



UNIVERSIDADE DA BEIRA INTERIOR

Faculdade de Engenharia

Departamento de Informática

Google Earth™ Search Engine:
Classificação de Imagens Aéreas

Susana Elisabete Pereira da Costa

Dissertação para obtenção do Grau de Mestre em

Engenharia Informática

(2º ciclo de estudos)

Orientador: Prof. Doutor Hugo Pedro Martins Carriço Proença

Covilhã, 22 Junho de 2013

Dedicatória

Dedicado ao meu pai (*in memoriam*)...

”A map of the world that does not include Utopia is not worth even glancing at, for it leaves out the one country at which Humanity is always landing. And when Humanity lands there, it looks out, and, seeing a better country, sets sail. Progress is the realisation of Utopias.”

Oscar Wilde

Agradecimentos

A viabilidade e execução deste projeto dependeu da contribuição e preciosa ajuda, direta ou indiretamente, de várias pessoas, sem as quais não teria sido possível a conclusão do mesmo e a consequente escrita desta tese. Por esse motivo, não posso deixar de expressar a minha profunda gratidão,

Em primeiro lugar, ao Prof. Doutor Hugo Proença, pela sua orientação em todo o projeto, pelos esclarecimentos e conhecimento transmitido, pelas opiniões e críticas, pela disponibilidade e paciência. O seu contributo foi determinante para a realização deste projeto. Por tudo, o meu muito obrigada;

Ao *Soft Computing and Image Analysis Laboratory* (SOCIA-Lab) pelas condições disponibilizadas para a realização deste projeto, bem como aos seus elementos pela ajuda, amizade e bom ambiente promovido;

Ao meu colega e amigo, Sílvio Filipe, pela incansável paciência e ajuda, pelo companheirismo e amizade reforçados ao longo destes meses;

Ao Prof. Doutor Paulo Rebelo, pela ajuda, disponibilidade e, acima de tudo, pela amizade;

Aos meus colegas de Engenharia Informática, em especial ao Miguel Frias, pela simpatia e amizade demonstradas ao longo desta etapa;

À minha mãe, pelo incentivo e apoio incondicional em todas as minhas decisões e escolhas, pela compreensão, paciência e amor sempre demonstrados;

À minha amiga Natália Cunha, pelo apoio e amizade nos meus bons e maus momentos;

Por último, mas não menos importante, agradeço ao meu namorado e amigo, Marco Portugal, pela amizade, força e incentivo para que fosse possível a realização deste trabalho.

Resumo

O reconhecimento de padrões com recurso a redes neuronais é cada vez mais utilizado, numa tentativa de dotar as máquinas computacionais de inteligência e capacidade de aprendizagem. Neste trabalho pretende-se reconhecer determinados elementos naturais como 'água', 'casas' ou 'estradas', em imagens aéreas provenientes do Google Earth™ e do Google Maps™, recorrendo à utilizando redes neuronais para o efeito. Foram realizadas experiências com quatro conjuntos de imagens utilizados para o treino da rede neuronal, com variação de número de neurónios, e foram analisados os erros de classificação testando cinco novos conjuntos de imagens.

Foram ainda realizadas diversas experiências quanto aos métodos de extração de características e à aplicação de operadores morfológicos com diferentes elementos estruturais visando o pós-processamento dos resultados obtidos.

Palavras-chave

Imagens Aéreas, Redes Neuronais Supervisionadas, Reconhecimento Padrão, Análise de Componentes Principais, Operadores Morfológicos

Abstract

Pattern recognition using neural networks is increasingly used in an attempt to provide machines computational intelligence and learning ability. This work aims to recognize certain natural elements like 'water', 'house' or 'roads' in aerial images from Google Earth™ and Google Maps™, resorting to using neural networks for the purpose. Experiments were performed with four sets of images used for training the neural network, with varying number of neurons, and analyzed the classification errors by testing five new sets of images.

Were also carried out several experiments on methods of feature extraction and application of morphological operators with different structural elements aimed at the post-processing of results.

Keywords

Aerial Images, Supervised Neural Networks, Pattern Recognition, Principal Components Analysis, Morphological Operators

Índice

1	Introdução	1
1.1	Motivação e Objetivos	1
1.2	Estrutura da tese	2
2	Estado da Arte	5
2.1	Classificação de Imagens Aéreas	5
2.1.1	Histograma, Textura e Matrizes de Pixeis com Redes Neurais	5
2.1.2	Características da Textura para Classificação de Imagens	8
2.1.3	Características baseadas na textura	9
2.1.4	Métodos <i>Maximum-Likelihood Supervised Classifier</i> e <i>Nearest Neighbour Classifier</i> em Fotografias Pancromáticas	10
2.1.5	Utilização de Dados LIDAR para classificação de Áreas Florestais	14
2.1.6	Utilização de Dados LIDAR para classificação de Áreas Agrícolas	15
2.1.7	Classificação com Base em Descritores Locais	18
2.1.8	Bag-of-Visual-Words	20
2.1.9	<i>Quaternion Orientation Difference</i>	23
2.1.10	Classificação usando Descritores de Gabor	25
2.1.11	Extração de Estruturas de Estacionamento - Interpolação, Extrapolação, Previsão de blocos e Filtragem	27
2.2	Conclusões	30
3	Método Proposto: Extração de Características	33
3.1	Cor/Intensidade	34
3.1.1	RGB	34
3.1.2	HSV	35
3.1.3	Escala de Cinzentos	35
3.2	Matriz de co-ocorrência de tons cinza	35
3.2.1	Contraste	36
3.2.2	Energia	36
3.2.3	Correlação	36
3.2.4	Homogeneidade	37
3.3	Função Entropia	37
3.4	Descritores de Gabor	37
3.5	Imagens 'Classe'	39
3.6	Conclusão	39
4	Método Proposto: Redução da dimensionalidade (PCA)	41
4.1	Normalização dos Dados	41
4.2	Análise de Componentes Principais	41
5	Método Proposto - Classificação: Reconhecimento de Padrões	45
5.1	Redes Neurais Artificiais para Reconhecimento de Padrões	45
5.2	MLP - <i>Multilayer Perceptron</i>	47
5.3	Aprendizagem não supervisionada <i>versus</i> Aprendizagem supervisionada	48

5.4	Dados não balanceados <i>versus</i> Dados balanceados	48
5.5	Rede neuronal e Parâmetros utilizados	49
6	Método Proposto - Pós-Processamento: Operadores Morfológicos	51
6.1	Erosão e Dilatação	51
6.1.1	Abertura e Fecho (<i>Open</i> e <i>Close</i>)	52
7	Experiências e Discussão	55
7.1	Experiências	58
7.1.1	Características - Cor/Intensidade	58
7.1.2	Características - Matriz de Co-Ocorrência e Função Entropia	62
7.1.3	Características - Filtros Gabor	66
7.1.4	Características - Filtros de Gabor, Matriz de Co-Ocorrência e Função Entropia	70
7.1.5	Características - Cor/Intensidade e Filtros de Gabor	74
7.1.6	Características - Cor, Matriz de Co-Ocorrência e Função Entropia	78
7.1.7	Características - Cor/Intensidade, Filtros de Gabor, Matriz de Co-Ocorrência e Função Entropia	82
7.2	Discussão	87
8	Conclusões e Trabalho Futuro	89
8.1	Conclusões	89
8.2	Trabalho Futuro	89
	Bibliografia	91

Lista de Figuras

2.1	(a) Área experimental, (b) Mapa de referência, (c) Mapa de classificação baseado na representação BOV	22
2.2	Análise <i>Low Level</i>	28
2.3	Análise <i>High-Level</i> - Interpolação (vermelho), Extrapolação (azul), Análise <i>Low-Level</i> (verde)	29
3.1	<i>Google Earth</i>	33
3.2	<i>Google Maps</i>	33
3.3	Maltlab	34
3.4	Exemplos da influência do parâmetro <i>orientation</i>	38
3.5	Exemplos da influência do parâmetro <i>wavelength</i>	39
5.1	Arquitetura de um MLP	47
6.1	Elemento estruturante <i>disk</i>	53
6.2	Elemento estruturante <i>line</i>	53
6.3	Elemento estruturante <i>square</i>	53
7.1	Conjunto número 1 (Imagem <i>Google Earth</i> , Imagem <i>Google Maps</i> , Imagem 'classe')	55
7.2	Conjunto número 2 (Imagem <i>Google Earth</i> , Imagem <i>Google Maps</i> , Imagem 'classe')	55
7.3	Conjunto número 3 (Imagem <i>Google Earth</i> , Imagem <i>Google Maps</i> , Imagem 'classe')	55
7.4	Conjunto número 4 (Imagem <i>Google Earth</i> , Imagem <i>Google Maps</i> , Imagem 'classe')	56
7.5	Conjunto número 5 (Imagem <i>Google Earth</i> , Imagem <i>Google Maps</i> , Imagem 'classe')	56
7.6	Conjunto número 6 (Imagem <i>Google Earth</i> , Imagem <i>Google Maps</i> , Imagem 'classe')	56
7.7	Conjunto número 7 (Imagem <i>Google Earth</i> , Imagem <i>Google Maps</i> , Imagem 'classe')	57
7.8	Conjunto número 8 (Imagem <i>Google Earth</i> , Imagem <i>Google Maps</i> , Imagem 'classe')	57
7.9	Conjunto número 9 (Imagem <i>Google Earth</i> , Imagem <i>Google Maps</i> , Imagem 'classe')	57
7.10	Gráfico - 'Água': Taxa de erro / Operadores morfológicos	59
7.11	Gráfico - 'Casas': Taxa de erro / Operadores morfológicos	59
7.12	Gráfico - 'Estradas': Taxa de erro / Operadores morfológicos	60
7.13	Imagens originais; Imagens 'classe'; Melhor classificação rede; Melhor classificação operadores pós-processamento; Imagens XOR	61
7.14	Imagens originais; Imagens 'classe'; Melhor classificação rede; Melhor classificação operadores pós-processamento; Imagens XOR	61
7.15	Imagens originais; Imagens 'classe'; Melhor classificação rede; Melhor classificação operadores pós-processamento; Imagens XOR	62
7.16	Gráfico - 'Água': Taxa de erro / Operadores morfológicos	63
7.17	Gráfico - 'Casas': Taxa de erro / Operadores morfológicos	63
7.18	Gráfico - 'Estradas': Taxa de erro / Operadores morfológicos	64
7.19	Imagens originais; Imagens 'classe'; Melhor classificação rede; Melhor classificação operadores pós-processamento; Imagens XOR	65
7.20	Imagens originais; Imagens 'classe'; Melhor classificação rede; Melhor classificação operadores pós-processamento; Imagens XOR	65

7.21	Imagens originais; Imagens 'classe'; Melhor classificação rede; Melhor classificação operadores pós-processamento; Imagens XOR	66
7.22	Gráfico - 'Água': Taxa de erro / Operadores morfológicos	67
7.23	Gráfico - 'Casas': Taxa de erro / Operadores morfológicos	67
7.24	Gráfico - 'Estradas': Taxa de erro / Operadores morfológicos	67
7.25	Imagens originais; Imagens 'classe'; Melhor classificação rede; Melhor classificação operadores pós-processamento; Imagens XOR	69
7.26	Imagens originais; Imagens 'classe'; Melhor classificação rede; Melhor classificação operadores pós-processamento; Imagens XOR	69
7.27	Imagens originais; Imagens 'classe'; Melhor classificação rede; Melhor classificação operadores pós-processamento; Imagens XOR	70
7.28	Gráfico - 'Água': Taxa de erro / Operadores morfológicos	71
7.29	Gráfico - 'Casas': Taxa de erro / Operadores morfológicos	71
7.30	Gráfico - 'Estradas': Taxa de erro / Operadores morfológicos	71
7.31	Imagens originais; Imagens 'classe'; Melhor classificação rede; Melhor classificação operadores pós-processamento; Imagens XOR	73
7.32	Imagens originais; Imagens 'classe'; Melhor classificação rede; Melhor classificação operadores pós-processamento; Imagens XOR	73
7.33	Imagens originais; Imagens 'classe'; Melhor classificação rede; Melhor classificação operadores pós-processamento; Imagens XOR	74
7.34	Gráfico - 'Água': Taxa de erro / Operadores morfológicos	75
7.35	Gráfico - 'Casas': Taxa de erro / Operadores morfológicos	75
7.36	Gráfico - 'Estradas': Taxa de erro / Operadores morfológicos	75
7.37	Imagens originais; Imagens 'classe'; Melhor classificação rede; Melhor classificação operadores pós-processamento; Imagens XOR	77
7.38	Imagens originais; Imagens 'classe'; Melhor classificação rede; Melhor classificação operadores pós-processamento; Imagens XOR	77
7.39	Imagens originais; Imagens 'classe'; Melhor classificação rede; Melhor classificação operadores pós-processamento; Imagens XOR	78
7.40	Gráfico - 'Água': Taxa de erro / Operadores morfológicos	79
7.41	Gráfico - 'Casas': Taxa de erro / Operadores morfológicos	79
7.42	Gráfico - 'Estradas': Taxa de erro / Operadores morfológicos	79
7.43	Imagens originais; Imagens 'classe'; Melhor classificação rede; Melhor classificação operadores pós-processamento; Imagens XOR	81
7.44	Imagens originais; Imagens 'classe'; Melhor classificação rede; Melhor classificação operadores pós-processamento; Imagens XOR	81
7.45	Imagens originais; Imagens 'classe'; Melhor classificação rede; Melhor classificação operadores pós-processamento; Imagens XOR	82
7.46	Gráfico - 'Água': Taxa de erro / Operadores morfológicos	83
7.47	Gráfico - 'Casas': Taxa de erro / Operadores morfológicos	83
7.48	Gráfico - 'Estradas': Taxa de erro / Operadores morfológicos	83
7.49	Imagens originais; Imagens 'classe'; Melhor classificação rede; Melhor classificação operadores pós-processamento; Imagens XOR	85
7.50	Imagens originais; Imagens 'classe'; Melhor classificação rede; Melhor classificação operadores pós-processamento; Imagens XOR	85
7.51	Imagens originais; Imagens 'classe'; Melhor classificação rede; Melhor classificação operadores pós-processamento; Imagens XOR	86

Lista de Tabelas

7.1	Matriz Confusão do melhor resultado para o elemento 'água'	60
7.2	Matriz Confusão do melhor resultado para o elemento 'casas'	60
7.3	Matriz Confusão do melhor resultado para o elemento 'estradas'	60
7.4	Matriz Confusão do melhor resultado para o elemento 'água'	64
7.5	Matriz Confusão do melhor resultado para o elemento 'casas'	64
7.6	Matriz Confusão do melhor resultado para o elemento 'estradas'	64
7.7	Matriz Confusão do melhor resultado para o elemento 'água'	68
7.8	Matriz Confusão do melhor resultado para o elemento 'casas'	68
7.9	Matriz Confusão do melhor resultado para o elemento 'estradas'	68
7.10	Matriz Confusão do melhor resultado para o elemento 'água'	72
7.11	Matriz Confusão do melhor resultado para o elemento 'casas'	72
7.12	Matriz Confusão do melhor resultado para o elemento 'estradas'	72
7.13	Matriz Confusão do melhor resultado para o elemento 'água'	76
7.14	Matriz Confusão do melhor resultado para o elemento 'casas'	76
7.15	Matriz Confusão do melhor resultado para o elemento 'estradas'	76
7.16	Matriz Confusão do melhor resultado para o elemento 'água'	80
7.17	Matriz Confusão do melhor resultado para o elemento 'casas'	80
7.18	Matriz Confusão do melhor resultado para o elemento 'estradas'	80
7.19	Matriz Confusão do melhor resultado para o elemento 'água'	84
7.20	Matriz Confusão do melhor resultado para o elemento 'casas'	84
7.21	Matriz Confusão do melhor resultado para o elemento 'estradas'	84

Lista de Acrónimos

ANN	Artificial Neural Network
ASM	Angular Second Moment
BOV	Bag-Of-Visual-Word
BOW	Bag-Of-Words
CHM	Canopy Height Model
CMAC	Cerebellar Model Articulation Controller
DEM	Digital Elevation Model
DMC	Digital Mapping Camera
DSM	Digital Surface Model
DTM	Digital Terrain Model
GIS	Geographic Information System
GLCM	Gay Level Co-occurrence Matrix
GPS	Global Positioning System
HSB	Hue Saturation Brightness
HSI	Hue Saturation Intensity
HSL	Hue Saturation Lightness
HSV	Hue Saturation Value
IDM	Inverse Difference Moment
JPEG	Joint Photographic Experts Group
LIDAR	Light Detection and Ranging
LPIS	Land Parcel Information System
LVQ	Learning Vector Quantizer
MATLAB	MATrix LABoratory
MLP	Multilayer Perceptron
MLP	MultiLayer Perceptron
OBIA	Object-Based Image Analysis
PCA	Principal Component Analysis
PNG	Portable Network Graphics
PNN	Probabilistic Neural Network
RBF	Radial Basis Function
RCE	Restricted Coulomb Energy
RGB	Red Green Blue
RMS	Root Mean Square
RTK	Real Time Kinematics
SIFT	Scale Invariant Feature Transform
SVM	Support Vector Machines
TIN	Triangulated Irregular Network
UBI	Universidade da Beira Interior
VHR	Very High Resolution
WTA	Winner-Take-All

Capítulo 1

Introdução

O *Google Earth*™, programa desenvolvido e distribuído pela empresa americana *Google*, apresenta um modelo tridimensional do globo terrestre, construído a partir de fotografias de satélite e GIS 3D. Desta forma, é possível identificar lugares, construções, cidades e paisagens, entre outros elementos.

O *Google Maps*™ é um serviço de pesquisa e visualização de mapas e imagens de satélite da Terra, fornecido e desenvolvido pela mesma empresa do *Google Earth*™.

Atualmente, os serviços disponibilizam mapas e rotas com imagens de satélite de todo o mundo, com possibilidade de zoom em muitos pontos do globo.

Este projeto visa implementar e validar um método de reconhecimento de elementos naturais em imagens aéreas provenientes do *Google Earth*™ e *Google Maps*™. Com potencial interesse para imobiliárias, visa encontrar, por exemplo, casas situadas numa determinada região, com características específicas, como um rio numa determinada distância, uma estrada a menos de 'x' quilómetros ou ambas as condições.

