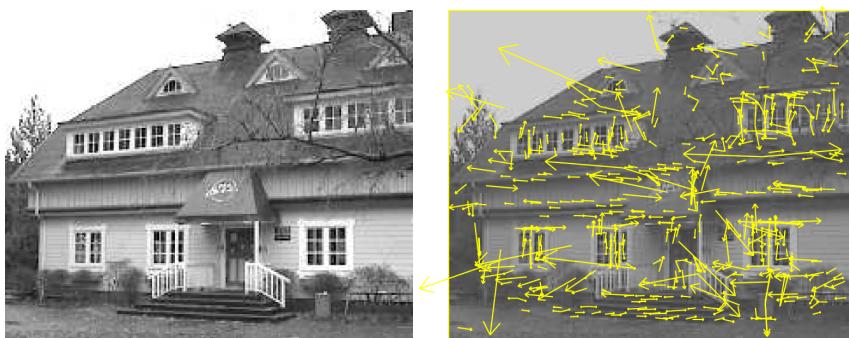


Figura 2.1: SIFT Keypoints [Low04].

O parágrafo anterior apresenta um resumo do método, contudo a identificação dos pontos de interesse é um processo que envolve várias técnicas sendo interessante reter principalmente as suas capacidades. Cada um dos pontos de interesse identificados pelo

¹<http://www.cs.ubc.ca/~lowe/keypoints/>

método é invariante à translação da imagem, à mudança de escala, à rotação, parcialmente invariante às mudanças de iluminação e invariante à distorção geométrica local. As características são distintivas sendo que uma única característica pode ser identificada eficazmente numa base de dados com uma elevada quantidade de características.

Além das vantagens enunciadas, existe também a vantagem de ser possível efectuar uma comparação (*matching*) eficaz entre os descritores, permitindo a identificação e reconhecimento de objectos. Além de propor uma técnica para extracção de características, David Lowe propôs também um método para *matching* de características baseado no método Best-Bin-First (BBF) [BL97]. Em resumo, o *matching* só será considerado se o ponto de interesse com a menor distância for maior do que uma percentagem da distância do segundo ponto de interesse mais próximo. Assim, garante-se que o *matching* só ocorre quando não há dois pontos de interesse muito parecidos com o ponto de interesse em teste, o que elimina falsos *matches*.

2.2.2 Gabor Filter

Este método foi proposto em 1946 por Dennis Gabor [Gab46] quando este propôs a representação de sinais complexos como a combinação de funções elementares. Em geral, este método traduz-se num filtro linear utilizado para detecção de padrões. As funções elementares de Gabor, principalmente as representações de frequência e orientação, são semelhantes às funções do sistema visual humano, sendo particularmente adequadas para representação e discriminação de textura. As características extraídas de imagens utilizando o Gabor Filter têm sido utilizadas para segmentação de textura, reconhecimento de escrita ou de impressões digitais [SBF07].

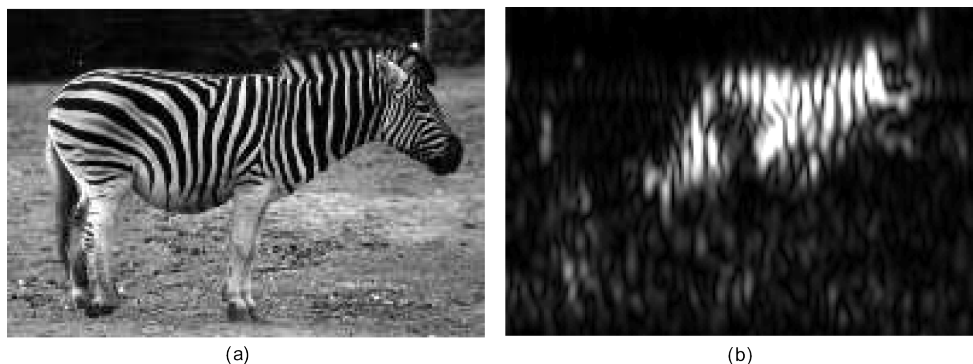


Figura 2.2: Visualização de características extraídas utilizando o Gabor Filter [PD05] (a) imagem digital, (b) Aplicação de um filtro de Gabor.

Mais especificamente, o Gabor Filter é definido como o produto de um *kernel* gaussiano com uma sinusóide. Esta técnica de extracção de características consiste na aplicação de um banco de filtros com várias escalas e orientações a uma imagem. Estes filtros procuram detectar na imagem diversos padrões. Estes padrões, podem ser, por exemplo, curvas capturadas através da textura da imagem (ver figura 2.2). Por vezes os filtros de Gabor confundem-se com as Gabor Wavelets, na medida em que as funções de Gabor,

normalizadas de certa forma, podem ser usadas como uma wavelet "mãe" para gerar uma família de Gabor Wavelets não ortogonais. Estas surgiram pelo facto de alguns neurologistas sugerirem que a estrutura espacial das células receptoras do córtex visual com diferentes tamanhos são virtualmente invariantes. Assim, Tai Sing Lee [Lee96] propôs que estas células eram melhor modeladas como uma família de Gabor Wavelets 2D mostrando no domínio da frequência na forma log-polar. Os autores de [FRDF01] utilizam como representação para as Gabor Wavelets um banco de filtros de Gabor normalizados para se obter respostas iguais a 0 e desenhados para diminuir a redundância na representação.

Ao longo dos anos este método tem sido muito utilizado para extração de características, sendo um dos principais métodos utilizados na análise de obras de arte. A utilização do banco de filtros de Gabor e das Gabor Wavelets pretende capturar as orientações principais da pincelada, identificando assim a pincelada característica de cada pintor. O trabalho apresentado em [BPvdH09] utiliza um banco de filtros de Gabor circular com o intuito de captar a orientação dominante da pincelada. A extração de características neste caso traduz-se nos contornos captados, sendo depois estes processados de modo a encontrar a pincelada dominante. Outro dos casos onde surgem os filtros de Gabor é no trabalho apresentado em [She09]. Neste trabalho são extraídos vários conjuntos de características, incluindo características de textura que são obtidas através da aplicação de um filtro de Gabor. Num dos principais trabalhos realizados sobre a obra de Van Gogh, a extração de características foi efectuada utilizando Gabor Wavelets de modo a captar as pinceladas mais hesitantes, sugerindo que este tipo de características poderia ser representativo de falsificações. Além da utilização de Gabor Wavelets, foram também utilizadas Complex Wavelets, uma aproximação mais simplificada das Gabor Wavelets, de modo a captar as diversas escalas onde determinados pormenores da pintura se destacam [JHB⁺08].

2.2.3 Bag-of-Features

A técnica Bag-of-Features surgiu a partir do método Bag-of-Words utilizado no processamento de documentos [BYRN99]. O método Bag-of-Words caracteriza-se pelo processamento de um documento e a construção de histogramas de ocorrências de cada uma das palavras diferentes existentes no documento. Alguns investigadores [NJT06], na área da computação multimédia, verificaram que este conceito poderia ser aplicado a imagens, sendo estas tratadas de forma análoga a documentos e as características extraídas efectuando a analogia com as palavras do documento.

Este método é principalmente utilizado para organização de características extraídas de imagens digitais com vista a futura classificação. Ao longo dos anos este método tem-se tornado cada vez mais popular na classificação de imagens de acordo com o conteúdo de cada uma delas, tendo obtido resultados positivos em trabalhos anteriores [NJT06]. O método divide-se em quatro fases: (1) extração de características das imagens, através

da utilização, por exemplo, da técnica SIFT; (2) aprendizagem do "vocabulário visual", isto é, construção do dicionário de palavras visuais possíveis; (3) quantificação das características utilizando o vocabulário de palavras construído, ou seja, atribuir um grau de importância a cada uma das palavras do vocabulário; (4) por fim, cada imagem é representada por um histograma de ocorrências de "palavras visuais", o que se traduz na frequência com que cada palavra ocorre nessa determinada imagem (ver figura 2.3). Face às fases do método descritas, é possível afirmar que este permite construir um descritor global para uma imagem a partir de um conjunto de descritores locais, como é o caso do conjunto de características extraídas utilizando a técnica SIFT.

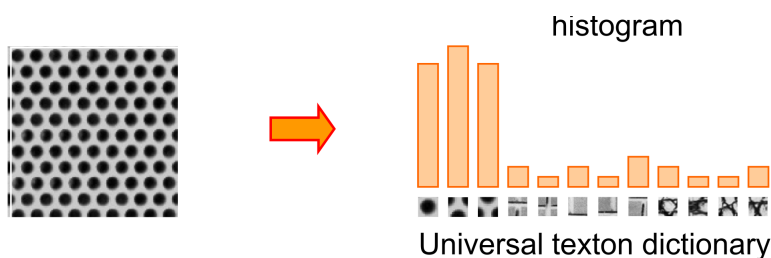


Figura 2.3: Exemplo de construção do histograma que representa uma imagem [FFLT].

2.3 Análise Computacional de Obras de Arte

Esta secção descreve alguns trabalhos de investigação em que a análise de obras de arte foi efectuada recorrendo a técnicas computacionais, abordando sobretudo metodologias para reconhecimento de padrões.

2.3.1 Técnica de Extracção da Orientação Dominante

Quando os peritos analisam obras de arte, as características espaciais são um dos principais focos de atenção na análise, isto porque o padrão da pincelada é como a impressão digital do artista. A orientação, a forma e a distribuição das pinceladas são pistas muito importantes para análise e nas quais esta técnica é baseada. Assim, a técnica POET [BPvdH09] é baseada na aplicação de um banco de filtros circular para detecção de contornos paralelos e numa fase de extracção dedicada apenas à orientação da pincelada. Esta técnica foi inspirada na observação e análise do desempenho humano na tarefa de análise de obras de arte, complementando essa análise com o poder computacional que os humanos não têm. Assim, esta técnica divide-se numa fase de filtragem e uma fase de extracção da orientação, ambas aplicadas a fragmentos de uma imagem.

A fase de filtragem (ver figura 2.4) caracteriza-se pela aplicação de um filtro circular a um fragmento, construído de modo a satisfazer dois critérios: (1) invariância à orientação, ou seja, o filtro deve captar todos os contornos orientados em todas as direcções; (2) captação de contornos dentro de uma determinada gama de frequências espaciais.

A fase de extracção da orientação (ver figura 2.4) transforma o fragmento resultante da filtragem num conjunto de objectos binários orientados que correspondem às pinceladas mais brilhantes. Para se obter uma imagem binária, aplica-se um simples método de *threshold* multi-nível, onde todos os valores obtidos da filtragem são mapeados num intervalo entre 0 e 1. O objectivo neste patamar é maximizar o número de objectos orientados numa imagem binária. Define-se assim como objecto orientado um grupo de pelo menos oito *pixels* diferentes de 0 e em que a elipse envolvente tem o maior comprimento de eixo que excede 70% o comprimento do seu menor eixo. Depois de seleccionado o valor do *threshold* que maximiza o número de objectos orientados, ordenaram-se esses mesmos objectos de acordo com o maior comprimento de eixo da elipse envolvente. O objecto com o maior valor de excentricidade da elipse envolvente é escolhido como orientação predominante do respectivo fragmento.

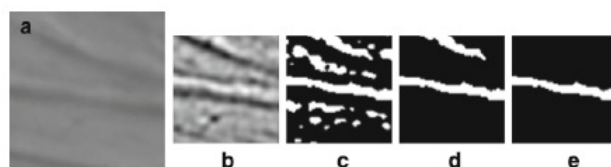


Figura 2.4: Fase de filtragem (a, b). Fase de extracção da orientação (c, d, e) [BPvdH09].

Esta técnica foi validada através de duas abordagens: comparação com os resultados obtidos por humanos e comparação com outras técnicas existentes no estado da arte. Através da comparação cruzada entre a POET e os vários peritos, concluiu-se que os resultados obtidos computacionalmente estão de acordo com a maioria dos resultados apresentados pelos peritos. Esta técnica foi também comparada com outras duas técnicas: os filtros adaptáveis de escala única (*single-scale steerable filters* [FA91], SF, designação inglesa) e a análise dos componentes principais multi-escala (*multi-scale principal components analysis* [FM02], MS-PCA). Os resultados obtidos com a POET foram significativamente melhores do que os resultados obtidos com as restantes técnicas, concluindo-se assim é uma técnica atractiva e eficaz para estimar a orientação predominante da pincelada em imagens digitais.

2.3.2 Classificação de Pinturas utilizando Características Locais e Globais

O processo de classificação de imagens tem, em geral, duas fases: extracção de características e classificação em classes. As características de uma imagem podem ser locais ou globais, diferindo em vários aspectos. No trabalho apresentado em [She09] são testadas várias configurações de características de modo a identificar qual o tipo de características que permite um melhor desempenho.

2.3.2.1 Extracção de Características Locais e Globais

As características extraídas podem ser locais ou globais. No trabalho descrito em [She09], entre as características globais extraídas destacam-se: (1) Histogramas de Cor, (2) Textura (filtros de Gabor), (3) Forma (histograma normalizado com as direcções dos contornos para cada imagem) e (4) Momentos de cor. As características locais aplicam-se a fragmentos de uma imagem tratados de forma individual. Uma imagem é fragmentada em 16 blocos e para cada um desses blocos constroem-se vectores de características baseados no cálculo das 2D Gabor Wavelets.

2.3.2.2 Classificação de Pinturas

Para classificar as pinturas em [She09], foi construída uma infraestrutura com duas grandes camadas: uma contendo os classificadores de características locais e globais, sendo estes baseados numa rede neuronal com uma função de base radial (radial basis function, RBF) e outra camada contendo uma fusão de esquemas de classificação de pontuação (*score*) baseados em regressão linear.

O classificador (ver figura 2.5) baseado na rede neuronal RBF foi criado, contendo três camadas: a camada de entrada de dados, a camada oculta e a camada de saída de dados. As características individuais são extraídas e o classificador é construído com base em cada característica. Para cada vector de características que entra na rede neuronal, uma pontuação é calculada.

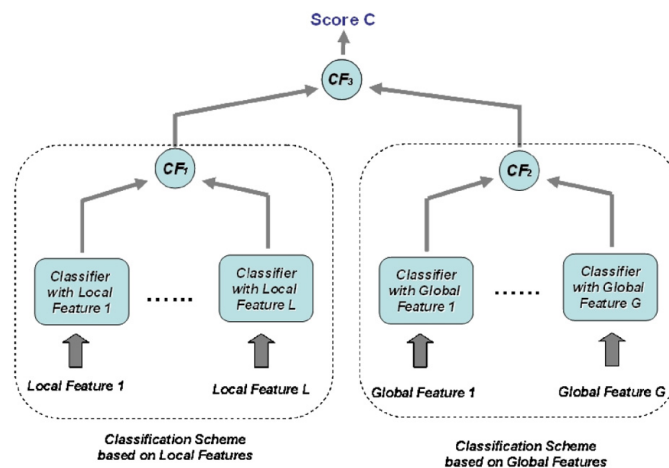


Figura 2.5: Arquitetura da infraestrutura construída [She09].

Na camada de classificação de pontuação, entram as pontuações calculadas na camada anterior, sendo este um processo incremental. De modo a melhorar a eficiência do modelo de classificação, foi construído um esquema cuja designação é Binary Output Codes with Penalty. Nesta camada, os dados passam por duas fases: a fase de treino onde é construída a palavra código única de cada pintor; e a fase de teste onde é calculada a distância de Hamming entre cada vector de saída da rede neuronal (obtido através dos vectores de características de entrada) e a palavra de código para cada pintor. Com

este esquema obtém-se uma medida de verosimilhança entre a palavra de código que representa cada artista e um vector de características.

Diferentes conjuntos de características têm diferentes graus de importância no cálculo da pontuação de uma determinada imagem. O esquema de fusão das pontuações é dividido em duas camadas: uma camada de propagação das redes, combinando os dados de saída de cada um dos classificadores; e uma camada com uma função de propagação (neste caso utilizou-se regressão linear) para combinar as pontuações mais relevantes obtidas na camada anterior e produzir a etiqueta final de uma pintura, indicando o artista que a pintou.

2.3.2.3 Validação da Técnica

Este trabalho foi validado através da comparação do seu desempenho com o desempenho de outras técnicas. Resultados dessas comparações relatados em [She09] mostram que a taxa de acerto para a técnica construída foi de 69.7%, corroborando assim a eficiência da técnica face às taxas de acerto obtidas com outras técnicas já implementadas e que revelaram percentagens obtidas inferiores a 55%. Outro dos aspectos importantes na validação da técnica foi identificar quais os tipos de características que conduziam a um melhor desempenho do sistema, verificando-se que uma combinação das características globais com as características locais (taxa de acerto obtida de 69.7%) era a melhor configuração do sistema invés de uma configuração com apenas um tipo de características (taxas de acerto de 60.3% e 58.7% para a configuração apenas com características locais ou globais, respectivamente).

2.3.3 Autenticidade de Pinturas Baseada na Utilização de Wavelets

A análise de autenticação de obras de arte é feita através da combinação de inúmeras técnicas, sendo uma delas a análise da presença da "assinatura" do artista, ou seja, a pincelada que o caracteriza exclusivamente. Isto sugeriu que a análise de imagens digitais de pinturas poderia ajudar os peritos em arte no processo de autenticação.

Para construir um modelo computacional suficientemente robusto de modo a classificar as pinturas quanto à sua autenticidade, primeiro é necessário perceber que critérios são tidos em conta na análise da pincelada efectuada por peritos em arte. Essa análise segue geralmente um conjunto de passos, tendo em conta os vários factores que afectam a qualidade da pincelada, sendo esses passos descritos em [JHB⁺08].

Em primeiro lugar, deve estabelecer-se que partes de uma pintura devem ser eliminadas da análise, ou seja, que não foram feitas pela mão do artista, como por exemplo a deterioração da cor ou o surgimento de rachas na pintura. No passo seguinte, descreve-se a pincelada original observada ao longo do quadro, isto é, procura-se o uso recorrente de determinados padrões de pincelada, envolvendo pinceladas rápidas e efectuadas de forma rítmica dispostas numa orientação em particular.

Por vezes, os analistas tentam também identificar o tamanho específico e a forma dos

pincéis utilizados em cada pintura através das marcas que estes deixaram na mesma. A fluidez da tinta é também um dos passos tidos em conta na análise pois é de facto importante analisar o impacto visual dos materiais de pintura utilizados.

Na sequência dos passos anteriores, é também realizada a análise da qualidade do óleo e das tintas utilizadas na pintura de modo a identificar por exemplo de onde são provenientes os materiais ou onde foram fabricados. Por fim, o último passo a ter em conta na análise é a análise da superfície onde a pintura se encontra, mais propriamente a tela em si. Isto porque por vezes as pinceladas são efectuadas sobre uma tela seca, outras sobre um tela humedecida ou até sobre figuras já pintadas na mesma tela.

Em [JHB⁺08] é descrito o trabalho de três grupos de investigadores com vista à verificação de autenticidade de obras de arte tendo em conta o conjunto de indicações apresentado anteriormente. Ambos os grupos desenvolveram técnicas utilizando o mesmo conjunto de dados, 101 imagens de pinturas de Van Gogh e possíveis falsificações. Cada grupo incidiu a sua investigação num tipo de wavelets diferente (ver figura 2.6).

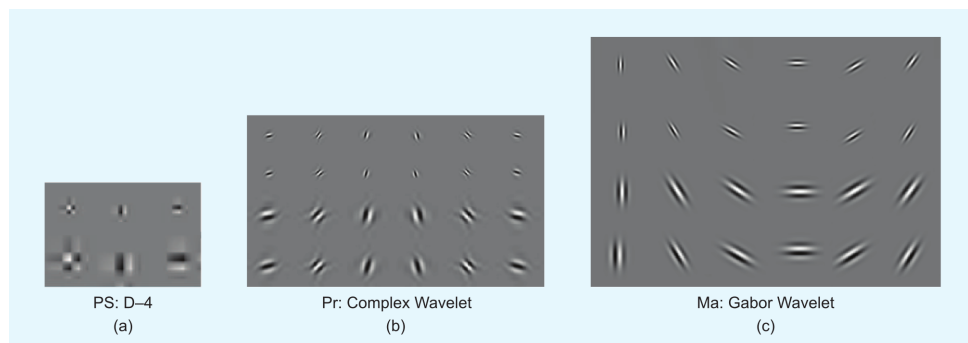


Figura 2.6: *Wavelet Templates* utilizados por cada grupo de investigação [JHB⁺08].

2.3.3.1 Análise da Modelação Geométrica da Pincelada

Inicialmente, este grupo de investigadores oriundos de Penn State extraiu do conjunto de dados 23 pinturas que eram, sem qualquer dúvida, pertencentes a Van Gogh e que representavam diferentes períodos na sua vida artística, sendo estas utilizadas como conjunto de treino [JHB⁺08].

Cada pintura foi dividida em *patches* de aproximadamente 512×512 *pixels* e dois tipos de características foram extraídas. O primeiro tipo era baseado na textura, sendo extraídas da transformada de *wavelet* ortonormal D-4 (ver figura 2.6, (a)). Os coeficientes extraídos reflectem as diferentes orientações e variações bruscas na imagem, restringindo-se a sua utilização apenas aos coeficientes com a escala mais pequena. O segundo tipo de características incluía as características geométricas da pincelada, sendo estas características de mais alto-nível e mais perceptíveis pelo olho humano do que a textura (ver figura 2.7). Entre as características geométricas calculadas destacam-se o comprimento, a orientação e a curvatura média de cada contorno da pincelada.

Para comparar as características extraídas, construíram-se modelos estatísticos para

ambos os tipos de características. No caso das características baseadas na textura, construiu-se um modelo espacial, mais especificamente, uma variação do 2-D HMM (Hidden Markov Model)[CRB00], isto porque neste caso a relação espacial entre as características era um factor importante. No caso das características baseadas na geometria da pincelada, construíram-se modelos independentes e identicamente distribuídos, sendo que a relação espacial das características era menos importante. Este modelo era baseado numa distribuição discreta pois o número de linhas de contorno da pincelada varia abruptamente ao longo dos vários *patches*.

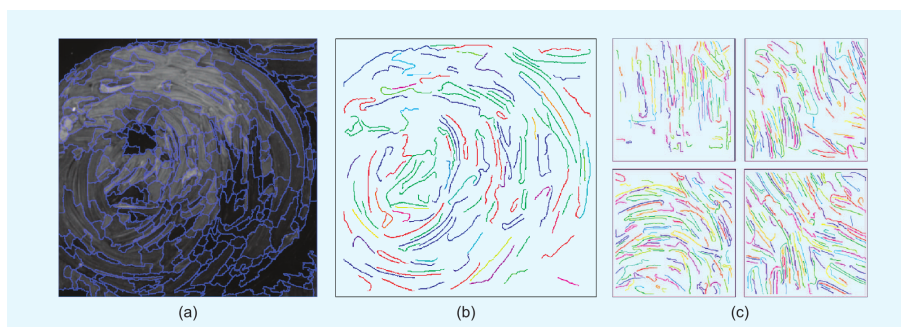


Figura 2.7: Extração das características geométricas da pincelada [JHB⁺08].

Depois de removidas as 23 pinturas para o conjunto de treino, foi calculada a distância entre as restantes imagens (78) e o conjunto de treino. Posteriormente, as 78 imagens foram ordenadas de acordo com as distâncias calculadas. Por exemplo, tendo apenas em conta a modelação das características da textura, nas cinco imagens mais semelhantes do conjunto de treino, quatro pertencem a Van Gogh enquanto uma pertence a Gauguin. Quanto às menos semelhantes ao conjunto de treino, três delas não pertencem mesmo a Van Gogh enquanto duas são sem dúvida autênticas de Van Gogh. Isto acontece pois o conjunto de treino que serve de base de conhecimento ao processo de aprendizagem pode não ser suficientemente representativo para cobrir todos os estilos de pincelada de Van Gogh.

2.3.3.2 Análise da Pincelada em Diferentes Escalas

Este grupo (oriundo de Princeton) decidiu utilizar apenas 76 imagens do total disponibilizadas, sendo 65 de Van Gogh, seis de outros artistas e cinco atribuídas a Van Gogh, mas que não foram definitivamente confirmadas [JHB⁺08].

Extração das Características Também neste caso as imagens foram divididas em *patches* de 512×512 *pixels*. Para cada *patch*, os coeficientes de *wavelet* (ver figura 2.6, (b)) são modulados por uma Hidden Markov Tree (HMT). Neste modelo, cada coeficiente é associado a um estado oculto, tendo um de dois valores, contorno ou não contorno, indicando se o coeficiente do *wavelet template* se sobrepõe a um contorno na imagem. Isto permitiu definir uma matriz de transições entre os estados ocultos, obtendo-se depois um

conjunto de vectores designados por vectores m -similaridade sobre os quais será possível calcular a semelhança entre duas imagens. As 108 características extraídas segundo o modelo anterior foram ordenadas e numeradas de acordo com a sua eficácia em distinguir *patches* provenientes das obras de Van Gogh e não Van Gogh. Verificou-se que as características que dominavam a ordenação permitiam identificar as escalas a que a informação detalhada se realça. Estas escalas características acabam por ser diferentes para diferentes orientações, parecendo ser uma característica do estilo de Van Gogh.

Classificação das Pinturas Para efectuar a classificação de pinturas em [JHB⁺08], um algoritmo de escala multidimensional foi construído para dispor os 76 pontos (um ponto por imagem) num espaço 3-D, de modo a mostrar as similaridades entre as 76 imagens. Nesta representação o centro C_{vg} do *cluster* de imagens pertencentes a Van Gogh foi determinado. A separação das imagens é feita usando um classificador: para um determinado raio r , os pontos mais perto de C_{vg} do que r representavam pinturas de Van Gogh, enquanto que as mais distantes de r são classificadas como não sendo de Van Gogh. Para teste do classificador efectuou-se uma validação cruzada *leave-one-out*, ou seja, cada imagem é classificada de acordo com o C_{vg} e o r das restantes 75 imagens do conjunto.

Identificar Pinturas Falsificadas O algoritmo anterior permitiu classificar as pinturas referentes a van Gogh ou não referentes, mas em termos de pinturas falsificadas não foi possível obter uma separação evidente, pois tratam-se de quadros semelhantes mas não autênticos. De modo a efectuar esta classificação, procedeu-se a uma análise mais detalhada, utilizando-se *patches* de 128×128 pixels. O objectivo era assim captar zonas onde existiam pinceladas hesitantes, causadas por uma redução na fluidez da tinta aquando da cópia de uma pintura para outra. Observou-se então que as pinturas falsificadas tinham coeficientes de *wavelet* com valores muito altos, o que sugeria um excesso de coeficientes com valores muito elevados nas escalas mais finas que o algoritmo utilizava. Assim calculou-se a mediana dos coeficientes de *wavelet* nas duas melhores escalas e isso permitiu separar as pinturas falsificadas das verdadeiras, pois as que não são autênticas possuíam um excesso de características de tamanho 0.25 - 0.5 mm principalmente provenientes de pinceladas mais hesitantes.

2.3.3.3 Análise de Pinturas Inspirada em Processos Biológicos

Este grupo, designado por Maastricht Group, guiou-se por três princípios básicos: (1) os contornos são importantes, (2) as imagens devem ser analisadas a várias escalas, e (3) as semelhanças entre as imagens são reflectidas na textura local (padrão da pincelada) [JHB⁺08].

Análise de Pinturas através de Gabor *wavelets* Os princípios definidos foram implementados aplicando filtros de Gabor *wavelet* (ver figura 2.6, (c)) orientados em várias

escalas e organizando os coeficientes obtidos em histogramas. Estes histogramas representam a decomposição da imagem em "valores de energia" (ver figura 2.8), um valor de energia para cada *pixel*, orientação e escala. Neste caso, os filtros são aplicados a *patches*, sendo que a energia total de uma imagem é calculada simplesmente somando os valores de energia de cada *patch* que a constitui, o que significa contar os contornos na pintura. A partir de uma escala escolhida para análise, a escala mais fina onde melhor as Gabor *wavelets* realçam as pinceladas, procedeu-se à análise dos valores de energia das várias pinturas. Verificou-se que, por exemplo, para uma das pinturas falsificadas foram obtidos maiores valores de energia do que para as pinturas verdadeiras, sugerindo a existência de pinceladas hesitantes.

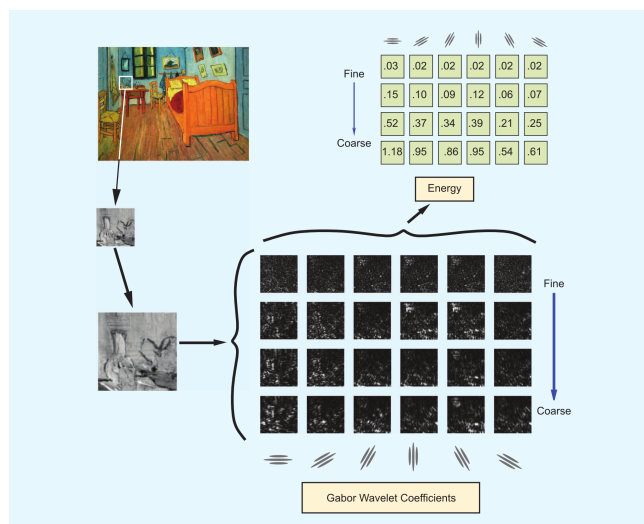


Figura 2.8: Extração dos valores de energia [JHB⁺08].

Para tornar este processo (descrito em [JHB⁺08]) mais rigoroso, construiu-se uma representação mais detalhada das imagens através da criação de um histograma multi-dimensional que captura a informação de configuração das frequências espaciais e orientações num único *patch*. Utilizando estes histogramas para as pinturas, este grupo incluiu na sua análise um modelo para efectuar classificação automática com auxílio de Support Vector Machines (SVM) [MMR⁺01]. Este requeria um conjunto de vectores de entrada e as respectivas *labels* representando as suas classes. Os vectores de entrada são assim os vectores multi-dimensionais enquanto que as *labels* representam a autenticidade dos quadros, Van Gogh e não Van Gogh. A seguir, foi efectuada uma validação cruzada *leaving-one-out* para treinar o desempenho geral da SVM construída com as 101 pinturas disponibilizadas.

Com a combinação de vectores multi-dimensionais, quatro das seis pinturas não Van Gogh foram identificadas, enquanto que as outras duas foram classificadas de forma errada como pertencentes a Van Gogh.

2.4 Análise de Assinaturas Digitais

Esta secção apresenta alguns dos trabalhos existentes no estado da arte sobre análise de assinaturas, mais especificamente para validação de autenticidade de assinaturas. No estado da arte existe um conjunto de trabalhos nesta área com resultados significativos, existindo diferentes tipos de características a extrair de assinaturas, e também diferentes tipos de classificadores. Em [LP94] é feita uma abordagem sobre o estado da arte na análise de assinaturas, enunciando alguns dos principais métodos aplicados nesta área de investigação. De referir a existência de dois tipos de métodos para verificação de autenticidade de assinaturas, sendo um designado *online* e outro *offline*. Um sistema *online* captura as características dinâmicas da assinatura como tempo, pressão e velocidade de escrita enquanto que um sistema *offline* analisa as imagens estáticas das assinaturas, sendo um processo mais complicado de realizar do que a verificação *online*.

2.4.1 Autenticação Baseada nas Características Geométricas das Assinaturas

A autenticação de assinaturas de forma automática continua a ser um dos grandes desafios devido às características próprias das assinaturas [DIL⁺04]. Determinados pontos nas assinaturas, tal como os pontos de começo e os pontos de fim da caneta ou pontos onde a trajectória muda transportam informações importantes. No trabalho apresentado em [jai02] estes pontos são considerados críticos, incidindo a investigação sobre as características geométricas da assinaturas. Os autores de [jai02] sugerem um método que segue várias fases de modo a efectuar a classificação de autenticidade.

Na fase de pré-processamento, aplica-se um Gaussian Filter de modo a tornar a assinatura mais uniforme, eliminando grande parte do ruído. Todos os elementos da escrita são combinados num só, isto é, se uma assinatura tiver várias palavras, elas ficam todas ligadas após este passo.

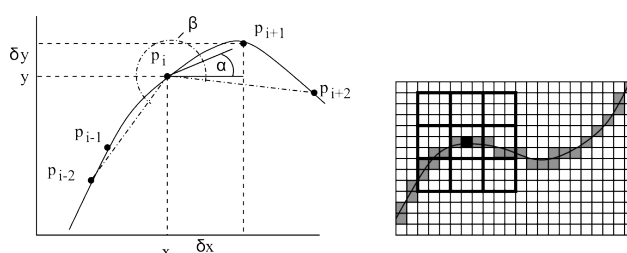


Figura 2.9: Extração de Características Espaciais [jai02]

Segue-se depois a fase de extração de características, onde se extraem características globais e locais. As características globais incidem sobre o número de palavras diferentes na assinatura. As características locais dividem-se em temporais e espaciais. As características espaciais são extraídas da forma da assinatura (ver figura 2.9), sendo elas as seguintes: a diferença entre coordenada x e y de dois pontos consecutivos; a coordenada y absoluta em relação ao centro da assinatura; o *seno* e o *coseno* do ângulo com o eixo

do x ; a curvatura da assinatura; os valores em escala de cinzentos para cada *pixel* numa vizinhança 9×9 . Em termos de características temporais, extraíram-se duas variantes, a velocidade absoluta e a velocidade relativa da escrita.