No que respeita à classificação de imagens aéreas, pretende-se o reconhecimento e a localização dos elementos naturais básicos de um mapa (água, estradas e casas). Para tal, foi desenvolvido um algoritmo, em ambiente Matlab®, que, a partir de várias características das imagens e recorrendo à utilização de redes neuronais, treina e aprende a identificar os elementos pretendidos.

O Matlab® é um importante ambiente de desenvolvimento e teste para processamento de imagem digital, grandemente utilizado em investigação, pelo que se torna um ótimo objeto de trabalho, por essa razão escolhido para a realização deste projeto.

1.1 Motivação e Objetivos

É indiscutível a importância dos satélites artificiais nos dias de hoje em todos os aspetos da nossa sociedade pois são eles que nos possibilitam, entre outras coisas, a transmissão de dados e o acesso a tecnologias como os telemóveis, televisão, Internet e rádio. Além disso, é inimaginável controlar o tráfego aéreo, prever a meteorologia ou usar o GPS do nosso carro, sem o auxílio de tais satélites.

As técnicas de interpretação de imagens de satélite e de fotografias aéreas são usadas para inúmeros fins, desde a atualização de mapas, planeamento territorial, mapeamento urbano e rural [28], delimitação de fronteiras e demarcação de propriedades, estudos altimétricos, ambientais [36], oceanográficos e de infraestruturas.

A necessidade de análise de imagens tem vindo a aumentar ao longo dos anos, principalmente quando este processo não pode ser efetuado pelo ser humano de forma precisa, ou é demasiado moroso. Por outro lado, o ser humano desde muito cedo que procura desenvolver tecnologia para o auxiliar e/ou facilitar nas tarefas quotidianas. O interesse na área de processamento de imagens tem vindo a crescer, principalmente em aplicações como a melhoria da qualidade de imagens e o reconhecimento dos dados nelas presentes.

Este projeto enquadra-se num objetivo mais abrangente que consiste em desenvolver um sistema automático de procura de determinadas regiões em mapas, com base em pesquisas feitas pelos utilizadores numa linguagem próxima da natural. O principal objetivo é o reconhecimento e localização, pelo agente computacional, de elementos naturais básicos de um mapa e seleção de regiões que respeitem as restrições pedidas pelo utilizador.

No que respeita à classificação de imagens aéreas, realizada neste projeto, é utilizado o *Google Earth™* para obter as imagens aéreas e o *Google Maps™* para obter as imagens complementares. Processar uma imagem consiste em decompô-la com o objetivo de retirar informação que nela está contida. A complexidade de processar uma imagem dependente do sistema ao qual a imagem se encontra associada, não havendo uma solução que permita resolver todos os problemas. Neste âmbito, a necessidade de um agente automático de reconhecimento de elementos naturais básicos de um mapa torna-se importante e com muita utilidade para encontrar o local ideal para viver, passar férias ou apenas visitar.

A primeira etapa deste projeto consistiu na aquisição de dados. Para isso, como já foi referido, foi utilizado o *Google Earth™*. A próxima etapa teve como finalidade a análise e a interpretação das imagens para extração de características que possibilitaram, mais tarde, a localização e distinção dos elementos de acordo com o pretendido. Neste trabalho, o processo de extração de informação baseou-se essencialmente nas características dos canais de cor RGB, níveis de cinza da imagem, componentes *Hue*, *Saturation*, *Value* (HSV), matriz de co-ocorrência de tons cinza para características relativas à textura, descritores de Gabor e de entropia. Na etapa seguinte, foram selecionadas as melhores características, recorrendo à análise das componentes principais (PCA) para eliminar redundância nos dados previamente obtidos. Na fase seguinte, foi utilizada uma rede neuronal para treinar e testar o reconhecimento nas imagens e assim obter um método automático de identificação de água, casas e estradas. Obtidos os resultados da simulação, foram aplicados operadores morfológicos de pós processamento como abertura e fecho e foram calculadas as taxas de erro, os falsos positivos e falsos negativos, para vários exemplos, sendo que foram efetuados sucessivos testes com alteração de alguns parâmetros respeitantes às fases anteriores, com vista à optimização do modelo.

1.2 Estrutura da tese

Esta tese encontra-se dividida em oito capítulos. Neste capítulo, para além desta secção, consta uma introdução onde se descrevem os objetivos e a motivação que envolvem o propósito deste projeto.

No capítulo 2, apresentam-se os principais fundamentos e métodos relacionados com a classificação de imagens aéreas, tais como a extração de características, utilização de descritores e de redes neuronais para o efeito.

O capítulo 3 é dedicado à extração de características, métodos e técnicas utilizados para obter as características das imagens utilizadas neste trabalho.

No capítulo 4 é descrito o processo de normalização dos dados e a importância da análise de componentes principais (PCA) para redução da dimensionalidade das características extraídas.

O capítulo 5 é dedicado à classificação utilizando redes neuronais.

No capítulo 6 abordam-se os operadores morfológicos de pós-processamento utilizados para efetuar melhorias nas imagens resultantes da classificação da rede neuronal.

O capítulo 7 resume as experiências realizadas e discute os resultados obtidos.

No capítulo 8 é feita uma conclusão do trabalho, numa tentativa de retirar algumas deduções

Google Earth™ Search Engine - Implementação e Validação do Método de Reconhecimento

sobre a realização deste projeto, bem como discutir aspetos que possam ser melhorados num trabalho futuro.

Capítulo 2

Estado da Arte

Com os recentes desenvolvimentos e atualizações de aplicações como *Google Earth*™ 7.1.1.1580 (beta) e *Google Maps*™ (incorporado no anterior), é-nos dada a possibilidade de visualizar um modelo tridimensional do globo terrestre, construído a partir de mosaico de imagens de satélite.

Com a disponibilidade de imagens aéreas, de todo o planeta, tornou-se possível modelar, com maior nível de detalhe, as características da cobertura do solo. A resolução da maioria das imagens é suficiente para visualizar detalhes, por exemplo, do telhado de uma casa, estradas, árvores, ruas e até pequenos cursos de água.

Nesse contexto, têm sido desenvolvidos diversos trabalhos utilizando métodos e técnicas automáticas de representação, interpretação e classificação de imagens aéreas. Esses métodos envolvem não somente modelos de análise e tratamento de dados, mas também englobam métodos de classificação da informação.

Este capítulo aborda o estado da arte relativamente a características e métodos de classificação de imagens aéreas. O capítulo inicia com abordagens relativas a características de textura para classificação de imagens e descreve, depois, várias técnicas e métodos para classificação de objetos em imagens aéreas como *Maximum likelihood*, *Nearest Neighbour classifier*, dados LIDAR¹, uso de descritores locais, BOV, SVMs, SIFT, Gabor e GIST.

2.1 Classificação de Imagens Aéreas

2.1.1 Histograma, Textura e Matrizes de Pixeis com Redes Neurais

Na tese de doutoramento de Dev Ashish [3], foram analisadas três abordagens para a preparação de entradas para uma rede neuronal artificial, para classificação de terrenos. Essas abordagens incluíam histograma de intensidade de píxeis, parâmetros de textura e matrizes de informação espacial.

Neste trabalho, foram utilizadas duas imagens aéreas em escala de cinzentos e multi-espectral. As imagens em escala de cinzentos foram obtidas no *site* da Câmara Geórgia, de todo o estado da Geórgia². As imagens utilizadas eram de Baker, Bibb, Clarke, Colquitt, Fulton, Houston, Macon, Mitchell, e Seminole. Essas imagens aéreas de alta resolução, em escala de cinza, eram geo-referenciadas e geo-corrigidas. A maioria dessas imagens foi fotografada em 1993.

As fotografias aéreas em escala de cinza foram obtidas em formato JPEG, a partir de dados GIS. Foram selecionadas, manualmente, sub-áreas de classes conhecidas, chamadas de 'janelas', utilizando o *software Paintshop Pro*®. As imagens tinham tamanhos diferentes, devido ao processo de seleção. Estas imagens foram separadas em seis classes diferentes: 'cidade', 'floresta', 'água', 'campo escuro', 'campo escuro médio' e 'campo de luz'. Cada uma das seis classes foi representada por 250 imagens para um total de 1500 imagens. Do total, 200 imagens

¹Tecnologia ótica de detecção remota que mede propriedades da luz refletida de modo a obter a distância e/ou outra informação a respeito de um determinado objeto distante

²<http://www.ganet.org/gis/chouse>

de cada classe foram utilizadas para o desenvolvimento do modelo, que incluiu 150 imagens para treino e 50 imagens para teste. As restantes 50 imagens de cada classe foram separadas para utilização posterior nos dados de avaliação estabelecidos para avaliação final do modelo de rede neuronal. O conjunto de treino consistiu de padrões usados para ajustar os pesos de modo a minimizar o erro, e o conjunto de teste foi utilizado periodicamente para determinar quando parar o treino e evitar sobre-ajuste.

Na primeira abordagem, os histogramas de intensidade de píxeis foram criados a partir dos valores dos píxeis da imagem. Foram normalizados em função do número de píxeis da imagem inteira, para permitir uma comparação de imagens de diferentes tamanhos. Para a segunda abordagem, a informação espacial de uma imagem foi extraída a partir da parte central da imagem. Esta informação corresponde à intensidade dos píxeis na ordem espacial específica para preservar a sua localização espacial na grelha. A porção central da imagem utilizada para a extração de informação pré-processada foi denominado de 'janela'. Na terceira abordagem, os parâmetros da textura foram calculados a partir dos dados dos píxeis da imagem de acordo com diferentes métodos. Os parâmetros da textura usados neste estudo foram o *Angular Second Moment (ASM)* e o *Inverse Difference Moment (IDM)*. Além disso, foram também calculados a média e desvio padrão da intensidade dos píxeis.

Depois de pré-processados, os dados foram tratados para a criação da rede neuronal. Nos testes preliminares, a Rede Neuronal Probabilística (PNN) mostrou uma maior precisão do que o algoritmo padrão de propagação e foi, por isso, aplicada em todo o desenvolvimento posterior do modelo. O número de neurónios da camada oculta era igual ao número de padrões no conjunto de treino. A camada de saída, na PNN, tinha tantos neurónios quanto o número de classes.

A classe 'cidade' consistia em imagens de edifícios e áreas vizinhas. A classe 'floresta' era composta por imagens tiradas de regiões de árvores. As imagens da classe 'água' foram tiradas de lagos, lagoas e rios. A classe 'campo escuro' consistia em imagens de áreas de vegetação exuberante, enquanto a classe 'campo médio escuro' consistia em imagens de áreas de vegetação relativamente menos densas. A classe 'campo luz' era composta por imagens de áreas de campo de pousio.

A melhor rede neuronal foi selecionada com base na maior precisão global para o conjunto de dados de avaliação. A exatidão global foi calculada a partir da matriz de erro dos resultados da classificação. A precisão global é a razão entre a soma dos padrões corretamente classificados de todas as classes mais o número total de padrões apresentados à rede neuronal. A matriz de erro também pode fornecer informações sobre a precisão do produtor e utilizador. A precisão do produtor é a razão entre o número de padrões corretamente classificados de uma classe sobre o total dos padrões de classe que apresenta para a rede. A precisão do utilizador é a razão entre o número de padrões corretamente classificados numa classe sobre o total dos padrões classificadas como sendo de classe.

Cada um dos três métodos foi considerado individualmente com os mesmos conjuntos de dados. Na abordagem relativa à textura, foram considerados quatro diferentes parâmetros utilizados como entrada. Foi realizado um estudo separado para determinar se alguma das entradas podia ser eliminada e ainda obter o mesmo nível de precisão. Depois de testar as três abordagens individualmente, foi realizado um estudo, considerando as várias combinações de abordagens para determinar se esta podia melhorar ainda mais a precisão.

Resultados:

- Histograma: A rede neuronal baseada no histograma apresentou bom desempenho, exceto

para o 'campo escuro' e classe 'água'. De 50 imagens de 'campo escuro', nove foram classificadas como 'floresta' e cinco como 'água'. Seis imagens de 'água' foram classificadas como 'campo escuro'. Este erro de classificação entre duas classes com intensidade pixel na mesma faixa pode ser devido à falta de informação espacial ou de textura. Foram observadas precisões relativamente mais baixa para as classes 'floresta', 'campo escuro' e 'água'.

- Espacial: O tamanho da 'janela' foi limitado a um máximo de 15x15 píxeis. Os tamanhos de 'janela' utilizados foram 1x1, 3x3, 5x5, 10x10, 15x15 píxeis. A rede neuronal baseada na 'janela' 3x3 píxeis resultou na mais alta precisão global de 66%. Estas experiências também mostraram que o desempenho global da rede foi reduzido para tamanhos de 'janela' maiores do que 3x3. No entanto, os resultados das classes individuais variou consideravelmente. A maior precisão para 'floresta' (58%) e 'campo médio escuro' (98%) foi para uma 'janela' de 5x5 píxeis e o mais alto para o 'campo de luz' (100%) foi com 'janelas' de 10x10 e 15x15 píxeis. As classes 'cidade', 'floresta' e 'campo escuro' foram as que apresentaram o pior desempenho. Catorze imagens da classe 'cidade' foram classificadas como 'campo de luz', possivelmente devido às intensidades semelhantes e distribuição espacial dos píxeis. Da mesma forma, as classes 'floresta' e 'campo escuro' também tinham semelhantes intensidade e distribuição espacial, o que provavelmente resultou em confusão, embora as texturas fossem diferentes.
- Parâmetros da textura: Foram comparados vários tamanhos de 'janela' para a abordagem do parâmetro textura. Foram utilizados os dois parâmetros da textura de Haralick, ASM e IDM, média e desvio padrão de intensidade dos píxeis. Os tamanhos de 'janela' que foram analisadas incluíram 3x3, 5x5, 10x10, 15x15 píxeis e a imagem completa. A imagem completa teve o melhor desempenho com uma exatidão de 92% para o conjunto de dados de avaliação. O 'campo escuro' (80%) e a classe 'água' (84%) apresentaram uma precisão relativamente baixa, em comparação com outras classes. Seis imagens de 'campo escuro' foram classificadas erroneamente como 'floresta' e cinco imagens de 'água' foram erroneamente classificadas como 'campo escuro'. Uma vez que o uso da característica textura demonstrou a maior precisão geral, foram realizadas outras comparações com diferentes combinações de parâmetros da textura, como ASM e IDM, e média e desvio padrão da intensidade do pixel, a partir de toda a imagem. A rede baseada numa combinação de ASM e IDM apresentou uma exatidão global de apenas 58%. O melhor desempenho, utilizando uma combinação de média e desvio padrão da intensidade dos píxeis pode ser atribuído ao facto de a intensidade média de pixel ter sido utilizada como um dos critérios para classificar manualmente algumas das imagens. A rede baseada em todos os quatro parâmetros da textura de toda a imagem proporcionou a maior precisão geral.
- Combinações de informações, Histograma da textura e Espacial: Após análise das três abordagens individualmente, elas foram avaliados em combinação. Na primeira combinação, foram avaliadas as abordagens da textura e de histograma. A informação estrutural foi obtida a partir de uma imagem completa e o histograma baseou-se em 128 células, pois apresentaram o melhor desempenho nos primeiros testes. A Rede Neuronal Artificial (ANN), com base numa combinação de parâmetros da textura e abordagens de histograma, obteve uma precisão global de 88% para o conjunto de avaliação. Esta precisão foi menor do que a precisão geral da ANN usando apenas parâmetros da textura (92%). O desempenho da ANN combinada para a classe de 'campo escuro' (54%) foi mais baixa do que para

a ANN com base nos parâmetros de textura (apenas 80%). Na segunda análise, o desempenho de informação da textura, em combinação com a informação espacial foi considerada. A ANN foi baseada em parâmetros de textura para a imagem inteira, em combinação com uma 'janela' de 3x3 píxeis de informação espacial. Esta ANN teve uma exatidão de 88% para o conjunto de avaliação, que foi novamente menor do que a precisão geral da ANN baseada apenas no parâmetro textura (92%). Na terceira análise, o desempenho de informação espacial foi testado em combinação com a informação do histograma. A ANN foi baseada numa rede, usando uma combinação de 128 valores de células de histograma e uma 'janela' 3x3 de informação espacial. Esta ANN teve uma exatidão de 89% para o conjunto da avaliação. Este foi novamente menor do que a precisão geral da ANN baseada em parâmetros de textura da imagem inteira (92%). Em última análise, foi avaliada uma ANN com base numa combinação de informação estrutural, espacial e histograma. A informação estrutural foi obtida a partir de toda a imagem, a informação espacial foi obtida a partir da 'janela' central, 3x3, e foi utilizado o histograma de 128 células. Esta ANN teve uma exatidão de 89% para o conjunto de avaliação, que foi novamente inferior à precisão geral da ANN baseada apenas em parâmetros da textura. A precisão para a classe 'campo escuro' foi muito menos (66%) do que a precisão no caso do uso apenas de parâmetros da textura, mas o desempenho para a classe 'água' melhorou (92%).

A ANN baseada apenas nos parâmetros da textura teve a maior precisão global de 92% para o conjunto de dados de avaliação. O melhor desempenho de abordagem relativa à textura, de imagens em escala de cinza em comparação com outras abordagens, reforça os resultados de estudos anteriores com base em abordagens de textura em imagens em escala de cinza e multi-espectral.

2.1.2 Características da Textura para Classificação de Imagens

O desenvolvimento de um conjunto de características para classificação e categorização de imagens tem como maior dificuldade a definição de um conjunto de características significativas para descrever a informação da imagem a partir de um bloco de células. Depois de definidas as características, os blocos de imagem podem ser categorizados usando técnicas de reconhecimento padrão.