Após extraídas as características para cada uma das assinaturas, foi realizada a comparação de assinaturas. Para comparar as características extraídas entre duas assinaturas foi construído um método designado String Matching (método proposto em [OKM00]). Cada assinatura é assim representada por uma *string*, englobando todos os seus vectores de características e é aplicado o método String Matching. String Matching é um método para comparar strings de diferentes tamanhos. Este encontra pontos de alinhamento entre as duas *strings* de tal modo que a soma das diferenças entre cada par de pontos é mínima. Depois de aplicado o método, a diferença entre o número de palavras das assinaturas é calculada e incorporada numa medida de dissimilaridade. Um *threshold* foi posteriormente definido sobre esta medida de modo a separar entre assinaturas autênticas e falsificações.

De forma semelhante ao trabalho apresentado anteriormente apresenta-se também o trabalho realizado em [MRB06], onde as características extraídas incidem também sobre as características geométricas da assinatura. Cada assinatura é representada por dois conjuntos de pontos que representam a distribuição dos *pixels* da escrita na imagem. Os pontos são calculados por dois métodos, um designado Vertical Splitting e outro Horizontal Splitting (ver figura 2.10). Antes de aplicados os dois métodos anteriores, a assinatura deve ser centrada na imagem de modo a otimizar o cálculo dos pontos.



Figura 2.10: (a) Vertical Splitting, (b) Horizontal Splitting [MRB06].

Após extracção de características, a classificação é efectuada com base na distância Euclidiana sobre os vectores de características extraídos, sendo definido um *threshold* sobre a distância calculada para definir uma assinatura como falsificação ou autêntica.

Por fim, de realçar ainda um trabalho de investigação descrito em [ABM06] com vista a autenticação de assinaturas no qual foram extraídas um conjunto de características como as direcções e as transições presentes numa assinatura. Foi também extraída uma característica designada "centroid feature". Esta última consistia no ângulo dominante da assinatura. A classificação foi efectuada com recurso a dois classificadores de redes neuronais, um designado Resilient Backpropagation (RBP) e outro designado Radial Basis Function (RBF).

2.4.2 Autenticação de Texto Utilizando Características de Textura

Além da autenticação de assinaturas, existe também a questão da autenticação de texto que surge um pouco no seguimento da questão anterior. Na secção anterior foram apresentados alguns dos trabalhos de investigação que efectuam classificação a partir das características geométricas da assinatura, contudo é também possível efectuar esse processo através de características de textura. O trabalho de investigação relatado em [SBT98] é a prova disso mesmo. Este apresenta um método de validação que segue três etapas (ver figura 2.11): normalização dos dados, extracção de características e identificação do escritor.

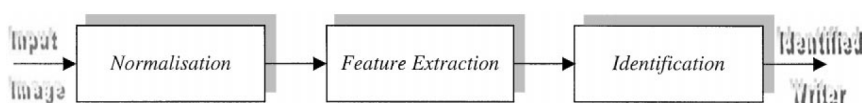


Figura 2.11: Diagrama de blocos da técnica apresentada em [SBT98].

A análise de textura não pode ser aplicada directamente sobre as imagens de escrita pois o espaçamento entre palavras e o espaçamento entre linhas afectam a textura. Para contornar estes factores, os autores deste trabalho aplicaram uma normalização, onde são efectuadas várias operações, destacando-se a fixação do espaçamento entre as palavras.

Após a normalização dos dados segue-se a fase de extracção de características. No trabalho proposto em [SBT98] utilizaram-se dois tipos de características, os filtros de Gabor multi-canal (MGF) e a matriz de co-ocorrências de tons de cinzento (GSCM).

Segue-se então a fase de identificação do escritor tendo como base as características extraídas. Dois classificadores foram considerados, o Weighted Euclidean distance (WED) e o K Nearest-Neighbour (K-NN). Para treino dos classificadores, foram extraídas um conjunto de características representativas de cada um dos escritores. No classificador WED, as características extraídas para um determinado bloco de texto são comparadas com as características representativas dos escritores, tendo em conta que determinadas características irão ter mais peso na distância final. O texto é atribuído ao escritor para o qual a distância euclidiana entre as características for menor. No caso do classificador do vizinho mais próximo (K-NN), um bloco de texto é comparado com todas as classes de características para cada escritor existente na base de dados. A distância Euclidiana é utilizada para medir as semelhanças entre as características, sendo que o escritor que tiver a distância mínima é considerado o autor do bloco de texto.

Os resultados relatados em [SBT98] verificaram que as características de Gabor (MGF) permitiram uma taxa de acerto significativamente mais elevada do que a matriz de co-ocorrências de tons de cinzento (GSCM). A nível de classificadores, o classificador WED revelou-se bem mais eficaz na classificação do que o classificador K-NN.

2.4.3 Identificação de Assinaturas Utilizando Pontos de Interesse

Muitos dos métodos de análise de assinaturas são baseados em características como a forma ou a análise de determinados pontos críticos na assinatura. Contudo, nos últimos anos tem-se desenvolvido algum trabalho no sentido de identificar assinaturas através da textura. O trabalho apresentado em [rui08] comprovou, mais uma vez, ser possível efectuar autenticação de assinaturas através de características de textura como o caso da técnica SIFT.

Os autores da proposta descrita em [rui08] defendem que as SIFT têm uma boa performance na extracção de características e no reconhecimento de objectos, contudo afirmam que o método de *matching* apresentado por Lowe é ineficiente a eliminar falsos *matches*. Baseiam a sua afirmação no facto do método apresentado por Lowe possuir a limitação de usar apenas uma simples fase de rejeição baseada em hipóteses probabilísticas, o que não é suficiente para reduzir os falsos positivos. Face a esta conclusão decidiram utilizar um método de *matching* apresentado por Loncomilla e Ruiz-del-Solar (L&R) no trabalho descrito em [Lds06]. Esse método engloba uma fase de rejeição rápida baseada em hipóteses probabilísticas, uma fase de verificação da correlação linear, uma fase de verificação da distorção geométrica, uma fase de verificação da correlação entre os *pixels*, um procedimento de transformação de fusão, o uso do algoritmo RANSAC [FB81] e um teste de restrições semi-local para eliminar os falsos *matches*.



Figura 2.12: Pontos de interesse extraídos nas assinaturas [rui08].

Depois de extraídos os descritores utilizando o método SIFT (ver figura 2.12), procede-se à aplicação do método L&R referido anteriormente para *matching* de pares de assinaturas. Um dos problemas na autenticação de assinaturas é a distorção não linear nas imagens das assinaturas, o que afecta razoavelmente o processo de *matching*. Contudo este método é suficientemente robusto quanto a isso, podendo verificar-se isso pela imagem 2.13 que apresenta o *matching* entre pares de assinaturas.

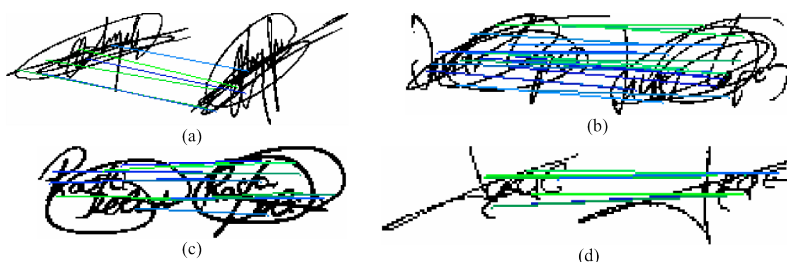


Figura 2.13: (a)-(d) *Matches* entre pares de assinaturas [rui08].

Após aplicação do *matching* a assinaturas em que a qualidade se alterava significativamente (por exemplo, através de mudanças na posição ou orientação), verificou-se que o número de falsos *matches* era elevado. Para colmatar este problema os autores deste trabalho utilizaram um classificador estatístico Naïve Bayes [Zha04] após o processo de *matching*. Este classificador foi treinado utilizando um conjunto de 12 características, destacando-se de entre elas o número de *matches* entre as assinaturas e o tempo de processamento do *matching* como medida da complexidade do *matching*.

Os resultados relatados em [rui08] revelam potencial para validação de autenticidade utilizando características de textura como é o caso das SIFT. Contudo, entre as assinaturas verdadeiras e falsificações feitas por profissionais não foi possível retirar ainda resultados positivos, sendo um dos grandes desafios ainda em aberto.

De forma semelhante foi também aplicado um classificador Naïve Bayes no trabalho descrito em [KSX04], sendo este baseado em características geométricas da assinatura e tendo revelado resultados significativos na autenticação de assinaturas.

2.4.4 Classificação de Assinaturas através de SVMs

Os métodos de análise de assinaturas apresentados anteriormente diferem principalmente nas características extraídas das assinaturas, algo que foi importante realçar. Contudo, a nível da classificação de assinaturas é possível enunciar alguns métodos cujo algoritmo de classificação importa referir. Em [BC97] é apresentado um trabalho onde a análise de assinaturas segue o processo habitual de um algoritmo de classificação. A extracção de características tem em conta as características geométricas da escrita, sendo extraídos um conjunto de pontos da assinatura. Em termos de classificação, é utilizado um conjunto de classificadores conjugados sendo estes classificadores representados por várias redes neuronais *feed-forward*. De forma semelhante no trabalho apresentado em [HY97] é também efectuada classificação utilizando redes neuronais.

Além de um sistema de classificação organizado em redes neuronais, é também possível efectuar a classificação utilizando Support Vector Machines (SVMs). No trabalho apresentado em [OSK05], as SVMs são utilizadas para efectuar a classificação de assinaturas em três grupos: *random*, falsificações geradas sem qualquer conhecimento da forma ou conteúdo da assinatura original; *simple*, falsificações construídas conhecendo o nome do detentor da assinatura, mas não a forma; e por fim *skilled*, geradas por pessoas tentando copiar o melhor possível a assinatura original.

O sistema apresentado em [OSK05] segue várias fases: pré-processamento dos dados, extracção de características e classificação das assinaturas. A fase de pré-processamento engloba eliminação do fundo da imagem, redução do ruído, normalização da largura e uma fase de *thinning* da mesma.

2.4.4.1 Extração de Características

As características extraídas são características globais, máscara e grelha (ver figura 2.14). As características globais fornecem informação acerca de casos específicos da forma da assinatura. São elas as seguintes: área da assinatura, relação altura-largura, histograma horizontal máximo e histograma vertical máximo, centro horizontal e vertical da assinatura, número local máximo da assinatura e os pontos de contorno da assinatura. As características máscara fornecem informação acerca das direcções das linhas da assinatura. As características de grelha fornecem uma informação geral acerca da aparência da assinatura.

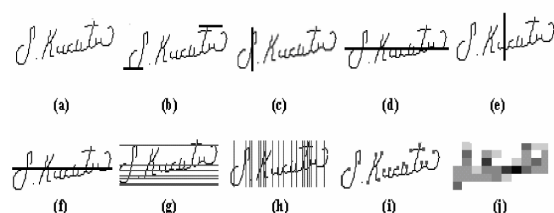


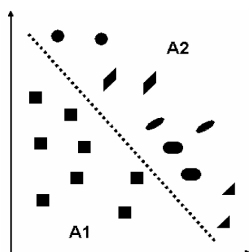
Figura 2.14: Extração de características: **(a)** pré-processamento **(b)** altura **(c)** histograma vertical máximo **(d)** histograma horizontal máximo **(e)** centro horizontal **(f)** centro vertical **(g)** número local horizontal máximo **(h)** número local vertical máximo **(i)** pontos de contorno **(j)** características de grelha [OSK05].

2.4.4.2 Treino e Teste do Classificador

A classificação pretendida é um problema multi-classe, contudo devido ao facto das SVM suportarem apenas duas classes foi necessário combinar várias SVMs de duas classes (ver figura 2.15) através de uma técnica designada one-against-all [GGL⁺03]. Para treino do classificador utilizou-se várias assinaturas pertencentes a cada uma das classes, sendo que foram também testadas várias funções *kernel* (linear, polinomial, função de base radial e sigmóide). Os testes com as várias configurações do classificador revelaram que os melhores resultados foram obtidos utilizando a função de base radial.

O classificador foi testado tendo como entrada um conjunto de assinaturas verdadeiras e falsas, calculando-se depois a taxa de acerto do classificador em distinguir as duas classes. Para comprovar a eficácia deste classificador comparou-se ainda este método com um classificador utilizando redes neuronais (ANN). Os resultados relatados em [OSK05] mostram a eficácia deste classificador, sendo isso visível pela imagem 2.16.

Continuando a temática da classificação de assinaturas utilizando SVMs, de referir um trabalho de investigação no qual são comparadas duas técnicas de classificação com vista à verificação de autenticidade de assinaturas [JBS05]. Nesse trabalho é efectuada uma comparação entre um método de classificação utilizando SVMs e outro utilizando Hidden Markov Models (HMM), uma técnica estatística. Em todos os testes efectuados sobre várias assinaturas utilizadas (quer para treino quer para teste), a classificação com



CLASS A1	CLASS A2
Signer 1 (8 genuine)	Signer 2 (4 skilled forgeries)
	Signer 3 (2 random forgeries)
	Signer 4 (2 random forgeries)
	.
	.
	Signer 41 (2 random forgeries)

	True Classification Ratio	False Classification Ratio
SVM	0.95	0.05
ANN	0.75	0.25

Figura 2.15: Estrutura da SVM construída [OSK05].

Figura 2.16: Resultados de aplicação do método [OSK05].

SVMs permitiu obter melhores resultados do que a classificação com HMM, revelando assim que esta poderá ser uma técnica promissora para resolver alguns dos problemas de autenticação de assinaturas.

Por fim, importa ainda realçar um outro trabalho de investigação [KY05] onde a verificação de autenticidade de assinaturas é efectuada não só recorrendo a SVMs, mas combinando estas com um classificador Naïve Bayes e um classificador linear usado em conjunção com o método Principal Component Analysis (PCA) [Pea01].

2.5 Aquisição e Análise de Imagens Hiperespectrais

Nesta secção irão ser abordados alguns trabalhos nos quais se aplicaram métodos de análise de imagens hiperespectrais com vista a mapeamento de materiais. É feita também uma introdução sobre aquisição de imagens hiperespectrais.

2.5.1 Aquisição de Imagens Hiperespectrais

A aquisição de imagens hiperespectrais consiste no registo de imagens em várias bandas espectrais estreitas e contínuas. Este processo é tipicamente tridimensional: duas dimensões espaciais e uma dimensão espectral (comprimentos de onda). Os sensores da câmara registam, em cada *frame*, as duas dimensões espaciais enquanto que a dimensão espectral é obtida por um varrimento sequencial.

O sistema multiespectral utilizado é constituído por câmara digital CCD monocromática, controlada computacionalmente e integrada com um filtro sintonizável de cristal líquido, LCTF (Liquid Crystal Tunable Filter), cuja transmitância espectral pode ser controlada electronicamente. Entre a câmara e o filtro encontra-se uma objectiva fotográfica. À frente do LCTF encontra-se um filtro de interferência que bloqueia os infravermelhos (IV). A cena que se pretende registar é colocada paralelamente ao CCD, iluminada por três lâmpadas Solux (ver figura 2.17) [dAeTdC04].

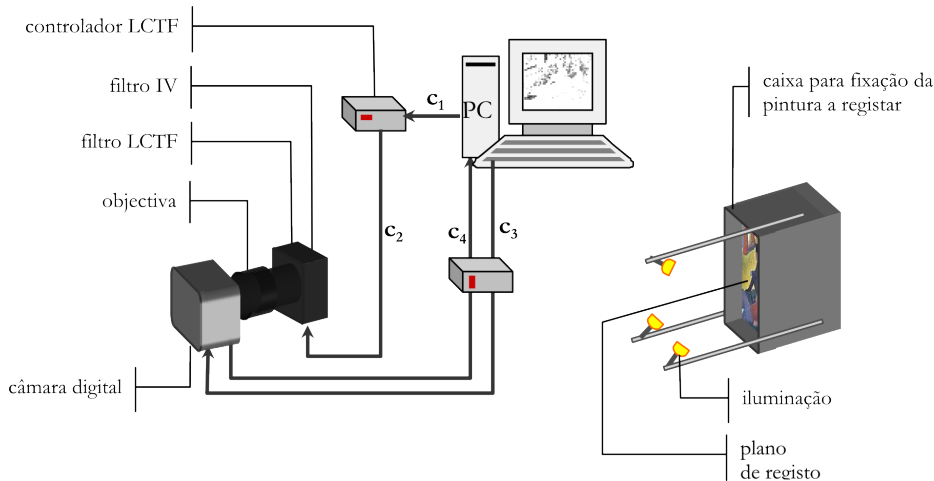


Figura 2.17: Esquema do sistema de captação de dados. (c1) controlo digital do filtro, (c2) controlo analógico do filtro (c3) controlo do *shutter* e tempo de exposição (c4) sinal digital da imagem [dAeTdC04].

Um computador pessoal (PC), com placa de aquisição digital de imagens, permite a aquisição dos dados, registando-os como ficheiros de imagem. O software utilizado sintoniza o filtro e cria a matriz de tempos de exposição, isto é, determina o tempo de exposição para cada comprimento de onda; esta matriz é determinada de modo a que o valor máximo do sinal do *pixel* seja inferior ao valor de saturação do sensor da câmara.

2.5.2 Objectivos para Visualização de Imagens Hiperespectrais

A visualização de imagens hiperespectrais tem sido um dos grandes desafios ao longo dos últimos anos. Uma pergunta se coloca: como pode a informação contida numa imagem hiperespectral ser apresentada de forma a ser interpretada pelos humanos sem requerer treino específico para tal? Existe um conjunto de trabalhos de investigação que tentam resolver a questão anterior, apresentando vários métodos para a análise de imagens hiperespectrais, contudo cada caso específico requer um método adequado de acordo com o domínio e os agentes do domínio a que se destina. Os autores de [JG05] apresentam um conjunto de indicações para construção de um método adequado para visualização de imagens hiperespectrais. Entre essas indicações destacam-se:

1. **Visualização consistente:** um espectro é sempre representado pela mesma cor de modo a ser facilmente identificado ao longo de várias imagens, facilitando assim a sua comparação;
2. **Preservação do conteúdo da imagem:** o conteúdo da imagem original deve ser representado fielmente ao original, mantendo os contornos dos vários objectos;
3. **Eficiência computacional:** a visualização deve ser mostrada rapidamente, permitindo interactividade real-time.

4. **Pequenas diferenças efectivas:** a visualização deve realçar as pequenas diferenças entre espectros semelhantes permitindo, por exemplo, identificar pigmentos semelhantes;
5. **Visualização sem características distractivas:** a visualização deve minimizar as características que distraem o utilizador sem razão aparente;
6. **Paleta natural:** a visualização deve usar um conjunto de cores que é consistente com a visualização de imagens naturais, tirando partido do facto do ser humano estar bem preparado para analisar cenas naturais;
7. **Igual peso dos comprimentos de onda:** todos os comprimentos de onda devem ter igual peso na visualização, funcionando assim o método para qualquer número de bandas e sendo independente do intervalo de frequências.

As indicações enunciadas servirão assim de objectivos de aceitação para os métodos construídos ao longo do desenvolvimento, auxiliando no caminho a seguir ao longo do mesmo.

2.5.3 Visualização de Imagens Hiperespectrais Através do Mapeamento de Três Comprimentos de Onda

Um dos métodos de visualização recorre ao mapeamento de três comprimentos de onda do cubo hiperespectral em cada um dos canais RGB. Em [TKDO03] é apresentado um trabalho de investigação onde se seleccionou três bandas igualmente espaçadas, sendo o valor de reflectância de cada *pixel* o valor de cada um dos canais RGB respectivamente. As bandas escolhidas devem cobrir a maioria do espectro de cada *pixel* e devem ser minimamente correlacionadas devido à sua separação espectral.

As bandas escolhidas pretendiam realçar a absorção em particular do mineral kaolinite. O mapeamento obtido (ver figura 2.18) destaca algumas características da imagem e segmenta a imagem em regiões com diferentes materiais. Contudo existe um enorme número de possibilidades para mapear três bandas específicas nos canais (R,G,B). O esquema de cores mudará para cada uma das configurações, não existindo *a priori* forma de identificar qual a intensidade de cor em que um determinado material será mapeado.

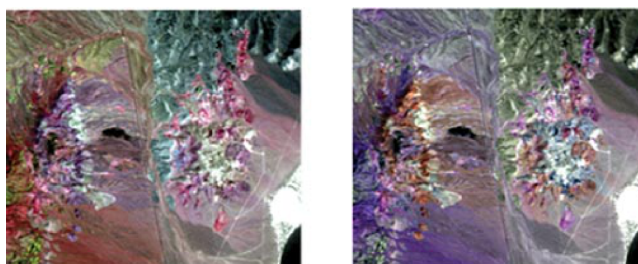


Figura 2.18: Mapeamento de dois conjuntos diferentes de comprimentos de onda nos canais RGB [TKDO03].

2.5.4 Visualização de Imagens Hiperespectrais *Principal-Components-Based*

Apesar do método anteriormente descrito ser utilizado em situações pontuais para identificar a presença de determinados elementos numa imagem hiperespectral, existem alguns métodos mais complexos para efectuar uma análise visual do conteúdo de uma imagem hiperespectral. O método Principal Components Analysis (PCA) [Pea01] tem sido aplicado ao longo dos anos na análise de imagens hiperespectrais. O PCA é efectivo a comprimir informação em conjuntos de dados multivariados através do cálculo das projecções ortogonais que maximizam a quantidade de variância dos dados. É tipicamente efectuado através da decomposição dos vectores próprios da matriz de covariância da dimensão espectral no cubo hiperespectral. No caso de imagens hiperespectrais, este método é utilizado para condensar a informação de um espectro num conjunto mais pequeno de variáveis.

2.5.4.1 Principal Components Analysis (PCA)

Quando aplicado na decomposição de imagens hiperespectrais, este método pode ser também designado por Multiway-PCA [JAMB08]. Antes da aplicação do método é necessário efectuar o *unfold* do cubo hiperespectral de modo obter-se uma matriz, em que o número de linhas da matriz é o número de *pixels* da imagem e o número de colunas é a dimensão do espectro. A figura 2.19 mostra os passos até à aplicação do método.

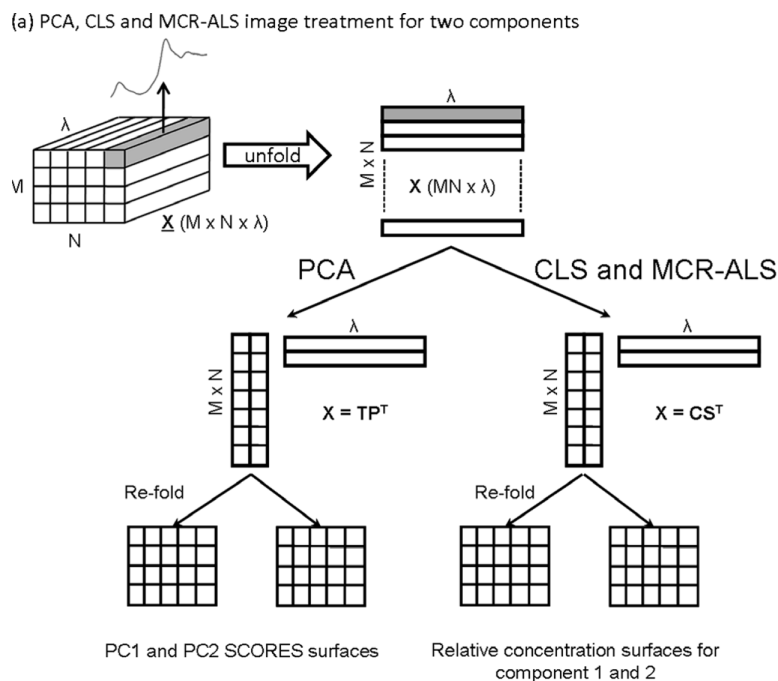


Figura 2.19: Aplicação de diferentes algoritmos a imagens hiperespectrais (a) PCA, CLS and MCR-ALS [JAMB08].

O PCA decompõe a matriz \mathbf{X} da seguinte forma:

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E} \quad (2.1)$$

onde \mathbf{T} ($MN \times F$) é a matriz de *scores*, \mathbf{P} ($\lambda \times F$) é a matriz de *loadings* e \mathbf{E} ($MN \times \lambda$) representa a matriz de resíduos para os F componentes principais.

Uma abordagem típica é a apresentação de alguns dos *scores* em imagens pseudo-cor através do mapeamento dos três primeiros *scores* nos canais RGB. As escolhas dos primeiros *scores* prende-se com o facto de que estes estão organizados em termos de representação da variabilidade dos dados decrescente. Isto é, por exemplo, os primeiros três *scores* produzidos representam aproximadamente 95% da variabilidade dos dados.

No estado da arte existe uma quantidade significativa de trabalhos que recorrem à aplicação do PCA para análise de imagens hiperespectrais. Normalmente as imagens hiperespectrais são obtidas através de imagens de satélite, sendo depois feita uma análise do conteúdo da imagem. Em [JAMB08] são apresentados alguns dos métodos para análise de imagens multivariadas, destacando-se o PCA e alguns algoritmos de *clustering* como o K-Means. Em [TKDO03] descreve-se a aplicação do PCA com vista ao mapeamento de materiais semelhantes (ver figura 2.20).

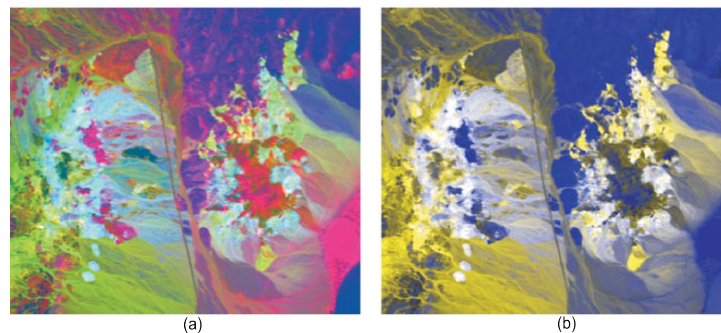


Figura 2.20: Imagens pseudo-cor obtidas através do mapeamento directo dos *scores* nos canais (R, G, B). **(a)** (P1, P2, P3) **(b)** (P2, P2, P3)[TKDO03].

Na proposta descrita em [TKDO03] a destacar ainda a utilização de scatter-plots para identificar *pixels* com materiais semelhantes, tal como a construção de imagens pseudo-cor com o mesmo fim. No artigo [JG05] são analisadas as vantagens e desvantagens da aplicação do PCA para fins de visualização. Além disso, os autores deste último trabalho apresentam também um novo método de visualização (Spectral Weighted Envelopes) baseado na percepção da cor pelo olho humano. No trabalho descrito em [ZMM06] os autores optaram por fazer uma comparação entre a aplicação do PCA e um método semelhante designado por Independent Component Analysis (ICA) [CS93] para efeitos de visualização. Também efectuaram o mapeamento em diversos espaços de cor, não se restringindo apenas ao espaço RGB. Outro dos trabalhos onde se pode encontrar uma comparação semelhante é em [ZVC11].

2.5.5 Visualização de Imagens Hiperespectrais Através de Classificação Supervisionada

Os dois métodos apresentados constituem alternativas para efectuar o mapeamento de materiais ou para identificar de regiões de cor semelhantes. Contudo, devido ao facto de não se conhecer *a priori* a cor com que determinado material irá ser representado é difícil a interpretação e análise visual. Face a esta desvantagem, surgiram inúmeros trabalhos onde o mapeamento de materiais é feito através de um algoritmo de classificação. A classificação supervisionada Spectral Angle Mapper (SAM) [YGB92] tem-se afirmado como um dos principais métodos de classificação nesta área. Além do SAM, existem ainda outras técnicas utilizadas em alguns trabalhos de investigação com vista a classificação que se verificaram não ser tão eficazes, destacando-se o uso da distância Euclidiana [HR04] ou da Spectral Correlation Similarity [HR04] como métricas de decisão.

2.5.5.1 Spectral Angle Mapper (SAM)

O SAM é um método para comparação rápida entre um espectro de teste e um espectro de referência. Normalmente descreve-se como uma técnica para classificação espectral que usa um ângulo n-dimensional para efectuar a *match* entre o espectro de um *pixel* e os espectros de referência. Este método determina a similaridade espectral através do ângulo entre os espectros, tratando-os como vectores n-dimensionais num espaço com dimensionalidade igual ao número de bandas. Os espectros de referência são normalmente adquiridos a partir do cubo hiperespectral em estudo, contudo podem também ser obtidos em laboratório ou através de medições no conteúdo da própria cena. Ângulos pequenos entre dois espectros indicam alto grau de similaridade. O ângulo calculado é invariante a mudanças de iluminação pois ignora o facto de um espectro ser o resultado de um outro multiplicado por uma constante, assim como não tem em conta os comprimentos de cada um dos vectores. Este é obtido através da seguinte fórmula:

$$\alpha = \cos^{-1} \left(\frac{\vec{a} \cdot \vec{c}}{\|\vec{a}\| \|\vec{c}\|} \right) \quad (2.2)$$

2.5.5.2 Classificação Supervisionada

Diversos trabalhos recorreram à utilização do SAM devido ao seu elevado desempenho em identificar espectros de materiais desejados. Todavia, a construção de uma base de dados de referência é um dos passos mais importantes no processo de classificação.

Construção da Base de Dados A construção da base de dados é muito importante pois o resultado da classificação vai depender directamente da qualidade dos espectros de referência. Estes podem ser obtidos em laboratório ou em amostras da cena contida na imagem hiperespectral, mas na maioria dos casos são obtidos através de amostragem da imagem hiperespectral em estudo. Assim garante-se que os espectros de referência

possuem as mesmas condições de ruído e iluminação que os espectros de teste, sendo o ruído um factor importante quando se compara espectros captados por instrumentos diferentes. No trabalho descrito em [PWLS07] pretendia-se identificar zonas contaminadas em carcaças de frangos. A base de dados foi construída a partir de regiões de interesse de aproximadamente 50 *pixels*, identificando-se uma região para cada um dos contaminantes. Depois o espectro médio foi calculado para cada região obtendo-se assim o espectro de referência para cada contaminante. A necessidade de cálculo do espectro médio surge pelo facto de ao longo de uma superfície com o mesmo material existir uma variabilidade significativa no espectro, sendo isso eliminado pelo espectro médio [MMS03]. À semelhança do método anterior, também nos trabalhos [GBHD04, 22604, SR02] a base de dados é construída através de amostragem de pequenos pontos ou zonas na própria imagem hiperespectral em estudo.

Construção do Classificador A classificação descrita em [GBHD04] é efectuada seguindo os seguintes passos:

1. Para cada *pixel* da imagem hiperespectral, calculou-se o ângulo entre o espectro desse *pixel* e cada um dos espectros de referência contidos na base de dados;
2. Atribuiu-se ao *pixel* a *label* correspondente ao material com o menor ângulo espectral de entre os presentes na base de dados. Por vezes define-se um *threshold* como medida para um *pixel* pertencer a um determinado material. Isto é, se o ângulo espectral mínimo for abaixo do *threshold*, então o *pixel* pertence a esse material. Caso contrário é lhe atribuída a etiqueta de *unclassified* por exemplo.

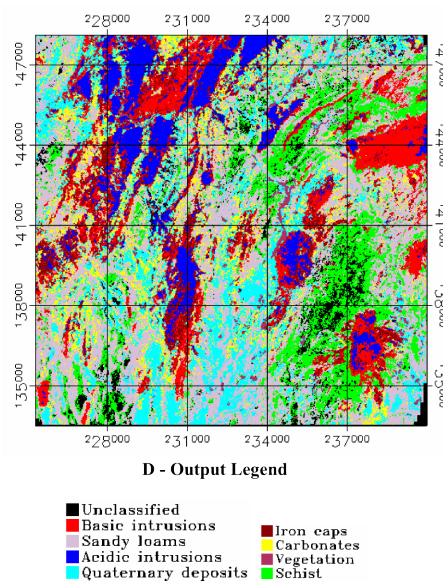


Figura 2.21: Classificação de uma imagem hiperespectral com a respectiva legenda de cores [GBHD04].

Para construir uma representação visual consistente basta apenas associar uma cor a cada um dos espectros de referência, sendo depois cada *pixel* representado pela cor do material a que foi atribuído. Um exemplo disso é a figura 2.21.

2.5.6 Visualização de Imagens Hiperespectrais Através de Classificação Não-Supervisionada

O mapeamento de materiais numa determinada cena pode ser efectuado de uma forma mais automática face ao método anterior, designando-se assim por classificação não-supervisionada. No método anterior, o utilizador deve fornecer um conjunto de espectros de referência de modo a construir uma base de dados para treino do classificador. Neste caso, a classificação segue um processo mais automático, não sendo necessário construir uma base de dados para treino. O agrupamento de *pixels* com o mesmo material é feito recorrendo a alguns dos métodos de *clustering* como o caso do K-Means. O K-Means é um método para agrupamento de dados, sendo normalmente a distância Euclidiana a principal métrica utilizada para agrupar dados semelhantes. Nos trabalhos descritos em [SMG99, SR02], o K-Means foi combinado com o Spectral Angle Mapper devido ao seu elevado desempenho em identificar espectros com características semelhantes.

O K-Means associado ao SAM tem revelado resultados positivos, contudo existem dois problemas inerentes ao mesmo. Em primeiro lugar, surge o problema da definição do número de *clusters* para a análise de uma determinada cena, sendo o número de *clusters* fundamental pois representa o número de materiais a mapear. Em segundo lugar, não existe forma *a priori* de identificar que material irá ser mapeado em cada um dos *clusters*. Para resolver este problema, o trabalho descrito em [SMG99] efectuou uma projecção de cada *cluster* num conjunto de espectros de referência. Ou seja, definiu-se um conjunto de espectros de referência, efectuou-se o *clustering* com um determinado número de *clusters* (número de *clusters* elevado permite uma melhor eficiência) e por fim, para cada um dos *clusters*, identificou-se o espectro de referência mais semelhante (utilizando o SAM), sendo esse o material que o *cluster* representa. Este método pode ser utilizado com vista a mapeamento automático de materiais, contudo devido aos problemas decorrentes do método para fins de visualização este acaba por se tornar menos automático.