Segundo Haralick, Shanmugam e Dinstein [14], as características fundamentais para a interpretação humana de fotografias coloridas são o espectro, a textura e as características contextuais. As características relativas ao espectro descrevem a variação de tonalidade em vários canais da porção visível ou infravermelha de um espectro eletromagnético, enquanto as características da textura contêm informação sobre a distribuição espacial da variação de tonalidade num canal. As características contextuais contêm informação derivada de blocos de dados da imagem circundantes à área analisada. A textura e a tonalidade tornam-se mais importantes quando pequenas áreas de imagens de uma fotografias a preto e branco são processadas independentemente. O conceito de tonalidade é baseado nos vários tons de cinzento das células de resolução na imagem fotográfica, enquanto a textura está relacionada com a distribuição espacial (estatística) de tons de cinza. A textura e a tonalidade são conceitos intrinsecamente relacionados. O contexto, a textura e a tonalidade estão sempre presentes na imagem, embora, por vezes, uma propriedade possa dominar a outra.

A textura pode ser avaliada com sendo fina, lisa ou grossa, ondulada, moldada, irregular ou linear. Ela contém informação importante que permite a discriminação entre diferentes tipos

de dados da imagem.

As funções de auto-correlação, espectros de potência, a malha restrita de primeira e segunda ordem de Markov e as frequências relativas de vários níveis de cinzento em imagens não normalizadas foram utilizados em estudos anteriores, com algum grau de sucesso. Mas a textura não foi definida especificamente, caracterizada ou modelada. Foram apenas utilizadas transformações matemáticas gerais que atribuíam números à imagem transformada, de modo não específico.

Rosenfeld e Troy, [47], apresentaram um conjunto de procedimentos para extrair das imagens algumas características da textura. Os seus procedimentos baseavam-se nas diferenças entre os valores dos tons de cinzento dos elementos adjacentes da imagem e na auto-correlação dos valores dos tons de cinzento da imagem. Rosenfeld e Thurston [33] criaram um procedimento para detetar limites de separação de regiões, as quais diferem na aspereza da textura. Rosenfeld e Troy [47] descreveram procedimentos de manipulação de níveis de cinzento antes de aplicar os algoritmos de Rosenfeld e de Troy. Os procedimentos para detetar propriedades da textura como linhas e pontos foram também sugeridos por outros investigadores. O procedimento é baseado na suposição de que a informação relativa à textura, numa imagem, é contida no total da relação espacial dos tons de cinzento. Essas características são calculadas num domínio espacial e a natureza estatística da textura é tida em conta. Foi processado um conjunto de tons de cinzento espacialmente dependentes em matrizes de distribuição de probabilidade para um dado bloco de uma imagem e foi sugerido um conjunto de catorze características da textura que podiam ser extraídas dessas matrizes. Essas características continham informações como homogeneidade, contraste, dependência linear dos tons de cinzento, número e limites presentes e complexidade da imagem. Também foi investigada a utilidade das características da textura para categorização ou classificação de blocos de imagem de três diferentes conjuntos de dados variados de fotomicrografias de alta resolução para imagens de satélite de baixa resolução.

2.1.3 Características baseadas na textura

As características tom e textura estão sempre presentes nas imagens, existindo uma relação entre elas: se existe uma grande variação de tons de cinza num pequeno fragmento de imagem, então a propriedade dominante é a textura, caso contrário, é o tom. Quando o pequeno fragmento é apenas do tamanho de uma célula de resolução, então existe lá apenas uma característica presente - o tom. Quando o número de características distinguíveis de tons de cinzento aumentam, a textura dominará. Quando não existe um padrão espacial e a variação de tons de cinzento é grande, temos uma textura fina. Se o padrão espacial se tornar mais definido e envolve mais células de resolução, temos uma textura grosseira.

O procedimento sugerido por Haralick et al., 1973, [14], para obter características da textura de uma imagem é baseado na suposição de que a informação relativa à textura, numa imagem, é contida na relação espacial global que os tons de cinzento têm uns com os outros. A informação da textura é especificada por um conjunto de matrizes com tons de cinzento espacialmente dependentes. A distância, ou mais precisamente, as quatro medidas intimamente relacionadas, das quais todas as características de textura derivam, são vetores denominados matrizes de tons de cinzento espacialmente dependentes, angularmente mais próximos. Para descrever esses vetores devemos ter a noção de adjacente ou vizinho mais próximo das células de resolução. É considerado que uma célula de resolução tem oito vizinhos mais próximos.

Neste estudo foram usados três conjuntos de dados, extraídos de microfotografias de diferentes

rochas, de fotografias aéreas de cenários naturais e artificiais e de fotografias de satélites de alta altitude. Foi utilizado o método da função linear por partes discriminadas, para categorizar os blocos de imagem nos primeiro e terceiro conjuntos de dados. Os conjuntos de dados foram divididos arbitrariamente em conjuntos de treino e teste. Para o conjunto de fotografias aéreas, o qual tinha poucas amostras por categoria, foi usada a regra de decisão *Min-Max* para categorizar as imagens baseadas nas suas características da textura. Os limites da decisão foram obtidos usando todas menos uma amostra do conjunto de dados, e a amostra deixada de fora, foi depois, classificada. O procedimento foi repetido para todas as amostras do conjunto de dados para obter a precisão global da classificação.

Resultados dos estudos da classificação de imagem:

- Conjunto de dados de microfotografias: foi usado, como classificador de entrada, um conjunto de oito variáveis compreendendo o significado e a variância das características da textura, calculadas com uma distância de uma matriz espacialmente dependente de um tom de cinzento. O conjunto de dados foi arbitrariamente dividido num conjunto de treino de 143 amostras e um conjunto de teste de 100 amostras. A precisão global do esquema de classificação no conjunto de teste foi de 89%.
- Conjunto de dados de fotografias aéreas: Foi usada a regra de decisão *Min-Max* para a classificação das 170 imagens em oito categorias. Para cada imagem foram calculados quatro tons de cinzento espacialmente dependentes. Foram calculadas onze características da textura produzindo 44 componentes para cada imagem. Foram calculados, estatisticamente, o mínimo e o máximo, para 169 imagens e a regra de decisão *Min-Max* foi usada para classificar a imagem que foi deixada de fora. Este procedimento foi repetido 170 vezes para classificar todas as imagens. Fora dessas 170 imagens, 140, ou seja, 82.3% das imagens, foram corretamente classificadas.
- Conjunto de dados das imagens de satélite: o método da função de discriminação linear por partes foi usada para identificar a categoria de cada bloco de imagem no conjunto de dados. As variáveis de entrada para a classificação consistiam na variância média das quatro características da textura e oito características do espectro em cada uma das quatro bandas de espectro. As 624 amostras no conjunto de dados foram divididas, arbitrariamente, num conjunto de treino de 314 amostras e num conjunto de teste de 310. O classificador foi treinado em 314 amostras do conjunto de treino e cada amostra no conjunto de teste foi atribuído a uma de sete possíveis categorias. A precisão geral do classificador no teste de amostras foi de 82.5%.

A classificação dos dados multi-espetrais, como os das imagens de satélite processadas, foi realizada usando apenas características do espectro. Foi tentada uma classificação de espectro nos dados das imagens de satélite e alcançada uma precisão de 74 a 77% no conjunto de teste. Este resultado, comparado com os 83.5% usando a combinação de características de espectro e de textura, mostra que a melhoria na classificação da precisão pode resultar da utilização das características da textura no classificador.

2.1.4 Métodos *Maximum-Likelihood Supervised Classifier* e *Nearest Neighbour Classifier* em Fotografias Pancromáticas

A aplicação de métodos computadorizados para análise de dados da vegetação das fotografias aéreas enfrenta várias dificuldades. A classificação de fotografias aéreas pancromáticas é um

caso especial de classificação de imagens de sensores remotos envolvendo apenas um canal de cor, o nível de cinza, o que diminui seriamente a precisão dos métodos de classificação. Outro problema é a iluminação, que muda ao longo da imagem.

O estudo das alterações da vegetação de médio a longo prazo (10 - 50 anos) prende-se com a compreensão de processos ecológicos, nomeadamente a degradação florestal, as alterações na distribuição animal, bem como para o planeamento de programas de gestão de vegetação e vida selvagem [6]. Uma das fontes de dados para esses estudos foram as fotos aéreas, com alta resolução espacial. As fotografias de larga escala (1:10 000 para 1:15 000) estão disponíveis para várias áreas, e podem ser identificados objetos de 10 a 20 centímetros. Nos últimos 50 a 60 anos, passou a ser possível encontrar fotografias pancromáticas (preto/branco), de alta qualidade, de muitas partes do mundo. Assim, as fotografias aéreas são comumente usadas para estudos das alterações da vegetação a médio prazo. Estudos anteriores empregavam, normalmente, interpretação manual de fotografias aéreas para extração de dados da vegetação. A interpretação manual de fotografias aéreas é assumida com sendo 100% correta, o que pode não ser verdade.

Seguindo o esquema de classificação de Tomaselli (1977) [29], o objetivo era criar um método computadorizado para gerar mapas precisos de vegetação mediterrânica, de alta resolução, usando fotografias aéreas pancromáticas. As três classes, baseadas na altura da vegetação eram: 'árvores' > 2.5 metros; 'vegetação lenhosa' < 2.5 metros, incluindo arbustos; semi-arbustos e árvores baixas; e 'vegetação herbácea', que incluía terra nua. Esta aproximação foi avaliada usando duas fotografias (de 1964 e 1992) da área testada em Monte Meron, Israel. O processo de pré-classificação incluiu a preparação de um foto-mapa e ajustes de iluminação. Foram aplicadas duas classificações de aproximação:

- Classificação supervisionada *Maximum-likelihood*: Algoritmo de classificação que tem em consideração os níveis de cinzento de píxeis individuais;
- *Nearest Neighbour classification*: tem em consideração as características dos seus vizinhos.

Os resultados das duas classificações foram comparados com uma extensiva base de dados do terreno.

2.1.4.1 Métodos

Área de Estudo

A área escolhida para o estudo abrangia 400 hectares da encosta norte do Monte Meron, Montanhas Upper Galilee, em Israel. A área era heterogénia em termos de topografia e estrutura da vegetação. A espécie de árvores dominantes, arbustos e semi-arbustos, eram o *Quercus calliprinos*, *Calicotome villosa* e a *Sarcopoterium* espinhosa. Toda a área foi objeto de pastagem intensiva e corte das árvores. Foram escolhidas duas fotografias aéreas (1964 e 1992) e obtido o mapeamento do centro de Israel. Ambas as fotografias foram tiradas no verão, ao meio-dia. A escala da primeira fotografia era de 1:14 000 e da segunda 1:12 000. Os diapositivos das fotografias foram digitalizados usando um *scanner* de alta resolução, onde cada pixel representava 15x15 centímetros de solo. Os elementos antropogénicos (estradas, estabelecimento e áreas agrícolas - menos de 15% da área de estudo) foram digitalizados manualmente e excluídos.

Preparação do foto-mapa

No processo de ortorretificação³, a inclinação, a radial e a distorção inerentes às fotografias aéreas foram removidas, e a fotografia foi registada para um sistema de coordenadas planimétricas. Os pontos de controlo em terra foram identificados em ambas as fotografias e no campo. A distância entre a estação de base e o recetor móvel não excedeu os quatro quilómetros. Os dados foram analisados para diferentes correções. A precisão do controlo dos pontos no solo foi estimada usando quatro pontos de triangulação independentes, na área estudada, para os quais as coordenadas precisas foram obtidas do mapeamento do centro de Israel.

O estudo DEM (*Digital Elevation Model*) da área foi preparado usando duas fotografias parcialmente sobrepostas, a partir de 1992, e foi preparado outro DEM da mesma área usando um correspondente par de fotografias de 1964. Ambos os DEMs foram construídos usando *software* ISMT da Integraph Inc., com uma resolução espacial de dez metros. O erro RMS (*Root Mean Square*) foi calculado separadamente para cada modelo estereoscópico. Cada foto-mapa foi depois preparado usando o seu próprio DEM, como objetivo de reduzir o erro espacial. Os foto-mapas foram importados para o Erdas Imagine, para processo de classificação. Foi utilizado um processo de degradação (ERDAS função degradar) para definir o tamanho final do pixel em 30 centímetros em ambos os foto-mapas.

Ampliando dez vezes a fotografia de 1992, foi possível identificar manchas de árvores, arbustos e vegetação herbácea no campo e definidos os seus limites na fotografia ampliada. Nesta escala, até os arbustos pequenos puderam ser identificados. As assinaturas das três classes de vegetação definidas foram obtidas a partir de uma variedade de regiões ao longo da área de estudo. Para a imagem de 1964 não foi possível identificar elementos de vegetação no campo. Os conjuntos de treino para a fotografia de 1964 foram preparados usando uma *plotter* estéreo digital IMD (Image-Station 6400, Integraph Inc.), o qual permitia, no ecrã, a medição de altura de arbustos e árvores individuais. Os elementos da vegetação identificados desta forma foram usados para definir assinaturas para a respetiva categoria.

Ajustes de iluminação

A principal fonte de iluminação irregular em imagens provenientes de sensores remotos é o efeito topográfico. Neste estudo foi usado o processo de normalização topográfica, baseado num modelo Minnaert [42], para corrigir a inclinação e efeitos de brilho. Os resultados preliminares indicaram que este processo melhorou a precisão da classificação em regiões específicas nas fotografias, mas diminuiu a precisão em outras regiões. O seu efeito global foi de uma ligeira redução na precisão da classificação, pelo que não foi aplicado às imagens. Observou-se, ainda, um aumento gradual nos níveis de cinzentos de objetos semelhantes (árvores, arbustos e rochas nuas), do centro para a periferia da foto, presumivelmente devido a efeitos de reflexos direcionais. Este efeito não poderia ser detetado numa inspeção visual da fotografia, mas teve um efeito crucial sobre os resultados da classificação. Os fenómenos aparecem em ambas as imagens, mas foi mais pronunciado na fotografia de 1992. Foram usados dois métodos *standard* (transformação de Fourier e ajustes de brilho) para corrigir este efeito, mas nenhum deles corrigiu a imagem com sucesso. Por isso, foi desenvolvido um algoritmo de correção simples, baseado na distância do centro de fotografia. Foram identificadas, no campo, 51 árvores de carvalho que estavam espalhados sobre toda a imagem 1992, e digitalizadas as suas fronteiras na imagem. Para cada árvore foi calculada a distância e o azimute do centro da imagem, e a média dos seus valores de níveis de cinzento (usando o editor IMAGINE). Um gráfico do nível de cinzento, medido em função da distância a partir do centro, indicou a existência de uma

³Processo de correção da imagem, pixel por pixel, das distorções decorrentes do relevo, transformando a imagem de uma projeção cónica para uma perspetiva ortogonal

relação linear entre estes parâmetros (relação não existente para distâncias superiores a 600 metros). Por meio de regressão linear, foi determinado o efeito da distância ao centro das suas médias do nível de cinza das árvores, para distâncias superiores 600 metros. Foram calculadas, separadamente, equações de regressão para as árvores em direções norte e noroeste e para todas as outras direções juntas. Os parâmetros de regressão foram usados para calcular um fator de ajuste para cada pixel com base na sua distância e direção do centro da imagem. Foi utilizada a linguagem Arco-Info AML para escrever um programa para essa correção, que utilizou o módulo GRID Arc-Info como plataforma. A imagem de 1964 foi corrigida de uma maneira semelhante.

Métodos de classificação

Neste estudo foram comparados dois métodos de classificação de imagens: O *Maximum-likelihood supervised classifier*, foi utilizado para as assinaturas de árvores, arbustos e vegetação herbácea, derivadas dos conjuntos de treino; o segundo método foi um algoritmo que responde também em função das características da vizinhança de píxeis individuais - *Nearest Neighbour classifier*.

Os histogramas das três classes de vegetação indicaram que, em geral, as árvores apareciam mais escuras do que os arbustos, os quais, por sua vez, eram mais escuras do que as ervas. No entanto, existia uma sobreposição considerável entre os valores de nível de cinza de classes vizinhas (por exemplo: árvores e arbustos), indicando que qualquer método utilizado resultaria em grandes erros de classificação do intervalo de sobreposição. A abordagem *Nearest Neighbour classifier* tenta reduzir esses erros, tendo em conta as informações dos píxeis com níveis de cinzento no alcance da sobreposição. Em primeiro lugar, a imagem foi classificada em três classes:

Classe 1: píxeis que pertencem a 'árvore';

Classe 2: píxeis pertencentes a 'árvore' ou 'não-árvore';

Classe 3 - píxeis 'não-árvore'.

Comparando a imagem classificada com o mapa de vegetação do campo base, concluiu-se que a maioria dos aglomerados de árvores na imagem (mais de 99%) tinha alguns píxeis da classe 1 dentro dos limites da copa da árvore. Foi, então, usada uma janela circular de foco mínimo, para marcar como píxeis de árvores de classe 2 os que estavam numa proximidade própria para píxeis de classe 1. A proximidade própria foi definida como o raio da copa de uma única árvore pequena. Foi aplicado o mesmo processo para distinguir arbustos e ervas. Os píxeis que não foram classificados como árvores ou arbustos foram, então, classificados como vegetação herbácea.

Base de dados do terreno

Em 1994, foram preparados mapas detalhados da vegetação do terreno, para cinco parcelas de 200x200 metros. Estas parcelas foram ampliadas dez vezes na imagem de 1992 para identificar os elementos da vegetação no campo e delinear os limites nas ampliações. Foi medido cada objeto da imagem 1964, e delineados os seus limites sobre a fotografia ampliada. Tal como na imagem de 1992, foram preparados mapas do terreno da vegetação em cinco parcelas de 200x200 metros.

Avaliação da precisão da classificação

A precisão de cada método de classificação foi avaliada utilizando uma amostragem aleatória de pelo menos 80 pontos em cada um dos cinco lotes de terreno. A unidade de amostragem

definida para a avaliação da exatidão foi um conjunto homogêneo de 3x3 píxeis (0.9x0.9 metros). Dados de todas as parcelas foram então agrupadas para a construção de uma matriz de erro único para cada ano. A matriz de erro, a precisão geral, e *k-statistic* (Congalton [7]) foram determinados para cada classificação. O teste exato de Fisher (Sokal et Rohlf [41]) foi aplicado a uma tabela de pontos corretamente *versus* incorretamente classificados, em cada método, para testar se as diferenças entre os dois métodos eram estatisticamente significativas.