2.6 Análise de Imagens Hiperespectrais em Obras de Arte

Esta secção aborda alguns dos trabalhos de investigação onde se efectuaram análises de obras de arte a partir do processamento de imagens hiperespectrais adquiridas sobre pinturas.

2.6.1 Identificação de Materiais em Obras de Arte

Identificar os materiais de uma determinada pintura é um passo crucial no processo de conservação. Quando o objectivo é preparar um plano de intervenção, é fundamental

saber que problemas o restaurador vai encontrar, aplicando-se a mesma ideia à validação de autenticidade. Para identificação dos materiais numa pintura o método mais popular é a análise de micro-amostras da camada da pintura. Contudo existe o problema deste ser um método invasivo e da análise ser apenas num ponto específico, não sendo generalizada para a sua vizinhança. Para colmatar estes problemas existe um conjunto de alternativas não invasivas, tal como a análise com raios-X de fluorescência (XRF), as câmaras hiperespectrais ou a espectroscopia Raman. A ideia base destes métodos é o facto de inferirem dados da imagem explorando o facto dos materiais reflectirem, absorverem e emitirem radiação electromagnética que depende directamente da sua composição molecular.

O trabalho apresentado em [PDMDRP08] tira partido de alguns dos métodos não invasivos. O objectivo principal do trabalho era efectuar o *clustering* de *pixels* que apresentassem semelhanças a nível da reflectância, da absorção e da emissão de radiação electromagnética. Na prática, o pretendido seria distinguir *pixels* com a mesma informação espectral, o que significa materiais semelhantes. Destacando assim materiais semelhantes poderia realçar áreas onde houve um retoque da pintura ou áreas com o material original. Um dos pontos a destacar no trabalho referenciado é o facto do algoritmo construído aceitar três tipos de dados captados com instrumentos diferentes: (1) reflectância visual, adquirida entre 400 e 750 nm com espaçamento de 50 nm (VIS), (2) reflectância por infravermelhos (IR), captada entre 800 e 1100 nm com espaçamento 50 nm, (3) Bandas visíveis UV captadas com um filtro UV entre 400 e 750 nm com espaçamento de 50 nm (UV).

2.6.1.1 Métricas para Comparação de Informação Espectral

Os autores de [PDMDRP08] utilizaram várias técnicas para comparação de informação associada a cada *pixel*: Spectral Angle Mapper (SAM) e a distância Euclidiana (ED), ambas métricas determinísticas; distância de Kolmogorov-Smirnov (KSD) e a Spectral Information Divergence (SID), ambas medidas probabilísticas. Tanto o Spectral Angle Mapper como a distância Euclidiana já foram abordados no capítulo anterior.

2.6.1.2 Classificação Baseada em Informação Espectral

O método apresentado em [PDMDRP08] baseia-se na metodologia de classificação efectuada de forma supervisionada. Isto é, também neste caso se definiu uma base de dados com espectros de referência construída por alguns peritos em arte, contendo assim várias amostras de materiais identificáveis nas pinturas em análise. O classificador foi construído em duas fases: uma fase de testes e uma fase de aplicação do algoritmo a obras de arte.

Os testes relatados em [PDMDRP08] permitiram tirar duas conclusões. Em primeiro lugar, as técnicas SAM e SID permitiram obter resultados mais consistentes do que as restantes duas técnicas, existindo também a vantagem que a definição de um *threshold*

sobre uma destas métricas é independente do material analisado. A segunda conclusão surge pelo facto de que o desempenho alcançado é mais elevado quando se combina a informação no esquema VIS + UV, sendo o desempenho mais baixo sempre que se considera a informação IR. Isto é facilmente explicável pelo facto de cada amostra ser proveniente de uma pintura que por norma possui uma estrutura em camadas, havendo a camada de preparação, a camada dos materiais e a camada dos óleos por exemplo, sendo as características dessas camadas captadas pela análise IR.

Após as conclusões retiradas dos testes anteriores procedeu-se à aplicação do método numa obra de arte com o objectivo de identificar por exemplo o material *lapis lazuli* (ver figura 2.22). Na imagem pode ver-se que uma zona que deveria ter sido classificada como *lapis lazuli* não foi identificada como tal. Isto acontece porque a zona em questão (na parte inferior da imagem) foi restaurada e devido a esse factor a sua reflectância não se assemelha à do material original *lapis lazuli*. Confirmou-se assim a potencialidade do método para extensão da análise pontual para uma análise na vizinhança de um determinado ponto com o intuito de efectuar um mapeamento dos materiais efectivo. Portanto, à semelhança da imagem apresentada, será possível efectuar a segmentação pretendida para cada um dos materiais presentes na pintura e com a respectiva amostra.

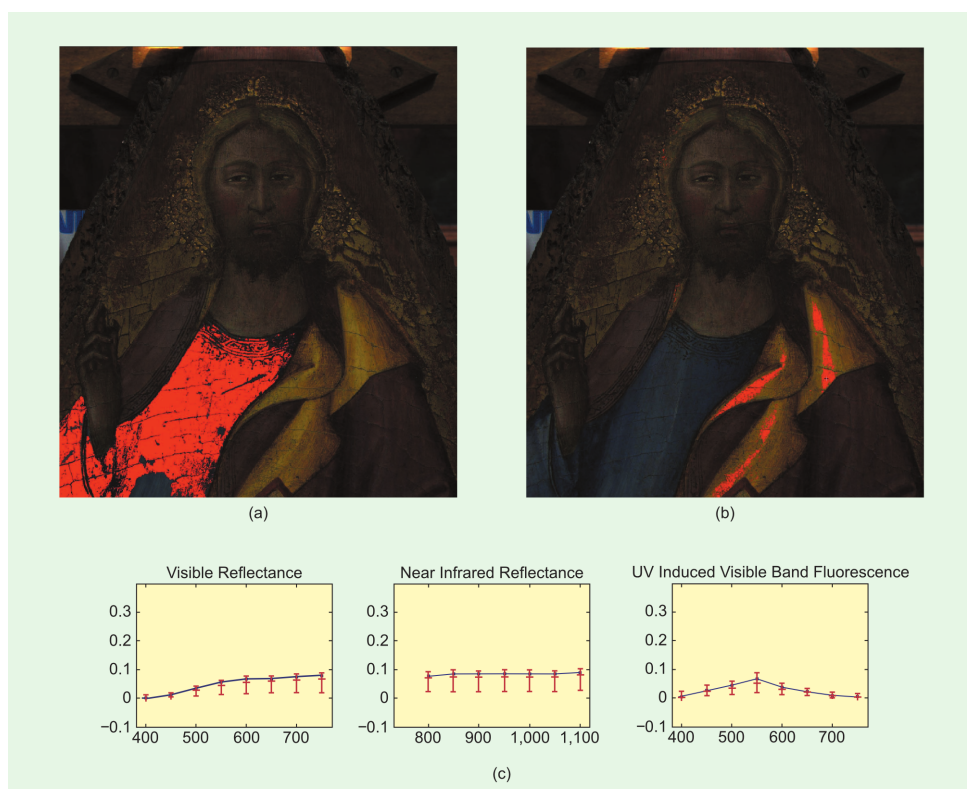


Figura 2.22: **(a)** A vermelho os *pixels* classificados como *lapis lazuli* (configuração VIS + UV). **(b)** A vermelho os *pixels* classificados como *lead-tin yellow* e *lead white*. **(c)** Exemplo de espectros adquiridos com cada um dos instrumentos [PDMDRP08].

2.6.2 Mapeamento e Identificação de Materiais Utilizados por Picasso

O trabalho apresentado em [DZT⁺10] descreve cada um dos métodos de aquisição da reflectância dos materiais existentes numa cena. Entre os métodos enumere-se: Hyperspectral Cameras, Luminescence Imaging Spectroscopy, Fiber-Optic Reflection Spectrometer (FORS), Polarized Light Microscopy, X-ray Fluorescence Spectrometry (XRF) e Scanning Electron Microscopy–Energy Dispersive Spectrometry. Os métodos enunciados foram utilizados para adquirir o conjunto de dados necessário para análise. Esses dados consistiam em duas imagens, uma hiperespectral e outra de infravermelhos, adquiridas na pintura *Harlequin Musician* do pintor *Pablo Picasso*, sendo depois combinadas numa única imagem contendo 260 comprimentos de onda (entre 441 e 1680 nm) e posteriormente processadas por algoritmos geométricos convexos.

Após a construção do cubo da pintura procedeu-se a uma redução do ruído através de um conjunto de normas. O processamento do cubo já tratado foi efectuado utilizando um programa de processamento de imagens espectrais designado ENVI². Na cadeia de processamento do cubo espectral, um conjunto de passos foram tidos em conta. Em primeiro lugar calcularam-se as imagens próprias usando o método Minimum Noise Fraction (MNF). Em seguida, calculou-se o índice de pureza do *pixel* e em último lugar, utilizou-se o ENVI Nd visualizer para encontrar a base de dados de espectros de referência através de *clustering*. Para garantir que os espectros de referência descrevem bem a superfície da pintura, utilizou-se o SAM para mapeamento dos materiais de acordo com os espectros de referência encontrados (ver figura 2.23).

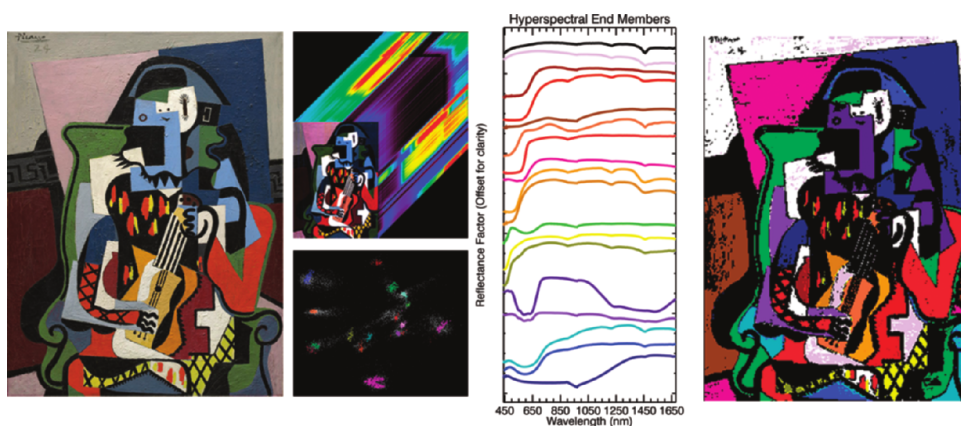


Figura 2.23: (Esquerda) Imagem da pintura *Harlequin Musician*. (Centro-Esquerda) Cubo hiperespectral e *clustering* de pixels. (Centro-Direita) Espectro dos materiais que constituem a base de dados. (Direita) Mapeamento dos materiais. [DZT⁺10]

Após o mapeamento anterior, utilizou-se o método Luminescence Imaging Spectroscopy para adquirir um novo cubo espectral no intervalo de frequências entre 650 e 900 nm com espaçamento 50 nm. Esta técnica permite identificar principalmente tons de amarelo, laranja ou vermelho. O mesmo tipo de processamento do cubo anterior foi

²<http://www.exelisvis.com/ProductsServices/ENVI.aspx>

também aplicado neste caso, revelando-se também uma eficácia generalizada no mapeamento dos diferentes materiais (ver figura 2.24).

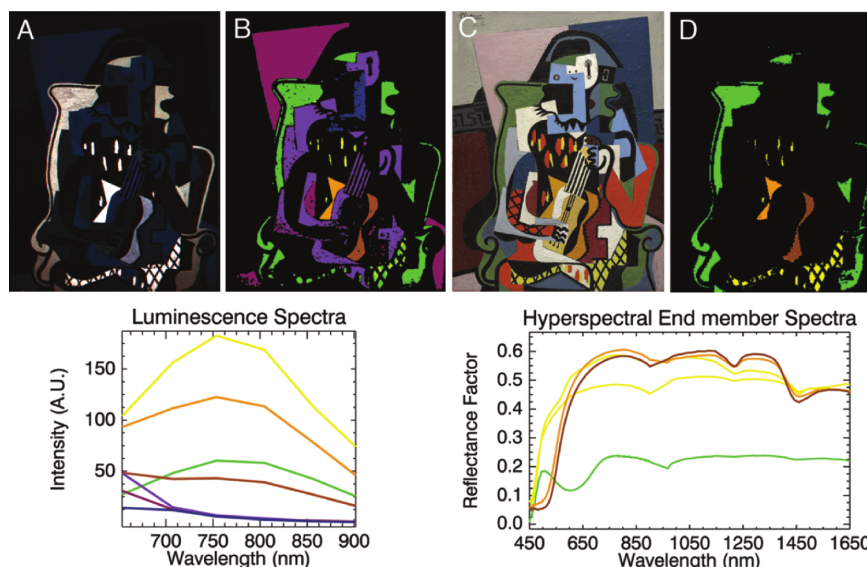


Figura 2.24: **(A)** Imagem de luminescência infravermelhos. **(B)** Mapeamento dos espectros de emissão primários. **(C)** Imagem da pintura *Harlequin Musician*. **(D)** Mapeamento dos espectros de referência que são luminescentes no cubo hiperespectral. Em baixo, os espectros obtidos utilizando Luminescence Imaging Spectroscopy e Hyperspectral Camera [DZT⁺10].

Por fim, efectuou-se a atribuição de cada um dos espectros encontrados de forma automática através de *clustering* a cada um dos materiais puros utilizados pelo pintor. Esta atribuição revelou vários aspectos. Em primeiro lugar, é possível concluir que os materiais puros utilizados na construção da pintura em análise podem ser mapeados e identificados utilizando a análise de imagens hiperespectrais. Complementando os dados de reflectância com a Luminescence Imaging Spectroscopy e a análise pontual XRF conduziu a um desempenho significativo na atribuição dos materiais, a nível de mapeamento e identificação. Os resultados obtidos no trabalho descrito em [DZT⁺10] revelaram ser significativos ao ponto de se considerar uma técnica para efectuar o mapeamento efectivo de materiais.

2.7 Discussão

Os trabalhos descritos neste capítulo ilustram o estado da arte associado ao tema da dissertação. Esta secção surge pela necessidade de efectuar uma comparação entre algumas das técnicas ou métodos apresentados anteriormente. A dissertação divide-se em duas partes que incidem sobre o mesmo domínio, análise de obras de arte.

A primeira parte consiste na análise das imagens digitais de assinaturas, sendo as primeiras três secções focadas na análise de alguns dos métodos para processamento e análise de imagem. A secção 2.2 apresenta uma visão geral de duas técnicas utilizadas

para extracção de características, assim como um método para organização das mesmas. Apesar de uma grande quantidade de trabalhos de investigação utilizar os filtros de Gabor para extracção de características, neste caso devido às características muito inconsistentes das assinaturas (constantes variações de escala, rotação, translação e da geometria) optou-se como técnica principal para extracção de características a SIFT pois é invariante aos problemas anteriores. O Gabor permite uma análise global da imagem, sendo as características extraídas a nível global enquanto que as SIFT extraem um conjunto de características locais, obtendo-se para cada imagem um conjunto de vectores de características que poderão ser organizadas utilizando a técnica Bag-of-Features.

A secção 2.3 enuncia alguns dos trabalhos de investigação com vista a análise de obras de arte, destacando-se principalmente a validação de autenticidade e atribuição de autor nos trabalhos descritos. Esta secção tem como principal objectivo estabelecer um contexto para a respectiva dissertação ao nível do trabalho já realizado no mesmo domínio. A maioria dos métodos apresentados efectua uma análise da pincelada característica do pintor, normalmente realizada através de filtros de Gabor ou Gabor Wavelets, contudo devido às razões apresentadas no parágrafo anterior, o Gabor foi preterido face às SIFT.

Ainda no âmbito da primeira parte da dissertação surge a secção 2.4, totalmente dedicada ao relato de vários trabalhos de investigação com vista a verificação de autenticidade de imagens de assinaturas digitais. Pode-se encontrar uma enorme diversidade de características extraídas das assinaturas e também de classificadores utilizados. Após análise do estado da arte, verificou-se que não existia uma ferramenta para validação de assinaturas de pintores, nem existia nenhuma ferramenta que conjugasse as SIFT com as Support Vector Machines para classificação de assinaturas. Assim, surge o interesse em desenvolver uma ferramenta que permita efectuar verificação de autenticidade e atribuição de autor de pinturas baseada nas características da assinatura do pintor em questão.

A segunda parte da dissertação consistia no mapeamento dos materiais de pinturas através da análise de imagens hiperespectrais. A nível do trabalho relacionado destacam-se duas secções, 2.5 e 2.6. A secção 2.5 pretende estabelecer um contexto para os dados que se pretende analisar. Para contextualização, é descrita a forma como os dados são captados e alguns dos métodos para análise dos dados adquiridos. De realçar um conjunto de objectivos para visualização de imagens hiperespectrais descritos em 2.5.2, sendo estes fundamentais no desenvolvimento da dissertação. Inicialmente foram abordados dois métodos em 2.5.3 e 2.5.4, sendo o segundo um método bastante utilizado no estado da arte para análise de imagens hiperespectrais. Contudo, devido ao facto destes não permitirem atingir alguns dos objectivos considerados como indispensáveis para construção de um novo método foram preteridos. Caminhou-se assim no sentido de alguns algoritmos de classificação, sendo que face aos resultados obtidos em vários trabalhos de investigação com a técnica SAM, considerou-se que esta seria a principal técnica a utilizar.

Por fim, a secção 2.6 aborda a aplicação de imagens hiperespectrais na análise de obras

de arte, destacando-se a utilização de alguns dos métodos para análise de cubos hiperespectrais. Face aos trabalhos analisados verificam-se de facto mapeamentos de materiais efectuados de forma efectiva, contudo pretende-se ir mais longe a nível do mapeamento de materiais através da construção de um método mais robusto, para efectuar também uma análise ao nível das misturas, algo em que o SAM é pouco vantajoso. Os métodos anteriores revelaram algumas dificuldades a nível da semelhança de espectros de materiais diferentes, ou seja, quando a curva espectral é semelhante em dois espectros mas estes pertencem a materiais diferentes. O desenvolvimento da dissertação pretende resolver este problema de forma a considerar estes casos no processo de mapeamento.



Análise de Obras de Arte

Este capítulo descreve o trabalho desenvolvido ao longo da dissertação com vista a realizar a solução proposta.

3.1 Introdução

Este capítulo encontra-se dividido em várias secções de acordo com os objectivos definidos. A figura 3.1 apresenta um diagrama estruturando todo o desenvolvimento da dissertação. Neste diagrama é possível visualizar as várias fases do desenvolvimento, assim como o objectivo principal de todo o desenvolvimento, verificação de autenticidade e atribuição de autor de pinturas. A primeira fase consiste na análise das imagens digitais de assinaturas, estando presente no diagrama no primeiro bloco, onde se pode destacar a identificação de períodos de assinaturas e a construção de um sistema de classificação de assinaturas em termos de atribuição e autenticidade. A segunda fase engloba a reconstrução parcial do classificador de pinturas baseado em características de pincelada com vista à integração no sistema de classificação combinada de pinturas. A terceira fase, representada pelo terceiro bloco engloba a análise de imagens hiperespectrais, tendo sido construídos vários métodos para análise de obras de arte. Construiu-se um método de classificação recorrendo a uma base de dados interna, um método automático de classificação, um método de classificação recorrendo a uma base de dados externa e um método para análise de misturas e validação de autenticidade baseada na análise dos materiais. Por fim, surge o último bloco onde se encontra o sistema combinado para verificação de autenticidade e atribuição de pinturas baseado em vários tipos de características, quer sejam características da pincelada do artista, quer características das assinaturas do pintor em análise.

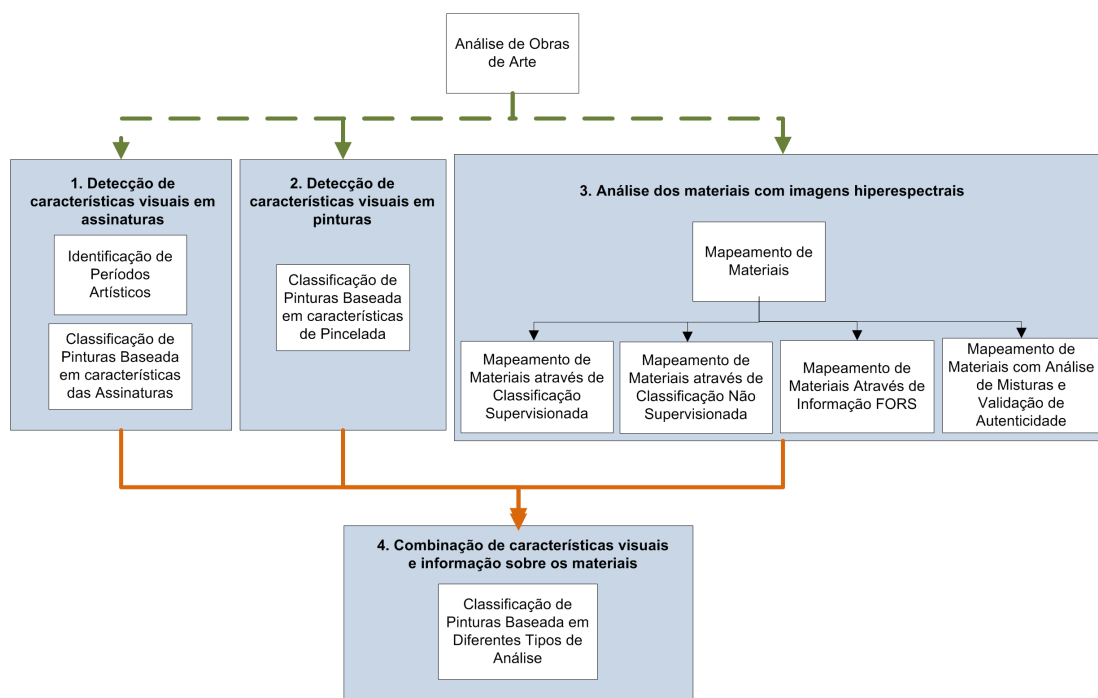


Figura 3.1: Diagrama do sistema desenvolvido para análise de obras de arte.

3.2 Análise de Assinaturas

A análise de assinaturas não consistiu no desenvolvimento de uma aplicação com uma interface gráfica definida, mas sim no desenvolvimento de um conjunto de ferramentas cujo maior interesse reside nos resultados produzidos. O pintor em estudo, Amadeo de Souza Cardoso, utilizou um conjunto de assinaturas para autenticar alguns dos seus quadros. O desenvolvimento foi efectuado sobre um conjunto de imagens digitais das assinaturas existentes.

3.2.1 Pré-Processamento dos Dados

Inicialmente existiam 140 imagens de assinaturas pertencentes a Amadeo de Souza-Cardoso, 21 imagens pertencentes a outros pintores (entre eles, Almada Negreiros, Mario Eloy), 10 imagens de assinaturas falsificadas em meio laboratorial e 2 imagens de assinaturas ainda por desvendar quanto à autenticidade. Verificou-se então uma escassez de dados ao nível de assinaturas de outros pintores, assim como também ao nível das falsificações. Para colmatar a falta de assinaturas de outros pintores, efectuou-se uma pesquisa online de assinaturas de outros pintores. Deste modo foi possível obter cerca de 56 assinaturas de outros pintores, destacando-se Degas, Miró, P. Cezanne, Magritte, entre outros. Ao nível das falsificações não foi possível obter mais dados.

Em primeiro lugar, redimensionaram-se todas as imagens de modo a que a comparação entre as mesmas fosse efectuada de forma equilibrada. Isto devido ao facto de existirem imagens muito grandes com resolução 900x500 *pixels* e imagens muito pequenas em

que a resolução era 150×100 *pixels*. Definiu-se um valor médio para a resolução (250×250 *pixels*) e converteram-se todas as imagens de acordo com esse valor sempre mantendo o *aspect ratio* (relação entre a altura e a largura da imagem) da imagem.

Ao extrair as características utilizando a SIFT, verificou-se que devido à textura bastante irregular dos quadros onde as assinaturas estavam contidas, grande parte das características era proveniente do fundo da imagem e não do conteúdo da assinatura (ver figura 3.2, (a)). Para colmatar esse aspecto, surgiram duas hipóteses: efectuar a extracção das SIFT em pontos regulares; ou tentar extrair o *background* das imagens de alguma forma de modo a eliminar as características provenientes do mesmo. Devido à constante rotação e translação nas imagens, seria praticamente impossível definir um conjunto de pontos que capturassem as características de igual modo em todas as assinaturas. Assim optou-se por se proceder à extracção do fundo da imagem.

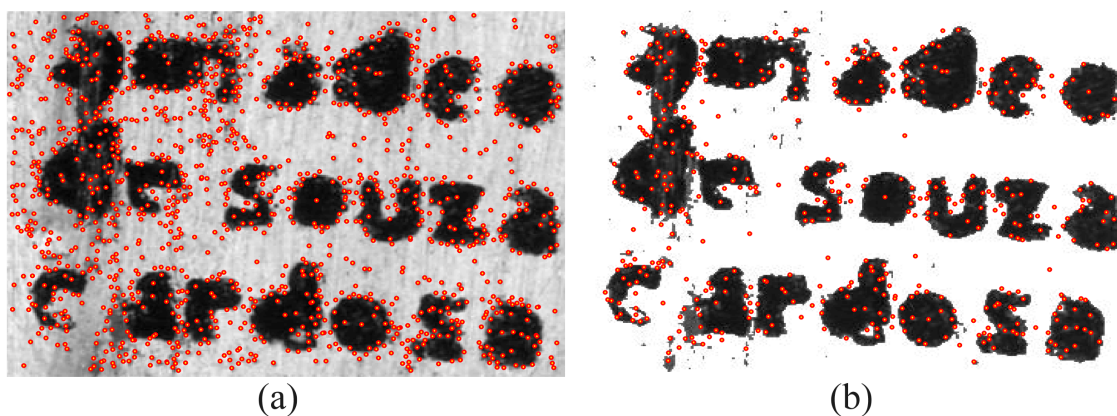


Figura 3.2: Extracção dos pontos de interesse utilizando a SIFT. (a) Aplicadas à imagem original. (b) Aplicadas à imagem após aplicação da extracção do *background*.

A extracção automática do fundo da imagem seguiu os seguintes passos:

1. A imagem da assinatura é convertida do espaço de cor RGB para uma imagem numa escala de cinzentos.
2. A imagem é convertida em valores binários 0 e 1, separando a assinatura do fundo. Para converter uma imagem em tons de cinzento para uma imagem binária, primeiro é necessário definir um *threshold* para que todos os *pixels* sejam mapeados nos valores 0 ou 1 de acordo com esse valor. Neste caso utilizou-se o método de Otsu [Ots79] para encontrar o valor do *threshold*. Este método calcula o *threshold* global para uma imagem que minimiza a variância intra-classe entre os *pixels* pretos e brancos. Após a obtenção do *threshold*, aplicou-se uma função que substituiu todos os *pixels* da imagem com uma luminância maior do que o *threshold* pelo valor 1 (branco) e substituiu todos os outros *pixels* com o valor 0 (preto). Assim obteve-se uma imagem apenas com valores de 0 e 1 (ver figura 3.3).

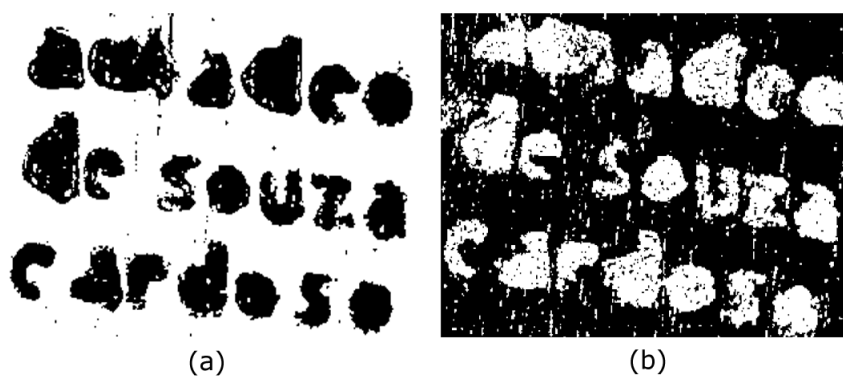


Figura 3.3: Exemplos de imagens binárias obtidas após o processo de segmentação.

3. Após a obtenção da imagem binária surgiu o problema de identificar a codificação que correspondia ao conteúdo da assinatura e do *background* de uma forma automática, se 0 ou 1. Face à dificuldade em encontrar uma condição eficaz, considerou-se que o valor dominante representava o fundo pois na grande maioria das imagens o fundo era o elemento mais presente. Assim, para cada imagem procedeu-se à identificação do valor dominante (0 ou 1), e selecciona-se todos os *pixels* com o valor oposto que modelariam o conteúdo da assinatura. Após a identificação dos *pixels* pretendidos, seleccionou-se cada um desses *pixels* na respectiva imagem em tons de cinzento. Esta condição não foi eficaz em alguns casos pontuais onde a assinatura era o elemento dominante da imagem, sendo necessário forçar o algoritmo a escolher um dos valores como dominantes para estes casos. O resultado pode ser comprovado na figura 3.2, (b).



Figura 3.4: Extracção dos *background* da imagem. (a) Resultado da extracção automática. (b) Resultado da extracção manual.

Apesar da eficácia do método anterior ter sido elevada, em casos onde a assinatura se confundia com o fundo, teve de se recorrer a um método bem menos automático, sendo necessário utilizar uma ferramenta de edição de imagem para extrair o restante fundo manualmente. Este foi um processo demorado de modo a que a extracção do fundo não afectasse o conteúdo da assinatura. O resultado pode comprovar-se na figura 3.4, (b).

De referir que devido à fraca qualidade das imagens e à ineficiente extracção do fundo em alguns casos, tanto de forma automática como de forma manual, algumas assinaturas foram excluídas do estudo pois a natureza das mesmas iria afectar de forma negativa os resultados. Resultou assim num total de 114 assinaturas de Amadeo de Souza-Cardoso e 67 pertencentes a outros pintores. Com as imagens das falsificações não houve qualquer problema.

3.2.2 Extracção e Organização de Características

Apesar da eficácia dos filtros de Gabor em inúmeros trabalhos de investigação anteriores, neste problema não foi a solução mais eficaz, como se pode constatar da experiência realizada e descrita mais à frente. Isto porque o Gabor aplica um banco de filtros à imagem, captando as direcções que se destacam. Devido ao facto das assinaturas apresentarem uma translação, rotação, mudança de escala e variação na geometria local em todas as imagens, o Gabor não seria tão eficaz a captar as características das assinaturas de Amadeo. Assim, devido à invariância das SIFT aos quatro factores apresentados anteriormente, optou-se por utilizar este tipo de características. O método SIFT foi relatado na secção 2.2.1, podendo observar-se um exemplo do conteúdo extraído na figura 3.5, (a).

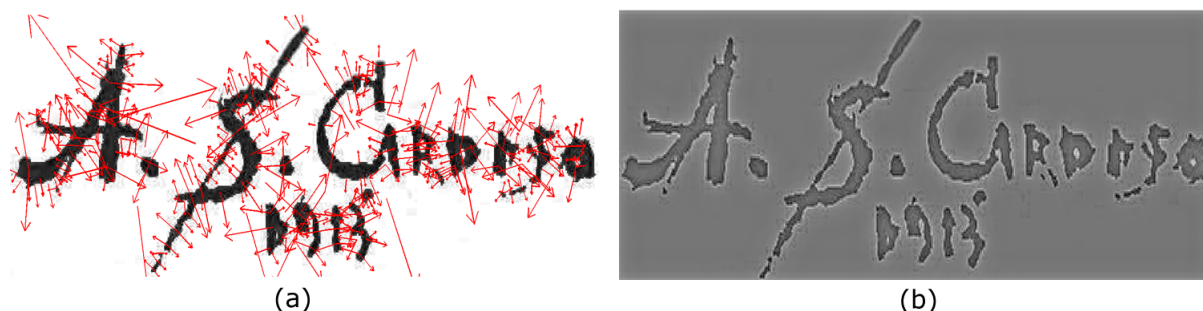


Figura 3.5: Extracção de características. (a) Pontos de interesse SIFT. (b) Aplicação dos filtros de Gabor.

Contudo um problema surgiu ao optar por este tipo de características. A utilização do Gabor permite obter um conjunto de características globais para cada imagem, contudo a utilização das SIFT extrai um conjunto de características locais à imagem, sendo cada uma delas representada por um vector de 128 valores. Por norma, as imagens utilizadas apresentam em média 600 pontos de interesse. Surgiu assim a necessidade de transformar o conjunto de medidas locais numa medida global para cada imagem de modo a facilitar o processo de comparação de imagens. Optou-se assim por uma técnica já referida anteriormente designada por Bag-of-Features, que indica o número de ocorrências de "palavras" visuais pertencendo a um vocabulário.

Considerando uma base de dados composta por K imagens $B_D = \{I_1, \dots, I_K\}$ e um vocabulário $V_p = \{p_1, \dots, p_N\}$ de N "palavras" extraídas da base de dados, a Bag-of-Features para uma imagem I_j é representada pelo histograma que conta as ocorrências,

$P_{t,j}$ de cada palavra p_t na imagem I_j ,

$$x_j(t) = \sum_{i=0}^{N-1} P_{i,j} \delta(t-i) \quad (3.1)$$

onde $\delta(t)$ é a função de impulso unitário.