2.1.4.2 Resultados

Na imagem de 1992, a precisão média global foi 81% e 89% para os *Maximum-likelihood Supervised classifier* e *Nearest Neighbour classifier*, respectivamente. A precisão média da imagem de 1964 foi de 54% e 82% para *Maximum-likelihood Supervised classifier* e *Nearest Neighbour classifier*, respectivamente. As diferenças de precisão geral entre as parcelas não foram significativas para todas as imagens e classificadores.

A classificação usando o *Nearest Neighbour classifier* foi significativamente superior à classificação usando o *Maximum-likelihood Supervised classifier*. A precisão da classificação para a imagem de 1992 foi significativamente mais elevada do que a de 1964. As matrizes de erro para a imagem 1992 revelaram que as maiores fontes de erro no método *Maximum-likelihood Supervised classifier* foram a classificação de árvores como arbustos e de arbustos como vegetação herbácea. O método *Nearest Neighbour classifier* reduziu eficazmente ambos os tipos de erro. A imagem de 1964 foi caracterizada por maiores sobreposições entre as assinaturas espectrais de classes de vegetação em comparação com a imagem de 1992. As maiores fontes de erro decorrente do uso do método *Maximum-likelihood Supervised classifier* da imagem de 1964 foram a classificação de arbustos como ervas e vice-versa. O método *Nearest Neighbour classifier* reduziu drasticamente esses erros de classificação.

2.1.5 Utilização de Dados LIDAR para classificação de Áreas Florestais

O processo de utilização de sensores remotos tem sido reconhecido como uma ferramenta eficiente para inventário florestal, pois permite observar dados precisos e prever dados futuros de grandes áreas florestais. *Scanners a laser* aerotransportados forneceram informações sobre a geometria da floresta e câmaras digitais aéreas forneceram informações sobre a sua radiometria.

Normalmente, podem ser distinguidas diferentes espécies de árvores utilizando várias bandas na parte visual e perto do infravermelho do espectro eletromagnético, tendo em conta que diferentes espécies têm características dos espectros de refletância⁴ diferentes. No entanto, os valores de refletância de duas plantas diferentes, por vezes, são muito semelhantes, o que levou ao problema em diferenciar certas classes de vegetação.

Diferentes espécies de madeira têm diferentes estruturas e geometrias. Plantas com refletância espectral semelhante podem ter uma estrutura diferente e a informação geométrica dos instrumentos LIDAR pode ser usada para melhorar os resultados da classificação.

Neste estudo feito por Juho Lumme [23], a digitalização a *laser* foi feito por um aparelho Op-*tech* ALTM 2033, em junho de 2004. Foram feitos vários voos e a densidade de pontos foi de mais de dez pontos por metro quadrado. Imagens aéreas foram tiradas utilizando um aparelho Vexcel UltraCam-D, capaz de detetar a radiação infravermelha e visual. A altura de voo foi de 600 metros e o tamanho do pixel foi cerca de vinte centímetros.

⁴Relação entre o fluxo de radiação que incide numa superfície e o fluxo de radiação que é refletido

Os dados do *scanner a laser* e das imagens aéreas da UltraCam foram investigados neste estudo com objetivo de combinar os diferentes conjuntos de dados e realizar uma classificação supervisionada usando ambos os dados (espectro e geometria).

2.1.5.1 Métodos

Área de teste:

Foram feitas medições aéreas com *scanner a laser* e câmara aérea em Espoonlahti, no sul da Finlândia, ordenada pelo Instituto Geodésico finlandês. A área de teste continha principalmente área urbana e florestas mistas.

Levantamentos de campo:

Foram obtidas fotografias e outros dados do campo terrestre durante o verão de 2005 e foram usados como uma base de dados de referência para a classificação. O inventário da área de treino incluiu diferentes características artificiais e espécies de madeira. Por exemplo, o abeto da Noruega, o pinheiro e a bétula.

Trabalhos preliminares:

Foram feitas correções geométrica de imagens aéreas usando o *software* PCI Geomatica. Foram anteriormente medidos pontos de controlo em terra pelo Instituto Geodésico finlandês. Os pontos do *scanner a laser* foram convertidos para dados rastreados e foram extraídos vários recursos diferentes das imagens *a laser*. Todas as imagens (*laser* e aéreas) originais foram examinadas para que as características coincidissem.

2.1.5.2 Resultados

Foram calculadas medidas de separabilidade entre cada classe usando a distância de Bhattacharya⁵. Os resultados das medições de separabilidade foram ligeiramente melhores quando se usaram características LIDAR com imagens aéreas. Houve problemas com árvores de folha caduca. A sua refletância e geometria eram semelhantes e as diferentes espécies misturaram-se. A classificação de imagens foi feita usando o algoritmo Bayesiano⁶. A área foi classificada nas seguintes classes: 'estrada', 'estrada de terra', 'construção', 'relva', 'areia', 'abeto da Noruega', 'pinheiro' e 'bétula'. A classificação foi realizada utilizando dois conjuntos de características diferentes. A primeira classificação foi realizada com quatro canais de imagem aérea e segunda classificação foi realizada utilizando quatro canais de imagem aérea e oito características de imagens LIDAR.

Foram calculadas as precisões médias e globais quando os dados foram classificados de acordo com quatro canais de imagem aérea e oito características de imagens LIDAR. A precisão e média global foram cerca de 10% melhores quando as características do espectro e geométricas foram usadas em conjunto.

2.1.6 Utilização de Dados LIDAR para classificação de Áreas Agrícolas

Com o principal objetivo de distinguir as árvores que são cultivadas numa determinada área, foi realizado um estudo [34] para classificação de três espécies (oliveira, amendoeira e alfarrobeira). Foram usadas imagens digitais aéreas e LIDAR para obter uma técnica eficiente para análise e atualização do LPIS (*Land Parcel Information System*) estabelecido pelo regulamento Council (EC) 1782/2003, para atribuição de subsídios à agricultura.

⁵Mede a semelhança entre duas distribuições de probabilidade discretas ou contínuas

⁶Expressa a probabilidade condicional de um acontecimento aleatório A dado B

Foram tiradas fotografias aéreas digitais com uma câmara de mapeamento digital (DMC) no inverno e no verão e foram adquiridos dados LIDAR com uma ALTM Optech 3025 no inverno, com uma densidade de um ponto por metro quadrado. Foram detetadas árvores individuais a partir de dados *laser* usando o método hidrográfico. As imagens derivadas de LIDAR, juntas, com 4 canais de imagem (R, G, B e IR), das duas épocas foram a entrada para uma série de testes de classificação de objetos feitos com o *software* de análise de imagem eCognition⁷. Em todos os ensaios, a segmentação do objeto foi baseada na imagem da altura das árvores (derivada de LIDAR). Os melhores resultados de classificação foram obtidos utilizando apenas imagens DMC ortorretificadas numa resolução reduzida, após a segmentação da árvore a partir de dados LIDAR. Os resultados usando apenas dados LIDAR também se mostraram aceitáveis. A área de estudo está localizada a sudoeste de Vila-Rodona (60 quilómetros a sudoeste de Barcelona) com uma extensão de 6x6 quilómetros. É uma região seca do Mediterrâneo, onde as culturas mais importantes são vinhas, oliveiras, alfarrobeiras e amendoeiras.

Os dados de imagens digitais multi-temporais foram adquiridos ao longo da área de estudo com o avião Cessna Caravan da Cartogràfic Institut de Catalunya (ICC). Todos os voos foram realizados em 2007. Para mapear a área do projeto foi usada uma DMC, em 17 de janeiro e em 19 de julho, 1500 metros acima do nível do solo, resultando em 70 imagens para cada época. Foram criados dois tipos de imagens a partir de dados DMC que combinam três dos quatro canais de cores, gerando o ficheiro de cor verdadeira (azul, verde e vermelho) e o ficheiro de cores IRC (canais verde, vermelho e infravermelho).

Nas primeiras etapas do processamento do conjunto de imagens DMC e dados LIDAR, notou-se que a ortorretificação tradicional das imagens não era satisfatória. À medida que a altura das árvores não foi tida em consideração, as coroas apareceram deslocadas e houve uma incompatibilidade entre imagens de épocas diferentes. O problema foi resolvido retificando as imagens com o modelo digital de superfície (DSM), incluindo modelos de árvore em que a árvore foi representada como um cilindro.

Houve um total de 24 faixas, dois campos de controlo e três áreas de teste com cobertura vegetal diferente. Após algumas correções de erros sistemáticos, os pontos LIDAR foram classificados em pontos terrestres e não-terrestres com o TerraScan e o TerraModeler de Terrasolid. Após a classificação automática, não foi realizada nenhuma alteração manual. A precisão do modelo de terreno resultante foi verificada com pontos medidos com GPS / RTK em três testes de áreas com cobertura diferente.

Um modelo digital de terreno (DTM) com um metro por célula de grelha foi calculado por interpolação no modelo TIN⁸ e foi calculado um modelo digital de superfície (DSM) a partir de todos os pontos LIDAR para cada célula da grelha na elevação do interior do maior ponto LIDAR. A partir da diferença entre o DSM e DTM, foi calculado um modelo da altura da copa (CHM) que foi utilizado para detetar as árvores individuais com o método hídrico. A maioria das árvores nesta área estava isolada e considerou-se que cada copa terminava quando a altura alcançada era de um terço da altura total da árvore. Deste modo, foi possível gerar uma imagem em que todos os píxeis dentro da coroa de uma árvore tinham o mesmo valor e que correspondia à altura da árvore. Neste modelo, cada árvore é representada como um cilindro generalizado. Cada árvore recebeu um identificador exclusivo. Na maioria dos casos, as árvores automaticamente detetadas correspondiam a árvores reais, mas, por vezes, a coroa de uma árvore dava a ilusão de serem duas e terem, então, sido considerados duas árvores diferentes pelo procedimento automático. Em todos os testes, a segmentação da imagem foi realizada com uma imagem

⁷*Software* cuja principal característica é a análise e classificação em grupos de píxeis vizinhos

⁸Representa uma superfície como um conjunto de triângulos contíguos, que não se sobrepõem

derivada a partir do CHM.

A partir da elevação dos pontos, foram calculados os valores mínimos e máximos, a média, o desvio padrão e o coeficiente de variação⁹. A partir da intensidade dos dados LIDAR foram também calculados a média, o desvio padrão e o coeficiente de variação.

A área de estudo era caracterizada pela existência de espécies mistas, mas também existiam plantações homogêneas. O conjunto das árvores foi dividido em dois grupos independentes: um para ser usado como o conjunto de treino e o outro para ser usado como um conjunto de teste para análise do desempenho do método. As amostras do campo concentraram-se numa pequena região e os testes de classificação foram restritos a uma área retangular.

A alturas das árvores a partir de medições de campo foram comparados diretamente com alturas LIDAR e verificou-se uma boa concordância com os valores de medição de campo, mas a dispersão foi mais elevada para as amendoeiras. Esta é uma árvore de folha caduca e, no momento da obtenção de dados LIDAR, não tinha folhas. Apesar disso, as amendoeiras foram detetadas somente com dados LIDAR.

Os dados LIDAR tenderam a superestimar a área da coroa de oliveiras e alfarrobeiras. As oliveiras e amendoeiras mostraram uma área um pouco maior a partir de dados LIDAR do que o valor real. A altura das árvores foi superestimada em catorze casos e subestimada em dezanove. A área da coroa foi superestimada em 21 casos e subestimado em 14.

2.1.6.1 Métodos

Um dos principais problemas foi a copa das árvores por serem compostas de píxeis discretos, cobrindo uma vasta gama de valores do espetro. Neste estudo, os objetos de imagem correspondentes às coroas foram apontados com eCognition. Num primeiro passo, chamado segmentação multi-resolução, a imagem foi dividida em regiões homogêneas com base num subconjunto de variáveis selecionadas pelo utilizador e em vários parâmetros definidos pelo utilizador. Uma vez segmentada a imagem, foi possível começar a classificação destes segmentos com base num conjunto de variáveis selecionado pelo utilizador. Depois de alguns testes, a ideia de individualizar as coroas como segmentos foi realizada utilizando dados LIDAR da altura das árvores como a única variável no processo de segmentação. Com essas configurações, a maioria das copas das árvores foram identificadas como um segmento independente da sua área.

Foram realizados 27 ensaios eCognition diferentes com várias combinações dos dados com dados de DMC: inverno, verão e multi-temporal. Cada um destes grupos foi analisado em diferentes tamanhos de píxeis: 0.25, 0.5 e 1 metro, foram agrupados em três modos diferentes: RGB, IRG (infravermelhos) e os canais todos juntos. Por último, foi analisada a utilização combinada de variáveis LIDAR e dados DMC. No entanto, em todos os testes, a segmentação da imagem foi baseada apenas na altura das árvores.

Para cada projeto eCognition, foram classificados todos os segmentos em três categorias de acordo com a variável altura da árvore: vegetação baixa (menos de 26 centímetros), árvores (26 centímetros) e objetos altos (mais de dez metros). Todos os segmentos da árvore foram classificadas como oliveiras, amendoeiras ou árvores de alfarroba de acordo com os seus valores para as mesmas variáveis, pelo vizinho mais próximo.

⁹Rácio entre o desvio padrão e a média de uma variável

2.1.6.2 Resultados

Num primeiro teste, os nove derivados de dados LIDAR foram considerados e foi realizada uma análise de otimização para detetar as variáveis mais explicativas. Tendo em consideração apenas as três melhores, os resultados foram próximos aos obtidos com todos os nove parâmetros. No entanto, os resultados de classificação foram pobres, ou mesmo muito pobres para amendoeiras. Provavelmente porque as amendoeiras não tinham folhas quando os dados LIDAR foram adquiridos e a sua estrutura não poderia ser adequadamente definida.

Tendo em conta todos os parâmetros dos dados LIDAR, o desempenho da classificação melhorou muito, as oliveiras foram corretamente identificados em 92% dos casos, mas para as outras espécies a taxa de sucesso foi inferior a 70%. A discriminação entre as espécies atingiu o seu máximo quando foram utilizadas dez variáveis. Os resultados foram bons para oliveiras.

O desempenho da classificação melhorou quando as imagens multi-temporais de DMC foram incorporadas aos dados LIDAR, especialmente com a combinação de canais infravermelhos, verde e vermelho de cada época. A melhoria foi mais significativa para as alfarrobas (de 58% para 74%) e para as amendoeiras (de 68% a 98%) e menor para oliveiras (de 89% para 91%), pois já tinham sido alcançados bons resultados com os dados LIDAR.

Os dados DMC também foram analisados separadamente dos LIDAR. Nesses testes foram usadas apenas as imagens da digitalização *laser* CHM durante a fase de segmentação. As imagens DMC são uma combinação de oito imagens diferentes: um mosaico de quatro imagens espectrais e quatro imagens de baixa resolução (Hinz, 1999 [16]).

O canal de infravermelhos foi decisivo para detetar alfarrobeiras e o canal azul foi importante para reduzir a confusão entre as oliveiras e as alfarrobas em imagens de verão.

Os resultados de apenas uma época eram mais pobres para as oliveiras do que aqueles obtidos a partir de imagens multi-temporais, mas as imagens de inverno foram fundamentais para discriminar amendoeiras.

2.1.7 Classificação com Base em Descritores Locais

O *Object-Based Image Analysis* (OBIA) é um método popular que recorre a sensores remotos de alta resolução. Esta técnica baseia-se na extração de muitos tipos diferentes de características dos objetos, tais como características do espectro, forma e textura. Neste artigo [49], é apresentado o *Bag-of-Visual-Words* (BOV) para o processamento de imagem recorrendo a sensores remotos para descrição do objeto e sua classificação.

O método BOV, originalmente para categorização de texto, atraiu muita atenção no campo da categorização genérica visual para construção das representações de nível médio, em vez de características de baixo nível. Na representação BOW, um documento de texto é codificado como um histograma do número de ocorrências de cada palavra selecionada. Da mesma forma, podemos caracterizar uma imagem por um histograma da contagem de palavras visuais. O vocabulário visual fornece um nível médio de representação.

A ideia básica do método BOV é extrair várias sub-regiões distintas (*patches*) utilizadas para calcular descritores locais. Tradicionalmente, estas sub-regiões podem ser detetadas pelos detetores de interesse do ponto escolhido, pela sua repetibilidade e propriedades invariantes em certa transformação ou grelhas uniformemente regulares a partir de toda a imagem. Qualquer que seja a estratégia usada, o objetivo é reduzir a grande quantidade de informação transportada pelas imagens e a concentrar-se sobre as partes importantes.

O BOV permite extrair o local bem visível através da simulação do sistema visual humano. Neste

trabalho são propostos os descritores de atenção baseados em locais para a modelagem das regiões mais importantes. A abordagem BOV pode ser processada por modelos de descritores baseados nos locais. Este método foi testado num conjunto de dados de imagem VHR (*Very High Resolution*) aérea, em estudo.

Extração de regiões salientes

O algoritmo *attention region* é baseado na implementação do modelo de Itti de atenção visual [17], o qual deteta as regiões salientes da imagem. O modelo de Itti extrai características para a cor, intensidade e orientação, e constrói um mapa de saliência com a integração de mapas intermédios de característica por competição das mesmas. Uma rede *winner-take-all* (WTA) de neurónios *integrate-and-fire* seleciona locais vencedores, e um mecanismo de inibição de retorno permite ao modelo atender a muitos locais sucessivamente. Primeiro define-se um limite do valor de saliência ϵ , e em seguida, extraem-se todas as regiões de atenção, com elevado valor de saliência. Neste trabalho, é considerada uma janela grande de sub-região saliente, e são extraídas grelhas uniformemente regulares onde cada grelha é composta por intervalos de 11x11 píxeis para a sub-janela. Os passos do algoritmo de extração de região saliente são os seguintes:

1. Construção das pirâmides multi-escala gaussiana da imagem de entrada;
2. Calculo as características de competição em três canais das pirâmides gaussianas e obtenção dos mapas de características;
3. Normalização e definição de limites para compor cada mapa de características e mapas visíveis no mapa de saliência;
4. Extração das grelhas uniformemente regulares no mapa de saliência.

2.1.7.1 Representação BOV

A representação BOV consiste em quatro etapas. Na sequência da deteção da região saliente, é usada uma combinação de características do espetro e de textura para descrever as regiões salientes. Ou seja, são geradas características para cada grelha regular de forma uniforme em duas categorias: média e desvio padrão calculado a partir do canal de i valores de 11x11 píxeis e texturas GLCM¹⁰ de três canais de cor. São calculadas quatro diferentes direções para gerar a matriz: 45, 90, 135 e 180. São, assim, extraídas 48 características da GLCM. No total, são calculadas 54 características em cada *patch*.