O vocabulário de palavras visuais V_p é obtido da seguinte forma:

1. Extraem-se as características utilizando a técnica SIFT, tendo resultado num conjunto médio de 600 pontos de interesse por cada imagem.
2. Integra-se as características de todas as imagens em estudo numa única estrutura de dados, aplicando-se depois o método K-Means com um determinado número de *clusters* que vai representar o número de palavras visuais. O centróide de cada um dos *clusters* constitui cada uma das "palavras" visuais (figura 3.6).

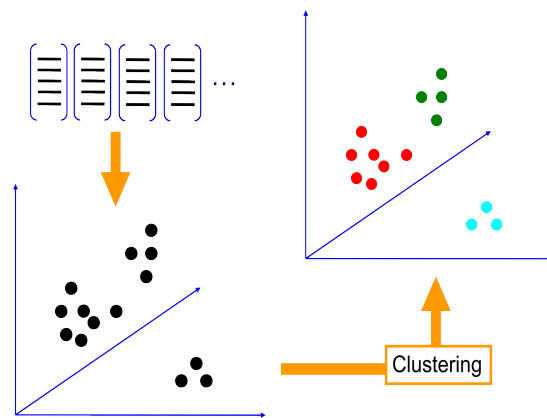


Figura 3.6: Esquema de construção do vocabulário de "palavras" visuais [FFLT].

Após a construção do vocabulário é necessário encontrar o histograma de ocorrências para cada imagem. Para cada imagem, a distância Euclidiana entre cada vector de características e o vocabulário V_p é calculada. Se a distância for menor do que um *threshold* definido, então a ocorrência de p_t é incrementada para a imagem I_j .

Após calculado o histograma para cada uma das imagens, é necessário normalizar o mesmo, utilizando a técnica Term Frequency–Inverse Document Frequency (TF-IDF) [PS03]. Em resumo, esta técnica traduz-se num valor numérico que reflecte o quanto importante uma "palavra" é para a base de dados. É baseada na contagem do número de documentos numa colecção que contém uma determinada "palavra" a pesquisar. Quando uma determinada "palavra" a pesquisar ocorre em muitos documentos (neste caso imagens) então não é um bom discriminador, e deve ser-lhe atribuída menos peso do que uma "palavra" que aparece num conjunto pequeno de documentos.

Com a representação da Bag-of-Features cada imagem I é representada por um histograma de ocorrências de palavras:

$$h_i = [P_{1,i}, P_{2,i}, \dots, P_{N_t,i}]^T \quad (3.2)$$

e para normalizar os vectores, é aplicado o TF-IDF:

$$P_{t,j} = \frac{N_{t,j}}{\sum N_{t,j}} \log \frac{K}{n_t} \quad (3.3)$$

onde n_t representa o número de imagens onde a "palavra" visual t está presente e $N_{t,j}$ é o número de ocorrências do termo t na imagem j .

Apesar do Gabor não ter sido utilizado como característica principal, utilizou-se o Gabor para demonstrar o que foi referido no início desta secção, o facto de no caso em estudo a técnica SIFT permitir obter resultados mais significativos do que o Gabor. O Gabor foi utilizado recorrendo a um banco de 24 filtros, com 4 escalas diferentes e 6 rotações, sendo que as características de textura representadas pelo Gabor traduzem-se num vector de 48 valores. A utilização deste tipo de características pressupõe uma abordagem global, sendo os filtros utilizados com o mesmo tamanho das imagens. No capítulo 2.2.2, foi relatado o funcionamento deste tipo de características, podendo observar-se um exemplo do tipo de informação extraída na imagem 3.5, (b).

3.2.3 Identificação de Períodos Artísticos

Esta secção relata um dos primeiros objectivos da dissertação, sendo este considerado um objectivo secundário no âmbito da análise de assinaturas. Com vista a desvendar ainda mais pormenores sobre a vida de um pintor, seria vantajoso possuir uma ferramenta que pudesse identificar períodos artísticos na vida de um pintor através da análise das características das assinaturas. A utilização de determinado padrão de assinatura pode ser a marca de um pintor aquando da estadia num determinado país, a identificação de uma fase com um determinado estilo de pintura ou uma marca para identificar determinadas pinturas mais importantes.

Ao nível do estado da arte não foi possível identificar uma quantidade de trabalhos significativa que efectuasse o agrupamento de imagens de acordo com as suas características visuais. Contudo, o trabalho apresentado em [ANP07] efectua um agrupamento baseado em características de faces através de um conjunto de medidas aplicadas às características SIFT.

Assim, procedeu-se no sentido de desenvolver um método, independente do artista em estudo, que permitisse efectuar um agrupamento de assinaturas baseado em características visuais semelhantes. Este agrupamento poderia estender-se ao longo de vários níveis de modo a encontrar sub-períodos, sendo assim o agrupamento hierárquico efectuado de uma forma menos automática. A figura 3.7 contém um diagrama de blocos demonstrando a estrutura da ferramenta.

Para que o agrupamento fosse efectuado de forma a agrupar características visuais semelhantes, procedeu-se em primeiro lugar ao pré-processamento dos dados tal como já foi explicado anteriormente e depois à extracção de características. A técnica SIFT tinha a limitação de extrair um conjunto de características locais, havendo necessidade de encontrar uma medida global para cada imagem, de modo a que o agrupamento fosse

efectuado ao nível da imagem. Recorreu-se assim à estratégia Bag-of-Features para construir um descritor global para cada imagem, sendo esses os dados de entrada do método de *clustering*.

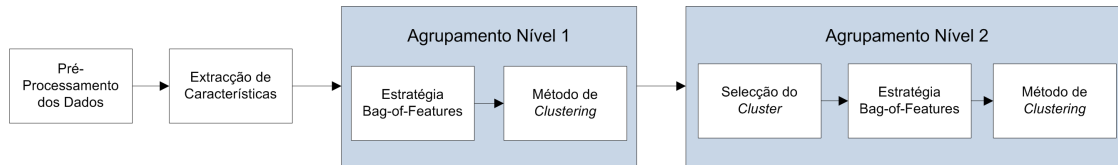


Figura 3.7: Diagrama de blocos do método de agrupamento de assinaturas.

O agrupamento pode ser efectuado de forma hierárquica de modo a encontrar sub-períodos dentro de períodos mais gerais. Assim no primeiro nível de agrupamento, para todas as imagens de entrada no algoritmo constrói-se a respectiva *bag*, seguindo-se um método que efectua o agrupamento com base na informação de entrada. Para agrupamento de padrões, optou-se pela utilização do método K-Means conjugado com a distância Euclidiana devido à sua eficácia no agrupamento de informação com padrões semelhantes. Assim, imagens com uma configuração semelhante da respectiva *bag* iriam ser agrupadas no mesmo *cluster*.

O método K-Means é utilizado para agrupamento de padrões, sendo um problema computacional difícil. Dado um conjunto de observações (x_1, x_2, \dots, x_N) onde cada observação é um vector de valores reais, este método agrupa as N observações em k conjuntos $(k \leq N)$ $S = \{S_1, S_2, \dots, S_k\}$ de modo que em cada conjunto a soma da distância ao quadrado seja minimizada,

$$\arg \min_s \sum_{i=1}^k \sum_{x_j \in S_i} \|x_j - \mu_i\|^2 \quad (3.4)$$

onde x_j representa cada um dos pontos presentes no *cluster* i e μ_i representa o centróide do *cluster* i .

O início do algoritmo requer a definição de um conjunto de dados para formação inicial dos *clusters*. Em grande parte das implementações do método, os *clusters* iniciais são encontrados aleatoriamente, contudo o resultado do algoritmo dependeria sempre da escolha aleatória. Neste caso considerou-se que os primeiros dados até ao número de *clusters* formariam os *clusters* iniciais, não havendo assim aleatoriedade. Em seguida segue-se um conjunto de iterações, em que cada imagem é atribuída ao *cluster* com o qual mais se assemelha, procedendo-se então à actualização de cada um dos centróides de cada *cluster*.

Contudo existe um problema relativamente à utilização do K-Means que é o facto de ser necessário indicar o número de *clusters* em que se pretende agrupar os dados. Na solução implementada optou-se por ser o utilizador o principal responsável por definir o número de partições que pretende. Uma das formas de avaliar a distribuição dos dados

pelos vários *clusters* é a utilização do índice de Silhouette [KR05], sendo este cálculo indicativo se o número de *clusters* é o mais adequado para a partição dos dados encontrada.

Após a aplicação do K-Means no primeiro nível obtém-se então a distribuição mais geral de assinaturas pelos vários *clusters* indicados. Terminado o primeiro nível de agrupamento, o utilizador pode prosseguir a sua análise tentando encontrar sub-períodos, seguindo-se o segundo nível de agrupamento. Após a selecção de um dos *clusters* anteriormente obtidos, repete-se o processo de construção da Bag-of-Features apenas com as imagens agrupadas nesse *cluster* no primeiro nível. A reconstrução da Bag-of-Features no novo nível de agrupamento surge devido ao facto do vocabulário de "palavras" visuais para este subconjunto de imagens ser diferente do vocabulário inicialmente calculado. Isto porque no primeiro nível, o vocabulário de "palavras" visuais foi construído com base em todas as imagens do conjunto de dados de entrada no algoritmo. Neste segundo nível, o vocabulário de "palavras" não deverá ser o mesmo pois apenas se pretende agrupar um conjunto restrito de imagens, daí o vocabulário e o peso de cada um das palavras do vocabulário não ser o mesmo que no início do algoritmo. Por esta razão procede-se à aplicação da estratégia Bag-of-Features apenas às imagens contidas no *cluster* que se pretende analisar.

É então aplicado de novo o método K-Means para agrupamento dessas imagens, obtendo-se um conjunto de sub-períodos. Este processo poderia continuar por vários níveis, contudo na solução apresentada apenas foram contemplados 2 níveis de agrupamento, estando preparado para ser extensível a mais níveis no futuro.

A construção de um método de agrupamento hierárquico menos automático face às implementações *hierarchical clustering* que já existem prende-se com a maior liberdade do utilizador, ficando a seu cargo decidir quando a análise deve prosseguir para um determinado *cluster* ou não, algo que não é controlado no método *hierarchical clustering* já existente. No capítulo seguinte serão apresentados alguns resultados decorrentes da aplicação da ferramenta.

3.2.4 Classificação de Pinturas Baseada em Características das Assinaturas

Um dos objectivos principais da dissertação era construir um classificador que permitisse classificar pinturas em termos da sua autoria de acordo com as características visuais das assinaturas contidas nas mesmas. Assim, o classificador construído deveria etiquetar cada imagem de entrada com uma etiqueta indicando se pertence ao pintor em estudo, podendo tratar-se de uma assinatura de outro pintor ou até mesmo de uma falsificação. No futuro, ao surgirem novas assinaturas para estudo, poderão então ser submetidas ao classificador já construído e aí este poderá ajudar a decidir com alguma certeza se as assinaturas pertencem ao pintor em estudo para o qual classificador foi treinado. O diagrama de blocos do método de classificação encontra-se na figura 3.8.

A classificação pode ser baseada em vários tipos de características. Tal como referido, optou-se por utilizar a técnica SIFT devido à sua eficácia face às características das

assinaturas, sendo estas organizadas utilizando a técnica Bag-of-Features.

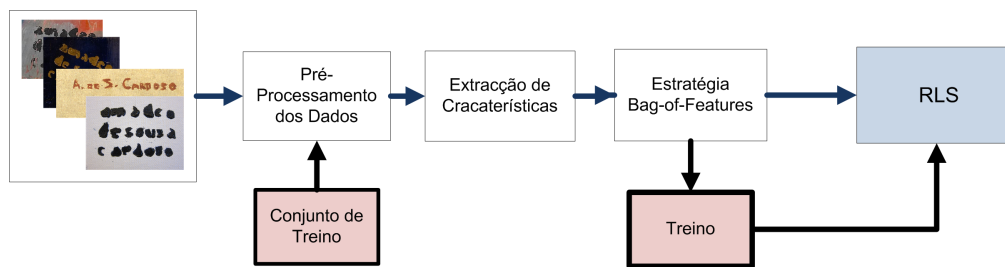


Figura 3.8: Diagrama de blocos do método de classificação de assinaturas.

Como algoritmo de classificação utilizou-se um classificador binário designado por Regularized Least Squares Classifier (RLSC) [PS03], uma versão simplificada das Support Vector Machines (SVM) [Wah99] e cujos resultados são equivalentes [RYP03]. Descreve-se em seguida o processo de classificação. Dado um conjunto de treino $T_n = \{(v_i, x_i)_{i=1}^n\}$ onde as etiquetas x_i tomam os valores $\{-1, 1\}$ e v_i é o vector de características da imagem. O RLSC encontra uma fronteira de decisão entre duas classes, por norma uma positiva e uma negativa, sendo esta obtida pela função discriminante,

$$f(v) = \sum_{i=1}^n c_i K(v_i, v) \quad (3.5)$$

onde $K(v_i, v)$ representa o *Kernel* Gaussiano $K(v_i, v) = e^{-\frac{\|v-v_i\|^2}{2\sigma^2}}$, n é o número de pontos no treino e $c = [c_1, \dots, c_n]^T$ é o vector de coeficientes estimado pelo método dos mínimos quadrados [PS03],

$$(n\gamma I + M)c = x \quad (3.6)$$

onde I é a matriz identidade, M é uma matriz quadrada definida positiva com os elementos $K_{i,j} = K(v_i, v_j)$, x é um vector com as coordenadas x_i e γ é um parâmetro de regularização. Os valores óptimos de σ e γ foram obtidos através de validação cruzada.

Por fim, para cada imagem contida no conjunto de teste, representada pela histograma de palavras visuais (Bag-of-Features) v_i , se $f(v_i) \leq 0$ então a assinatura é etiquetada como pertencente à classe negativa, ou seja, por exemplo "Não Amadeo" ($x_i = -1$). Caso contrário se $f(v_i) > 0$, então a assinatura é etiquetada como pertencente à classe positiva, ou seja, pertencente à classe "Amadeo" por exemplo ($x_i = 1$).

A construção deste classificador recebe como entrada um conjunto de imagens de assinaturas, sendo uma parte pertencente ao pintor sobre o qual se pretende construir o classificador e outra não pertencendo ao pintor. Selecciona-se um conjunto de imagens para treinar o classificador, sendo obrigatoriamente uma parte dessas imagens pertencentes ao pintor e outra parte não pertencente. As restantes imagens que não estão incluídas no treino são então utilizadas para teste de classificador e avaliação do seu desempenho.

Com o classificador construído, é possível introduzir novas imagens de assinaturas

no classificador de modo a obter a sua classificação com um determinado grau de certeza. Este classificador pode ser utilizado para verificação de autenticidade em que as assinaturas positivas pertencem a um pintor em estudo e as assinaturas negativas constituem falsificações de assinaturas do pintor. Pode também ser utilizado para atribuição do artista, em que as assinaturas positivas são assinaturas pertencentes ao pintor em estudo e as assinaturas negativas pertencem a outros pintores, designando o classificador se uma determinada assinatura pertence ao pintor em estudo ou não. Isto constitui assim uma ferramenta importante para verificação de autenticidade e atribuição do artista baseada nas características existentes nas assinaturas. Os resultados obtidos utilizando este classificador, mas com diversas configurações de imagens de entrada, serão descritas no capítulo seguinte.

3.3 Classificação de Pinturas Baseada em Características de Pincelada

Anteriormente no âmbito do projecto em que a dissertação se realiza já havia sido construído um classificador que resolvia o problema da atribuição de pinturas baseada nas características da pincelada do artista. Esse trabalho encontra-se relatado em [MJC⁺12], tendo sido obtido resultados significativos.

A forma, a orientação e a distribuição das pinceladas são marcas distintivas deixadas pelo artista na superfície da pintura. O classificador construído capta isso mesmo, as características da pincelada, sendo a atribuição do artista feita a partir daí. Com o objectivo da construção de um classificador combinado de pinturas, foi necessário voltar a construir este classificador a fim de integração final com o classificador de assinaturas.

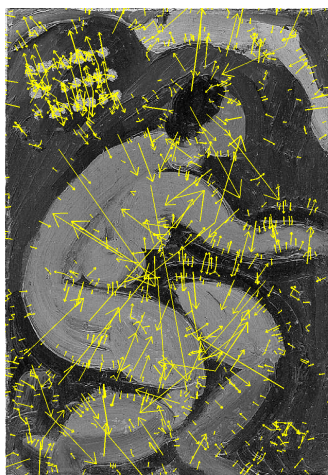


Figura 3.9: Pontos de interesse extraídos utilizando a técnica SIFT numa pintura de Amadeo.

A estrutura deste classificador segue a mesma estrutura do classificador de assinaturas (ver figura 3.8). As características das pinturas são extraídas utilizando a técnica

SIFT (ver figura 3.9), sendo depois organizadas utilizando a estratégia Bag-of-Features, resultando assim num vector de características para cada imagem. Esses vectores de características constituem os dados de entrada do classificador, tendo sido neste caso também utilizado o Regularized Least Squares Classifier (RLSC), à semelhança do caso anterior. No trabalho descrito testaram-se várias configurações para o número de imagens no treino e no teste, assim como também vários tipos de características extraídas. No âmbito da dissertação esse estudo não era relevante, tendo sido dada especial atenção apenas aos resultados finais do classificador com vista à construção do classificador combinado.

3.4 Mapeamento de Materiais

Esta secção relata todo o trabalho realizado no âmbito do mapeamento de materiais. O objectivo era construir uma ou mais ferramentas que permitissem mapear os materiais existentes numa pintura recorrendo a análise de uma imagem hiperespectral. Aplicaram-se vários métodos para efectuar o respectivo mapeamento, contudo devido ao facto de não ser possível atingir alguns dos objectivos considerados como fundamentais (ver secção 2.5.2) acabaram por ser preteridos. A visualização de imagens hiperespectrais utilizando o PCA tem sido utilizada em inúmeros trabalhos anteriores, mas devido ao facto de não permitir uma visualização consistente ao analisar várias imagens não se considerou a aplicação deste método. Descrevem-se em seguida quatro métodos construídos para efectuar o respectivo mapeamento.

3.4.1 Organização e Pré-Processamento dos Dados

Inicialmente os dados consistiam em quatro imagens hiperespectrais: duas de quadros autênticos de Amadeo de Souza-Cardoso e duas de pinturas falsificadas do mesmo pintor. Os métodos desenvolvidos incidiram principalmente sobre estas quatro pinturas embora tenham surgido, na parte final da dissertação, mais cinco imagens hiperespectrais de pinturas autênticas do pintor.

A imagem hiperespectral é constituída por uma sequência de 33 imagens de uma dada cena (neste caso, uma cena é constituída por um quadro autêntico de Amadeo de Souza-Cardoso ou por uma pintura falsificada do mesmo pintor), correspondentes a diferentes bandas espectrais, de 10 nm de intervalo, na gama de frequências entre 400 nm e 720 nm (ver figura 3.10). A câmara utilizada possui uma resolução de 1344 x 1024 *pixels*, existindo para cada *pixel* um espectro de reflectância, constituído pela amostragem na cena de uma zona em 33 comprimentos de onda.

Antes de iniciar o processo de mapeamento dos materiais foi necessário efectuar um tratamento da imagem para eliminar a moldura e a estrutura onde esta estava colocada a pintura. Isto devido ao facto do conteúdo das imagens hiperespectrais não se cingir apenas ao conteúdo da pintura, mas conter também elementos pertencentes à estrutura

onde a pintura estava colocada.

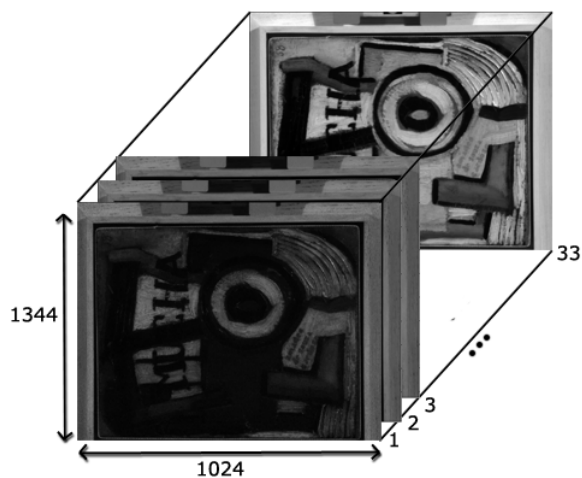


Figura 3.10: Organização da imagem hiperespectral.

Para que a análise se cingisse apenas ao conteúdo da pintura procedeu-se ao corte da imagem hiperespectral de modo a eliminar as zonas que pertenciam à estrutura onde a pintura estava alojada (exemplo na figura 3.11).

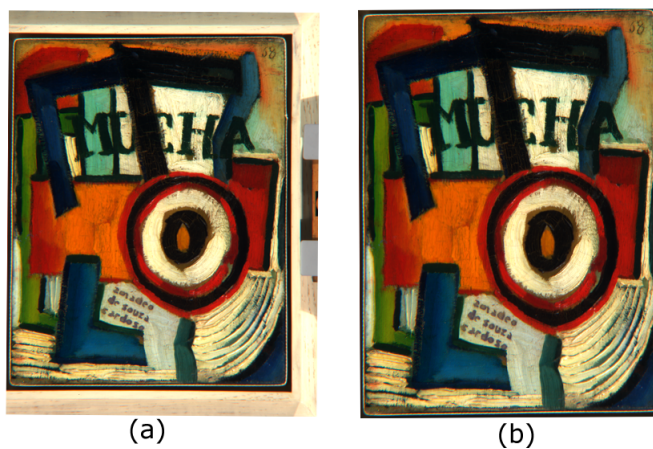


Figura 3.11: Pré-Processamento das imagens hiperespectrais. (a) Imagem original. (b) Imagem recortada.

3.4.2 Mapeamento de Materiais Através de Classificação Supervisionada

O mapeamento de materiais através de classificação supervisionada assenta na definição de uma base de dados de referência construída a partir da imagem hiperespectral em análise, sendo a classificação directamente dependente do conteúdo dessa base de dados (ver secção 2.5.5). Este processo consiste em três fases: construção da base de dados, algoritmo de classificação e visualização dos resultados.

3.4.2.1 Construção da Base de Dados de Referência

A classificação supervisionada surge quando a base de dados de espectros de referência é construída através de amostragem da imagem hiperespectral em estudo (figura 3.12). Por norma a definição da base de dados é efectuada por peritos.

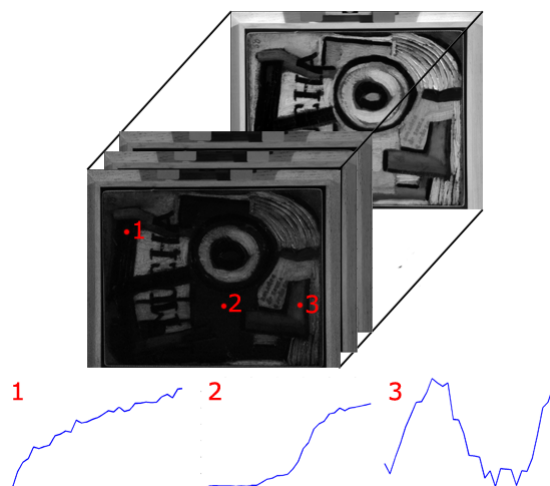


Figura 3.12: Processo de construção da base de dados de referência

No caso da ferramenta construída, o utilizador deve seleccionar um conjunto de pontos na imagem hiperespectral sobre os quais irão ser recolhidas as amostras. A escolha destes pontos deve garantir que numa área quadrada de 13×13 *pixels* não existe outro material diferente.

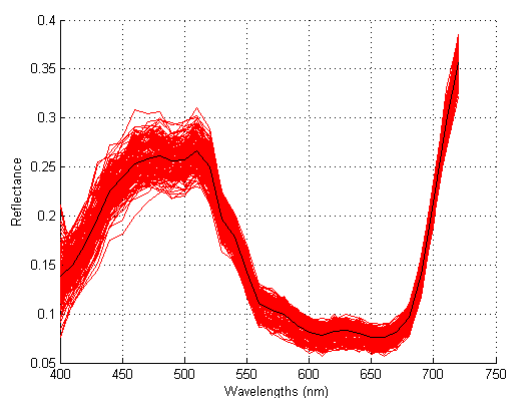


Figura 3.13: Exemplo de variabilidade de um material ao longo de uma mesma superfície. A preto o espectro médio calculado.

Para cada ponto escolhido, define-se uma área quadrada de 13×13 *pixels* à sua volta de modo a captar um conjunto de amostras mais significativas para um mesmo material. Devido à variabilidade do mesmo material ao longo da mesma superfície (ver figura 3.13), a captação de uma única amostra poderia conter demasiado ruído e não ser suficientemente representativa do material em questão. Assim, define-se uma área contendo

169 amostras, sendo o espectro de referência para um dado material calculado através da média de todos os espectros da mesma área. Repete-se este processo para todos os pontos seleccionados de modo a construir a base de dados de espectros de referência para cada um dos materiais seleccionados.

Para efeitos de visualização, atribui-se um valor (R,G,B) a cada um dos espectros de referência de modo a ser possível construir um *output* visual onde cada zona da pintura é colorida com a cor do material que lhe foi atribuído através do algoritmo de classificação.

3.4.2.2 Algoritmo de Classificação

Com a base de dados de espectros de referência construída, procede-se à classificação. Em termos de classificação, muitos dos trabalhos de investigação na área da análise de imagens hiperespectrais recorreram ao Spectral Angle Mapper (SAM, ver secção 2.5.5.1) devido à sua eficácia em comparar espectros. Este método é invariante a mudanças de iluminação e não tem em conta os comprimentos de cada um dos espectros. Contudo, por vezes, determinados problemas podem surgir com a sua utilização. De uma forma simplificada pode afirmar-se que o SAM efectua comparações considerando principalmente a forma dos espectros, ignorando as mudanças de iluminação. A distância Euclidiana, uma das métricas também utilizadas na comparação de espectros é, contrariamente ao SAM, sensível a mudanças na intensidade dos espectros, isto é, várias tonalidades de um mesmo pigmento podem ser consideradas como pertencentes a pigmentos diferentes.

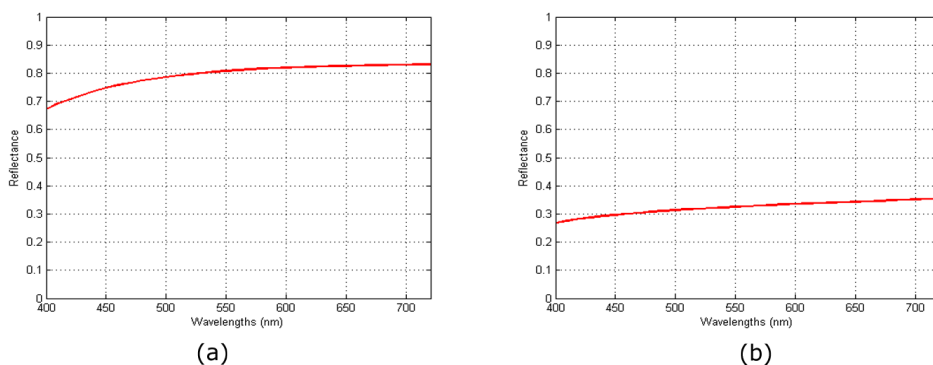


Figura 3.14: Espectros com forma semelhante mas pertencentes a materiais diferentes (a) Base de suporte de uma pintura (b) Pigmento cinzento.

Porém, em alguns casos acontece que dois espectros têm a mesma forma, mas um deles pode ter uma reflectância maior que o outro, pertencendo a pigmentos diferentes (ver figura 3.14). No caso do SAM, iria considerar os espectros como pertencendo ao mesmo material, influenciando negativamente o resultado da classificação. Para colmatar este aspecto, construiu-se um método que combina o SAM com a distância Euclidiana, suportando-se no método de *matching* proposto por David Lowe em [Low04] para *matching* de *keypoints* ou pontos de interesse. O algoritmo segue os seguintes passos:

1. Calcular o ângulo $\alpha_{p,r}$ entre o *pixel* p da imagem hiperespectral e cada espectro de referência r contido na base de dados:

$$\alpha_{p,r} = \cos^{-1} \left[\frac{\sum_{i=1}^n R_{p,i} R_{r,i}}{\sqrt{\sum_{i=1}^n R_{p,i}^2 \sum_{i=1}^n R_{r,i}^2}} \right], \quad (3.7)$$

onde $\alpha_{p,r}$ é o ângulo espectral entre o *pixel* p e o espectro de referência r , n é o número de comprimentos de onda (neste caso, 33), $R_{p,i}$ é o valor de reflectância do *pixel* p no comprimento de onda i e $R_{r,i}$ é o valor de reflectância para o espectro de referência r no comprimento de onda i .

2. Ordenar os vários ângulos $\alpha_{p,r}$ para cada um dos materiais da base de dados por ordem crescente do ângulo obtido. Se o menor ângulo α_{p,r_1} calculado for menor do que uma percentagem (d_x) do segundo menor ângulo α_{p,r_2} (método de *matching* proposto por David Lowe, ver equação 3.8), então o *pixel* p é atribuído ao material de referência com menor ângulo (r_1). No caso da ferramenta construída é permitida alguma flexibilidade neste passo de tal modo que se pode definir um *threshold* sobre o ângulo espectral. Se o ângulo mínimo calculado for maior ou igual do que o *threshold*, então o *pixel* contém um material que não está presente na base de dados e é etiquetado como "Unclassified", enquanto que nos restantes casos é etiquetado com o material de referência para o qual o ângulo espectral é mínimo.

$$\alpha_{p,r_1} < d_x \alpha_{p,r_2} \quad (3.8)$$

3. Caso a condição anterior não se verifique, procede-se ao cálculo da distância Euclidiana $Ed_{p,r}$ para o *pixel* p e cada um dos espectros de referência r :

$$Ed_{p,r} = \sqrt{\sum_{i=1}^n (R_{p,i} - R_{r,i})^2} \quad (3.9)$$

Atribui-se ao *pixel* p a etiqueta do material de referência r para o qual se obteve a menor distância Euclidiana de entre os presentes na base de dados de espectro de referência.

O método de classificação construído pretende sobretudo eliminar alguns dos casos onde existem dois ângulos espectrais muito próximos, sugerindo o facto dos espectros terem formas semelhantes mas um deles ter mais intensidade que outro, algo que pode não se dever apenas a mudanças de iluminação.

3.4.2.3 Visualização dos Resultados

A visualização permite apresentar os resultados num ecrã para análise visual. Aquando da definição da base de dados de espectros de referência, o utilizador deve introduzir um

valor de RGB para cada um dos espectros de referência. Para cada *pixel* teste de entrada no algoritmo de classificação, este é classificado como pertencente ao material com o espectro de referência mais próximo de acordo com a técnica utilizada. Esta classificação pressupõe também que a cada *pixel* é atribuído o valor RGB correspondente ao material com que este foi classificado. O utilizador pode optar por visualizar todos os *pixels* de uma imagem hiperespectral, podendo facilmente identificar zonas com o mesmo material e pode também visualizar apenas todos os *pixels* classificados como pertencentes a um determinado material contido na base de dados.

3.4.3 Mapeamento de Materiais Através de Classificação Não Supervisionada

O método anterior requer a amostragem de um conjunto de pontos para construir a base de dados de referência. A amostragem dos pontos influenciará sempre o resultado da classificação, ficando a cargo do utilizador a escolha dos mesmos. Apesar da eficácia do método anterior, o factor descrito pode consistir numa limitação para os utilizadores. Para colmatar esta limitação procedeu-se no sentido de desenvolver um método mais automático de classificação não supervisionada. Esta classificação apresenta duas fases: algoritmo de *clustering* (que por si acaba por efectuar também a respectiva classificação) e visualização dos resultados.

3.4.3.1 Algoritmo de Clustering

A classificação é efectuada de forma automática através da utilização de um algoritmo de *clustering*, neste caso o K-Means combinado com o Spectral Angle Mapper. O algoritmo de classificação segue os seguintes passos:

1. Selecção de k *pixels* da imagem hiperespectral que irão ser utilizados como centróides iniciais dos k *clusters*.
2. Calcular o ângulo espectral $\alpha_{p,c}$ entre o *pixel* p e a média de cada *cluster* μ_c :

$$\alpha_{p,c} = \cos^{-1} \left[\frac{\sum_{i=1}^n R_{p,i} \mu_{c,i}}{\sqrt{\sum_{i=1}^n R_{p,i}^2 \sum_{i=1}^n \mu_{c,i}^2}} \right], \quad (3.10)$$

3. Para cada *pixel* p , encontra-se o *cluster* c de tal modo que $\alpha_{p,c}$ é mínimo e atribui-se o *pixel* a esse *cluster*.
4. Recalcular o centróide (ou média) do *cluster*,

$$\mu_{c,i} = \frac{\sum_{j=1}^N x_{j,i,c}}{N} \quad (3.11)$$

onde $\mu_{c,i}$ é o valor médio do *cluster* c no comprimento de onda i , N é o número de *pixels* no *cluster* c e $x_{j,i,c}$ é o valor do *pixel* j no comprimento de onda i no *cluster* c .

5. Repetir os 3 passos anteriores até as iterações atingirem um *threshold* de convergência. Quando esse *threshold* for atingido, a classificação termina.

Mais um vez, com a utilização do K-Means surge a desvantagem de se ter de indicar o número de *clusters*, ficando a sua indicação a cargo do utilizador.