Dada a coleção de manchas detetadas a partir dos objetos de treino de todas as categorias e os seus descritores locais, aprende-se o vocabulário visual. A nova representação de cada objeto é calculada pela contagem de histogramas. Nesta construção do histograma, é usada a distância euclidiana para medir a similaridade entre descritores e palavras de correção visual. Cada *patch* é atribuído à palavra visualmente mais semelhante e, por conseguinte, uma característica nova é gerada para cada objeto através da contagem do número de ocorrências de cada palavra visual. Em detalhe, uma imagem extrai uma grande quantidade de pontos-chave pelo *DoG's detector*, enquanto o modelo saliência apenas seleciona três regiões interessantes contendo a informação mais significativa na imagem.

2.1.7.2 Classificação com *Support Vector Machine*

Uma vez que os descritores de baixo nível foram atribuídos ao vocabulário visual para formar os novos vetores de características, o problema da categorização de objeto através de sensores

¹⁰Matriz de co-ocorrência de níveis de cinza

remotos é reduzido ao de multi-classe supervisionado. Neste trabalho, foi utilizado o método SVMs (*Support Vector Machine*), uma vez que apresenta um bom desempenho em problemas de alta dimensão.

2.1.7.3 Experiências e Resultados

Foram realizadas as experiências e avaliado o desempenho do método proposto na imagem aérea VHR. As imagens aéreas eram compostas por três bandas multi-espectrais (RGB) com 0.25 metros de resolução. Todas as imagens foram corrigidas geometricamente antes da segmentação. Na experiência, foi possível distinguir quatro tipos de cobertura do solo que dominam a área de estudo: 'culturas' (417), 'árvores' (518), 'lagoa' (448) e 'área residencial' (499). Foi usada a interpretação manual como uma referência para o algoritmo de classificação. Objetos dentro de cada classe foram divididos aleatoriamente em conjunto de treino e conjunto de testes. Cada experiência foi repetida cinco vezes, com divisão aleatória, e relatada a média dos resultados obtidos ao longo de cinco conjunto de teste diferentes. Foi possível extrair os *patches attention-based* do conjunto de treino e utilizar os descritores locais para representar esses *patches*. Foi construído um vocabulário com 450 palavras visuais, e depois calculado um histograma contando o número de ocorrências das palavras visuais.

As experiências mostram que, quando aumenta o número de objetos de treino, a exatidão de classificação de SVMs com base em três algoritmos diferentes aumenta conforme o esperado. A melhor exatidão geral da linha de base 2 é de 88.35%, enquanto o algoritmo *attention-based* de descritores locais é de 92,88%. A comparação mostra que a abordagem aqui descrita supera o método BOV padrão.

É de notar que esta abordagem extrai a região saliência como *patches* em vez das grelhas uniformemente regulares em toda a imagem, o que, obviamente, pode reduzir o custo de processamento. Esta abordagem pode obter melhores resultados com menos tempo e complexidade.

2.1.8 Bag-of-Visual-Words

O método *Bag-of-Visual-Words* (BOV) foi inspirado na abordagem *Bag-Of-Words* (BOW) para categorização de texto, como referido anteriormente. No método BOW, um documento de texto é codificado como um histograma do número de ocorrências de cada palavra selecionada. Um texto é representado como uma coleção não-ordenada de palavras. Da mesma forma, pode caracterizar-se uma imagem por um histograma de número de palavras visuais. O vocabulário visual fornece uma representação de nível médio, que ajuda a colmatar a enorme lacuna semântica entre as características de baixo nível extraídas de uma imagem e os conceitos de alto nível a serem categorizados. Zhu et al., [20] estendeu o livro de códigos de palavras-chave a partir do domínio de texto para domínio da imagem e introduziu a quantização vetorial de pequenas janelas de imagens quadradas, chamado *keyblocks*. Com este artigo [50], Sheng Xu, Tao Fang, Deren Li e Shiwei Wang mostram que este método pode produzir melhores resultados semanticamente orientados do que as tradicionais características de baixo nível, como espectro e textura.

Abordagem de classificação BOV: Dado um objeto a partir do conjunto de dados de amostra, a detecção do *patch* e a descrição são usados para formar um conjunto de vetores com as características de um objeto. Na fase de treino, o método *k-means* é aplicado para treinar *k* classes cujo centro é nomeado como palavras visuais e, em fase de testes, uma única palavra virtual é criada para representar todos os *patches* implausíveis que não estão perto o suficiente para

justificar a representação por quaisquer palavras visuais relevantes. Com base nas palavras visuais, o histograma é gerado por contagem de números das suas ocorrências. O histograma é definido como uma representação BOV. Esta nova característica é utilizada para a classificação.

2.1.8.1 Detecção da Região Local e Descritores Locais

As regiões locais podem ser extraídas de duas maneiras diferentes:

Uniformemente em grelhas regulares: são extraídos em diferentes escalas, onde cada grelha é um intervalo de 11x11 píxeis para um determinado objeto. O tamanho do *patch* é de amostragem aleatória entre a escala de 10 a 30 píxeis;

Detetor de *DoG Lowe*, onde é extraído um conjunto de regiões locais (*patches*) estáveis e invariantes em escalas diferentes usando o detetor *DoG*. Assim, os pontos notáveis estão localizados, e suas vizinhanças são consideradas como as marcas detetadas para descrição adicional.

O conteúdo de objetos geoespaciais em OBIA é frequentemente descrito usando vários tipos de características de espectro e de textura. Nesta experiência, também é apresentada uma combinação de característica espectral e textura uniformemente distribuídas em grelhas regulares como a entrada da abordagem BOV. O recurso combinado é composto por duas categorias de componentes: as médias e desvios padrão de três canais (RGB) e 48 texturas processadas a partir de 12 matrizes de níveis de cinza de co-ocorrência (GLCMs) que são gerados a partir de quatro direções diferentes (45, 90, 135 e 180) em cada canal RGB, respetivamente. Todas as funções foram normalizados para o intervalo [0, 1] por normalização mínimo-máximo.

2.1.8.2 Construção de Vocabulário Visual

A questão-chave do método BOV centra-se em como construir as palavras visuais automaticamente a partir de um conjunto de treino. O vocabulário visual oferece uma maneira de construir um vetor de características inovadoras para classificação relativa de novos descritores em objetos de características de consulta de baixo nível do conjunto de treino. Com base na deteção e descrição da região local, os objetos são separados num conjunto de marcas, cada uma das quais é descrita por um vetor de características. Assim, todo o objeto é descrito por um conjunto de descritores de *patch* e, em seguida, todos os descritores de *patch* dos objetos formam um conjunto de dados. As classes de agrupamento k são aprendidas a partir do conjunto de dados de treino, e os seus centros são definidos como as palavras visuais. A construção visual de vocabulário é um método não supervisionado, sem qualquer referência ao rótulo da classe de cada *patch*, que faz a geração de etiquetas de palavras visuais livres.

2.1.8.3 Palavra Visual e Histograma

Com base no algoritmo de agrupamento, um vocabulário visual é construído para descrever o conteúdo do objeto. Cada marca é atribuída à palavra visual mais próxima usando a distância euclidiana, e um objeto pode ser representado como um histograma através da contagem do número de ocorrência das palavras visuais. Este método baseia-se no pressuposto de que os dados de teste devem pertencer a uma de todas as classes de formação, e cada *patch* é bem representado pela sua única palavra visual mais próxima. No entanto, se os dados de teste são de classes de amostra, os *patches*, muitos deles discrepantes, não podem ser atribuídos a qualquer candidato adequado no vocabulário. É apresentado um limite para reduzir o impacto das discrepâncias no processo de teste, e as marcas, cuja distância à palavra visual mais próxima é maior do que este princípio adaptativo, irá ser atribuído a uma única palavra visual virtual. Usando

este princípio de processamento de histograma, os dados de teste podem gerar características distintas. A ocorrência da palavra virtual é muito mais frequente em valores discrepantes do que em outros. Durante o treino, com base na suposição anterior, estes limites foram definidos para a distância máxima de todos os *patches* no mesmo *cluster* para o seu centro.

2.1.8.4 Resultados Experimentais

Neste artigo, são utilizados SVMs para a classificação, com base no método BOV. O método *one-against-all* é explorado para resolver o problema multi-classe, e o algoritmo *leave-one-out* é usado para selecionar os parâmetros do *kernel* da *Radial Basis Function* (RBF). A exatidão geral e o coeficiente *Kappa* com base na matriz de confusão, é utilizada para avaliar os desempenhos de classificação. Foram utilizadas imagens aéreas, adquiridas a partir do Condado de Shanghai, China, em junho de 2006, compostas por três bandas multi-espetrais (RGB) com 0.25 metros de resolução. São distinguidos quatro tipos de cobertura do solo que dominam a área de estudo: 'área residencial', 'culturas', 'floresta', 'lagoa', num total de 1.882 objetos dessas quatro classes, segmentada pelo *software eCognition*. Foi usada interpretação manual como referência para o algoritmo de classificação. Os objetos dentro de cada classe foram divididos aleatoriamente num conjunto de treino e num conjunto de testes. Cada experiência foi repetida cinco vezes, com divisão aleatória e relato dos resultados médios obtidos sobre o conjunto dos cinco testes diferentes.

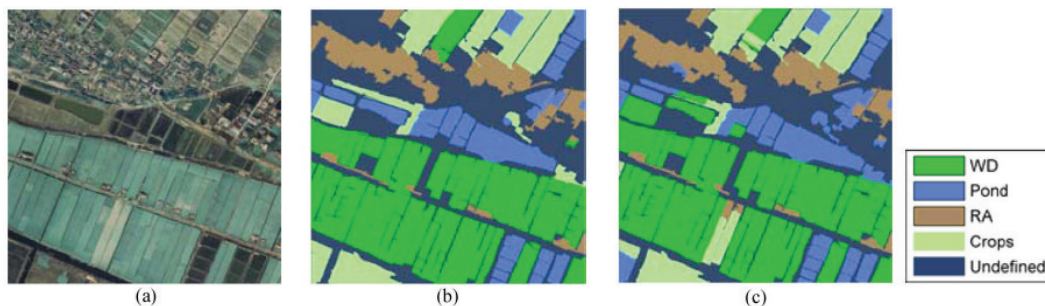


Figura 2.1: (a) Área experimental, (b) Mapa de referência, (c) Mapa de classificação baseado na representação BOV

2.1.8.5 Resultados baseados na classificação BOV

A abordagem de base utiliza 54 características combinadas sobre o nível do objeto, em vez de uma representação com base no histograma de palavras visuais.

As melhorias da precisão de classificação e coeficiente *Kappa* são principalmente notórias nas classes 'área residencial' e 'floresta'. Por exemplo, existe uma melhoria de 2.4% na classe 'área residencial' e uma melhoria de 3.73% na classe 'floresta'. Estes detalhes verificam que a melhoria da exatidão global é óbvia nesses objetos complexos e compostos. A representação BOV proposta pode descrever os objetos composto e complexos de forma mais eficaz.

2.1.8.6 Influência do Tamanho do Conjunto de Treino

A precisão de classificação com base na representação BOV e as características de baixo nível têm tendência decrescente similar quando o tamanho da amostra de formação em cada categoria reduz de 250 para 50. No entanto, a exatidão da classificação baseada na BOV supera o das

características de baixo nível. O desempenho com base na representação BOV é mais confiável quando o número do conjunto de dados de treino é pequeno.

2.1.8.7 Influência da seleção de recursos na representação BOV

Este método combina características de espectro e de textura para descrever os *patches* e construir a representação do histograma, em vez do SIFT (*Scale Invariant Feature Transforms*), normalmente utilizado. Para uma análise completa do efeito das diferentes características de baixo nível, foi comparado o desempenho em quatro diferentes recursos, isto é, as características do espectro, GLCM, características combinadas e SIFT. Observou-se que, quando o número do conjunto de treino é 250 e o número de vocabulário visual é 450, a precisão global obtida pela representação base BOV nas características combinadas, as características espectrais, GLCM, e SIFT, é de 93.81%, 86.60%, 91.40%, e 87.61%, respectivamente. Os resultados indicam que a combinação de características do espectro e de textura é a melhor escolha para a abordagem proposta, ao passo que com o SIFT não é possível obter um resultado satisfatório. O ponto fraco da SIFT é provavelmente devido ao facto de ser apenas calculada a orientação e gradiente de pontos salientes, o que nem sempre é útil para detecção remota de imagens. Da mesma forma, apenas as características do espectro e de textura não oferecem informações suficientes para a classificação exata. Assim, podemos afirmar que o recurso combinado é mais adequado.

2.1.8.8 Influência do parâmetro k

O mais importante para a abordagem proposta é a construção do vocabulário visual usando o *clustering* não supervisionado, como o método de *k-means clustering*. Nesta abordagem, cada palavra visual é definida como um centro do *cluster*. Assim, a influência do tamanho das palavras visuais sobre a representação BOV na classificação é estimada, e é definido o parâmetro no *k-means* para 100, 150, 200, 250, 300, 350, 400, e 450, respectivamente. Quando aumenta o número de *clustering centers*, a exatidão da classificação geral aumenta também e os resultados variam de acordo com os valores iniciais. Esta instabilidade é porque a convergência do *k-means* não é garantida para a melhor solução, e as palavras visuais variam entre diferentes configurações iniciais. No entanto, o intervalo de confiança de 95% diminui claramente conforme o parâmetro k aumenta. Isso implica que a estabilidade da classificação será melhorada quando o parâmetro k aumenta. Assim, pode verificar-se que a estabilidade da representação BOV na classificação pode ser melhorada usando mais palavras visuais.

2.1.8.9 Influência das *outliers*

Foram, intencionalmente, adicionados vinte objetos de classes não definidas, para os dados de teste definidos, para avaliar a sua influência na classificação. Apenas três *outliers* são classificadas incorretamente e os outros dezassete objetos implausíveis são rejeitados.

2.1.9 Quaternion Orientation Difference

O propósito geral da classificação de imagens com base nos recursos locais, tais como (SIFT) são comuns hoje em dia. Isto é compreensível visto que as características locais têm algumas propriedades desejáveis, que permitem, em certa medida, para a compensação dos efeitos adversos, tais como variações da iluminação e perspectiva e oclusão. As características locais

também dão bons resultados na classificação da textura. No entanto, em domínios especializados, outros descritores podem mostrar melhores resultados.

A textura desempenha um papel fundamental na análise de imagens provenientes de sensores remotos. Foi demonstrado recentemente que os descritores de textura de Gabor produzem um desempenho razoável em grandes conjuntos de dados de imagens de textura.

As principais contribuições deste trabalho [31] são os descritores de orientação obtidos a partir das respostas de filtros de Gabor em orientações diferentes e na mesma escala, e a sua extensão para imagens coloridas obtidas com o *quaternion framework* para a representação da cor da imagem. Foram testados os descritores propostos num conjunto de dados com vinte e uma imagens aéreas e demonstrado que os resultados melhoram consideravelmente em comparação com descritores de imagens locais e globais.

2.1.9.1 Descritor *Quaternion Orientation Difference*

O objetivo foi projetar um descritor de imagem colorida usando a estrutura *quaternion* para a representação de imagens coloridas. A parte da resposta *simplex* corresponde à resposta do filtro de Gabor para o sinal de luminância, e parte *perplex* corresponde à sua resposta ao sinal de crominância.

2.1.9.2 Resultados

Para a classificação foi usado o SVM com uma variante do *kernel* da RBF, que utiliza métrica de distância em vez de métrica euclidiana: $K(Fa; Fb) = \exp[-d(Fa; Fb)]$

A classificação multi-classe é obtida por formação de um SVMs *one-vs-all* para todas as classes e classifica uma amostra de teste para a classe que corresponde à resposta máxima SVM.

Foi testado o descritor proposto num conjunto de dados de imagens aéreas. Todas as imagens eram multi-espetrais (RGB), de 256x256 píxeis. Foram manualmente classificadas em vinte e uma classes, correspondentes a cobertura do solo e vários tipos de uso de solo: 'agricultura', 'campo de basebol', 'praia', 'edifícios', 'chaparral', 'zona residencial', 'floresta', 'estrada', 'campo de golfe', 'porto', 'parque', 'viaduto', 'estacionamento', 'rio', 'pista', etc. Cada classe continha 100 imagens. Para a extração de características em todas as experiências foram utilizados filtros Gabor em quatro escalas e seis orientações. Foi primeiro testado o descritor *Orientation difference* para escala de cinza. Consequentemente, todas as imagens foram convertidas em tons de cinza. Para fins de comparação, foram incluídos também os resultados para descritores de textura, assim como médias e desvios padrão das diferenças entre as respostas de filtro em diferentes escalas e mesma orientação. Foi feita a média das taxas de classificação a cinco conjuntos de dados aleatórios e diferentes. Foi possível verificar que os descritores com base nas diferenças entre as respostas de filtro para diferentes orientações superaram todos os outros descritores.

Para fins de comparação, foram também calculados dois descritores de textura e cor. O primeiro é o descritor *color opponent*, aumentado com desvios padrão das diferenças de respostas de filtro em diferentes escalas e a mesma orientação. O segundo foi o descritor *orientation difference* para imagens de cor.

Os resultados foram comparados com os resultados de classificadores baseados em BOV, bem como os BOV com os *spatial cooccurrence kernel*. Foi realizada validação cruzada (cinco vezes) em 80% das imagens no conjunto de treino, e em 20% das imagens no conjunto de teste. O descritor *orientation difference* para tons de cinza superou as abordagens baseadas em BOV,

também baseados em tons de cinza. Incluindo informações de cor, o melhor desempenho foi obtido com os descritores *quaternion orientation difference* e *color orientation difference*. Os descritores de imagem baseados em correlações cruzadas entre as respostas dos filtros de Gabor em diferentes orientações têm um bom desempenho na classificação de imagens aéreas e incluindo informação de cor verifica-se uma melhoria considerável.