3.4.3.2 Visualização dos Resultados

A classificação agrupa conjuntos de *pixels* semelhantes, contudo com esta abordagem surge um problema. Não existe forma *a priori* de identificar o que modela cada um dos *clusters* encontrados, sendo difícil atribuir uma cor a cada um dos *clusters* que permita visualizar os materiais de uma forma apropriada. Existem duas alternativas para construir uma visualização consistente a partir desta abordagem:

- A primeira consiste em projectar cada um dos centróides de cada *cluster* numa base de dados de espectros de referência. Utilizando o Spectral Angle Mapper, calcula-se o ângulo espectral entre cada *cluster* e cada um dos espectros da base de dados de referência. O espectro mais próximo é atribuído ao *cluster* em teste. Esta técnica foi utilizada em dois trabalhos anteriores (em [SMG99, SR02]), contudo torna todo o processo menos automático pois é necessário definir uma base de dados de espectros de referência.
- A segunda alternativa consiste em calcular a cor de um conjunto de *pixels* a partir do centróide de cada *cluster* e atribui-se essa mesma cor a todos os *pixels* desse mesmo *cluster*. Isto é possível através da escolha de um iluminante (D65), retirando-se primeiramente do espectro as componentes no espaço XYZ e depois convertendo-se essas componentes para RGB¹. Esta foi a opção utilizada, permitindo assim ter um método automático para mapeamento de materiais, podendo até efectuar-se um estudo de quais os materiais que se realçam à medida que o número de *clusters* aumenta.

3.4.4 Mapeamento de Materiais Através de Classificação com Base de Dados Externa

Os dois métodos anteriores efectuam o mapeamento recorrendo apenas ao conteúdo da imagem hiperespectral. No entanto, o mapeamento de materiais pode também ser efectuado recorrendo a informação externa, como o caso do Fiber-Optic Reflection Spectrometer (FORS)², pois este fornece informação com um sinal semelhante à câmara hiperespectral e que pode ser integrada. Construíram-se assim dois métodos que incorporavam

¹http://personalpages.manchester.ac.uk/staff/david.foster/Tutorial_HSI2RGB/Tutorial_HSI2RGB.html

²Trata-se de uma técnica não-invasiva que permite efectuar medições em objectos. Cada medição pontual resulta na espectroscopia reflectida na região do visível, isto é, tomando a utilização mais usual, normalmente pinturas, o resultado é um espectro de reflectância que fornece informação acerca do pigmento utilizado ou da análise colorimétrica (<http://fors.ifac.cnr.it>).

a informação obtida com esta técnica, sendo que o segundo efectua uma análise mais aprofundada recorrendo a outros tipos de informação.

3.4.4.1 Mapeamento de Materiais Através de Informação FORS

No estudo de obras de arte por peritos com recursos na área da análise química, estes recorrem normalmente a uma análise utilizando a técnica FORS a fim de analisar com mais detalhe determinadas zonas da pintura. Todavia esta análise é feita pontualmente em determinados pontos do quadro, o que não resulta numa análise global do mesmo. Esta informação pontual pode assim ser utilizada para efectuar o mapeamento de materiais, à semelhança do trabalho descrito em [DZT⁺10]. O método construído assemelha-se ao mapeamento de materiais através de classificação supervisionada descrito anteriormente em 3.4.2, apenas com a única diferença ao nível da construção de base de dados de espectros de referência.

Construção da Base de Dados e Classificação Neste caso, a base de dados é definida a partir de amostras do próprio quadro mas a informação é adquirida com uma técnica diferente, a FORS. A análise FORS é efectuada por peritos, sendo efectuada em pontos específicos da pintura, não eliminando a variabilidade de um material ao longo de uma mesma superfície. Esta análise é dependente de cada pintura pois cada pintura possui a sua análise FORS única. No âmbito da dissertação, existia a análise FORS apenas para as duas pinturas falsificadas, existindo em alguns casos várias análises pontuais para um mesmo material. Nesses casos, com a ajuda de um perito para construir a base de dados, para cada material calculou-se o espectro médio das análises pontuais existentes para esse mesmo material, tentando assim eliminar a variabilidade de um mesmo material ao longo da pintura.

Após a definição da base de dados, procedeu-se à classificação e à visualização dos resultados, sendo ambos os processos idênticos à classificação supervisionada (ver secção 3.4.2).

3.4.4.2 Mapeamento de Materiais com Análise de Misturas e Validação de Autenticidade

Para as pinturas verdadeiras de Amadeo de Souza-Cardoso para as quais havia sido construída uma imagem hiperespectral não havia qualquer análise FORS sobre estas pinturas. Isto levou a um obstáculo, o método anterior não poderia ser aplicado pois não havia uma base de dados externa de espectros de referência. Porém, a fim de ultrapassar este obstáculo obteve-se uma base de dados de espectros amostrada a partir da paleta do pintor (figura 3.15), ou seja, dos materiais puros que este utilizou. Na pintura os materiais não aparecem apenas no seu estado puro devido a vários factores: mistura de pigmentos, sobreposição de camadas pictóricas, presença da camada de preparação, entre outros factores. Esta base de dados poderia ser uma base de dados geral a todas as

pinturas deste pintor, contudo a presença apenas de materiais puros iria condicionar os resultados, esperando-se uma classificação pouco assertiva.



Figura 3.15: Paleta de tintas do pintor Amadeo de Souza-Cardoso.

Face aos factos anteriores caminhou-se no sentido de se efectuar um mapeamento de materiais complementando com uma análise de misturas de pigmentos a partir da base de dados de materiais puros, tentando também efectuar a verificação de autenticidade de zonas da pintura. Vários trabalhos de investigação incidiram sobre um método de análise de misturas designado por Spectral Mixture Analysis [DHR04], contudo este método incide principalmente sobre a resolução dos *pixels* contidos na imagem hiperespectral e o facto de um *pixel* poder conter vários materiais com espectros diferentes, o que resulta num espectro anormal para esse *pixel*. Este método obriga também que todos os *pixels* sejam automaticamente considerados uma mistura de materiais, o que neste caso não seria correcto afirmar pois existem zonas das pinturas onde se podem encontrar apenas materiais puros.

Dados de Suporte ao Algoritmo Após vários testes efectuados com as métricas anteriores verificou-se não ser possível identificar correctamente grande parte dos pigmentos puros presentes nas pinturas. Isto devido à diferença apresentada entre os espectros puros da base de dados de referência e os espectros presentes na pintura, algo que era considerável tendo em conta as condições de ruído e iluminação, assim como a presença de vernizes aplicados sobre as pinturas. Este facto condicionou a abordagem do problema pois era expectável que fosse possível efectuar a análise de pinturas de Amadeo de Souza-Cardoso a partir desta base de dados de referência. Face aos obstáculos encontrados foi necessário combinar vários tipos de informação num esquema de classificação combinada semelhante a um fluxograma de decisão (ver figura 3.16).

Para dar suporte ao fluxograma de decisão construído, teve de ser fornecido um conjunto de informação complementar além da imagem hiperespectral de uma pintura. Apresenta-se em seguida os vários dados necessários e respectiva notação matemática definida para cada tipo de modo a descrever o algoritmo matematicamente:

- Em primeiro lugar foi necessário calcular a cor na norma CIELab para cada um dos *pixels* da imagem hiperespectral. Para um determinado pixel i da imagem hiperespectral considere-se a notação Lab_i a sua cor na norma CIELab.

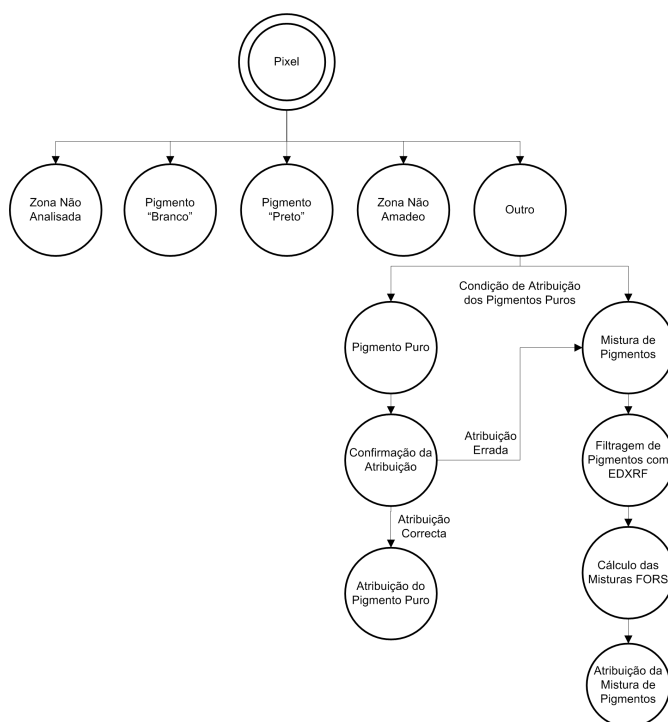


Figura 3.16: Fluxograma de decisão construído para classificação.

- Além da cor de cada *pixel* foi também necessário fornecer a base de dados de espectros puros adquiridos com a técnica FORS num conjunto de amostras dos tubos de tinta pertencentes ao pintor (figura 3.15), sendo esta constituída por 17 pigmentos (cuja designação habitual é a seguinte: Red Vermillion, Red Carmine, Raw Sienna, Terra Rossa, Orange Cd, Yellow Ochre, Yellow Cr, Violet, Cerulean Blue, Cobalt Blue, Prussian Blue, Green Cd, Viridian, Ultramarine, White, Black e Esmerald). Essa base dados é representada pela notação

$$P_{es} = \{es_1, es_2, \dots, es_N\} \quad (3.12)$$

onde es_N representa o espectro de cada um dos materiais puros contidos na base de dados, N é o número de materiais puros na base de dados e P_{es} representa a base de dados ou conjunto de espectros contidos na mesma. Para complementar a informação dos espectros puros, forneceu-se também para cada material puro uma lista dos elementos químicos presentes (ver 6.1) sendo designada por,

$$P_{el} = \{el_1, el_2, \dots, el_N\} \quad (3.13)$$

onde $el_N = \{sq_1, sq_2, \dots, sq_K\}$ representa a lista de K elementos químicos de cada um dos materiais puros contidos na base de dados, N é o número de materiais puros na base de dados e P_{el} representa a base de dados ou conjunto de listas de elementos químicos contidos na mesma.

- Por fim, de modo a colmatar as fragilidades da técnica FORS e tirando partido das vantagens de uma outra técnica, utilizou-se um conjunto de amostras pontuais EDXRF (Energy Dispersive X-Ray Fluorescence)³ captadas na superfície da pintura e para cada amostra, a cor respectiva na norma CIELab do ponto onde a amostra foi adquirida (ver figura 3.17). Normalmente uma amostra EDXRF caracteriza-se por um espectro que indica os elementos presentes através de uma análise mais completa, contudo o facto de se trabalhar com o próprio espectro apresentaria algumas limitações quando uma pintura possui uma camada de preparação, o que adiciona a cada amostra elementos estranhos aos pigmentos em estudo. Assim, invés de se utilizar o próprio espectro, utilizou-se apenas a lista de elementos químicos presentes para cada amostra, o que permite integrar vários tipos de informação.

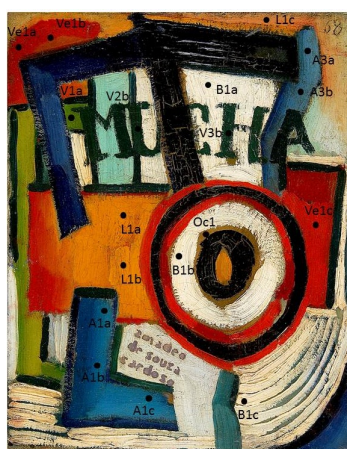


Figura 3.17: Análise pontual EDXRF efectuada na pintura Mucha

Considere-se a seguinte notação para representar este tipo de dados,

$$XRF_{el} = \{el_1, el_2, \dots, el_M\} \quad (3.14)$$

onde $el_M = \{sq_1, sq_2, \dots, sq_K\}$ representa a lista de K elementos químicos para cada uma das amostras EDXRF efectuada na pintura, M é o número total de amostras EDXRF e XRF_{el} representa a base de dados de amostras. Além da lista de elementos químicos para cada amostra, forneceu-se também a cor na norma CIELab para cada amostra. Considere-se a seguinte notação,

$$XRF_{lab} = \{xlab_1, xlab_2, \dots, xlab_M\} \quad (3.15)$$

onde $xlab_1$ representa a cor na norma CIELab para cada uma das amostras EDXRF efectuada na pintura, M é o número total de amostras EDXRF e XRF_{lab} representa

³Trata-se de uma técnica analítica utilizada para análise elementar ou caracterização química de uma amostra. Baseia-se na análise da interacção entre algumas fontes de raios-X e uma amostra. Esta interacção produz um espectro que através da análise dos vários picos do mesmo permite identificar os elementos químicos que a amostra contém.

a base de dados de cor das amostras. Neste ponto, e para evitar futuros erros de classificação, é também necessário indicar as amostras de cor correspondentes ao pigmentos branco Brn_{lab} e as amostras correspondentes ao preto Prt_{lab} .

Metodologia de Classificação A classificação assenta no fluxograma de decisão apresentado na figura 3.16. Descreve-se em seguida cada um dos passos da classificação de forma pormenorizada, recorrendo sempre que possível a notação matemática para os descrever:

1. Um dos primeiros problemas encontrados foi o facto da análise EDXRF ser uma análise pontual na pintura, o que impossibilitava o facto de se ter uma análise por *pixel*. Para tornar esta análise global a toda a pintura, recorreu-se a algumas das técnicas utilizadas por historiadores e conservadores para tornar a análise global, intervalos de tolerância de cor. Isto é, para cada amostra EDXRF existia também a respectiva amostra de cor na norma CIELab. Definindo um intervalo em que a cor poderá ser a mesma que a cor dessa amostra, todos os *pixels* cuja cor esteja dentro desse intervalo de tolerância de cor são considerados como pertencentes à amostra respectiva, generalizando assim cada amostra a conjuntos de *pixels* (ver figura 3.18).

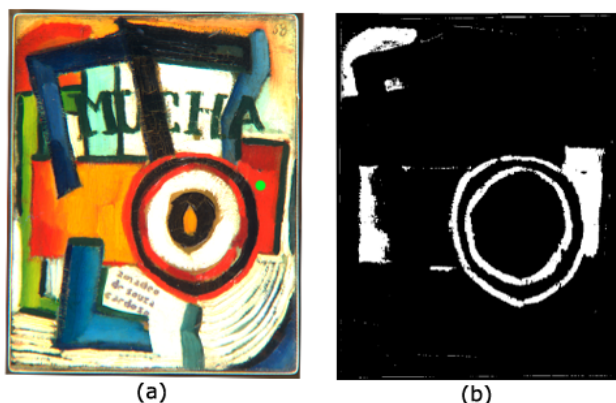


Figura 3.18: Segmentação através de intervalos de tolerância de cor. (a) Ponto na pintura onde foi efectuada a amostra. (b) A branco todos os *pixels* identificados como pertencentes à amostra.

Cada *pixel* da imagem hiperespectral é classificado individualmente, resultando numa determinada designação para o mesmo. O algoritmo começa por identificar para o *pixel* teste i qual ou quais as amostras EDXRF correspondentes em que a cor é a mesma segundo as condições anteriores. A condição simplificada, pois a norma CIELab tem três componentes, é a seguinte,

$$\exists x \in XRF_{lab} : Lab_i \geq (x - c) \wedge Lab_i \leq (x + c) \quad (3.16)$$

onde a condição se deve verificar para as três componentes da norma CIELab e onde c é uma constante definida como margem do intervalo de tolerância de cor.

Se existir uma ou mais amostras que verifiquem a condição, então obtém-se um conjunto de amostras sendo designado por M_{elem} (listas de elementos químicos) e M_{lab} (amostras de cor) que representam o subconjunto de amostras identificadas com os intervalos de tolerância de cor. Após essa identificação, a classificação é dividida nos seguintes passos:

- (a) Se não for encontrada nenhuma amostra dentro do intervalo de tolerância de cor estabelecido, isto é, que não verifique a condição descrita em 3.16, então o *pixel* é imediatamente classificado como pertencente a uma zona onde não foi efectuada qualquer análise EDXRF, sendo designado como "Zona Não Analisada" e terminando o processo de classificação para esse *pixel* (ver exemplo na figura 3.19, c).

$$Out_i = \text{"ZonaNaoAnalisada"} \quad (3.17)$$

- (b) Caso não seja identificado como uma zona não analisada, através do conhecimento das amostras de pigmento preto e branco efectuadas na pintura é possível identificar se o *pixel* pertence a algum destes materiais. A identificação destes dois materiais surge numa fase inicial da classificação pela razão que na identificação dos *pixels* pertencentes aos materiais puros, essa será feita com recurso à primeira derivada aplicada a cada um dos espectros. Como os espectros dos pigmentos preto e branco apresentam zonas horizontais, isto seria um problema computacional pois iriam surgir divisões por zero no cálculo das derivadas e muitos erros associados à classificação, daí a identificação dos *pixels* pertencentes a estes dois materiais num passo inicial do algoritmo (ver exemplo na figura 3.19, (a) e (b)). O resultado é descrito pelas condições seguintes:

$$\exists x \in M_{lab} : x \in Brn_{lab} \implies Out_i = \text{"PigmentoBranco"} \quad (3.18)$$

$$\exists x \in M_{lab} : x \in Prt_{lab} \implies Out_i = \text{"PigmentoPreto"} \quad (3.19)$$

- (c) Neste primeiro nível de classificação surge ainda a condição para verificação de autenticidade, sendo esta efectuada com recurso à informação fornecida pela análise EDXRF. A condição para que um *pixel* de uma pintura não pertença a um material de Amadeo é determinada pela presença de elementos característicos de Amadeo. Isto é, para o *pixel* em análise identificam-se todas as amostras correspondentes pelo intervalo de tolerância de cor M_{elem} . Para cada lista de elementos de cada amostra, percorrem-se todos os pigmentos puros da base de dados P_{es} e procuram-se esses elementos em cada uma das listas individuais dos materiais puros. Caso haja um elemento contido numa

das amostras que não esteja contido em nenhuma das listagens de elementos dos materiais puros, então o *pixel* é automaticamente classificado como "Não Amadeo". Por exemplo, considere que uma amostra contém o elemento "Titânio" (Ti)⁴, comparando com as listagens de elementos dos materiais puros (ver 6.1) verifica-se não existir o elemento Titânio em nenhuma delas, logo neste caso o *pixel* seria classificado como "Não Amadeo",

$$Out_i = "NaoAmadeo" \quad (3.20)$$

- (d) Caso o *pixel* não seja classificado nos passos anteriores, segue-se então a possibilidade de ser pertencente a um material puro ou de ser uma mistura de materiais, sendo a classificação efectuada noutra nível.

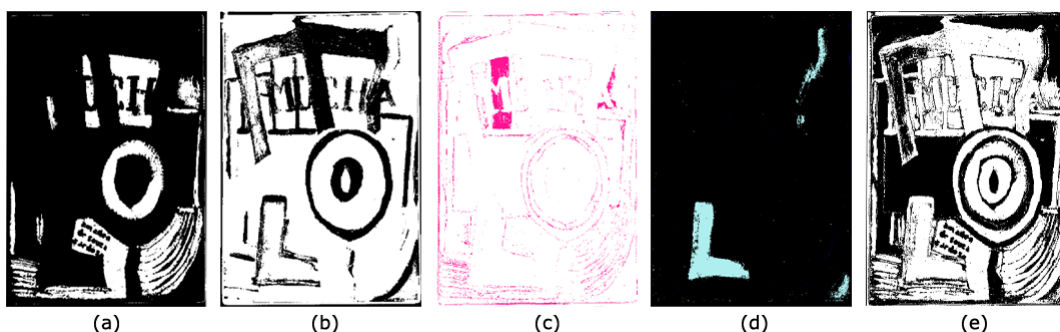


Figura 3.19: Exemplos de classificação. (a) Identificação do pigmento branco (A branco). (b) Identificação do pigmento preto (A preto). (c) Identificação das zonas sem análise EDXRF (A roxo). (d) Identificação das zonas com pigmentos puros. (e) Identificação das zonas com misturas de pigmentos (A preto).

2. No segundo nível de classificação surgiram dois grandes desafios: identificar *pixels* pertencentes a materiais puros, algo que nos testes não tinha sido possível com as métricas anteriores; e definir uma condição que permitisse identificar quando um *pixel* pertence ou não pertence a um material puro. Para resolver o primeiro problema recorreu-se a uma técnica utilizada pelos químicos na análise de espectros muito semelhantes, a utilização da primeira derivada. Esta técnica realça pequenas diferenças entre os materiais amplificando-as para uma análise e diferenciação mais eficaz, destacando principalmente os pontos de inflexão de cada uma das curvas dos espectros. Contudo não seria suficiente calcular a derivada do espectro de cada *pixel* devido ao ruído presente na imagem hiperespectral, o que iria conduzir a muita variação na derivada. Para reduzir a quantidade de ruído, antes de se aplicar a derivada procedeu-se à aplicação do *smooth* ao espectro do *pixel* em análise. Para tal efeito utilizou-se o método Savitzky–Golay [SG64], sendo este baseado em cálculos polinomiais do método dos mínimos quadrados. De referir ainda que em

⁴Elemento químico que constitui o Branco de Titânio (TiO_2), introduzido como tinta para artistas a partir de 1920, data posterior à morte de Amadeo de Souza-Cardoso.

vez de estimar uma curva de mínimos quadrados para o comprimento do espectro, este método estima o espectro pedaço por pedaço de acordo com um tamanho definido pelo utilizador. Este método tem sido abordado em inúmeros trabalhos devido à sua eficiência, podendo destacar-se o trabalho descrito em [Vai06]. A figura 3.20 mostra um exemplo do resultado obtido da aplicação deste método. Por fim calculou-se a primeira derivada do espectro do *pixel* em análise, tendo sido calculada também para cada um dos espectros de materiais puros da base de dados de espectros de referência.

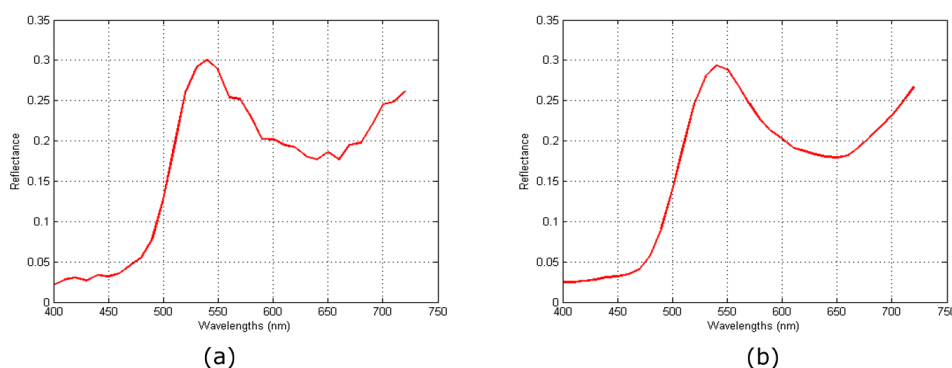


Figura 3.20: **(a)** Espectro original da imagem hiperespectral. **(b)** Espectro após aplicação de filtro Savitzky–Golay de ordem binomial e com tamanho do filtro 9.

Após os passos anteriores procedeu-se ao cálculo do ângulo espectral (SAM) entre o espectro com a primeira derivada do *pixel* teste i e cada um dos espectros da base de dados de materiais puros (com a primeira derivada), técnica já utilizada em outros trabalhos de investigação [DHD⁺06],

$$\alpha_{i,r} = \cos^{-1} \left[\frac{\sum_{k=1}^n D_{i,k} D_{r,k}}{\sqrt{\sum_{k=1}^n D_{i,k}^2 \sum_{k=1}^n D_{r,k}^2}} \right], \quad (3.21)$$

onde $\alpha_{i,r}$ é o ângulo espectral entre o *pixel* i e o espectro de referência r pertencente a P_{es} , n é o número de comprimentos de onda (neste caso, 33), $D_{i,k}$ é o valor da derivada do espectro do *pixel* i no comprimento de onda k e $D_{r,k}$ é o valor da derivada do espectro de referência r no comprimento de onda k .

Ordenaram-se os ângulos obtidos por ordem crescente e neste passo surgiu o segundo desafio, definir uma condição sobre este conjunto de ângulos para identificar se o *pixel* teste pertence a um material puro ou a uma mistura de materiais. Optou-se por utilizar uma condição já definida anteriormente pela equação 3.8, baseada na condição de *matching* proposta por David Lowe. Ou seja, se o menor ângulo obtido for menor do que uma percentagem do segundo menor ângulo, então o *pixel* é classificado como pertencente a um material puro, caso contrário trata-se de uma mistura de materiais puros. Esta condição apresenta vantagens sobre a definição

de um único *threshold* sobre a os ângulos, pois neste caso irão ocorrer menos falsas atribuições de materiais puros. Esta condição leva assim a duas opções:

- (a) Caso o *pixel i* seja identificado como pertencente a um material puro, a classificação não termina, seguindo-se um processo de confirmação da classificação. Como sugestão de alguns peritos envolvidos no projecto em que a dissertação se integra, decidiu-se construir um processo para confirmação da classificação efectuada. O processo de validação é baseado na análise dos elementos químicos, à semelhança da condição de verificação de autenticidade. Para o *pixel i* em análise, identificou-se no primeiro nível a amostra EDXRF (ou várias quando existem várias para a mesma zona) correspondente (representada por M_{elem}). Essa amostra é representada por uma lista de elementos químicos à semelhança de cada um dos materiais puros presentes na base de dados. Compara-se assim a lista de elementos da amostra EDXRF para o *pixel* de teste com a lista de elementos do material puro candidato (exemplos descritos em 3.1). Se as listas de elementos forem iguais, então a atribuição foi efectuada correctamente e o *pixel* é definitivamente classificado como sendo pertencente ao material puro (ver 3.22) anteriormente definido como candidato (exemplo figura 3.19, (d)).

$$Out_i = \text{''PigmentoPuro''} \quad (3.22)$$

Caso contrário, a atribuição baseada na informação FORS foi efectuada de forma incorrecta e surge a necessidade de corrigir essa mesma atribuição. Assim o que acontece é que o *pixel* pode eventualmente ter sido mal incluído como *pixel* de um material puro através da condição de definição dos materiais puros, partindo-se para uma análise das misturas para esse *pixel* específico, sendo esta definida no próximo passo.

Amostra EDXRF <i>Pixel</i> Teste	Material Puro Candidato	Resultado
Cr, Hg	Hg (Red Vermillion)	Classificação incorrecta
Cd	Cd (Orange Cd)	Classificação correcta

Tabela 3.1: Exemplo da metodologia de validação da classificação dos materiais puros.

- (b) A análise das misturas de pigmentos para um determinado *pixel* é efectuada quando este não é englobado na condição dos *pixels* puros ou quando a classificação dos puros foi efectuada de forma errada e surge a necessidade de corrigir esse facto. Este tipo de análise é efectuada em vários passos, conjugando a informação obtida com a técnica FORS com a informação obtida com EDXRF para aumentar a eficácia de classificação. Partindo da amostra EDXRF atribuída ao *pixel* teste de acordo com o intervalo de cor (M_{elem}), obtém-se a sua listagem de elementos. Uma mistura de pigmentos apenas pode ser

constituída por pigmentos cujos todos os seus elementos estejam presentes na mistura, ou seja, considerando uma mistura com Crómio (Cr) e Ferro (Fe), um pigmento puro com os elementos Cobalto (Co) e Crómio (Cr) não poderá fazer parte da mistura pois esta não tem o elemento Cobalto (Co). Partindo da mistura (formada pela amostra), percorre-se todas as listagens de elementos de pigmentos puros (P_{el}) e procuram-se aqueles em que todos os seus elementos estão presentes na mistura. Resulta assim num conjunto de pigmentos puros possíveis para a mistura final.

Veja-se um exemplo concreto da selecção de pigmentos que poderão fazer parte de uma mistura. Considere-se uma determinada amostra captada num dos quadros de Amadeo contendo a seguinte listagem de elementos: Cr (Crómio) e Hg (Mercúrio). Percorrendo as listagens de elementos presentes nos pigmentos puros (ver 6.1), obtêm-se como possíveis hipóteses para integrar a mistura:

- Red Vermillion (Hg)
- Yellow Cr (Cr)
- Viridian (Cr)

De modo a tornar o mais claro possível, enuncia-se outro exemplo. Considere-se uma amostra cuja listagem de elementos presentes é: Cr (Crómio), Fe (Ferro). Os pigmentos puros que poderão fazer parte desta mistura são:

- Yellow Ochre (Fe)
- Yellow Cr (Cr)
- Prussian Blue (Fe)
- Viridian (Cr)

Após a identificação de todos os pigmentos puros possíveis, procede-se à construção dos espectros das misturas através do cálculo de todas as combinações possíveis de pigmentos puros seleccionados e, para cada combinação, calcula-se o espectro FORS da combinação através de uma soma linear dos espectros dos pigmentos puros presentes na combinação. Apenas se constroem combinações admissíveis de modo a não introduzir erro na classificação. Considere-se o último exemplo apresentado em que se tratava de uma mistura de Cr (Crómio) e Fe (Ferro) e os elementos puros identificados. Não faria sentido calcular a combinação entre Yellow Ochre (Fe) e Prussian Blue (Fe), assim como entre Yellow Cr (Cr) e Viridian (Cr) pois os elementos presentes nestas misturas seriam diferentes da mistura real. Assim podem combinar-se Yellow Ochre (Fe) com Viridian (Cr), Yellow Ochre (Fe) com Viridian (Cr) e Yellow Cr (Cr), exemplificando assim o comportamento do método.

Para incrementar o desempenho do algoritmo, calcula-se a soma linear de espectros dos pigmentos puros individuais multiplicados por uma percentagem,

incidindo sobre todas as combinações de percentagem em passos de 10 em 10 pontos percentuais. Assim, garantia-se que existia um espectro com 90% de um pigmento verde e 10% de um pigmento amarelo, um espectro com 80% de pigmento verde e 20% de pigmento amarelo e assim sucessivamente (ver exemplo figura 3.21). De realçar que as combinações de pigmentos foram restringidas apenas a misturas de 3 pigmentos pelo facto de não se saber que existam misturas de quatro pigmentos. Este passo consiste numa aproximação bastante alargada do que se pode considerar uma mistura de pigmentos ou colorantes. Isto porque ainda que se possa considerar esta soma linear ao nível de curva espectral. Em termos colorimétricos esta é uma aproximação algo grosseira conduzindo por vezes a uma cor diferente do que se esperaria obter.

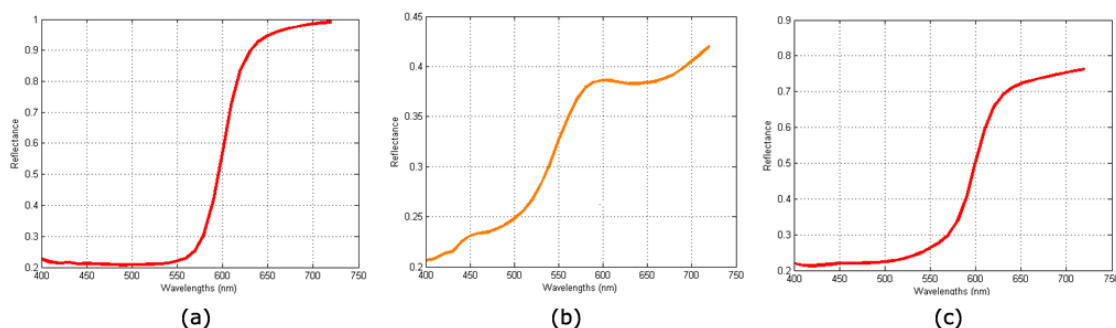


Figura 3.21: Exemplo de soma linear de dois espectros. (a) Espectro Red Vermillion. (b) Espectro Yellow Ochre. (c) Espectro soma 60% Red Vermillion e 40% Yellow Ochre.

Calculados os espectros para todas as combinações possíveis, procede-se à identificação da mistura respectiva, atribuição que é efectuada pela informação FORS. Procede-se ao cálculo do ângulo espectral (SAM) entre o espectro do *pixel* teste i directamente extraído da imagem hiperespectral e sem recurso a qualquer derivada, e todas as misturas calculadas. A mistura para qual o ângulo calculado for menor é atribuída ao *pixel* em questão (ver 3.23), terminando a classificação neste ponto (exemplo figura 3.19, (e)). Neste último passo não se utilizou a derivada pois a sua utilização anterior deveu-se à dificuldade em identificar os pigmentos puros, sendo o SAM neste caso a alternativa mais eficiente em termos de classificação.

$$Out_i = \text{''MisturadePigmentos''} \quad (3.23)$$

Independentemente do ramo escolhido, seja análise dos pigmentos puros ou análise das misturas, o resultado poderá ser sempre uma mistura de pigmentos. Sempre que o resultado é uma mistura, o utilizador pode visualizar a percentagem de cada um dos elementos presentes na mistura, sendo esta sempre uma percentagem

aproximada da realidade, algo que é impossível de identificar, mesmo em laboratório com os recursos mais sofisticados. No capítulo seguinte apresentam-se alguns dos resultados experimentais obtidos com o método de classificação relatado.

Visualização dos Resultados Tal como será referido no capítulo de conclusões e trabalho futuro, não foi ainda construído um método de visualização eficaz para esta análise por várias razões. Para testes utilizou-se a mesma lógica dos métodos anteriores, atribuir uma cor a cada um dos espectros presentes na base de dados de pigmentos puros e para as misturas procedeu-se ao cálculo da cor a partir do espectro da mistura seleccionada tal como foi efectuado no método de classificação não supervisionada.

3.5 Classificação de Pinturas Baseada em Diferentes Tipos de Análise

Importa referir que o classificador de pinturas baseado nas características das assinaturas surgiu no âmbito do projecto com o intuito de uma integração posterior com o classificador baseado nas características de pincelada de modo a construir um classificador combinado. Assim após a construção do classificador de assinaturas que revelou resultados significativos, procedeu-se à reconstrução do classificador com base nas características de pincelada e combinou-se a informação produzida por ambos os RLSC (ver figura 3.22). De realçar que a verificação de autenticidade e atribuição de autor de pinturas utilizando a análise de materiais deveria ter sido incluída no classificador mas no capítulo das conclusões serão enunciadas as razões que levaram a que tal não fosse concluído.

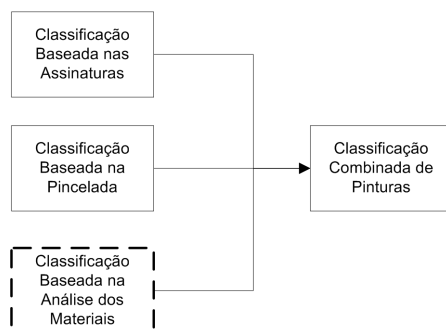


Figura 3.22: Diagrama de blocos do método de classificação combinada de pinturas conjugando vários tipos de informação.

Para converter a saída do classificador (RLSC) num valor de probabilidade, várias estratégias têm sido propostas em trabalhos anteriores. Em [Wah99] é proposto um método baseado em regressão logística para produzir valores probabilísticos a partir de valores de classificação obtidos por métodos baseados em *Kernel*. Contudo esta parece não ser a melhor alternativa, sendo proposto em [Pla99] um modelo paramétrico para estimar directamente o valor de probabilidade *a posteriori*. Em ambos os classificadores anteriores foi aplicada a segunda estratégia enunciada, respectivamente adaptada ao RLSC.

Consideram-se duas variáveis: w representando a classe positiva (por exemplo "Amadeo") e z representando a pertença à classe negativa (por exemplo "Não Amadeo"). A probabilidade de cada uma das variáveis anteriores dada uma determinada imagem representada por v_x (vector de características) pode ser obtida utilizando o resultado da classificação obtido por $f(v_x)$ e a função sigmóide de forma semelhante ao efectuado em [Pla99]. A saída da classe positiva (por exemplo "Amadeo") do classificador pode ser obtida,

$$p(w|v_x) = \frac{1}{1 + e^{-Af(v_x)+B}}, \quad (3.24)$$

enquanto que a saída negativa (por exemplo "Não Amadeo") é obtida,

$$p(z|v_x) = \frac{e^{-Af(v_x)+B}}{1 + e^{-Af(v_x)+B}}, \quad (3.25)$$

onde A e B são parâmetros para ajuste da função sigmóide, obtidos de forma empírica, embora possam ser estimados a partir de um conjunto de métodos existentes. A função $f(v_x)$ representa o discriminante do classificador RLS, sendo enunciada em 3.5.

Para combinar a informação produzida por ambos os classificadores, um modelo conjunto foi construído baseado apenas na probabilidade da classe positiva (por exemplo, "Amadeo") do classificador em ambos os casos. O modelo é obtido da seguinte forma,

$$A(v_x) = \alpha p(w_p|v_x) + (1 - \alpha) p(w_s|v_x), \quad (3.26)$$

em que α é um factor de peso associado a cada um dos classificadores, $p(w_s|v_x)$ é o valor de probabilidade da classe positiva para uma dada imagem utilizando o classificador de assinaturas, $p(w_p|v_x)$ é o valor de probabilidade da classe positiva para uma dada imagem utilizando o classificador de pincelada e por fim, $A(v_x)$ é a percentagem de pertença à classe positiva para uma dada imagem conjugando a informação pesada de ambos os classificadores. Para o classificador combinado construído atribuiu-se o mesmo peso a cada um dos classificadores individuais, $\alpha = 0.5$.

Para cada valor obtido de $A(v_x)$, se $A(v_x) \leq 0.5$ então a pintura é classificada como negativa, ou seja, não pertencente ao pintor para qual o classificador foi treinado. Caso contrário, se $A(v_x) > 0.5$ então a pintura é considerada positiva (por exemplo pertencente à classe "Amadeo"). Este processo é repetido para cada imagem do conjunto de teste, obtendo-se então uma etiqueta para cada imagem baseada na conjugação da informação dos dois classificadores.

Este classificador pode ser utilizado tanto para verificação de autenticidade (entre um pintor e respectivas falsificações), como para classificação de atribuição de autor (entre um pintor e outros pintores) ou para ambas as classificações conjuntas (entre um pintor e não pertencendo ao pintor). Os resultados da aplicação do classificador construído serão relatados no próximo capítulo.

3.6 Ambiente de Desenvolvimento

Esta secção descreve as ferramentas utilizadas na construção dos métodos relatados anteriormente. Todo o desenvolvimento foi efectuado no ambiente MATLAB 7.12.0 (R2011a), sendo a última versão estável aquando do início do desenvolvimento. A máquina utilizada no desenvolvimento possuía o Windows 7 32 bits instalado, Intel Core 2 Duo 2.5 GHz e 3GB de memória RAM instalada. A linguagem de programação utilizada foi o MATLAB. Devido à complexidade de alguns algoritmos no processamento de dados foi necessário recorrer a uma máquina com um sistema 64 bits pois a 32 bits não era possível endereçar todos os dados em memória.

3.7 Discussão

Neste capítulo foram descritas as principais ferramentas construídas, embora se tenham efectuado alguns testes com alguns métodos que acabaram por ser descartados face à sua ineficácia. Relatou-se a construção de um conjunto de métodos, não existindo qualquer interface gráfica para interacção construída sobre os mesmos. A sua utilização num futuro próximo incide principalmente sobre os resultados produzidos, podendo o utilizador adaptar alguns dos métodos de acordo com as suas necessidades. Este é um aspecto bastante interessante a referir, que de acordo com uma determinada análise, o utilizador pode aproveitar a flexibilidade dos métodos e direccionar a análise num determinado sentido de acordo com o problema que pretende resolver.

4

Resultados

Neste capítulo são analisados alguns resultados experimentais resultantes da aplicação dos métodos desenvolvidos ao longo da dissertação. De realçar a presença de vários peritos em análise de obras de arte envolvidos no projecto para auxílio da construção dos métodos e validação dos mesmos. As técnicas construídas incidem sobre o pintor de Amadeo de Souza-Cardoso, contudo são independentes do pintor em estudo.

4.1 Resultados da Análise de Assinaturas

Esta secção apresenta os resultados obtidos na análise de assinaturas, seja em termos de identificação de períodos de assinatura ou verificação de autenticidade e atribuição de autor.

4.1.1 Identificação de Períodos Artísticos

O objectivo deste método é descobrir períodos na vida do pintor através das características presentes nas assinaturas. Após o tratamento dos dados, existiam 114 assinaturas pertencentes ao pintor. O método teve como entrada as imagens das assinaturas e com o auxílio de um perito, definiu-se que no primeiro nível o mais adequado seria apenas dois *clusters*. Isto porque ao longo da sua vida Amadeo de Souza-Cardoso teve principalmente dois tipos de assinatura, um tipo em que a assinatura era feita com uma estampa sendo designado pelos peritos por "pouchoir" (1914-1917) e no segundo tipo designado por "mão livre", o autor autenticava as pinturas com uma representação do seu nome, variando de caso para caso (1908-1914).

A figura 4.1 mostra as imagens que após o agrupamento ficaram inseridas no primeiro



Figura 4.1: Imagens presentes no *cluster* 1

cluster, estando estas ordenadas por ordem crescente de distância ao centróide do mesmo. Após alguma análise com a ajuda de um perito em história de arte verificam-se dois aspectos. Neste *cluster* não existe nenhuma assinatura correspondente ao "pouchoir" e pela correspondência com as datas de cada uma das assinaturas conclui-se que este *cluster* engloba as assinaturas do pintor no período de 1908-1914, uma das fases da sua carreira artística.

A figura 4.2 mostra as imagens inseridas no segundo *cluster* ordenadas de acordo com a menor distância ao centróide, sendo este dominado pelas assinaturas do "pouchoir". Além do "pouchoir", existe também um conjunto de assinaturas que em termos de características locais possuem características muito semelhantes ao "pouchoir", mas a nível global estas diferem significativamente das restantes imagens contidas no *cluster*. Esperava-se que este *cluster* contivesse apenas as assinaturas do estilo "pouchoir" pois diferem significativamente de todas as outras assinaturas do pintor. Salvo as exceções referidas, este *cluster* contém principalmente as assinaturas construídas pelo artista no período 1914-1917, sendo esse período marcado pela utilização do "pouchoir".

Apesar de se esperar obter apenas um *cluster* com a assinatura "pouchoir", isso não foi possível. Foi feita a tentativa de identificar sub-períodos dentro dos dois super-períodos encontrados. Tanto a nível documental como a nível visual não foi possível tirar conclusões da identificação de sub-períodos. A nível visual, independentemente das configurações do algoritmo testadas (principalmente fazendo variar o número de palavras visuais e o número de *clusters*), não foi possível separar as assinaturas que não se assemelhavam ao "pouchoir" num único sub-período. Isto sugere algo já enunciado, o facto destas possuírem características locais semelhantes ao estilo "pouchoir". Ao nível do outro *cluster* (figura 4.1), a identificação de sub-períodos também não revelou padrões visuais significativos, devendo-se principalmente ao facto das assinaturas apresentarem características

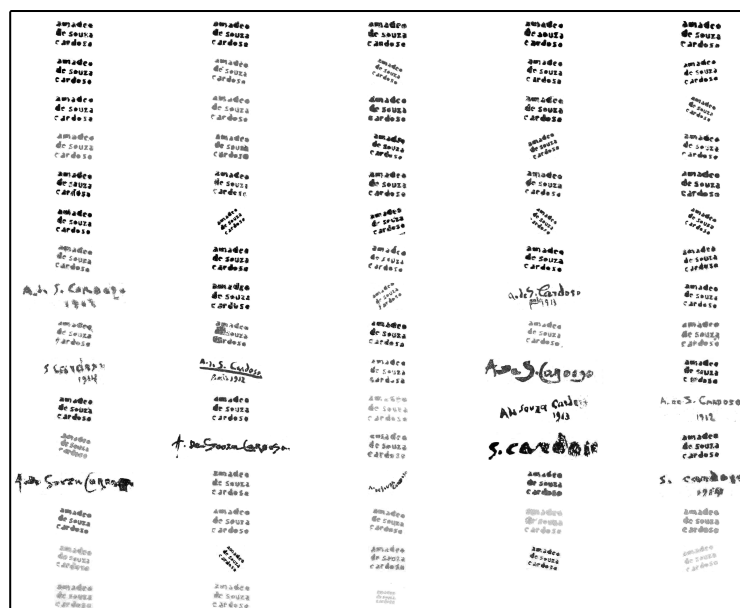


Figura 4.2: Imagens presentes no *cluster* 2

bastante distintas entre elas. Face à ausência de um estudo sobre as assinaturas deste pintor, não foi possível retirar conclusões mais robustas.

Para destacar a eficácia do método construído, foi experimentado também o *clustering* utilizando os filtros de Gabor mas estes não revelaram qualquer padrão ao longo de um mesmo *cluster*, não separando mesmo as assinaturas "pouchoir" num único *cluster*. Não foi assim possível obter uma separação evidente entre as assinaturas do pintor com este tipo de características.

O método construído carece sem dúvida de validação com outro conjunto de dados. Em determinado ponto do projecto procuraram-se assinaturas de outro pintor para validar este método de modo a permitir a construção de uma ferramenta com maior grau de fiabilidade. Contudo até à data da escrita do documento da dissertação, não foi possível obter quaisquer dados.

4.1.2 Classificação de Pinturas Baseada em Características das Assinaturas

A classificação de assinaturas assentou em 3 classificadores construídos sobre o mesmo método, apenas com a diferença dos dados utilizados em cada um dos casos. O primeiro efectuava apenas a atribuição do autor de assinatura, efectuando a separação entre as classes "Amadeo" e "Outros Artistas"; o segundo efectuava a verificação de autenticidade, entre as classes "Amadeo" e "Falsificação"; e por fim, o último classificador juntava os dois conceitos anteriores, diferenciando entre "Amadeo" e "Não Amadeo", englobando esta última classe falsificações ou assinaturas pertencentes a outros pintores. Para demonstrar a validade do método construído face a técnicas já existente no estado da arte, efectuou-se o mesmo tipo de classificação mas utilizando também os filtros de Gabor para extracção de características.

Para testar de facto a robustez dos classificadores construídos procedeu-se a uma aproximação do método *10-fold cross validation*. Em vez dos classificadores serem testados apenas uma vez, procedeu-se à validação num conjunto de 10 iterações em que em cada iteração, um conjunto de imagens é seleccionado aleatoriamente para o conjunto de treino, enquanto as restantes formam o conjunto de teste. A precisão global do classificador é obtida através da média da precisão obtida em cada uma das 10 iterações. A classificação utilizando os filtros de Gabor não foi efectuada de forma aleatória. De forma a ser comparável de forma equilibrada, a classificação utilizando os filtros de Gabor utilizava os mesmos conjuntos de treino e teste que o método construído para as 10 iterações.

As tabelas para avaliação de cada um dos classificadores contêm um conjunto de medidas para a precisão do classificador. Descreve-se de seguida como se procedeu ao cálculo de cada uma dessas medidas.

A medida Positivos Correctos ou acerto de imagens positivas reflecte a percentagem de imagens positivas classificadas de forma correcta,

$$P_{PositivosCorrectos} = \frac{N_{pc}}{T_p} \times 100, \quad (4.1)$$

onde N_{pc} é o número de imagens positivas classificadas de forma correcta, ou seja, classificadas como positivas; T_p é o número total de imagens positivas no teste; e $P_{PositivosCorrectos}$ é a percentagem de *hits* nas imagens positivas do teste. A percentagem de falsos negativos é dada pela percentagem de imagens positivas classificadas de forma errada como negativas, podendo ser obtida da seguinte forma,

$$P_{FalsosNegativos} = 100 - P_{PositivosCorrectos}, \quad (4.2)$$

A medida Falsos Positivos traduz-se na percentagem de imagens negativas classificadas de forma errada como sendo positivas,

$$P_{FalsosPositivos} = \frac{N_{ne}}{T_n} \times 100, \quad (4.3)$$

onde N_{ne} é o número de imagens negativas classificadas de forma errada, ou seja, classificadas como positivas; T_n é o número total de imagens negativas no teste; e $P_{FalsosPositivos}$ é a percentagem de falsos positivos nas imagens do conjunto de teste.

Por fim, a *Accuracy* ou desempenho geral do classificador é calculada de acordo com o número de imagens correctamente classificadas no conjunto de teste. Esta pode ser obtida da seguinte forma:

$$P_{Accuracy} = \frac{N_{pc} + N_{nc}}{T_p + T_n} \times 100, \quad (4.4)$$

onde N_{pc} é o número de imagens positivas classificadas de forma correcta, ou seja, classificadas como positivas; N_{nc} é o número de imagens negativas classificadas de forma

correcta, ou seja, classificadas como negativas; T_p é o número total de imagens positivas no conjunto de teste e T_n é o número total de imagens negativas no conjunto de teste.

4.1.2.1 Atribuição de Assinaturas

Este primeiro classificador foi construído com o intuito de resolver o problema de atribuição do autor de uma pintura baseando-se na assinatura, ou seja, se pertence a Amadeo ou a outro artista. Como classe positiva do classificador definiu-se a classe "Amadeo" e como classe negativa "Outros Pintores". A figura 4.3 contém alguns exemplos das imagens utilizadas em cada uma das classes.

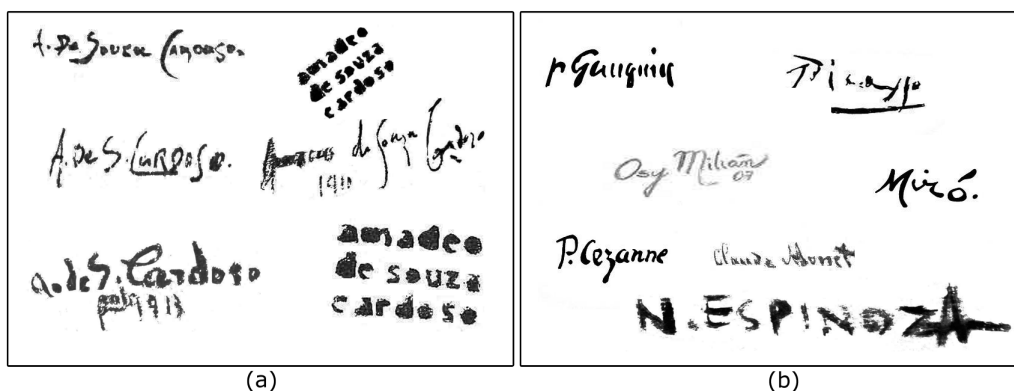


Figura 4.3: Exemplos de imagens de entrada no classificador (a) Classe "Amadeo". (b) Classe "Outros Artistas".

A tabela 4.1 mostra a distribuição dos dados para efeitos de classificação, havendo 114 imagens digitais de assinaturas de Amadeo e 67 de outros pintores, sendo cada um dos conjuntos repartido pelo treino e pelo teste.

(#)	Treino	Teste	Total
Amadeo	60	54	114
Outros Pintores	34	33	67

Tabela 4.1: Distribuição dos dados para atribuição de assinaturas.

De acordo com a distribuição dos dados anteriores, obtiveram-se os resultados descritos na tabela 4.2.

(%)	Positivos Correctos	Falsos Positivos	Falsos Negativos	Accuracy
SIFT	99.3	12.4	0.7	94.8
Gabor	91.4	20.7	8.6	86.9

Tabela 4.2: Resultados obtidos na classificação de atribuição de assinaturas.

Como é possível constatar pelos resultados observados, o classificador construído baseado nas características SIFT foi mais eficaz revelando uma taxa global de desempenho mais elevado do que com os filtros de Gabor. Tanto a nível dos falsos positivos como

dos falsos negativos, existe uma diferença percentual de aproximadamente 8%, sendo o classificador com as características SIFT o mais eficiente ao nível da classificação.

4.1.2.2 Autenticação de Assinaturas

Ao nível do estado da arte, os problemas mais difíceis de resolver ao nível da verificação de assinaturas prendem-se com a autenticação, ou seja, classificar assinaturas como autênticas ou falsificações construídas a partir das autênticas. Este classificador pretende resolver isso mesmo, isto é, diferenciar entre a classe positiva "Amadeo" e a classe negativa "Falsificações" de assinaturas do pintor. Neste caso, o classificador encontra-se desequilibrado pois a quantidade de assinaturas da classe negativa é bastante reduzida. Foram construídas algumas assinaturas da classe negativa em laboratório a partir das autênticas mas o número continua a ser insuficiente para construir um classificador totalmente robusto. A figura 4.4 contém alguns exemplos das imagens utilizadas em cada uma das classes.

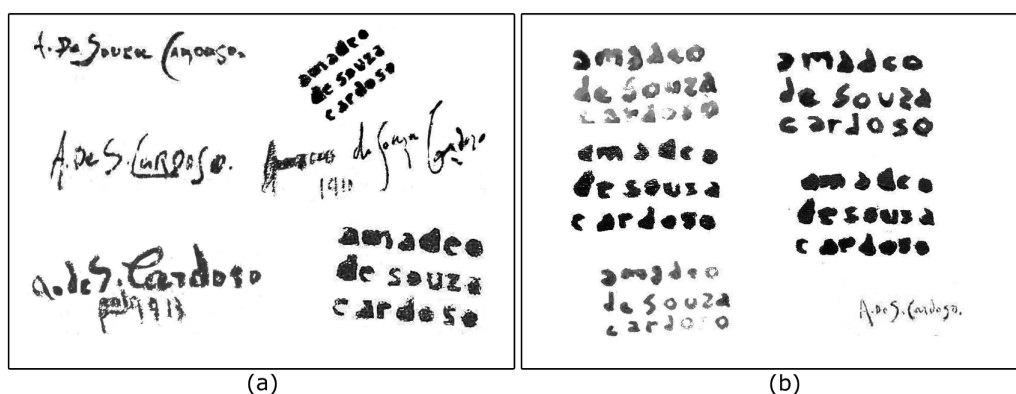


Figura 4.4: Exemplos de imagens de entrada no classificador (a) Classe "Amadeo". (b) Classe "Falsificações".

A tabela 4.3 mostra a distribuição dos dados para efeitos de classificação, havendo 114 imagens digitais de assinaturas de Amadeo e 10 de assinaturas falsificadas, sendo cada um dos conjuntos repartido pelo treino e pelo teste.

(#)	Treino	Teste	Total
Amadeo	57	57	114
Falsificações	5	5	10

Tabela 4.3: Distribuição dos dados para autenticidade de assinaturas.

De acordo com a distribuição dos dados anteriores, obtiveram-se os resultados descritos na tabela 4.4.

A análise dos resultados sugere algo que havia já sido referido, o desequilíbrio do classificador. Isto pois os dados positivos são suficientes para treinar a classe positiva, revelando-se uma taxa de acerto de 100% enquanto que a taxa de acerto para a classe

(%)	Positivos Correctos	Falsos Positivos	Falsos Negativos	Accuracy
SIFT	100.0	27.8	0.0	97.7
Gabor	93.5	42.0	6.5	90.6

Tabela 4.4: Resultados obtidos na verificação de autenticidade de assinaturas.

negativa não vai além de 72.2%, revelando uma taxa alta de falsos positivos. Face ao conjunto de teste reduzido, o classificador revela que em média 2 em 5 assinaturas negativas são classificadas de forma errada como positivas. A escassez de dados foi um grande obstáculo neste caso.

4.1.2.3 Atribuição e Autenticação de Assinaturas

Ambos os classificadores anteriores resolviam problemas distintos, sendo significativamente precisos a fazer uma classificação correcta entre as duas classes. Todavia, optou-se por juntar ambos os classificadores num único, tentando resolver ambos os problemas numa única ferramenta. Assim qualquer assinatura, seja pertencente a Amadeo, falsificação ou pertencente a outro pintor pode ser submetida ao classificador de modo a perceber a qual das classes pertence. A classe positiva designa-se a classe "Amadeo" e a classe negativa é designada "Não Amadeo", contemplando qualquer tipo de assinatura que não tenha sido feita pela mão do artista. A figura 4.5 contém alguns exemplos das imagens utilizadas em cada uma das classes.



Figura 4.5: Exemplos de imagens de entrada no classificador (a) Classe "Amadeo". (b) Classe "Não Amadeo".

A tabela 4.5 mostra a distribuição dos dados para efeitos de classificação, havendo 114 imagens digitais de assinaturas de Amadeo e 77 de outros pintores e falsificações, sendo cada um dos conjuntos repartido pelo treino e pelo teste.

De acordo com a distribuição dos dados anteriores, obtiveram-se os resultados descritos na tabela 4.6.

Novamente se verifica pelos dados a eficácia significativa do classificador construído com as características SIFT face ao classificador com os filtros de Gabor. À semelhança

(#)	Treino	Teste	Total
Amadeo	60	54	114
Não Amadeo	39	38	77

Tabela 4.5: Distribuição dos dados para atribuição e autenticidade de assinaturas.

(%)	Positivos Correctos	Falsos Positivos	Falsos Negativos	Accuracy
SIFT	98.3	11.9	1.7	94.1
Gabor	88.7	18.2	11.3	85.8

Tabela 4.6: Resultados obtidos na classificação de atribuição e autenticidade de assinaturas.

dos casos anteriores, também neste caso é possível verificar uma taxa de falsos negativos relativamente baixa, o que indica que o classificador consegue distinguir bem as imagens positivas.

4.2 Classificação de Pinturas Baseada em Características de Pincelada

Como já havia sido referido no capítulo anterior, já tinha sido construído no âmbito do projecto um classificador para verificação da atribuição de pinturas baseada em características de pincelada. Apesar dos resultados positivos já relatados no estado da arte [MJC⁺12], foi necessário voltar a refazer o classificador com vista à construção de um classificador combinado. No trabalho anterior tinham sido destacados alguns estudos sobre o número de imagens no conjunto de treino e teste, assim como o desempenho do classificador de acordo com o tipo de características extraídas.

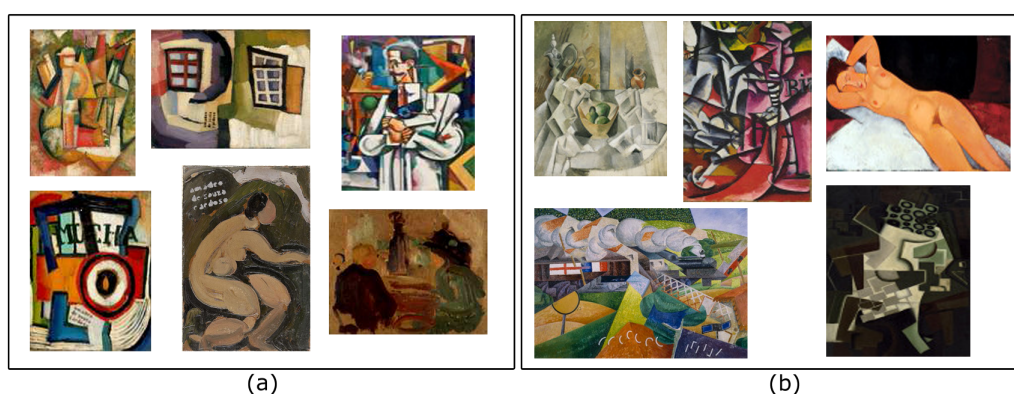


Figura 4.6: Exemplos de imagens de entrada no classificador (a) Classe "Amadeo", imagens cedidas pelo Centro de Arte Moderna. (b) Classe "Não Amadeo", imagens retiradas de <http://www.guggenheim.org>.

Na construção do classificador composto optou-se pela utilização da técnica SIFT, sendo este o tipo de características extraídas, tanto para a classificação com base nas

assinaturas como para a classificação com base na pincelada. A figura 4.6 contém alguns exemplos das imagens utilizadas em cada uma das classes.

Devido à escassez de falsificações para construir um classificador dedicado exclusivamente à autenticidade de pinturas com base nas características de pincelada optou-se por construir apenas um classificador para teste, resolvendo o problema da atribuição de autor e autenticidade simultaneamente. O classificador foi testado de forma semelhante ao classificador anterior com base nas assinaturas, sendo os resultados obtidos através da média de 10 iterações do classificador. Na construção do classificador anterior à dissertação, o conjunto de treino e teste eram fixos, havendo apenas uma iteração do algoritmo. Neste caso aplicou-se também a aproximação *10-fold cross validation* de modo a explorar a robustez da classificação independentemente dos conjuntos de treino escolhidos. A tabela 4.7 mostra a distribuição dos dados para efeitos de classificação, existindo 200 imagens de pinturas de Amadeo e 110 de outros pintores e falsificações, optando-se por uma junção de ambos os problemas.

(#)	Treino	Teste	Total
Amadeo	125	75	200
Não Amadeo	60	50	110

Tabela 4.7: Distribuição dos dados para atribuição e autenticidade de pinturas com base nas características de pincelada.

De acordo com a distribuição dos dados anteriores, obtiveram-se os resultados descritos na tabela 4.8.

(%)	Positivos Correctos	Falsos Positivos	Falsos Negativos	<i>Accuracy</i>
SIFT	99.7	16.0	0.3	93.4
Gabor	84.9	52.2	15.1	70.0

Tabela 4.8: Resultados obtidos na classificação de pinturas com base nas características de pincelada.

Tal como já tinha sido descrito na publicação [MJC⁺12], a SIFT a permitir obter resultados bastante positivos, enquanto que os filtros de Gabor se revelaram menos eficientes. A construção deste classificador revelou também alguma ineficácia ao nível dos falsos positivos, principalmente devido à escassez de dados da classe negativa. Contudo a reconstrução deste classificador não vale apenas pelos resultados individuais pois já tinha sido comprovada a sua eficácia, vale sim pela integração num classificador combinado de pinturas, integrando vários tipos de informação.

4.3 Resultados do Mapeamento de Materiais

Esta secção descreve os resultados obtidos na aplicação dos métodos desenvolvidos no âmbito da análise de imagens hiperespectrais.

4.3.1 Mapeamento de Materiais através de Classificação Supervisionada

Este foi um dos primeiros métodos a ser construído para efectuar uma análise de regiões de cor semelhantes numa mesma pintura a partir da reflectância emitida pelos pigmentos. Por norma, a base de dados é construída individualmente para cada quadro, contudo devido à flexibilidade do método pode construir-se a base de dados através da amostragem de vários quadros, sendo essa uma base de dados comum a várias pinturas. Apresenta-se em seguida um caso de estudo sobre o qual se aplicou o método construído, sendo este processo auxiliado por um perito em análise de obras de arte.

4.3.1.1 Caso de estudo: *Mucha*

A pintura *Mucha* é uma pintura bastante conhecida de Amadeo de Souza-Cardoso e sobre a qual foi adquirida uma imagem hiperespectral. A base de dados de entrada para o classificador foi construída com a ajuda de um perito, tentando captar pontos específicos para cada um dos pigmentos existentes na pintura (ver figura 4.7). Para cada pigmento seleccionado, seja puro ou uma mistura, calculou-se o espectro médio ao longo de uma zona de 13x13 *pixels* definida em redor de cada ponto e atribuiu-se um valor de RGB a cada pigmento seleccionado de modo a construir um *output* visual.



Figura 4.7: Construção da base de dados a partir de amostras retiradas da imagem hiperespectral e respectiva paleta de cores atribuída a cada um dos pigmentos.

Depois de construída a base de dados procedeu-se à classificação individual de cada um dos *pixels* de acordo com a base de dados de referência. Utilizou-se o valor de 0.85 para a percentagem d_x referida no capítulo anterior onde o método é descrito.

O resultado final da classificação pode ser obtido de várias formas. Na figura 4.8 (a), o resultado consiste na visualização de todos os materiais da base de dados mapeados numa única imagem. Esta é uma das formas de visualização mais frequentes em alguns

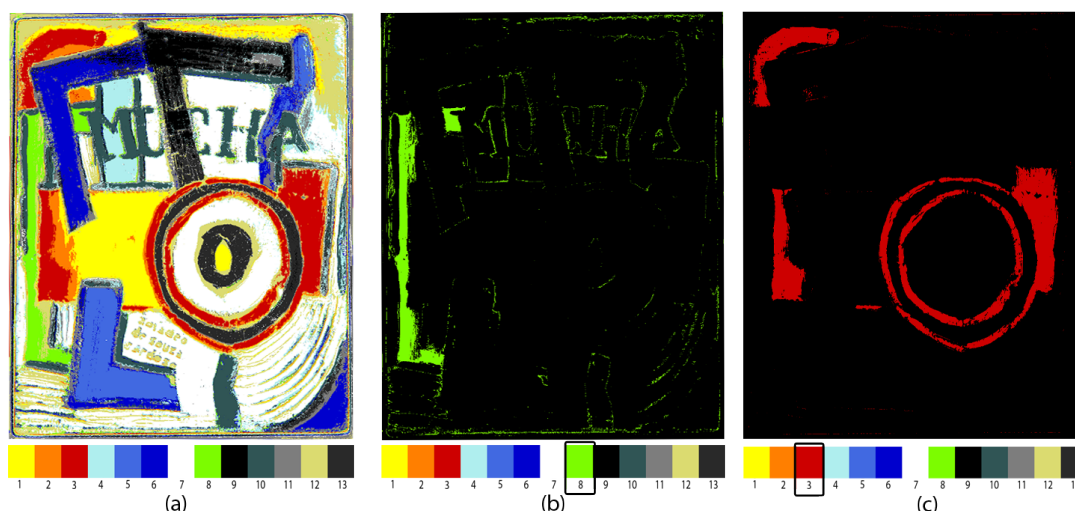


Figura 4.8: Resultado visual da classificação. (a) Todos os pigmentos mapeados. (b) Apenas mapeado o pigmento 8. (c) Apenas mapeado o pigmento 3.

dos trabalhos de investigação analisados, permitindo uma análise visual mais fácil para identificação de zonas com o mesmo material. Contudo duas variantes podem surgir de modo a analisar um pigmento mais específico. Em 4.8 (b) mapeou-se apenas o pigmento 8 da base de dados numa única imagem. Neste caso não se definiu nenhum *threshold* para identificar quando um determinado espectro pertence à base de dados ou não pertence, sendo classificado como "Unclassified". Por essa razão, aparecem muitos *pixels* isolados como atribuídos ao material mas que na realidade são *pixels* fronteira entre materiais e que realmente não pertencem ao pigmento 8. A figura 4.8 (c) pretende mostrar este factor. Neste caso mapeou-se o pigmento 3 da base de dados, mas utilizando um *threshold* sobre o ângulo espectral de 10° , isto pois o ângulo é calculado em graus. Todos os *pixels* cujo ângulo espectral mínimo entre o *pixel* de teste e a base de dados seja acima de 10 graus são classificados como "Unclassified", eliminando assim parte dos *pixels* fronteira.

A análise visual revelou um método bastante efectivo em distinguir os materiais existentes na base de dados, classificando de uma forma geral correctamente todos os materiais definidos. Todavia surgiu a necessidade de encontrar uma forma numérica de avaliar o classificador construído.