Para o descritor proposto, focado na textura, os resultados obtidos para classes orientadas à textura (por exemplo: 'praia', 'chaparral', 'floresta') provaram ser muito bons. No entanto, para classes orientadas a objetos (como 'edifícios' ou 'campos de ténis'), as taxas de classificação são mais baixas.

2.1.10 Classificação usando Descritores de Gabor

Existem muitas aplicações de sensores remotos de imagens, tanto militares como civis. As aplicações civis incluem o planeamento do uso da terra, previsão do tempo, estudo das mudanças climáticas a longo prazo, monitorização de culturas, planeamento da cidade, e muitos outros. Estas aplicações requerem o desenvolvimento de meios eficazes para a aquisição, transmissão, processamento, armazenamento, análise e recuperação de imagens.

O resultado de uma segmentação ao nível dos píxeis é um mapa temático em que a cada píxel é atribuído um rótulo predefinido a partir de um conjunto finito. No entanto, as imagens provenientes de sensores remotos são muitas vezes multi-espectrais e de alta resolução o que torna a sua segmentação semântica detalhada excessivamente exigente. Esta é a razão pela qual alguns pesquisadores decidiram classificar blocos de imagens, em vez de píxeis individuais. Esse procedimento foi também adotado na abordagem [30] a seguir descrita e onde foram avaliados os classificadores com base no estado dos descritores da imagem e SVMs, que têm mostrado bons resultados em tarefas de classificação de imagem aérea. Foram avaliados dois descritores de imagem, ambos com base em filtros de Gabor. Existe uma longa tradição na utilização de descritores de Gabor no processamento de imagem. Posteriormente, os descritores de Gabor têm sido usados para várias tarefas, incluindo segmentação de textura, reconhecimento de imagem, reconhecimento de íris e registo e controlo de movimento.

2.1.10.1 Descritor de Gabor

O descritor de Gabor para uma imagem é calculado pela passagem da imagem por meio de uma base de dados de filtros de Gabor. O filtro de Gabor é um filtro linear cuja resposta ao impulso é definida como uma função gaussiana modulada com uma senoide complexa.

2.1.10.2 Descritor GIST

Oliva e Torralba [26] propuseram os descritores GIST para representar o *spatial envelope*¹¹ da imagem. O *spatial envelope* é um conjunto de propriedades holísticas que podem ser utilizadas para deduzir a categoria semântica da imagem, sem a necessidade de reconhecimento dos objetos. O descritor de GIST de uma imagem é calculado pela filtragem da imagem por uma base de filtros de Gabor, e depois calculando a média das respostas dos filtros pertencentes a cada bloco de uma grelha 4x4. Comparando este descritor com o descritor de Gabor, vemos que o descritor GIST é essencialmente uma disposição espacial de texturas. Note-se que os desvios padrão e a distribuição de respostas de filtro não são utilizados. Apesar da sua simplicidade, este descritor mostra resultados muito bons em tarefas de classificação de imagens.

¹¹Representação da imagem em dimensões muito reduzidas

2.1.10.3 Classificador *Support Vector Machine*

A partir de distâncias de descritores de Gabor foi construída uma função de *kernel*. Esta função é essencialmente baseada numa norma ponderada e satisfaz a condição de Mercer. Para o descritor GIST foi usado o SVM com o *kernel* de função base radial. Foi construído um classificador multi-classe utilizando N (correspondendo ao número de categorias) de *one-vs-all* SVMs e selecionada a classe de saída SVM máxima.

2.1.10.4 Conjuntos de dados e resultados experimentais

As imagens foram testadas com classificadores em dois conjuntos de dados (imagens aéreas). O primeiro conjunto de dados (conjunto de dados *In-House*) continha imagens de parte de Banja Luka, na Bósnia e Herzegovina. O segundo conjunto de dados continha imagens usadas anteriormente para a classificação imagem aérea, que foram incluídas para fins de comparação.

Conjunto de dados *In-House*

Foi utilizada uma imagem multi-espectral (RGB) com 4500x6000 píxeis da parte de Banja Luka, na Bósnia e Herzegovina para a avaliação dos classificadores. Esta imagem continha uma grande variedade de estruturas, tais como edifícios, fábricas e armazéns, campos, árvores e rios. A imagem foi dividida em quadrados de 128x128 píxeis, e foi usado um total de 606 imagens na experiência. Todas as imagens foram classificadas manualmente em seis categorias: 'casas', 'cemitério', 'indústria', 'campo', 'rio' e 'árvores'. Metade das imagens foi usada para treino e a outra metade para teste. Foram calculados os descritores de Gabor em oito escalas e oito orientações para todas as imagens do conjunto de dados. Foram procuradas também outras combinações de números de escalas e orientações e foram escolhidos os que apresentavam melhor desempenho. Foram calculados os descritores de Gabor para as três bandas espectrais da imagem, e concatenados os vetores obtidos, que produziram $3 \times 8 \times 8 \times 2 = 384$ descritores. Para fins de comparação, também foram calculados os descritores de Gabor para tons de cinza (pancromática), $8 \times 8 \times 2 = 128$.

Quanto aos descritores GIST, foram obtidos os melhores resultados com a configuração padrão, ou seja, um conjunto de filtros em 4 escalas e 8 orientações. Para estes descritores também foi calculado o tom de cinza, $4 \times 8 \times 16 = 512$, e de cor, o que resulta em $3 \times 4 \times 8 \times 3 = 1536$ descritores. Para testar os classificadores foi utilizada a validação cruzada (dez vezes), com partições aleatórias diferente do conjunto de dados, e a média dos resultados.

De salientar que o descritor Gabor, muito mais simples, com dimensionalidade quatro vezes menor, produz um desempenho semelhante ao do descritor de GIST. Para imagens em tons de cinza (pancromática) o descritor Gabor supera o GIST. Quando as imagens em tons de cinza são consideradas, os desvios padrão de respostas de filtro de Gabor fornecem informações mais importantes sobre a textura da imagem, daí o seu melhor desempenho. Outra conclusão é que a disposição espacial das respostas do conjunto de filtros não tem influência benéfica sobre o desempenho do classificador de imagem aérea.

Matriz confusão - Descritores de Gabor: As confusões surgem principalmente entre as categorias que podem ser de difícil diferenciação, mesmo para os seres humanos. Não há muitas confusões entre as categorias naturais ('rio', 'árvores', 'campo') e artificiais ('casas', 'cemitério', 'indústria').

Matriz confusão - Descritores GIST: As observações para a matriz de confusão dos descritores Gabor também são válidas aqui.

Conjunto de dados do Reino Unido

Para a segunda experiência foi escolhido um conjunto de dados do Reino Unido. Este conjunto

de dados contempla 1040 imagens aéreas de 64x64 píxeis, que foram manualmente classificadas nas seguintes oito categorias: 'construção', 'estrada', 'rio', 'campo', 'erva/relva', 'árvore', 'barco', 'veículo'. O conjunto foi dividido em conjunto de treino e teste de 520 imagens cada. Para este conjunto de dados, as imagens foram processadas pelos descritores de Gabor em oito escalas e oito orientações, bem como pelo descritor GIST, e formado, depois, um classificador multi-classe. Verificou-se, novamente, alguma vantagem para os descritores de Gabor, tal como na primeira experiência.

Na utilização do Descritor Gabor ocorrem erros comuns de classificação em casos que também podem potencialmente confundir os seres humanos, como a 'construção' com 'veículo' e 'campo' com 'relva'. Raramente ocorrem erros de classificação entre categorias naturais e artificiais.

2.1.11 Extração de Estruturas de Estacionamento - Interpolação, Extrapolação, Previsão de blocos e Filtragem

A extração automática de estruturas de estacionamento é um desafio, porque as suas formas são apenas imagens aproximadamente regulares. As formas de estacionamentos nas imagens parecem semelhantes, mas, na realidade, são um pouco diferentes. A estrutura de cada lote de estacionamento tem de ser analisada separadamente. O ruído nas imagens faz com que os lugares de estacionamento tenham uma aparência inconsistente devido à ocupação do veículo, oclusões por outras estruturas, tais como árvores e edifícios adjacentes, ou iluminações diferentes (por exemplo, sob a sombra dos edifícios).

A fim de lidar com esses problemas de forma eficaz, Young-Woo Seo, Nathan Ratliff e Chris Urmson [37], propuseram uma abordagem hierárquica para a geração e filtragem de hipóteses candidatas. As variações de iluminação, a oclusão e a geometria do local de estacionamento ditam o uso de abordagens de aprendizagem por meio de treino em exemplos extraídos da imagem de interesse. Esta abordagem é implementada em duas camadas: Uma camada de baixo nível que extrai e compila a meta-informação geométrica para facilmente encontrar vagas de estacionamento, com uma elevada precisão e serve como principal fonte de exemplos de auto-supervisão; Uma camada de alto nível que usa as saídas da camada de baixo nível para prever hipóteses candidatas plausíveis para locais pontuais de estacionamento mais difíceis e depois filtra estas hipóteses utilizando *self-trained learners*.

2.1.11.1 Método

A estrutura de um estacionamento numa imagem aérea é caracterizada pela disposição de um conjunto de blocos e pontos de estacionamento. A extração da estrutura de estacionamento é a deteção de todos os lugares de estacionamento visíveis. O algoritmo proposto parametriza cada vaga de estacionamento individual pela sua altura, largura, orientação e localização em coordenadas de imagem. Considera-se um bloco de estacionamento como uma fila de lugares de estacionamento. Cada bloco de estacionamento é caracterizado pela distância entre os pontos de estacionamento adjacentes no bloco. Os blocos de estacionamento estão relacionados uns com os outros por duas medidas de distância, a distância entre pontos de estacionamento conjugados e a distância entre os blocos.

Análise Low-Level: Deteção Canonical

As características geométricas das imagens diferem entre estacionamentos. Na análise de baixo nível são extraídos facilmente pontos para serem utilizados pela análise de alto nível como se-

mentes para a geração adicional de hipótese. A camada de baixo nível realiza várias etapas de processamento de imagem: a extração de linha, agrupamento de linha e previsão do bloco. As linhas retas são importantes para compreender a forma de um estacionamento. A extração das linhas foi feita usando a abordagem proposta por Kahn et al., 1990, [18]. A abordagem processa derivados de imagem para obter gradientes de intensidade em cada pixel e quantifica as instruções de gradiente usando intervalos predefinidos. É então utilizado um algoritmo para grupos de píxeis atribuídos à mesma direção de modo a formar regiões de linha de suporte. O primeiro principal vetor próprio de uma região da linha de apoio determina a direção da linha. Embora a maioria das linhas extraídas poderem alinhar com marcações do estacionamento, algumas delas vêm de regiões da imagem tais como faixas rodoviárias ou contornos de edifícios adjacentes. Apenas interessam as linhas alinhadas com as marcações do parque de estacionamento, por isso é necessário remover as linhas que não pertencem à estrutura de estacionamento. Para isso, deve preservar-se o grupo de linhas extraídas com base nas suas orientações e remover as linhas que são ou muito curtas ou muito longas.

Para a estimativa de parâmetros, primeiro deve estimar-se a altura nominal de um estacionamento no local mediante o cálculo do modo de cada linha no conjunto selecionado. A seguir, construir uma matriz de distância euclidiana entre todos os pares de linhas possíveis, quantificar as distâncias e calcular o modo de obter a largura e altura de vagas de estacionamento dentro de um lote. Finalmente, quantificar as orientações das linhas e calcular novamente o modo a estimar a orientação da extremidade aberta dos lugares de estacionamento. A conclusão dessas etapas de processamento de imagem resulta numa geração de poucos mas altamente precisos pontos de estacionamento verdadeiros.

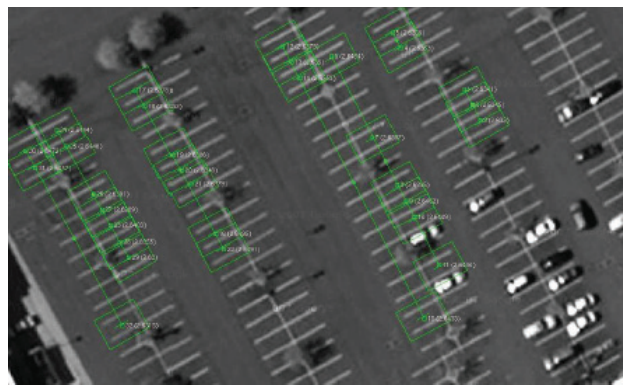


Figura 2.2: Análise *Low Level*

Análise *High-Level*: Interpolação, Extrapolação, Previsão de blocos e Filtragem

A camada de alto nível é destinada a detetar todos os lugares de estacionamento visíveis numa imagem. Para isso, primeiro é necessária a hipótese de lugares de estacionamento no local com base nos parâmetros obtidos pela camada de baixo nível. Em seguida, filtram-se essas hipóteses, classificando os *patches* retangulares em torno destas hipóteses usando classificadores auto-supervisionados.

A hipótese de vaga de estacionamento representa uma localização na imagem de uma potencial vaga de estacionamento. A hipótese de um *patch* retangular em torno da imagem é avaliada para determinar se as características locais da imagem são semelhantes às de um ponto de estacionamento real. Para cobrir o conjunto das regiões da imagem que possivelmente contêm verdadeiros pontos de estacionamento, são usadas as coordenadas da imagem dos pontos de

estacionamento iniciais para cada um dos blocos de estacionamento descobertos. Geram-se, então, as hipóteses de vaga de estacionamento por seleção de locais de imagem através de três processos: interpolação, extrapolação, e previsão do bloco. O procedimento de interpolação escolhe coordenadas de imagem entre dois pontos finais de estacionamento num bloco de estacionamento, o procedimento de extrapolação estende hipóteses além da fronteira dos dois pontos de estacionamento. A largura estimada de um lugar de estacionamento é usada como o intervalo espacial entre hipóteses do local de estacionamento. Finalmente, a previsão de bloco visa descobrir os blocos de estacionamento em falta.

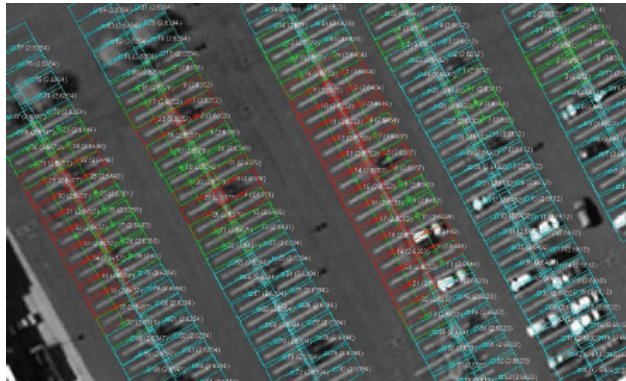


Figura 2.3: Análise *High-Level* - Interpolação (vermelho), Extrapolação (azul), Análise *Low-Level* (verde)

2.1.11.2 Resultados

O objetivo destas experiências era verificar o quão bem funcionam estes métodos para realizar a detecção de todos os pontos de estacionamento visíveis numa imagem.

Foram usadas treze imagens aéreas recolhidas a partir do *Google Maps*™. Há, em média, 147 lugares de estacionamento visíveis em cada imagem individual e um total de 1.912 lugares de estacionamento em todas as imagens aéreas.

O desempenho é calculado por meio da fusão de tabelas de contingência entre as treze imagens diferentes e então usada a tabela resultante para calcular as medidas de desempenho. Uma vez que a proposta inicial tinha uma taxa de falso positivo¹² menor do que 3%, as suas estimativas para a vaga de estacionamento são usadas como exemplos positivos para treino de todos os métodos de filtragem. Um número igual de exemplos negativos são gerados aleatoriamente. Apesar de ter quase zero falsos positivos, os pontos de estacionamento detetados pela análise de baixo nível cobre apenas 37.65% dos pontos de estacionamento verdadeiros (720 de 1.912 lugares de estacionamento verdadeiros.). Usando informações fornecidas pela análise de baixo nível, o alto nível de análise e geração de hipóteses reduz a taxa de falso negativo de 63.73% para 11.23%. No entanto, aumenta a taxa de falsos positivos para 38.12%. O estágio de filtragem corrige essa mudança na taxa de falsos positivos, eliminando hipóteses errôneas. No entanto, a taxa de falsos negativos na fase de geração de hipótese de análise de alto nível foi geralmente baixa e não prejudicou significativamente a precisão.

No primeiro cenário, foram treinados os métodos de filtragem, utilizando um conjunto de exemplos auto-supervisionados a partir da imagem, que consistiam na análise dos exemplos auto-rotulados positivos e negativos, gerados aleatoriamente. No segundo cenário, foram treinados esses métodos usando exemplos auto-supervisionados de todas as outras imagens não incluindo a imagem de destino. Finalmente, no último cenário foram treinados os métodos usando exem-

¹²local de não estacionamento, classificado como um lugar de estacionamento

plos auto-supervisionados de todas as imagens. Os exemplos negativos gerados aleatoriamente foram amostrados durante a execução de cada um desses cenários. Adicionalmente, foram geradas manualmente 1.079 marcas de vagas de estacionamento em todas as treze imagens. Foram refeitas as experiências usando esses locais de marcas manualmente extraídas do estacionamento. Valores positivos na coluna precisão indicaram melhorias de auto-aprendizagem sobre aprendizagem supervisionada. Surpreendentemente, o algoritmo realizado mostrou ser um pouco pior, quando treinado usando os exemplos de maior precisão gerados manualmente. Isto ocorre provavelmente porque a distribuição dos pontos de estacionamento de teste foi criada por esta abordagem de geração de hipóteses.

Idealmente, um método poderia alcançar duas menores taxas de falsos positivos e negativos, mas, na prática, é difícil de alcançar os dois simultaneamente. Os resultados mostram que a abordagem MRFs demonstra um desempenho superior na maioria das experiências. A abordagem MRFs utiliza interações de alto nível para melhorar a precisão da previsão.

2.2 Conclusões

Pela tese de doutoramento de Dev Ashish [3], é possível concluir que podem obter-se bons resultados utilizando uma metodologia de classificação baseada em ANN de imagens aéreas em escala de cinza. Na experiência, a maior precisão global (92%) foi conseguida com base nos parâmetros de textura para a classificação, em comparação com as abordagens com base em informações de histograma e espacial. Este estudo estabeleceu ainda a importância da utilização de parâmetros de textura para a classificação de imagens de sensores remotos. O erro de classificação de 'água' na classe de 'campo escuro' pode ser resolvido com a análise de imagens multi-espetrais, análise essa que também forneceu melhores resultados para outras classes.