Validação do Classificador O classificador foi validado de forma semelhante aos trabalhos descritos em [PWLS07, Kuc07, MGW⁺05]. Recorrendo a ajuda de um perito neste tipo de análise, seleccionou-se uma região de interesse para validação de cada pigmento contido na base de dados. A região de tamanho 13×13 *pixels* definida para cada pigmento apenas poderia conter *pixels* que se sabia *a priori* pertencerem ao pigmento em questão, ou seja, este passo pretendia encontrar as "ground truth" para cada pigmento. Para o caso de estudo em análise seleccionaram-se as regiões que se podem visualizar na figura 4.9.

Para avaliar o desempenho do classificador, analisou-se em primeiro lugar para cada

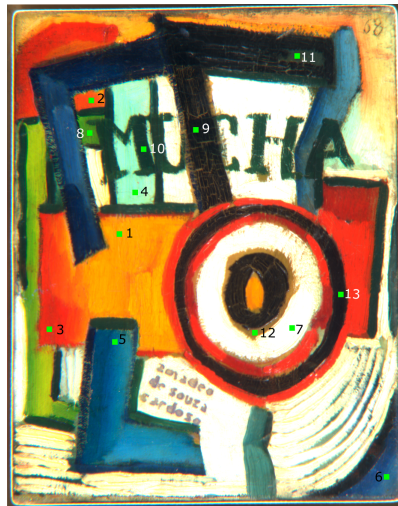


Figura 4.9: Amostragem das regiões de interesse para validação do classificador

pigmento o número de *pixels* classificados correctamente, avaliando-se depois o desempenho geral do classificador tendo em conta todas as regiões de interesse seleccionadas. Além desta medida de desempenho do classificador (*Accuracy*), calculou-se também o coeficiente kappa. Este coeficiente é bastante utilizado ao nível do estado da arte para comparar o desempenho de vários classificadores e para quantificar de forma sucinta a matriz de confusão produzida por cada classificador (por exemplo em [HR04, 22604, PWLS07]). Este toma sempre valores menores ou iguais a 1, sendo que o valor 1 implica uma classificação perfeita ao longo de todas as classes enquanto que valores superiores a 0.8 são considerados muito bons para uma classificação com várias classes. Surge neste trabalho com o intuito de tornar a validação do classificador mais robusta. O coeficiente kappa pode ser obtido através de,

$$k = \frac{C \sum_{i=1}^n d_{i,i} - \sum_{i=1}^n d_{i,\Sigma} d_{\Sigma,i}}{C^2 - \sum_{i=1}^n d_{i,\Sigma} d_{\Sigma,i}} \quad (4.5)$$

onde C é o número de regiões de interesse presentes na classificação, $\sum_{i=1}^n d_{i,i}$ é o somatório das diagonais da matriz de confusão, $d_{i,\Sigma}$ é o número de *pixels* em cada região de interesse e $d_{\Sigma,i}$ é o número de *pixels* classificados como pertencentes a cada uma das regiões de interesse.

Para uma análise mais concisa, apresenta-se na tabela 4.9 os resultados obtidos na avaliação deste classificador, sendo a matriz de confusão resumida pelo cálculo do coeficiente kappa.

Segundo o método de avaliação definido, o classificador revelou uma classificação eficaz com um desempenho geral de 98% para as regiões de interesse seleccionadas. De realçar que as regiões de interesse para avaliação foram escolhidas ao mesmo tempo que se definiu a base de dados de referência, numa fase anterior à classificação. Inúmeras localizações para as diferentes regiões de interesse poderiam ter sido escolhidas, contudo para avaliação do classificador optou-se pela definição das regiões apresentadas.

Região Nº	1	2	3	4	5	6	7	8	9	10	11	12	13	Total
PC (#)	169	168	169	169	169	169	169	169	168	169	153	169	144	2154
T (#)	169	169	169	169	169	169	169	169	169	169	169	169	169	2197
A (%)	100	99.4	100	100	100	100	100	100	99.4	100	90.0	100	85.2	98.0

Tabela 4.9: Resultados obtidos na classificação supervisionada com base de dados interna de acordo com as áreas definidas na figura 4.9. **PC** - Número de *pixels* classificados correctamente. **T** - Número total de *pixels*. **A** - *Accuracy* da classificação.

O coeficiente kappa para a classificação em questão foi de 0.925, um valor consideravelmente elevado, o que face ao que foi referido anteriormente mostra uma classificação bastante precisa. Isto é possível afirmar pelo facto de se ter obtido um valor bastante próximo de 1, tendo em conta uma validação com 13 classes.

Este método pode assim auxiliar os historiadores, identificando grupos de *pixels* semelhantes numa mesma pintura com a definição de um ou mais espectros de referência.

4.3.2 Mapeamento de Materiais através de Classificação Não-Supervisionada

Este método é mais automático do que o método anterior, não requerendo a definição de uma base de dados de referência, sendo esse processo feito automaticamente. Este permite, à semelhança do método anterior, agrupar *pixels* pertencentes ao mesmo material, facilitando a análise visual e a identificação de zonas com o mesmo material. Considerando novamente como caso de estudo a pintura Mucha de Amadeo de Souza-Cardoso, apresentam-se os resultados da classificação automática utilizando este método.

4.3.2.1 Caso de estudo: *Mucha*

Tentando analisar o comportamento do método, optou-se por definir um número de *clusters* igual ao número de pigmentos amostrados para a classificação supervisionada, no caso em questão, 13 *clusters*. À semelhança da classificação supervisionada, neste caso também é possível ver todos os *clusters* mapeados numa única imagem (figura 4.10, (a)) ou apenas um *cluster* mapeado de modo a analisar as zonas pertencentes a esse *cluster* (figura 4.10, (a) e (b)).

A análise dos resultados indica que o método é bastante sensível a pequenas diferenças no espectro de um mesmo material, muitas vezes existindo dois *clusters* para modelar esse material. Neste caso é também possível verificar a inexistência de qualquer pigmento preto, por ser em termos espectrais, bastante semelhante ao pigmento branco, tendo sido incorporado num dos restantes *clusters* obtidos. Devido à incerteza associada ao que irá ser mapeado em cada um dos *clusters*, não é possível construir um método de avaliação para o classificador, sendo esta unicamente visual.

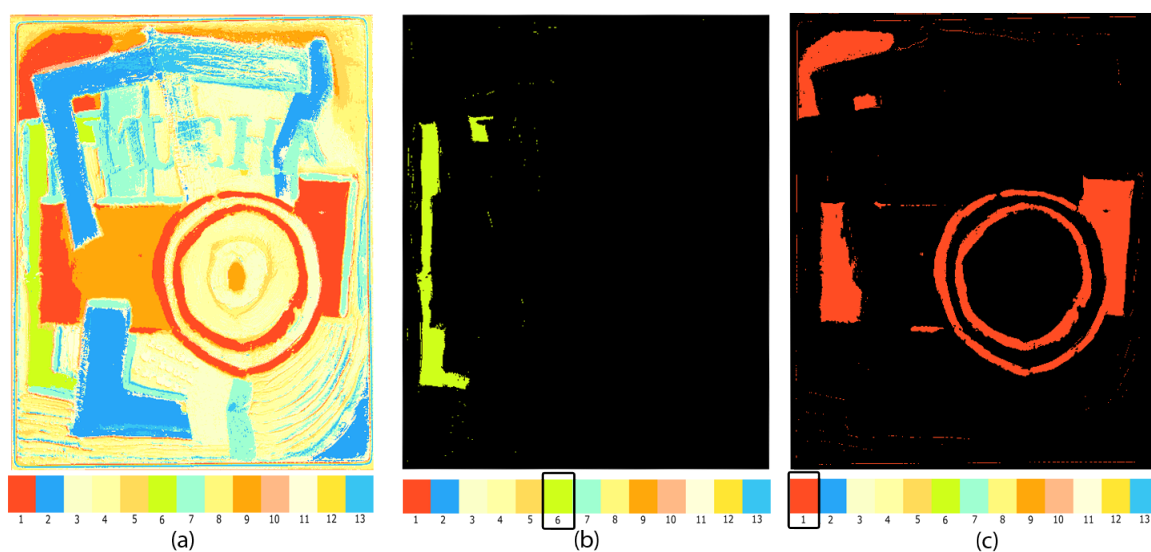


Figura 4.10: Resultado visual da classificação. (a) Todos os *clusters* mapeados. (b) Apenas mapeado o *cluster* 6. (c) Apenas mapeado o *cluster* 1.

4.3.3 Resultados do Mapeamento de Materiais através de Classificação com Base de Dados Externa

A classificação com a base de dados externa conduziu a dois tipos de análise, semelhantes na forma como a base de dados é adquirida e diferindo no processo de classificação. Apresentam-se em seguida os dois métodos construídos onde se aplicou este tipo de informação, sendo o primeiro aplicado a uma pintura falsificada do pintor, enquanto o segundo se aplica a algumas pinturas autênticas do pintor.

4.3.3.1 Mapeamento de Materiais através de Informação FORS

Caso de estudo: Falsificação Das várias imagens hiperespectrais existentes, duas delas pertenciam a falsificações de pinturas do pintor Amadeo de Souza-Cardoso. Para ambas as pinturas existia uma análise FORS captada em várias zonas do quadro. A figura 4.11 mostra os pontos onde foi efectuada a análise FORS para a falsificação considerada neste caso de estudo.

Com a ajuda de um perito neste tipo de análises identificaram-se os pontos existentes para cada um dos pigmentos respectivos. Para os pigmentos com vários pontos associados, calculou-se o espectro médio dos vários pontos. Nos restantes casos em que existia apenas um ponto para um dado pigmento, o espectro de referência desse pigmento era igual ao espectro do ponto.

Este método apenas difere do método de classificação supervisionada relatado em 3.4.2 na definição da base de dados de espectros de referência, sendo o restante processo efectuado da mesma forma, quer a nível de classificação como a nível de visualização dos resultados. Proceda-se então à atribuição de um valor RGB a cada um dos espectros de

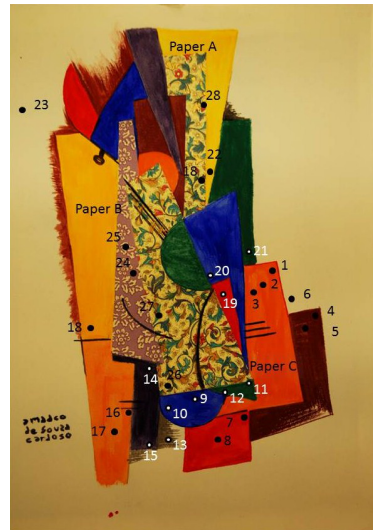


Figura 4.11: Pontos da pintura onde foi efectuada a análise FORS

referência na base de dados e procede-se à classificação através do método que conjuga o SAM com a distância Euclidiana. O resultado da classificação pode ser visto na figura 4.12.

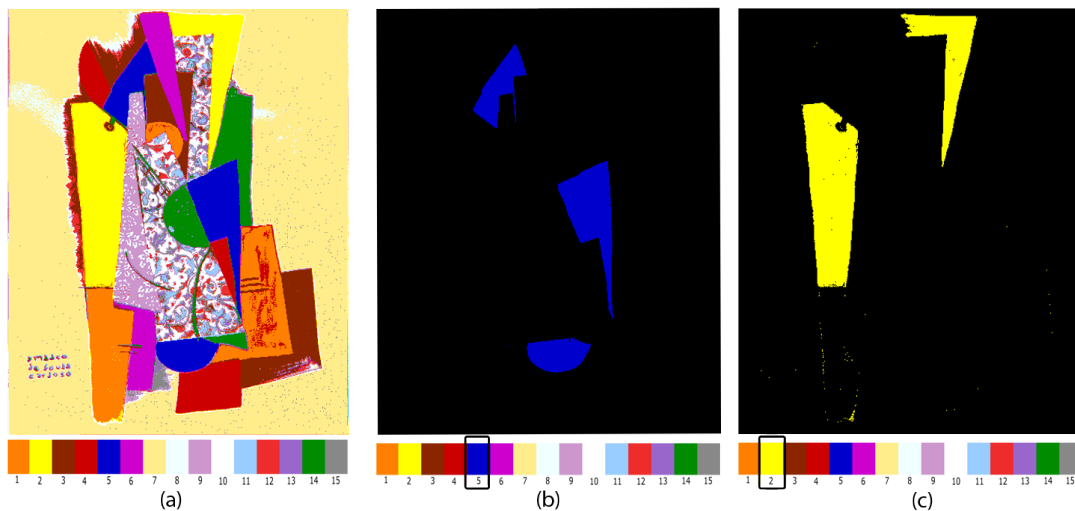


Figura 4.12: Resultado visual da classificação. (a) Todos os pigmentos de referência mapeados. (b) Apenas mapeado o pigmento 5. (c) Apenas mapeado o pigmento 2.

À semelhança de alguns dos testes anteriores, o resultado final da classificação pode ser obtido de várias formas. Na figura 4.12 (a) são mapeados simultaneamente todos os pigmentos contidos na base de dados de referência. Na figura 4.12 (b), é apenas mapeado o pigmento número 5 da base dados numa opção de visualização de um único pigmento. Por fim, na figura 4.12 (c) é mapeado o pigmento número 2 com a definição de um threshold de 7° sobre o ângulo espectral de modo a eliminar *pixels* fronteira que tenham sido classificados de forma errada como pertencentes a este pigmento. Uma

primeira análise visual revela que a informação produzida pela câmara hiperespectral é comparável com a informação adquirida com o FORS.

De modo a avaliar os resultados obtidos no mapeamento das várias regiões de cor, seguiu-se o processo de validação anteriormente utilizado para a classificação supervisionada, definindo-se uma região de interesse para cada um dos pigmentos presentes na base de dados. Para avaliar o desempenho do classificador analisou-se em primeiro lugar para cada pigmento o número de *pixels* classificados correctamente, avaliando-se depois o desempenho geral do classificador tendo em conta todas as regiões de interesse seleccionadas. De realçar que para a pintura em questão não foi possível avaliar o desempenho da classificação para alguns dos pigmentos da base de dados visto não ser possível definir uma área com 13×13 *pixels* que englobasse exclusivamente o respectivo pigmento na pintura. Os resultados da avaliação encontram-se descritos na tabela 4.10.

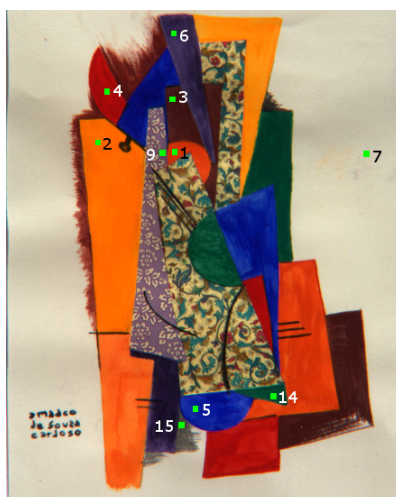


Figura 4.13: Amostragem das regiões de interesse para validação do classificador

Região N ^o	1	2	3	4	5	6	7	9	14	15	Total
PC (#)	150	169	169	169	169	169	169	169	169	64	1566
T (#)	169	169	169	169	169	169	169	169	169	169	1690
A (%)	88.8	100	100	100	100	100	100	100	100	37.9	92.7

Tabela 4.10: Resultados obtidos na classificação com recurso a base de dados externa. **PC** - Número de *pixels* classificados correctamente. **T** - Número total de *pixels*. **A** - *Accuracy* da classificação.

Os resultados presentes na tabela permitem ter uma ideia geral do desempenho alcançado pelo classificador. Este tipo de classificação revela que os sinais obtidos pelas duas técnicas (câmara hiperespectral e FORS) são aparentemente equivalentes e que podem ser integrados de modo a construir métodos mais completos e robustos. Para este classificador, além do desempenho geral calculou-se também o coeficiente kappa, de modo a ter uma medida complementar de desempenho de acordo com o número de classes e amostras utilizadas. Neste caso de estudo, o valor deste coeficiente foi de 0.918, o que

face aos dados anteriores revela uma eficácia de classificação elevada, pelo facto de ser um valor acima de 0.8, um valor considerado muito bom.

4.3.3.2 Mapeamento de Materiais com Análise de Misturas e Validação de Autenticidade

Para demonstrar alguns dos resultados obtidos com este método apresenta-se em seguida a classificação efectuada sobre as algumas pinturas de Amadeo para as quais tinham sido adquiridas imagens hiperespectrais. Nas figuras 4.14, 4.15 e 4.16 podem visualizar-se as imagens RGB de três das pinturas em análise, assim como alguns dos resultados experimentais obtidos para esses três casos. Utilizou-se o valor de 15 pontos para definir o intervalo de tolerância de cor CIELab para quando duas cores são semelhantes de modo a identificar a amostra EDXRF respectiva para cada *pixel*. Em termos de condição de identificação dos pigmentos puros utilizou-se o valor de 0.75 para a percentagem da distância entre os dois ângulos mínimos. Estes dois valores foram gerais a todas as pinturas nas quais foi aplicado o algoritmo.

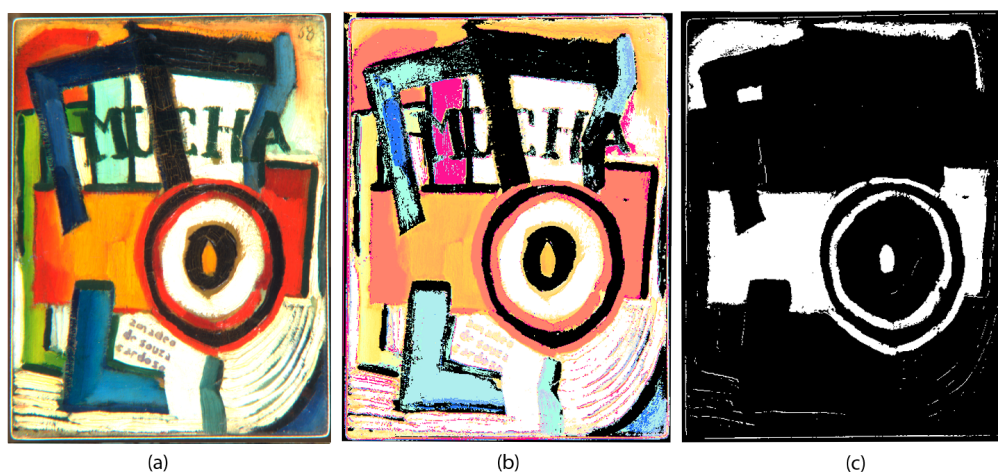


Figura 4.14: (a) Imagem RGB da pintura Mucha. (b) Resultado geral da classificação para teste. (c) Visualização individual de uma mistura de Red Vermillion e Yellow Cr, dois pigmentos puros

Análise dos Resultados Tal como foi referido no capítulo anterior, a visualização é o menos importante neste método, sendo o seu ponto forte a análise de diferentes zonas da pintura, incidindo essa análise na identificação de misturas e na validação de autenticidade de zonas da pintura. Aos historiadores e conservadores, interessa extrair também informação em termos percentuais: a percentagem de *pixels* classificados como "Zona Não Analisada" (P_{ZN}), ou seja, onde não houve uma análise EDXRF; a percentagem de *pixels* classificados como "Não Amadeo" (P_{NA}), ou seja, não pertencentes a Amadeo; a percentagem de *pixels* pertencentes a pigmentos puros (P_{PP}); e por fim a percentagem de *pixels* pertencentes a misturas de pigmentos (P_{MP}). Em todos os casos, o que se pretendia



Figura 4.15: **(a)** Imagem RGB da pintura Doutor. **(b)** Resultado geral da classificação para teste. **(c)** Visualização individual de uma mistura de Red Vermillion e Yellow Ochre, dois pigmentos puros

obter era o resultado da divisão entre o número de *pixels* classificados em cada caso e o número de *pixels* total da imagem hiperespectral.

As figuras 4.14 **(b)**, 4.15 **(b)** e 4.16 **(b)** mostram o resultado geral da classificação com todos os *pixels* mapeados para três dos casos de estudo analisados. As zonas estranhas podem identificar-se pela cor rosa em cada uma das pinturas. Contudo, a análise pode efectuar-se de outras formas, perspectivando já a construção de uma interface gráfica que englobe o resultado do método no futuro.

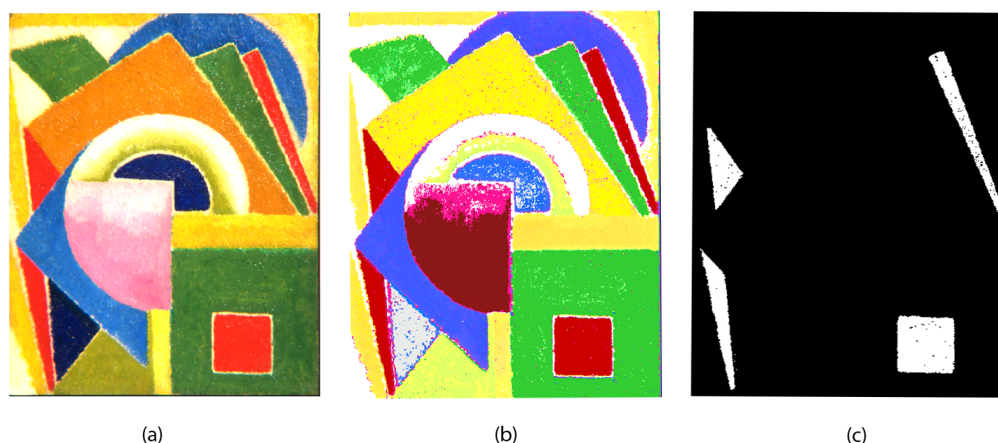


Figura 4.16: **(a)** Imagem RGB da pintura Geometrico. **(b)** Resultado geral da classificação para teste. **(c)** Visualização individual de um pigmento puro, o Red Vermillion

O utilizador pode desejar visualizar apenas as áreas com um determinado pigmento puro ou com uma mistura de pigmentos individualmente, surgindo mapeados apenas os *pixels* pertencentes a essa escolha. As figuras 4.14 **(c)** e 4.15 **(c)** mostram a identificação de

todos os *pixels* com uma dada mistura. No caso da pintura Mucha, trata-se de uma mistura de Red Vermillion com Yellow Cr, enquanto que na pintura Doutor se trata de uma mistura de pigmento Red Vermillion e Yellow Ochre, todos pigmentos existentes na base de dados de Amadeo. Esta identificação efectuada revelou-se substancialmente correcta pois na realidade a grande maioria das zonas identificadas correspondia uma mistura real destes dois pigmentos. Na figura 4.16 (c), pode visualizar-se todas as zonas identificadas como pertencentes ao pigmento puro Red Vermillion, sendo um identificação bastante eficaz.

Na pintura Doutor é de realçar, por exemplo, a identificação correcta de algumas misturas de Viridian com Yellow Ochre, assim como a identificação de alguns pigmentos puros como o caso do Esmerald (realçado por um verde mais carregado na figura 4.15 (b)) ou do Laranja de Cádmio, identificável pela cor laranja na figura 4.15 (b). No caso da pintura em questão é possível ainda realçar outro factor. A rosa estão identificados todos os *pixels* classificados como pertencentes à "Zona Não Analisada" verificando-se de facto várias pequenas zonas onde a análise elementar não foi efectivamente efectuada. O facto desta pintura apresentar uma enorme variabilidade de pigmentos e misturas torna a análise EDXRF mais restrita quando é efectuada num determinado ponto.

Outra das opções é o utilizador escolher um ponto para análise e o programa retornar quais os pigmentos ou misturas de pigmentos presentes naquele ponto, assim como a respectiva percentagem dos pigmentos presentes na mistura quando tal é possível. Por exemplo, tomando como referência a figura 4.15 (a), é facilmente identificável a zona imediatamente acima do ombro direito do doutor pintada com uma cor verde. Seleccionando um ponto dessa zona o programa retorna como resultado uma mistura de Yellow Ochre e Viridian, sendo esta mistura uma abordagem simplificada da mistura de pigmentos tal como já foi referido.

Em termos de percentagens requisitadas pelos historiadores, para os casos de estudo em que o método foi aplicado apresentam-se na tabela 4.11. O método foi aplicado a cinco pinturas autênticas, todavia para duas pinturas apresentam-se apenas as percentagens calculadas.

(%)	P_{ZN}	P_{NA}	P_{PP}	P_{MP}
Caso de Estudo: Mucha	4.8	0.0	51.4	43.8
Caso de Estudo: Doutor	11.6	0.0	51.5	36.9
Caso de Estudo: Guitarra	5.4	0.0	23.1	71.5
Caso de Estudo: Janelas	2.8	0.0	28.2	69.0
Caso de Estudo: Geometrico	3.7	0.0	31.4	64.9

Tabela 4.11: Resultados da classificação aplicada aos vários casos de estudo.

Analisando as percentagens obtidas para os cinco casos de estudo conclui-se que existe uma quantidade de pequenas zonas onde a análise EDXRF não foi efectuada de facto ou cujas amostras não englobam essas zonas em ambas as pinturas. Neste caso destaca-se a pintura Doutor onde a percentagem de zonas não analisadas foi maior,

devendo-se à sua enorme variabilidade de pigmentos e tonalidades.

Em termos de zonas "Não Amadeo" não foi identificada qualquer zona em qualquer uma das pinturas pois todos os elementos presentes nas amostras EDXRF pertenciam a Amadeo. Pode-se verificar que a condição implementada permite identificar zonas na pintura que não tenham sido feitas pela mão do pintor. Contudo a nível da solução não se pode afirmar que esta condição permita classificar pinturas em termos de autenticidade de modo a ser integrada no classificador combinado de pinturas. Talvez no futuro, através da definição de um *threshold* sobre a percentagem de zonas "Não Amadeo" numa pintura, seja possível construir um classificador de pinturas com base nesta condição. Esse classificador não englobou parte da solução proposta na dissertação pelo facto de existirem poucos dados para construir algo robusto e consistente, que pudesse ser suficientemente fiável. Assim, esta condição é apenas útil para identificar *pixels* contidos numa imagem hiperespectral que contenham materiais que não coincidam com a base de dados de Amadeo.

Em relação aos pigmentos puros, existem várias zonas identificadas em todas as pinturas, o que era de todo espectável de encontrar, destacando-se a identificação do Laranja de Cádmio e do Esmeralda na pintura Doutor. A pintura Geometrico contém várias zonas onde foram identificados pigmentos puros, como o caso do Esmerald, do Red Carmine e do Red Vermillion. De referir ainda a presença de uma quantidade significativa de pigmento branco em todos os casos de estudo, sendo essa identificação mais significativa no caso de estudo da pintura Mucha e da pintura Guitarra.

Por fim, é possível verificar a existência de grande parte das pinturas cobertas por misturas de pigmentos. No caso da pintura Guitarra, além da identificação de algumas zonas com o pigmento branco, todo o restante da pintura é coberta por misturas de pigmentos, o mesmo se passando praticamente com o caso de estudo das Janelas.

Porém, devido à diferença já referida entre os espectros dos pigmentos puros e os espectros contidos na imagem hiperespectral, algumas misturas não foram identificadas correctamente assim como alguns pigmentos puros que deveriam ter sido identificados e foram atribuídos a misturas de pigmentos. Este tipo de erros já era espectável, contudo quando uma zona é dominada por um pigmento na ordem dos 90% para 10% do outro pigmento presente na mistura, isto pode levar à conclusão que a presença do outro pigmento é muito reduzida ou quase inexistente. De referir ainda que os dados apresentados são dependentes do valor atribuído a cada um dos *thresholds* do algoritmo. Alterando qualquer um dos *thresholds* a percentagem obtida será certamente diferente em cada um dos casos. Os valores escolhidos foram os mesmos para todas as pinturas onde o método foi aplicado.

4.4 Classificação de Pinturas Baseada em Diferentes Tipos de Análise

Este classificador tinha como objectivo integrar dois tipos de análise sobre uma pintura com vista a verificação de autenticidade ou atribuição da mesma. Ambos os classificadores anteriores foram construídos de forma independente. Isto é, no caso da assinatura, existiam assinaturas no treino que não se conhecia a pintura a que pertenciam. Também no caso do classificador com base na pincelada, existia uma quantidade significativa de pinturas que não possuía assinatura. Face aos resultados obtidos com cada um dos classificadores procedeu-se à sua integração num único classificador.

Ao nível do treino, devido à quantidade algo reduzida de quadros com assinatura, assumiu-se que cada um dos classificadores poderia ser treinado de forma independente com os respectivos dados para cada um dos casos. Para ser possível integrar os dados num classificador conjunto, o conjunto de teste deverá ser o mesmo para cada um dos classificadores. Assim, invés do conjunto de teste ser escolhido aleatoriamente, um perito em arte seleccionou um conjunto de pinturas com assinatura para o conjunto de teste, sendo este fixo. Ao contrário dos classificadores anteriores, neste caso o teste do classificador consiste apenas numa iteração pois existe apenas um conjunto de teste fixo. Tendo em conta as configurações referidas construíram-se dois classificadores, diferindo, à semelhança dos casos anteriores, nos dados de entrada do algoritmo. Cada tipo de características tem um peso ao nível da classificação, todavia neste caso atribui-se o peso de 50% a cada um dos tipos de características, assim tanto a análise de assinatura como a análise da pincelada contam de igual modo para a classificação final.

4.4.1 Atribuição de Pinturas

A atribuição do autor de pinturas consiste no problema de classificação entre duas classes "Amadeo" e "Outros Pintores". Isto é, dada uma determinada pintura, o classificador deve identificar se pertence ao pintor em estudo, neste caso Amadeo, ou a um outro pintor. A tabela 4.12 mostra a distribuição dos dados para efeitos de classificação. O conjunto de treino no classificador de assinaturas é constituído por 111 imagens divididas em duas classes, enquanto que no caso do classificador de pincelada são 248 imagens de pinturas no total. Ao nível do teste, este é o mesmo para ambos os classificadores, sendo composto por um total de 67 imagens divididas em duas classes.

(#)	Treino Assinaturas	Treino Pincelada	Teste Conjunto
Amadeo	55	144	56
Outros Pintores	56	104	11

Tabela 4.12: Distribuição dos dados para atribuição de pinturas com base no classificador conjunto de pinturas.

De acordo com a distribuição dos dados anteriores, obtiveram-se os resultados descritos na tabela 4.13.

(%)	Assinatura	Pincelada	Classificador Combinado
Positivos Correctos	98.2	98.0	100
Falsos Positivos	0.0	9.0	0.0
Falsos Negativos	1.8	2.0	0.0
Accuracy	98.5	97.0	100

Tabela 4.13: Resultados obtidos na classificação de atribuição pinturas com base no classificador combinado utilizando a SIFT.

Os resultados revelam que ambos os classificadores se compensaram mutuamente, conseguindo obter o nível máximo de eficácia. De realçar que os classificadores não erraram ao classificar as mesmas imagens, daí o facto de a *accuracy* de nenhum deles ser 100%, mas depois a nível final ser de 100%, acabando cada um deles por corrigir o erro do outro. Para demonstrar a eficácia deste método face a outros já existentes apresenta-se em seguida os resultados obtidos com os filtros de Gabor para a mesma configuração de treino e teste 4.14.

(%)	Assinatura	Pincelada	Classificador Combinado
Positivos Correctos	87.5	83.9	89.3
Falsos Positivos	54.5	36.4	36.4
Falsos Negativos	12.5	16.1	10.7
Accuracy	80.6	80.6	85.1

Tabela 4.14: Resultados obtidos na classificação de atribuição pinturas com base no classificador combinado utilizando os filtros de Gabor.

Os resultados obtidos com os filtros de Gabor demonstram mais uma vez a menor eficácia em resolver o problema pretendido, sendo o desempenho geral do classificador combinado significativamente inferior que o classificador combinado utilizando as características SIFT.

4.4.2 Atribuição e Autenticação de Pinturas

Devido ao facto de só existirem cinco pinturas falsificadas com assinatura, não foi possível construir um classificador dedicado exclusivamente à autenticidade de pinturas. Assim, à semelhança da análise de assinaturas, construiu-se um classificador que resolve ambos os problemas, atribuição de autor e autenticidade de pinturas com base num classificador combinado. Ou seja, como classe positiva define-se a classe das pinturas pertencentes a "Amadeo" e como classe negativa define-se a classe de pinturas que não pertencem a Amadeo ("Não Amadeo"), podendo ser falsificações ou pertencentes a outro pintor. A tabela 4.15 mostra a distribuição dos dados para efeitos de classificação. O conjunto de treino no classificador de assinaturas é constituído por 116 imagens divididas

em duas classes, enquanto que no caso do classificador de pincelada são 256 imagens de pinturas no total. Ao nível do teste, este é o mesmo para ambos os classificadores, sendo composto por um total de 72 imagens divididas em duas classes.

(#)	Treino Assinaturas	Treino Pincelada	Teste Conjunto
Amadeo	55	144	56
Não Amadeo	61	112	16

Tabela 4.15: Distribuição dos dados para atribuição e autenticação de pinturas com base no classificador conjunto de pinturas.

De acordo com a distribuição dos dados anteriores, obtiveram-se os resultados descritos na tabela 4.16.

(%)	Assinatura	Pincelada	Classificador Combinado
Positivos Correctos	96.4	98.2	100
Falsos Positivos	0.0	6.3	0.0
Falsos Negativos	3.6	1.8	0.0
Accuracy	97.2	97.2	100

Tabela 4.16: Resultados obtidos na classificação de atribuição e autenticidade pinturas com base no classificador combinado utilizando a SIFT.

Tal como o caso anterior, também neste caso o classificador combinado permitiu obter a eficiência máxima possível em distinguir as assinaturas de cada uma das classes. A quantidade de dados de teste da classe negativa é reduzida pela dificuldade em encontrar pinturas de outros pintores com a respectiva assinatura. À semelhança dos casos anteriores, também neste caso se efectuou a mesma classificação mas utilizando os filtros de Gabor como características de entrada no classificador. Os resultados obtidos para este classificador encontram-se descritos na tabela 4.17, revelando novamente a eficácia inferior deste tipo de características face à utilização da SIFT.

(%)	Assinatura	Pincelada	Classificador Combinado
Positivos Hit	92.8	83.9	91.1
Falsos Positivos	56.3	31.3	43.7
Falsos Negativos	7.2	16.1	8.9
Accuracy	81.9	80.5	83.3

Tabela 4.17: Resultados obtidos na classificação de atribuição e autenticidade pinturas com base no classificador combinado utilizando os filtros de Gabor.

Face aos resultados obtidos e após a comparação com uma técnica com inúmeros resultados obtidos ao nível do estado da arte, pode-se afirmar que o classificador construído é robusto e pode ser submetido a casos de estudo com vista à obtenção de novas conclusões.

4.4.2.1 Casos de estudo

Construído o classificador anterior que resolve os problemas da atribuição e autenticidade de pinturas submeteu-se o classificador ao teste de três casos de estudo. Os casos de estudo eram constituídos por duas pinturas assinadas com uma assinatura semelhante ao "pouchoir" de Amadeo e uma terceira com uma assinatura com o nome do pintor. Ainda não existe um veredicto final quanto à autenticidade das mesmas, contudo a terceira das pinturas foi alvo de uma peritagem que revelou um estilo muito semelhante ao de Amadeo, podendo mesmo pertencer ao pintor.

Para analisar estes casos utilizou-se uma das percentagens produzidas pelo classificador, mais concretamente a percentagem de pertencer à classe positiva (classe "Amadeo"). De referir que abaixo ou igual a 50% a pintura em teste é considerada como pertencente à classe negativa, caso contrário é assinalada como pertencente à classe positiva. Considerando as mesmas condições do classificador anterior, obtiveram-se os resultados expressos na tabela 4.18.

(%)	Assinatura	Pincelada	Classificador Combinado
Caso de Estudo 1	99.3	99.9	99.6
Caso de Estudo 2	98.0	82.6	90.3
Caso de Estudo 3	71.0	99.8	85.9

Tabela 4.18: Resultados obtidos para os três casos de estudo submetidos ao classificador combinado de pinturas.

Os resultados do classificador combinado são inequívocos, todos os casos de estudo foram classificados como positivos com uma percentagem razoavelmente elevada, o que sugere que este não tem qualquer dúvidas em atribuir cada um dos casos à classe positiva, ou seja, "Amadeo". Tanto o classificador de assinaturas como o classificador de pincelada corroboram o resultado do classificador final sem qualquer dúvida. Apesar de no caso de estudo 3 uma peritagem ter concluído com forte probabilidade que a pintura poderá ser mesmo autêntica do pintor, o que revela que o classificador poderá estar efectivamente correcto, para os restantes dois casos não há ainda um estudo efectivo e conclusivo.

4.5 Discussão

Neste capítulo foram apresentados os vários testes aos métodos construídos, permitindo efectuar vários tipos de análise. No geral pode-se afirmar que os métodos apresentam um desempenho geral bastante significativo, ficando alguns dos métodos por validar de uma forma robusta devido a vários factores.

No âmbito da análise de assinaturas podem-se realçar várias evidências:

- Inicialmente o desenvolvimento iniciou-se com uma quantidade bastante reduzida

de dados em termos de assinaturas falsificadas e de outros pintores. Em termos de assinaturas falsificadas procedeu-se à construção em meio laboratorial de algumas assinaturas falsificadas para tornar os dados das falsificações mais completos. Em termos de assinaturas de outros pintores recorreu-se a pesquisas na web de modo a encontrar novas assinaturas.

- O tratamento dos dados foi indispensável para a obtenção de resultados positivos. Inicialmente verificou-se que, devido à textura dos quadros, muitas características eram provenientes do fundo. A extracção do fundo não foi tão eficaz em alguns casos, continuando as porções de fundo existentes a interferir nos resultados dos métodos. Assim, apesar de nada automático e bastante trabalhoso, tornou-se necessário extrair o *background* das imagens utilizando uma ferramenta de edição de imagem.
- O método para agrupamento de assinaturas apresentou resultados positivos, embora se tenham verificado algumas falhas pontuais. Estudos sobre Amadeo de Souza-Cardoso revelam que o pintor tinha principalmente dois períodos artísticos. Ainda com algumas excepções, o método permitiu encontrar uma separação entre esses períodos. Contudo, seria necessário um outro conjunto de dados para validar o mesmo.
- O classificador baseado nas características de assinaturas permitiu resolver vários problemas de acordo com o tipo de assinatura. Para demonstrar a robustez do método construiu-se um classificador dedicado à atribuição de autor de assinaturas, um para verificação de autenticidade e um outro que juntava estes dois conceitos. Os resultados obtidos foram positivos, revelando um desempenho relativamente elevado em todos os casos. Para demonstrar a eficácia do método construído, comparou-se o mesmo método com a utilização dos filtros de Gabor, revelando uma eficácia relativamente menor. Pelos testes efectuados considera-se um classificador robusto, podendo ser aplicado a outros conjuntos de dados.

Em relação ao classificador com base nas características de pincelada, não havia interesse em testar a eficácia do mesmo pois este já tinha revelado resultados significativos no âmbito do projecto. Porém, anteriormente os resultados obtidos tinham como base um conjunto de treino e teste fixo escolhido por historiadores. Os resultados descritos neste capítulo apresentam uma validação ao longo de 10 iterações, sendo em cada uma delas escolhido um conjunto de treino e teste aleatório. Novamente é de realçar a eficácia do classificador, sendo comparado também com o mesmo método mas utilizando os filtros de Gabor, o qual revelou uma eficácia bem menor.

A nível do mapeamento dos materiais é possível tirar várias conclusões com base nas análises efectuadas:

- O primeiro método construído consistia num algoritmo de classificação supervisionada. Apesar deste método requerer a definição de uma base de dados de espectros de referência por um perito, a classificação pode ser adaptada de acordo com o objectivo da análise. Este método destaca-se pela flexibilidade, permitindo vários tipos de classificação e respectiva visualização dos resultados. A análise dos resultados por um perito em análise de obras de arte revelou uma classificação bastante eficaz. O mapeamento dos *pixels* correspondentes apenas a um único espectro de referência da base de dados permitiu uma análise direccionada a determinadas zonas do quadro, permitindo aos historiadores dividir a pintura em várias zonas.

A visualização por um perito era de certo modo eficiente para avaliar os objectivos pretendidos, contudo surgiu a necessidade de uma avaliação mais eficaz. Com base em alguns processos de avaliação relatados no estado da arte, construiu-se um método de avaliação baseado em regiões de interesse para cada material. Este método de avaliação tem a fragilidade de não ser global a toda a pintura, restringindo-se a 169 valores para cada um dos espectros de referência. Contudo é praticamente impossível definir um método global para toda a pintura, pois a imagem hiperespectral captada possui milhões de *pixels*. Ainda assim, tendo como referência o método de avaliação utilizado verificou-se um desempenho elevado na classificação.

- Face ao aspecto menos automático do método anterior, os historiadores e peritos mostraram interesse em possuir um método mais automático. Para isso procedeu-se ao desenvolvimento de um método baseado no agrupamento de *pixels* pelo algoritmo K-Means. O desenvolvimento deste método fez emergir alguns problemas, um deles o facto de atribuir uma cor a cada *cluster* de modo a construir um resultado visual. No estado da arte existia já um método semelhante que projectava cada um dos *clusters* numa base de dados de espectros de referência através da utilização do SAM, contudo esse passo tornava o método novamente pouco automático. Assim optou-se por atribuir uma cor aos *pixels* de cada *cluster* baseada no centróide do respectivo *cluster*, recorrendo à extracção da cor a partir da informação do próprio espectro. Este método possui alguma flexibilidade ao permitir mapear cada um dos *clusters* individualmente, todavia parece ser ineficaz a detectar alguns pigmentos, como o caso do pigmento preto, que para o caso de estudo foi inexistente. Para este método não foi possível construir qualquer método de avaliação devido à incerteza gerada pelo número de *clusters* que o utilizador escolhe.
- A classificação com base de dados externa baseou-se em duas análises. A primeira análise efectuada, utilizando um conjunto de amostras adquiridas numa pintura falsificada com a técnica FORS, revelou uma classificação bastante eficaz. Este método é semelhante à classificação supervisionada, apenas com a diferença na forma como a base de dados de referência é definida. A classificação e a visualização de

resultados apresentavam a flexibilidade já enunciada, sendo a classificação avaliada recorrendo ao mesmo método já utilizado para a classificação supervisionada. Apesar de para alguns pigmentos não ser possível definir regiões de interesse para avaliação, os resultados obtidos com as restantes regiões revelaram uma eficácia significativa.

No mapeamento de materiais com análise das misturas e validação de autenticidade verificaram-se resultados que criam várias perspectivas para o futuro, tanto a nível da identificação correcta das misturas, como de zonas não analisadas ou de zonas não pertencentes a Amadeo. Apesar de em termos de visualização não ser um método eficaz, em termos de resultados permitiu aos historiadores e conservadores identificar mais rapidamente zonas de misturas com maior rapidez e eficácia. Por vezes, as misturas atribuídas a essas zonas não foram as mais correctas, algo que é aceitável pelo facto da soma linear de espectros ser apenas uma aproximação que não pode ser tomada como regra geral para todos os casos. Inicialmente o objectivo principal não seria identificar as misturas, mas sim as zonas da pintura com misturas, não sendo necessário indicar quais os pigmentos presentes. Contudo face ao desafio corrente de resolver também essa questão, procedeu-se à construção de uma abordagem aproximada para a sua resolução.

Por fim surge o classificador combinado de pinturas. Este classificador permitiu resolver dois problemas: atribuição de autor de pinturas e atribuição/autenticidade de pinturas. Devido à escassez de pinturas falsificadas com assinatura, não foi possível construir um classificador dedicado exclusivamente à verificação de autenticidade de pinturas. Este classificador era um dos grandes objectivos do projecto em que a dissertação está inserida, revelando resultados positivos. De realçar que apesar das quantidades reduzidas de pinturas não pertencentes a Amadeo de Souza-Cardoso, ambos os classificadores revelaram uma eficácia de 100%. Submeteram-se ao classificador construído três pinturas sobre as quais não existe ainda um veredicto se pertencem a Amadeo de Souza-Cardoso ou não. Os resultados destes três casos de estudo podem ajudar os historiadores a tirar algumas conclusões. Porém, por vezes, podem surgir situações inesperadas, como o caso de uma pintura autêntica do pintor sobre a qual foi colocada uma assinatura falsa. Estes casos extremos exigem uma análise dos resultados obtidos com ambos os classificadores individuais (assinatura e pincelada), podendo mesmo o resultado final revelar uma classificação da pintura no limiar entre pinturas falsas e verdadeiras. A integração da análise dos materiais nesta classificação poderia ser uma solução para o problema.



Conclusões e Trabalho Futuro

Neste capítulo serão apresentadas as conclusões retiradas do trabalho efectuado, assim como algumas perspectivas para o futuro em termos de trabalho a realizar.

5.1 Conclusões

Em termos de conclusões pode enumerar-se um conjunto significativo de resultados que abrem boas perspectivas para o desenvolvimento futuro.

Ao método para identificação de períodos artísticos faltou uma validação mais eficaz com outro conjunto de dados, o que o tornaria mais robusto. Contudo os resultados obtidos são razoáveis apesar de pela análise se identificarem falhas pontuais no agrupamento. Essas falhas não comprometem de todo a sua eficácia já que foram identificados os principais períodos na vida de Amadeo de Souza-Cardoso.

O método de classificação de pinturas baseado em características das assinaturas era um dos grandes objectivos da dissertação. Com o classificador construído foi possível resolver dois problemas: atribuição e autenticidade. Para demonstrar a eficácia do classificador, utilizaram-se outro tipo de características, os filtros de Gabor. A configuração com as características SIFT revelou nos três problemas uma eficácia geral do classificador acima dos 90%, em alguns casos perto dos 100%, o que revela uma classificação bastante eficaz. A classificação utilizando os filtros de Gabor foi relativamente menos assertiva, o que revela neste caso em particular a adequação das características SIFT para resolver o problema.

O método de classificação de pinturas baseado em características de pincelada já havia sido construído anteriormente, tendo sido feitos apenas testes para demonstrar que o classificador é realmente eficaz com uma filosofia diferente de validação, relativamente

aos conjuntos de treino e teste fixos que haviam sido construídos anteriormente. Estes revelaram uma percentagem de acerto perto dos 95%, o que mais uma vez constitui um resultado significativo.

O mapeamento de materiais através de classificação supervisionada combinando duas das técnicas mais utilizadas ao nível do estado da arte revelou resultados bastante positivos. A construção da base de dados a partir de amostragem da própria imagem hiperespectral da pintura fez com que a base de dados de espectros de referência tivesse as mesmas condições de iluminação e ruído que a imagem hiperespectral, o que melhorou a classificação. A análise do classificador não é totalmente ilustrativa do seu desempenho total mas é impossível no caso de pinturas construir um método que permita validar todos os resultados do classificador, pois estes são na ordem dos milhões de *pixels* classificados.

O mapeamento de materiais através de classificação não-supervisionada permitiu identificar algumas das zonas com materiais semelhantes, todavia a sua aplicação nunca identificou o pigmento preto, sendo este incorporado nos pigmento branco e nos restantes *clusters* obtidos, o que constituiu uma limitação. Exceptuando esse factor, pode constituir um método que permita aos historiadores analisar quais os materiais que se realçam à medida que o número de *clusters* aumenta.

Os resultados da classificação com base de dados externa utilizando a informação captada com a técnica FORS permitiram efectuar um mapeamento eficaz e bastante efectivo para qualquer uma das situações em que foi aplicado, quer seja para o caso de estudo relatado como para a outra pintura falsificada onde foi também aplicada. Tal como já foi referido anteriormente, esta eficácia reforça a ideia de que a câmara hiperespectral e a técnica FORS produzem informação bastante semelhante podendo ser comparada.

O mapeamento de materiais com análise das misturas e validação de autenticidade permitiu analisar com mais pormenor as obras de arte do pintor em estudo, não se restringindo a uma base de dados fixa. A análise das misturas, à excepção de alguns casos, foi razoavelmente assertiva relativamente às informações que se conheciam das pinturas. De realçar que apesar deste ser um dos problemas que o método se propõe resolver, a metodologia escolhida para análise das misturas consiste numa abordagem que não se aplica em termos colorimétricos, e que em termos de curvas espectrais por vezes conduz a erros ao nível do classificação. A condição de verificação de autenticidade revelou ser bastante eficaz nos casos testados, tendo sido de 0% em ambas as pinturas autênticas nas quais foi aplicada. Este método carece de validação devido à inexistência de uma análise com o método EDXRF nas duas pinturas falsificadas para as quais existiam as imagens hiperespectrais restantes, ficando esse teste para o futuro. Para o futuro fica também um teste mais robusto da condição de verificação de autenticidade com vista à integração deste tipo de informação no classificador combinado de pinturas. Devido à escassez de imagens hiperespectrais para construção de um método mais robusto, esta condição deverá ser testada no futuro sendo integrada num classificador de pinturas.

Por fim, surge o classificador combinado de pinturas baseado em diferentes tipos de

análise. Devido à impossibilidade de integrar a análise dos materiais na classificação tal como era objectivo, este foi construído apenas com a informação da pincelada e das assinaturas. A escassez de imagens hiperespectrais não permitiu assim incluir a análise dos materiais no classificador combinado pois algumas pinturas para as quais havia imagem hiperespectral não tinham assinatura, o que impossibilitava a integração. Ainda assim, o classificador combinado permitiu resolver dois problemas: (1) atribuição e (2) atribuição e autenticidade conjuntamente. Não foi possível construir um classificador apenas dedicado à resolução do problema autenticidade devido à escassez de falsificações com assinatura. Porém os resultados mostram uma eficácia de 100% nos dois casos, o que revela uma classificação sem qualquer falha. Enquanto os classificadores individuais se equivocaram na classificação de algumas pinturas, acabaram por se corrigir mutuamente e o resultado final é positivo tendo em conta a quantidade de dados existente. A classificação utilizando os filtros de Gabor revelou-se mais uma vez não adequada para resolver o problema em questão. Este classificador poderá constituir no futuro uma ferramenta fiel para verificação da autenticidade e atribuição de autor de pinturas de Amadeo de Souza-Cardoso, um dos principais objectivos do projecto.

5.2 Trabalho Futuro

Em termos de trabalho futuro no âmbito do projecto em que a dissertação está integrada é possível enunciar dois objectivos a realizar num futuro mais próximo. Em primeiro lugar, seria vantajoso construir uma interface gráfica que incidisse sobre o método de análise das misturas e validação de autenticidade de modo a ser acessível a todos os utilizadores sem qualquer experiência de programação. Este não era um objectivo da dissertação mas seria uma grande vantagem para os historiadores ter uma ferramenta que os auxiliasse nas deslocações aos museus, permitindo efectuar uma análise rápida sobre a pintura através de uma aplicação com uma interface gráfica acessível e de fácil usabilidade, mesmo que esta fosse direccionada numa primeira fase para um único pintor.

Outra das perspectivas de futuro encontra-se no classificador combinado de pinturas, faltando no momento a integração da análise dos materiais nesse mesmo classificador. Até ao término da dissertação tal não foi possível pela razão da quantidade de pinturas com imagem hiperespectral ser reduzida, o que impossibilitou a combinação dos resultados da classificação de pinturas e de assinaturas com o resultado obtido com a análise dos materiais. De realçar também que as pinturas estudadas neste trabalho e para as quais existia uma análise EDXRF não estavam presentes no conjunto de pinturas utilizadas no conjunto de teste do classificador combinado, daí a impossibilidade para juntar a informação, sendo esta uma perspectiva para trabalho futuro.

Bibliografia

- [22604] Optimal classification methods for mapping agricultural tillage practices. *Remote Sensing of Environment*, 91(1):90–97+, Maio 2004.
- [ABM06] Stephane Armand, Michael Blumenstein, e Vallipuram Muthukkumaramsamy. Off-line signature verification based on the modified direction feature. In *International Conference on Pattern Recognition (ICPR 2006)*, pág. 509–512, Hong Kong, 2006.
- [ANP07] P. Antonopoulos, N. Nikolaidis, e I. Pitas. Hierarchical face clustering using SIFT image features. In *Computational Intelligence in Image and Signal Processing, 2007. CIISP 2007. IEEE Symposium on*, pág. 325–329, Honolulu, Hawaii, 2007. IEEE Computer Society.
- [BC97] Reena Bajaj e Santanu Chaudhury. Signature verification using multiple neural classifiers. *Pattern Recognition*, 30(1):1–7, Janeiro 1997.
- [BL97] Jeffrey S. Beis e David G. Lowe. Shape indexing using approximate nearest-neighbour search in high-dimensional spaces. In *Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition (CVPR '97)*, CVPR '97, pág. 1000–, Washington, DC, USA, 1997. IEEE Computer Society.
- [BPvdH09] Igor Bereznoy, Eric O. Postma, e H. Jaap van den Herik. Automatic extraction of brushstroke orientation from paintings. *Mach. Vis. Appl.*, 20(1):1–9, 2009.
- [BYRN99] Ricardo A. Baeza-Yates e Berthier Ribeiro-Neto. *Modern Information Retrieval*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1999.
- [CRB00] Hyeokho Choi, Justin Romberg, e Richard Baraniuk. Hidden markov tree modeling of complex wavelet transforms. In *International Conference*

- on Acoustics, Speech, and Signal Processing*, pág. 133–136, Istanbul, Turkey, 2000. IEEE.
- [CS93] Jean-François Cardoso e Antoine Souloumiac. Blind beamforming for non gaussian signals. *IEE Proceedings-F*, 140:362–370, 1993.
- [dAeTdC04] Maria João de Alvarenga e Teles do Carvalho. Digitalização de Pintura Artística com Imagiografia Hiperespectral. Tese de Mestrado, Universidade do Minho - Escola de Ciências, Braga, Portugal, 2004.
- [DHD⁺06] B. Deronde, R. Houthuys, W. Debruyn, D. Fransaer, V. Van Lancker, e J. P. Henriët. Use of Airborne Hyperspectral Data and Laserscan Data to Study Beach Morphodynamics along the Belgian Coast. *Journal of Coastal Research*, 22:1108–1117, 2006.
- [DHR04] P.E. Dennison, K.Q. Halligan, e D.A. Roberts. A comparison of error metrics and constraints for multiple endmember spectral mixture analysis and spectral angle mapper. *Remote Sensing of Environment*, 93(3):359–367, 2004.
- [DIL⁺04] G. Dimauro, S. Impedovo, M. G. Lucchese, R. Modugno, e G. Pirlo. Recent advancements in automatic signature verification. In *Frontiers in Handwriting Recognition, 2004. IWFHR-9 2004. Ninth International Workshop on*, pág. 179–184, Los Alamitos, CA, 2004. IEEE Computer Society.
- [DZT⁺10] J. K. Delaney, J. G. Zeibel, M. Thoury, R. Littleton, M. Palmer, K. M. Morales, E. R. de La Rie, e A. Hoenigswald. Visible and infrared imaging spectroscopy of picasso’s harlequin musician: Mapping and identification of artist materials in situ. *Applied spectroscopy*, 64(6):584–594, 2010.
- [FA91] William T. Freeman e Edward H. Adelson. The design and use of steerable filters. *IEEE Trans. Pattern Anal. Mach. Intell.*, 13(9):891–906, 1991.
- [FB81] Martin A. Fischler e Robert C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.
- [FFLT] Rob Fergus Fei-Fei Li e Antonio Torralba. Bag-of-features models. http://www.cs.unc.edu/~lazebnik/spring09/lec18_bag_of_features.pdf.
- [FM02] Xiaoguang Feng e Peyman Milanfar. Multiscale principal components analysis for image local orientation estimation. In *Proceedings of the 36th Asilomar Conference on Signals, Systems and Computers*, pág. 478–482, Pacific Grove, California, 2002.

- [FRDF01] Ricardo J. Ferrari, Rangaraj M. Rangayyan, J. E. Leo Desautels, e Annie France Frère. Analysis of asymmetry in mammograms via directional filtering with gabor wavelets. *IEEE Trans. Med. Imaging*, 20(9):953–964, 2001.
- [Gab46] Dennis Gabor. Theory of communication. *J. Inst. Elect. Eng.*, 93:429–457, 1946.
- [GBHD04] G. Girouard, A. Bannari, A. Harti, e A. Desrochers. Validated spectral angle mapper algorithm for geological mapping: comparative study between QuickBird and Landsat-TM. In *XXth ISPRS Congress, Geo-Imagery Bridging Continents*, Istanbul, Turkey, July 2004.
- [GGL⁺03] H. Gomez-Moreno, P. Gil-Jimenez, S. Lafuente-Arroyo, R. Vicen-Bueno, e R. Sanchez-Montero. Color images segmentation using the support vector machines. *Recent Advances in Intelligent Systems and Signal Processing*, pág. 151–155, 2003.
- [HR04] S. Homayouni e M. Roux. Hyperspectral image analysis for material mapping using spectral matching. In *XXth ISPRS Congress, Geo-Imagery Bridging Continents*, Istanbul, Turkey, July 2004.
- [HY97] Kai Huang e Hong Yan. Off-line signature verification based on geometric feature extraction and neural network classification. *Pattern Recognition*, 30(1):9–17, Janeiro 1997.
- [jai02] On-line signature verification. *Pattern Recognition*, 35(12):2963–2972, Dezembro 2002.
- [JAMB08] M. Bautista S. MasPOCH J. Coello J.M. Amigo, J. Cruz e M. Alcalá M. Blanco. Study of pharmaceutical samples by nir chemical-image and multivariate analysis. *Trends in Analytical Chemistry*, 27(8):696 – 713, 2008.
- [JBS05] Edson Justino, Flavio Bortolozzi, e Robert Sabourin. A comparison of SVM and HMM classifiers in the off-line signature verification. *Pattern Recognition Letters*, 26(9):1377–1385, Julho 2005.
- [JG05] Nathaniel P. Jacobson e Maya R. Gupta. Design goals and solutions for display of hyperspectral images. In *International Conference on Image Processing (ICIP 2005)*, pág. 622–625, Genoa, Italy, September 2005. IEEE.
- [JHB⁺08] C Richard Johnson, Ella Hendriks, Igor Bereznoy, Eugene Brevdo, Shannon Hughes, Ingrid Daubechies, Jia Li, Eric Postma, e James Z Wang. Image processing for artist identification. *IEEE Signal Processing Magazine*, 25(4):37–48, July 2008.

- [KR05] Leonard Kaufman e Peter J. Rousseeuw. *Finding Groups in Data: An Introduction to Cluster Analysis (Wiley Series in Probability and Statistics)*. Wiley-Interscience, Março 2005.
- [KSX04] Meenakshi K. Kalera, Sargur Srihari, e Aihua Xu. Offline signature verification and identification using distance statistics. *International Journal of Pattern Recognition and Artificial Intelligence*, 18(7):1339–1360, 2004.
- [Kuc07] S. Kuching. The performance of maximum likelihood, spectral angle mapper, neural network and decision tree classifiers in hyperspectral image analysis. *Journal of Computer Science*, 3(6):419–423, 2007.
- [KY05] Alisher Kholmatov e Berrin Yanikoglu. Identity authentication using improved online signature verification method. *Pattern Recognition Letters*, 26(15):2400–2408, Novembro 2005.
- [LdS06] Patricio Loncomilla e Javier Ruiz del Solar. A fast probabilistic model for hypothesis rejection in sift-based object recognition. In *11th Iberoamerican Congress on Pattern Recognition (CIARP 2006)*, pág. 696–705, Cancun , Mexico, June 2006. Springer.
- [Lee96] Tai Sing Lee. Image representation using 2d gabor wavelets. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 18:959–971, 1996.
- [Low04] David G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60:91–110, 2004.
- [LP94] F. Leclerc e R. Plamondon. Automatic signature verification: the state of the art-1989-1993. *Intl. Journal of Pattern Recognition and Artificial Intelligence*, 8(3):643–660, 1994.
- [MGW⁺05] J. T. Mundt, N. F. Glenn, K. T. Weber, T. S. Prather, L. W. Lass, e J. Pettin-gill. Discrimination of hoary cress and determination of its detection limits via hyperspectral image processing and accuracy assessment techniques. *Remote Sensing of Environment*, 96(3):509–517, 2005.
- [MJC⁺12] C. Montagner, R. Jesus, N. Correia, M.J. Melo, M. Villarigues, R. Macedo, e H. de Freitas. Unveiling the hand of a 19th century artist with binary image classification and bag-of-features. In *Systems, Signals and Image Processing (IWSSIP), 2012 19th International Conference on*, pág. 201–204, Vienna, Austria, April 2012. IEEE.
- [MMR⁺01] K.-R. Müller, S. Mika, G. Rätsch, S. Tsuda, e B Schölkopf. An introduction to kernel-based learning algorithms. *IEEE Transactions on Neural Networks*, 12(2):181–202, 2001.

- [MMS03] D. Manolakis, D. Marden, e G. A. Shaw. Hyperspectral image processing for automatic target detection applications. *Lincoln Laboratory Journal*, 14(1):79–116, 2003.
- [MRB06] B. Majhi, Y. S. Reddy, e D. P. Babu. Novel features for off-line signature verification. *International Journal of Computers, Communications & Control*, 1(1):17–24, 2006.
- [NJT06] Eric Nowak, Frederic Jurie, e Bill Triggs. Sampling strategies for bag-of-features image classification. In *European Conference on Computer Vision*, Graz, 2006.
- [OKM00] T. Ohishi, Y. Komiya, e T. Matsumoto. On-line signature verification using pen-position, pen-pressure and pen-inclination trajectories. In *15th International Conference on Pattern Recognition*, pág. 547–550, Barcelona, Spain, September 2000. IEEE Computer Society.
- [OSK05] E. Özgündüz, T. Sentürk, e M. Karşligil. Off-Line Signature Verification And Recognition By Support Vector Machine. In *13th European Signal Processing Conference (EUSIPCO 2005)*, Antalya, Turkey, Setembro 2005.
- [Ots79] Nobuyuki Otsu. A Threshold Selection Method from Gray-level Histograms. *IEEE Transactions on Systems, Man and Cybernetics*, 9(1):62–66, 1979.
- [PD05] V. S. N. Prasad e J. Domke. Gabor filter visualization. Relatório técnico, University of Maryland, 2005.
- [PDMDRP08] A. Pelagotti, A. Del Mastio, A. De Rosa, e A. Piva. Multispectral imaging of paintings. *Signal Processing Magazine, IEEE*, 25(4):27–36, 2008.
- [Pea01] K. Pearson. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 2(6):559–572, 1901.
- [Pla99] John C. Platt. Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods. In *Advances in Large Margin Classifiers*, pág. 61–74. MIT Press, 1999.
- [PS03] Tomaso Poggio e Steve Smale. The mathematics of learning: dealing with data. *Notices Amer. Math. Soc.*, 50(5):537–544, 2003.
- [PWLS07] B. Park, W. R. Windham, K. C. Lawrence, e D. P. Smith. Contaminant classification of poultry hyperspectral imagery using a spectral angle mapper algorithm. *Biosystems engineering*, 96(3):323–333, 2007.
- [rui08] Offline Signature Verification Using Local Interest Points and Descriptors. In *13th Iberoamerican congress on Pattern Recognition: Progress in Pattern Recognition, Image Analysis and Applications (CIARP '08)*, pág. 22–29, Havana, Cuba, June 2008.

- [RYP03] R. Rifkin, G. Yeo, e T. Poggio. Regularized least-squares classification. *Nato Science Series Sub Series III Computer and Systems Sciences*, 190:131–154, 2003.
- [SBF07] LinLin Shen, Li Bai, e Michael Fairhurst. Gabor wavelets and general discriminant analysis for face identification and verification. *Image Vision Comput.*, 25(5):553–563, Maio 2007.
- [SBT98] H. E. S. Said, K. D. Baker, e T. N. Tan. Personal identification based on handwriting. In *Fourteenth International Conference on Pattern Recognition*, pág. 1761–1764 vol.2, Brisbane, Australia, August 1998. IEEE.
- [SG64] A. Savitzky e M. J. E. Golay. Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 36:1627–1639, 1964.
- [She09] Jialie Shen. Stochastic modeling western paintings for effective classification. *Pattern Recognition*, 42(2):293–301, 2009.
- [SMG99] Y. Sohn, E. Moran, e F. Gurri. Deforestation in North-Central yucatan(1985-1995)- mapping secondary succession of forest and agricultural land use in sotuta using the cosine of the angle concept. *PE & RS- Photogrammetric Engineering & Remote Sensing*, 65(8):947–958, 1999.
- [SR02] Y. Sohn e N. S. Rebelló. Supervised and unsupervised spectral angle classifiers. *Photogrammetric engineering and remote sensing*, 68(12):1271–1282, 2002.
- [TKDO03] J. Scott Tyo, Athanasios Konsolakis, David I. Diersen, e Richard Christopher Olsen. Principal-components-based display strategy for spectral imagery. *IEEE T. Geoscience and Remote Sensing*, 41(3):708–718, 2003.
- [Vai06] C. Vaiphasa. Consideration of smoothing techniques for hyperspectral remote sensing. *ISPRS journal of photogrammetry and remote sensing*, 60(2):91–99, 2006.
- [Wah99] Grace Wahba. Support vector machines, reproducing kernel Hilbert spaces and the randomized GACV. In B. Schölkopf, C. J. C. Burges, e A. J. Smola, editores, *Advances in Kernel Methods: Support Vector Learning*, pág. 69–88. MIT Press, Cambridge, MA, 1999.
- [YGB92] R. H. Yuhas, A. F. H. Goetz, e J. W. Boardman. Discrimination among semi-arid landscape endmembers using the spectral angle mapper (SAM) algorithm. In *Summaries of the Third Annual JPL Airborne Geoscience Workshop*, volume 1, pág. 147–149, 1992.

- [Zha04] Harry Zhang. The optimality of naive bayes. In *Seventeenth International Florida Artificial Intelligence Research Society Conference (FLAIRS 2004)*, Miami Beach, Florida, USA, 2004. AAAI Press.
- [ZMM06] Hongqin Zhang, David W. Messinger, e Ethan D. Montag. Perceptual display strategies of hyperspectral imagery based on pca and ica. In *Proceedings of SPIE, Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XII*, volume 6233, Orlando, Florida, April 2006.
- [ZVC11] Yingxuan Zhu, Pramod K. Varshney, e Hao Chen. Ica-based fusion for colour display of hyperspectral images. *Int. J. Remote Sens.*, 32(9):2427–2450, Maio 2011.

6

Apêndice

6.1 Lista de Elementos Químicos Materiais Puros

Material	Elementos
Red Vermillion	Hg
Red carmine	Sr
Raw Sienna	Fe, As
Terra Rossa	Fe, As
Orange Cd	Cd
Yellow Ochre	Fe
Yellow Cr	Cr
Violet	Co
Cerulean Blue	Co, Sn, Ni
Cobalt Blue	Co
Prussian Blue	Fe
Green Cd	Cd, Se
Viridian	Cr
Ultramarine	Al, Si
White	Pb
Black	-
Esmerald	Cu, As

Tabela 6.1: Elementos químicos presentes em cada um dos materiais puros de Amadeo de Souza Cardoso.