Ainda relativamente à textura, as suas características são baseadas em estatísticas que resumem a distribuição da frequência relativa, ou seja, que descrevem o quão frequente um tom de cinzento aparece numa determinada relação espacial para outro tom de cinzento, na imagem. Essas características foram usadas por Haralick et al [14] nas tarefas de identificação de categorias de três diferentes tipos de imagens. A precisão de identificação nos conjuntos de testes independentes foi de 89% para o conjunto de microfotografias (cinco categorias), 82% para fotografias aéreas (oito categorias), e 83% para imagens de satélite (sete categorias).

De [6], de Carmel e Ronen, pode concluir-se que o *Nearest Neighbour classifier* é menos sensível à sobreposição entre as classes do que o classificador *Maximum-likelihood Supervised*, e, portanto, pode ser especialmente útil nos casos de uma única banda de imagens, onde existem grandes sobreposições de espectro entre as classes. A classificação computadorizada de fotografias aéreas pancromáticas é raramente realizada. Short & Short [38] utilizou a análise de imagem para identificar árvores de carvalho numa fotografia pancromática. O método de classificação não foi especificado, mas parece ter contornos simples. A resolução espacial foi de 3 metros, a precisão espacial e a precisão de classificação não foram avaliadas. Neste estudo, foram aplicados ajustes de iluminação, que compensam o gradiente de valores cinzentos nas fotografias, antes de classificações computadorizadas das fotografias aéreas. O modelo Minnaert usado para corrigir efeitos de inclinação e aspeto sobre a radiação não melhorou os resultados da classificação. Este estudo demonstra que o uso de classificação da vegetação computadorizada de fotografias aéreas permite produzir mapas de vegetação com alta resolução espacial, de alta precisão espacial e cobrir uma extensão espacial relativamente grande.

Em [23], de Juho Lumme, as medidas de separabilidade provaram que diferentes espécies de

madeira são mais fáceis de distinguir, quando as imagens aéreas e os dados LIDAR são utilizados em conjunto. Os resultados após a classificação bayesiana usando conjuntos de características diferentes, foram semelhantes. As características espectrais e geométricas, juntas, conduzem a melhores resultados.

Relativamente a [34], também realizado por Juho Lumme, é possível concluir que, com o emprego dos parâmetros e densidade de pontos, os dados LIDAR não foram suficientes para discriminar as espécies. Os resultados obtidos usando apenas a câmara de mapeamento digital foram semelhantes ou mesmo melhores. A altura obtida pelos dados LIDAR permitiu individualizar as árvores com elevada fiabilidade, provavelmente porque as árvores foram bem isoladas no campo. A combinação dos dados com imagens multi-espectrais simplificou a análise e evitou a confusão entre as árvores e outros com resposta espectral semelhante.

Em [49], Sheng Xua et al. introduziram um descritor local *attention-based* para representar o objeto. Este modelo computacional foi biologicamente inspirado, e extraiu uniformemente grelhas regulares das regiões mais importantes. Com base nas manchas detetadas, foi usada uma combinação de características espectrais e de textura. Os resultados experimentais sugerem que o melhor desempenho de tais imagens aéreas de alta resolução foi obtido utilizando a representação BOV.

Em [50], também de Sheng Xua, é introduzida uma representação simples e útil para conteúdo VHR de imagens aéreas. A representação BOV é gerada para superar o problema de como descrever com precisão objetos complexos em imagens de resolução muito alta. Os resultados experimentais sugerem um bom desempenho dessa representação em imagens aéreas VHR, em comparação com as características de baixo nível nos resultados de classificação. Além disso, o estudo sobre a classificação da imagem aérea mostra que a representação BOV, numa combinação de características de espectro e de textura, supera significativamente a precisão da classificação em comparação com a representação com base SIFT. Os resultados experimentais mostram também que a representação BOV é insensível ao impacto de *outliers*.

No trabalho realizado por Vladimir Risojevic e Zdenka Babic, 'Orientation Difference Descriptor for Aerial Image Classification' [31], é proposto um descritor de orientação a partir das respostas de filtros de Gabor em orientações diferentes e na mesma escala. Os resultados obtidos foram bastante melhores do que os obtidos utilizando o *Bag-of-Visual-Words*, embora esta abordagem possa não ser a melhor estratégia para domínios especializados.

No trabalho 'Gabor Descriptors for Aerial Image', também de Vladimir Risojevic e Zdenka Babic, em conjunto com Snjezana Momic, foram avaliados dois descritores de imagem, Gabor e GIST, que mostraram resultados semelhantes ou melhores do que as abordagens mais complexas. Ambos os descritores já haviam demonstrado bons resultados em tarefas de classificação de textura. Foi demonstrado que, para imagens multi-espectrais de menores dimensões, os descritores de Gabor apresentam desempenho semelhante ou melhor desempenho que GIST, enquanto que, para imagens pancromáticas, os descritores de Gabor superam os de GIST. Esta diferença deve-se ao fato de a distribuição espacial não ser tão forte para classificação semântica das imagens aéreas, mas as suas regiões de textura são bastante homogêneas. Além disso, os descritores de Gabor usam desvios padrão de respostas do conjunto de filtros, e esta representação é mais rica e é uma outra razão para o seu melhor desempenho.

Relativamente à abordagem de Young-Woo Seo, Nathan Ratliff e Chris Urmson [37], esta propõe um algoritmo de duas camadas hierárquicas para análise da estrutura de parques de estacionamento vistos em imagens de satélite aéreas. A camada de análise de baixo nível extrai um conjunto de lugares de estacionamento facilmente detetados e blocos de estimativas de estacionamento utilizando a deteção de linha e técnicas de agrupamento. A análise de alto nível, em

seguida, estende esses pontos usando características geométricas das estruturas típicas do lote de estacionamento para interpolar e extrapolar novas hipóteses e usa uma máquina de aprendizagem auto-supervisionada de técnicas para filtrar falsos positivos nas hipóteses propostas. As experiências mostraram que o treino dos classificadores usando o conjunto de lugares de estacionamento auto-supervisionado se adapta, com sucesso, à fase de filtro para as características específicas da imagem em análise. Dados anteriores recolhidos do local de estacionamento em várias outras imagens aéreas fornecem informações importantes de oclusões e variações de iluminação não encontradas nos pontos de estacionamento extraídos pela análise de baixo nível, resultando num aumento de desempenho.

Capítulo 3

Método Proposto: Extração de Características

Este trabalho utiliza imagens *truecolor*, em formato PNG, obtidas a partir das aplicações *Google Earth™*, 3.1 e *Google Maps™*, 3.2, com 690 píxeis de largura e 546 de altura (690x546), a uma altitude de visualização entre um e dois quilómetros. Todas as imagens utilizadas pertencem ao território de Portugal continental.

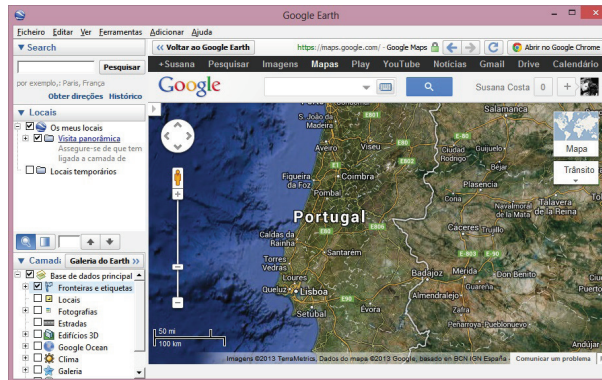


Figura 3.1: Google Earth

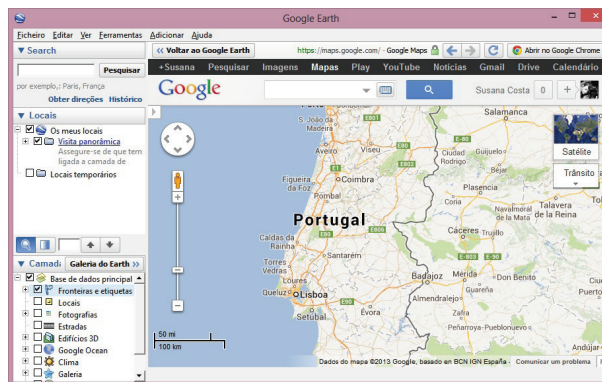


Figura 3.2: Google Maps

Uma imagem digital é representada como uma função bidimensional $f(x, y)$. A intensidade da luz refletida ou emitida por uma cena é representada por uma função f associada a uma determinada posição (x, y) da imagem. Geralmente, estas informações são guardadas em matrizes, em que (x, y) representa a posição linha e coluna na matriz e o valor associado a este ponto, o pixel correspondente.

O Matlab®¹, 3.3 disponibiliza uma linguagem de programação própria e fornece um ambiente de computação com excelentes capacidades gráficas e com um vasto conjunto de funções, organizadas segundo diversas áreas científicas [10]. O elemento de dados básico é a matriz e,

¹MATrix LABoratory - software interativo de alta performance voltado para o cálculo numérico

como tal, o armazenamento de imagens é feito em matrizes onde cada elemento da matriz corresponde a um único pixel na imagem.

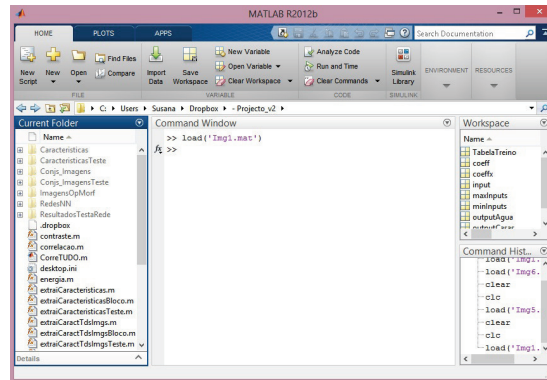


Figura 3.3: Maltlab

O Matlab® é uma importante ferramenta para efetuar o processamento da imagem digital e muito utilizada em investigação, por essa razão o escolhido para a realização deste trabalho.

3.1 Cor/Intensidade

3.1.1 RGB

A percepção da cor é uma experiência complexa e envolve aspetos que são alvo de estudo em várias disciplinas. Sempre que se observa um objeto e se distingue uma determinada cor, o olho humano não está necessariamente a detetar apenas uma frequência da luz. Acontece geralmente que um objeto está a refletir ou transmitir várias frequências da luz em simultâneo, com intensidades distintas, mas o cérebro interpreta estas frequências, que atingem o olho humano, como se fosse, por exemplo, a cor amarela.

A percepção humana da cor deriva da resposta dos cones (células que constituem o olho humano) à presença de três diferentes tipos de comprimento de onda. Existe um tipo de cone que deteta e responde à cor vermelha, um outro que deteta e responde à cor verde e um terceiro que deteta e responde à cor azul. Qualquer outra cor pode ser produzida através da mistura das quantidades adequadas de luz vermelho, verde e azul. Por isso, as cores R, G e B designam-se por cores primárias aditivas. O modelo de cores RGB é, baseado na teoria de visão colorida tricromática, de Young-Helmholtz [13], e no triângulo de cores de Maxwell [43].

RGB é, então, a abreviatura do sistema de cores aditivas formado por Vermelho (Red), Verde (Green) e Azul (Blue). O propósito principal do sistema RGB é a reprodução de cores em dispositivos eletrónicos como monitores de computador e televisões. Este modelo adequa-se à geração de qualquer cor quando há emissão de luz.

Na prática, a grande maioria das cores visíveis no mundo real estão presentes na gama de cores RGB, pelo que este modelo proporciona uma forma útil, simples e eficiente de as representar. As cores são, então, representadas por três valores (RGB), que correspondem à proporção de luz pura de vermelho, verde e azul que, quando adicionadas, constituem a cor resultante. Por exemplo, se exprimirmos as proporções de luz pura por intermédio de percentagens:

(100%, 0%, 0%) - Cor primária vermelha pura;

(100%, 100%, 100%) - Cor branca;

(0%, 0%, 0%) - Cor preta.

Utilizando o Matlab®, cada imagem do *Google Earth™* foi decomposta nos seus três canais R, G, B, dando assim origem a três vetores com a informação de cada um dos canais.

3.1.2 HSV

Este sistema de cores foi proposto por Alvy Ray Smith (1974) [39] e caracteriza-se por ser uma transformação não-linear do sistema de cores RGB. Outros sistemas de cores relacionados incluem o HSL (L de *luminosity* ou luminosidade) e o HSI (I de *intensity* ou intensidade).

HSV é a abreviatura para o sistema de cores formadas pelas componentes *Hue* (matiz), *Saturation* (saturação) e *Value* (valor). O HSV também é conhecido como HSB (*Hue, Saturation e Brightness*) matiz, saturação e brilho, respetivamente). Esse sistema de cores define o espaço de cor, utilizando os três parâmetros:

- Matiz (tonalidade): Verifica o tipo de cor, abrangendo todas as cores do espectro, desde o vermelho até o violeta e o magenta. Atinge valores de 0 a 360, mas para algumas aplicações, esse valor é normalizado de 0 a 100%.
- Saturação: Também chamado de 'pureza'. Quanto menor for esse valor, com mais tom de cinza aparecerá a imagem. Quanto maior o valor, mais 'pura' é a imagem. Atinge valores de 0 a 100%.
- Valor (brilho): Define o brilho da cor. Atinge valores de 0 a 100%.

No Matlab®, procedeu-se à decomposição dos três canais, H, S e V, dando, assim, origem a um vetor com a informação relativa à matiz, um para a saturação e um para o brilho.

3.1.3 Escala de Cinzentos

Numa imagem digital em escala de cinzentos o valor de cada pixel corresponde a uma única amostra de um espaço de cores. Estas imagens são compostas por diversos tons de cinza, variando entre o preto como a menor intensidade e o branco como a maior intensidade.

As imagens em escala de cinzentos são geralmente resultado do cálculo da intensidade da luz em cada pixel em cada faixa do espectro electromagnético. Computacionalmente, as imagens monocromáticas são armazenadas utilizando-se oito bits por pixel, o que permite 256 intensidades possíveis, normalmente numa escala não linear.

Neste trabalho, a imagem do *Google Maps™* foi convertida em escala de cinzentos, e colocada num vetor a respetiva informação.

3.2 Matriz de co-ocorrência de tons cinza

Neste trabalho, as características relativas à textura são extraídas a partir de uma matriz de co-ocorrência ou GLCM proposta por Haralick et al. [14]. Haipeng Yu et al. [51] e Zhu et al. [20] são exemplos de autores que utilizam a matriz de co-ocorrência para calcular a textura das imagens, através de características como contraste, energia, correlação e homogeneidade. GLCM é uma matriz bidimensional, gerada a partir da contagem das ocorrências das combinações entre os tons de cinza do pixel analisado e um pixel vizinho. Para cada uma das combinações possíveis entre o pixel e o pixel vizinho (normalmente o pixel à direita, por defeito), é criado um elemento na matriz de co-ocorrência de tons de cinza, de forma que o tamanho da matriz de coocorrência será de $t \times t$, onde t é o número de tons presentes na imagem original.

Neste trabalho, foi utilizada a matriz de co-ocorrência de tons de cinza, aplicada em blocos de três por três píxeis, para cada um dos canais RGB da imagem do *Google Earth*™, e para o canal cinza da imagem do *Google Maps*™, com o objetivo de obter os vetores de características relativos ao contraste, à energia, à correlação e à homogeneidade. Resultaram deste operação dezasseis vetores, quatro para cada uma das características.

3.2.1 Contraste

A função contraste retorna uma medida do contraste entre as intensidades de um pixel analisado e do pixel vizinho. A comparação é realizada em todos os píxeis da imagem. Para uma imagem constante (mesmo tom de cinza em toda a extensão), o contraste é zero. O contraste pode ser calculado através da fórmula 3.1

$$\sum_{i,j} (i - j)^2 p(i, j) \quad (3.1)$$

Onde $p(i, j)$ é a probabilidade de ocorrência de pares de píxeis da imagem, i o nível de cinza do pixel de interesse da matriz e j o nível de cinza do pixel vizinho ao nível de interesse. Da aplicação da função contraste aos canais RGB da imagem do *Google Earth*™ e à imagem do *Google Maps*™ em escala de cinzentos, resultaram quatro vetores de características.

3.2.2 Energia

Conhecida também como Segundo Momento Angular, calcula a uniformidade da textura na matriz de co-ocorrência de tons de cinza, ou seja, avalia as repetições de pares de píxeis em tons de cinza de uma imagem. A energia possui valor 1 para uma imagem constante, onde estão presentes apenas píxeis similares, ou 0 para uma imagem heterogênea, quando há grande disparidade entre os píxeis.

Em Matlab®, a função energia, com a fórmula matemática em 3.2, retorna a soma dos elementos elevados ao quadrado dentro da matriz de co-ocorrência de tons de cinza.

$$\sum_{i,j} p(i, j)^2 \quad (3.2)$$

A função energia foi aplicada, em blocos de 3x3 píxeis, a cada um dos canais R, G e B da imagem do *Google Earth*™ e à imagem do *Google Maps*™ em escala de cinzentos, resultando num total de quatro vetores de características.

3.2.3 Correlação

Em Matlab®, a função correlação, com a fórmula matemática em 3.3, retorna uma medida de quão correlacionado está um pixel com o seu vizinho. A comparação é realizada em todos os píxeis da imagem. A correlação é 1 para uma imagem totalmente correlacionada ou -1 para uma completamente descorrelacionada.

$$\sum_{i,j} \frac{(i - \mu_i)(j - \mu_j) \vec{p}(i,j)}{\sigma_i \sigma_j} \quad (3.3)$$

Onde μ_i e μ_j , representam as médias nas direções i e j , respectivamente, e $\sigma_i \sigma_j$ representam os desvios padrão.

Os valores relativos à correlação foram extraídos em blocos de 9x9 píxeis de cada um dos canais R, G e B da imagem do *Google Earth™*, resultando num total de três vetores de características.

3.2.4 Homogeneidade

Em Matlab®, a função homogeneidade, com a fórmula matemática 3.4, retorna um valor que representa a proximidade da distribuição dos elementos em relação à diagonal da matriz de coocorrência dos tons de cinza. Um valor de homogeneidade 1 representa uma matriz diagonal de co-ocorrência de tons de cinza.

$$\sum_{i,j} \frac{p(i,j)}{1 + (i - j)} \quad (3.4)$$

A função homogeneidade foi aplicada, em blocos de 3x3 píxeis, a cada um dos canais R, G e B da imagem do *Google Earth™* e à imagem do *Google Maps™* em escala de cinzentos, resultando num total de quatro vetores de características.

3.3 Função Entropia

Em Física, a entropia de um sistema é a medida da sua 'desordem', [25].

A entropia de uma imagem pode ser definida como um número quantificador da randomicidade da imagem, ou seja, quanto maior for este número, mais irregular, atípica ou não-padronizada será a imagem analisada. Por exemplo, para uma imagem com um único tom de cinzento em toda a sua extensão, temos que sua entropia é igual a zero.

No Matlab®, o cálculo da entropia é feito através da função $E = entropy(I)$, onde I é uma imagem em escala de cinzentos.

A entropia é definida pela equação:

$$\sum (p \cdot \log_2(p)) \quad (3.5)$$

Sendo p a probabilidade para uma dada intensidade.

Os valores relativos à entropia foram extraídos em blocos de 3x3 píxeis de cada um dos canais R, G e B da imagem do *Google Earth™* e da imagem do *Google Maps™* em escala de cinzentos, resultando num total de quatro vetores de características.

3.4 Descritores de Gabor

Desenvolvidos por Dennis Gabor (1946) [11], consistem numa variante local da transformada de Fourier, são um conjunto de funções Gaussianas moduladas por funções sinusoidais complexas,

bidimensionais que apresentam certas propriedades extremamente úteis para fins de classificação de imagens. Os filtros ou descritores de Gabor são bastante eficientes no processo na análise de textura, por terem a capacidade de caracterizar um sinal simultaneamente nos domínios temporal e das frequências, que são limitados pela relação de incerteza conjunta, ou princípio de Heisenberg:

$$(\Delta t)(\Delta \omega) \geq \frac{1}{4}\pi \quad (3.6)$$

onde, Δt e $\Delta \omega$ representam incerteza nos domínios temporal e das frequências respetivamente. Gabor determinou a família de funções que atingem este limite inferior de incerteza conjunta, como sendo:

$$f(t) = \exp\left[-\frac{1}{2}\left(\frac{t}{\sigma}\right)^2 + i\omega t\right] \quad (3.7)$$

Essencialmente, a função determinada por Gabor descreve uma onda sinusoidal com frequência ω modulada por um envelope Gaussiano com duração σ . O conjunto original de filtros proposto por Gabor foi estendido para o caso bidimensional por Daugman [8], podendo, portanto ser aplicado para dados do tipo imagem. O filtro bidimensional é representado pela função:

$$f(x, y, u_0, v_0, \sigma_x, \sigma_y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left\{-\frac{1}{2}\left[\left(\frac{x}{\sigma_x}\right)^2 + \left(\frac{y}{\sigma_y}\right)^2\right]\right\} \exp\{2\pi i(u_0 + v_0)\} \quad (3.8)$$

Os filtros de Gabor são, então, bastante úteis para a extração de características de texturas [12], por isso, neste trabalho, a partir desta técnica, foram obtidos vinte e oito vetores de características. A partir de uma função que implementa o filtro desenvolvido por Daugman [8], foram calculadas as características fazendo variar a combinação de alguns dos parâmetros existentes no filtro, nomeadamente, os valores relativos à orientação (*orientation*) e os valores relativos ao comprimento de onda (*wavelength*). Os valores de orientação utilizados foram: 0, $\pi/4$, $\pi/2$ e $3\pi/4$; os valores para o comprimento de onda foram: 0.7, 1.4, 2, 4, 8, 16 e 32.

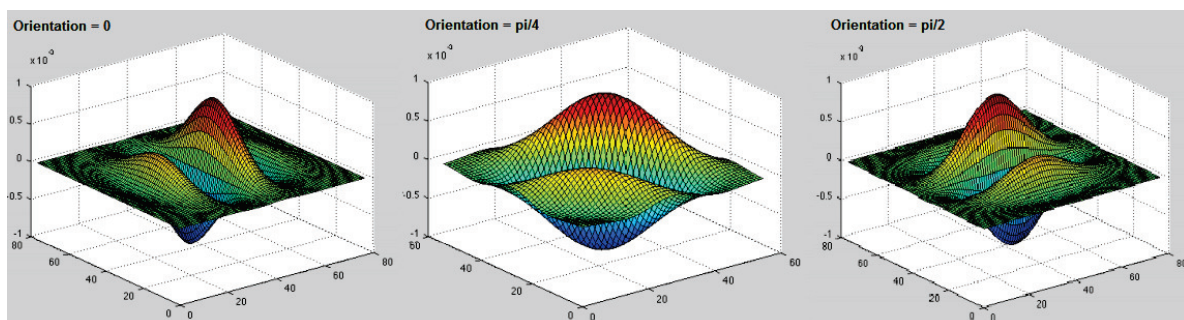


Figura 3.4: Exemplos da influência do parâmetro *orientation*

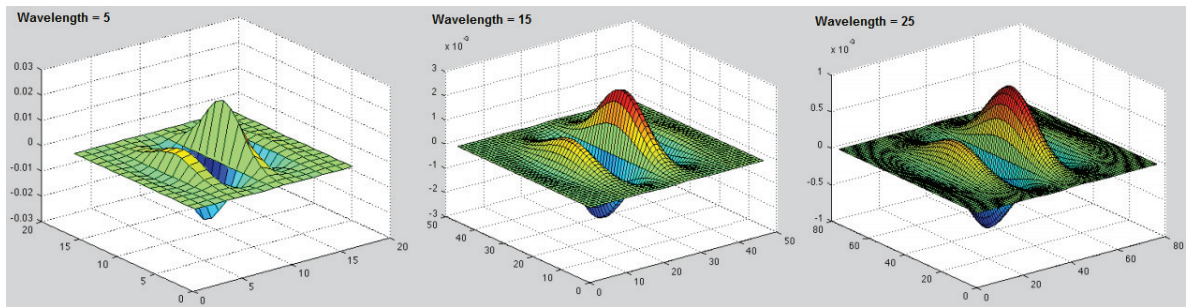


Figura 3.5: Exemplos da influência do parâmetro *wavelength*

3.5 Imagens 'Classe'

Para cada conjunto de imagens, foi criada, recorrendo à ferramenta *Photoshop*®, uma nova imagem, a partir da imagem do *Google Earth*™, onde a detecção de elementos foi feita manualmente, para ser usada como entrada na a rede neuronal.

A cada um dos elementos a ser detetado foi atribuído um tom de cinzento diferente. Posteriormente, no *Matlab*®, foram criados vetores com a informação da imagem, onde os píxeis sem interesse foram marcados com o valor zero e os píxeis com 'água', 'casas' ou 'estradas' com o valor um, resultando, assim, em três vetores, um para a classe 'água', um para a classe 'casas' e um outro para a classe 'estradas'.

3.6 Conclusão

Para cada conjunto de imagens, composto por uma imagem do *Google Earth*™ e uma imagem do *Google Maps*™, foi extraída uma matriz de características com 376740 linhas (número de píxeis de cada imagem) e 53 colunas quando utilizados todos os métodos mencionados neste capítulo. Para cada imagem 'classe' foram criados três vetores, um com a informação dos píxeis marcados manualmente como 'água', outro para os píxeis relativos à classe 'casas' e um terceiro com a informação dos píxeis marcados como 'estrada'. A dimensão de cada vetor é de 376740x1.

Capítulo 4

Método Proposto: Redução da dimensionalidade (PCA)

4.1 Normalização dos Dados

Uma vez extraídas as características definidas no capítulo anterior, é necessário pré-processar os dados, pois os intervalos entre as diversas características é diferente e corre-se o risco de haver características predominantes sobre as outras, devido a essa variabilidade relativa. Sendo assim, todos os dados constantes da matriz resultante da extração das características, foram enquadrados no intervalo [0 1], seguindo a regra $y = \frac{x - m}{M - m}$, onde y é o valor normalizado, x o valor inicial e M e m os valores Máximo e Mínimo da característica.

Além disso, segundo LeCun (1998), [21], a convergência do processo de classificação será mais rápida se o valor médio de cada característica, calculado sobre todo o conjunto de treino, for próximo de zero.

4.2 Análise de Componentes Principais

A Análise de Componentes Principais ou *Principal Component Analysis* (PCA) é um dos métodos estatísticos mais usados quando se pretendem analisar dados multivariados. O procedimento matemático utiliza uma transformação ortogonal para converter um conjunto de observações de variáveis possivelmente correlacionadas num conjunto de valores de variáveis linearmente descorrelacionadas, chamadas componentes principais. O número de componentes principais é menor ou igual ao número de variáveis originais. Esta transformação é definida de forma a que o primeiro componente principal tenha a maior variância possível, e cada componente seguinte, por sua vez, tenha a máxima variância sob a restrição de ser ortogonal a, não-correlacionado com os componentes anteriores.

A PCA foi proposta por Karl Pearson [27]. Atualmente, é mais usado como uma ferramenta de análise exploratória de dados e para fazer modelos preditivos. Segundo Kendall [19], a PCA pode ser utilizada para várias tarefas, nomeadamente para a simplificação do conjunto de dados, ou seja, encontrar uma forma simplificada de representar o universo de estudo, podendo ocorrer através da transformação, sendo uma combinação linear ou não, de um conjunto de variáveis autónomas num outro conjunto autónomo e de inferior dimensão.

A redução da dimensionalidade consiste em diminuir o conjunto de características, sendo possível porque algumas das características podem não ser relevantes para a tomada de decisão, tornando possível identificar o número mínimo de características suficiente para descrever uma região ou identificar um objeto na imagem. Existem duas abordagens para a redução da dimensionalidade, a extração e a seleção de características. A extração cria novas características a partir de combinações ou transformações do conjunto original, enquanto a seleção de características visa determinar um subconjunto de características capazes de discriminar objetos através de regras previamente estabelecidas.

O algoritmo da análise de componentes principais pode ser descrito da seguinte forma:

1 - Transformação da matriz dos dados originais, através da operação de redução das variáveis iniciais:

$$x_{ij} = \frac{1}{\sqrt{n}} \frac{q_{ij} - \bar{q}_j}{s_j} \quad (4.1)$$

Onde n é o número de indivíduos, q_{ij} representa o valor da j -ésima variável do indivíduo i , \bar{q}_j a média aritmética dos valores da variável j e s_j o desvio padrão de j .

2 - Cálculo da matriz de correlações R , cujo elemento genérico é dado por:

$$r_{ij} = \sum_{i=1}^n x_{ij} x_{ij'} = \frac{1}{n} \sum_{i=1}^n \frac{(q_{ij} - \bar{q}_j)(q_{ij'} - \bar{q}_{j'})}{s_j s_{j'}} \quad (4.2)$$

3 - Diagonalização da matriz de correlações de que resultam p valores próprios λ_α e p vetores próprios u_α .

4 - Cálculo das coordenadas das variáveis nos eixos fatoriais, dadas por:

$$f_{j\alpha} = \sum_{i=1}^n x_{ij} v_{i\alpha} \quad (4.3)$$

Onde v é um vetor de média nula e variância unitária e α a componente principal.

5 - Cálculo das projeções dos indivíduos nos eixos fatoriais, dadas por:

$$w_{i\alpha} = \sum_{j=1}^p x_{ij} u_{j\alpha} \quad (4.4)$$

6 - Seleção da dimensão do sub-espço, cuja inércia acumulada explique uma percentagem suficiente da inércia total.

A necessidade de utilização da análise de componentes principais, neste trabalho, está relacionada com a grande dimensionalidade da matriz de características extraídas das imagens (376740x53) e o risco de ver comprometidos o desempenho e a exatidão do classificador. Para tal, foi usada seleção para determinar um subconjunto de características capazes de discriminar objetos através de regras previamente estabelecidas.

Depois de extraídas as características das imagens e normalizados os seus dados, foi, então, aplicada a função PCA à respetiva matriz:

[coeff, score, latent, tsquared, explained] = pca(Matriz);

Onde 'coeff' retorna os principais coeficientes de componentes para dados da Matriz, [44]; 'score' corresponde à pontuação da componente principal [44]; 'latent' corresponde à variância da componente principal, [44]; 'tsquare' retorna a estatística *T-squared* de *Hotelling's* para cada observação da 'Matriz', [44]; 'explained' corresponde à percentagem de variância por cada componente principal, [44]; entenda-se por 'Matriz', a matriz onde constam todas as características extraídas.

Através destes resultados é possível obter as características mais relevantes, para isso podem ser usadas duas abordagens, uma consiste em obter as percentagens dos valores existentes em variância e a partir de uma certa percentagem obter k componentes principais; ou pode definir-se logo um valor k que representa o número de componentes principais que se pretende. Neste trabalho foi utilizada a primeira abordagem, tendo sido definida como 95% dos componentes principais, ou seja, são somadas as percentagens respeitantes os valores existentes na

Google Earth™ Search Engine - Implementação e Validação do Método de Reconhecimento

variância, desde o primeiro valor até sejam atingidos os 95%. Quando esta percentagem é atingida verifica-se o número de componentes já calculadas que será, então, o k . Esse valor permite selecionar as primeiras k ocorrências existentes em 'coeff'. A partir do resultado obtido, depois de escolher o número de componentes que se pretende, cada linha da matriz que contem todas as características é multiplicada pelo resultado obtido na PCA.

Capítulo 5

Método Proposto - Classificação: Reconhecimento de Padrões

Segundo Bittencourt, 2006 [4], Inteligência Artificial é o conjunto de técnicas para resolver problemas complexos, isto é, problemas que, apesar de não terem solução algorítmica (modelo matemático), são solucionados por seres humanos. O objetivo central da Inteligência Artificial é a criação de modelos para a capacidade cognitiva e a construção de sistemas computacionais baseados nesses modelos.

A necessidade do ser humano comunicar com máquinas computacionais através de linguagens naturais e o interesse na ideia de projetar e construir 'máquinas inteligentes' que possam realizar certas tarefas com habilidades comparáveis à performance humana direcionam-nos para as redes neurais e, mais especificamente, para o reconhecimento de padrões.

5.1 Redes Neurais Artificiais para Reconhecimento de Padrões

Uma definição interessante foi dada em 1973 por Duda & Hart [9]: "Reconhecimento de padrões é um campo interessado no reconhecimento por máquinas de regularidades significativas em ambientes ruidosos ou complexos". Reconhecimento de padrões é uma área do conhecimento onde um dos objetivos é a classificação de objetos (padrões) em categorias (classes) (Theodoridis; Koutroumbas, 2006, [45]).

Os estudos sobre ANN (Artificial Neural Network - Redes Neurais Artificiais) baseiam-se no reconhecimento que o cérebro humano realiza, no entanto, quando essa tarefa é realizada computacionalmente, ela é feita de forma totalmente diferente. Uma rede neuronal é um processador distribuído, fortemente paralelizado, feito de unidades simples de processamento, denominadas neurónios, que têm uma tendência natural para armazenar conhecimento experimental e fazê-lo disponível para uso. Esta rede neuronal é semelhante ao cérebro em dois aspectos (Haykin, 1999 [15]):

- O conhecimento é adquirido pela rede por meio de um processo de aprendizagem;
- Os pesos de conexão entre neurónios, conhecidos como pesos sinápticos, são utilizados para armazenar o conhecimento adquirido.

A principal função do reconhecimento de padrões é a classificação de padrões de entrada em categorias ou classes. Uma rede neuronal pode receber uma coleção de padrões de entrada com a categoria a que cada padrão particular pertence. Apresentando um novo padrão à rede, esta irá classificar devidamente esse novo padrão. Este tipo de classificação pode ser resolvido por uma aprendizagem supervisionada. A vantagem de usar uma rede neuronal para realizar a classificação de padrões é que ela pode construir fronteiras de decisões não lineares entre as diferentes classes e oferecer um método prático para resolver problemas de elevada complexidade de classificação de padrões.

É certo que para que o processo de classificação de padrões tenha sucesso é necessário que exista o mapeamento dos padrões de entrada num espaço de classificação. Este espaço é definido por regiões de decisão que separam os padrões, atribuindo cada padrão a uma classe.

Baseado na definição de funções discriminantes que definem as regras de decisão, é possível diferenciar os classificadores em quatro grupos.

- Classificadores Probabilísticos: Este grupo fundamenta-se numa rigorosa base probabilística assumindo que a estrutura e distribuição dos dados no espaço do problema são a princípio bem conhecidas (caracterizam classes pela sua função densidade de probabilidade no espaço de entrada). Os parâmetros da distribuição são tipicamente estimados usando treino supervisionado onde supõe que todos os dados de treino estão disponíveis simultaneamente. A construção dos classificadores baseia-se na conversão da probabilidade *a priori* das classes dos padrões de treino em medidas de probabilidade condicionada *a posteriori* de um dado padrão pertencer a uma dada classe de treino (Todesco, 1995 [46]).

- Classificadores de Hiperplano: As funções discriminantes usadas nos classificadores de hiperplano são capazes de formar regiões de decisão complexas que usam nós para formar hiperplanos como fronteiras de decisão no espaço de entradas transformado. A transformação do espaço de entradas é feita sobre nós formados por uma soma ponderada das entradas e que passam esta soma através da não-linearidade da sigmóide ou tangente hiperbólica. O Perceptron multi-camadas treinado com o *Backpropagation* pode ser um exemplo de classificadores de hiperplanos.

- Classificadores de *Kernel*: Em classificadores de *Kernel* existe uma transformação do espaço de entradas para um espaço maior. Os classificadores de *Kernel* criam regiões de decisão complexas baseados nos campos receptivos formados pelos nodos de função-*kernel*. O treino dos classificadores de kernel são relativamente rápidos, principalmente quando se utiliza treino não supervisionado para a determinação dos centros. Estes classificadores tem requerimentos de memória e de computação intermédios, quando comparados aos de hiperplanos. "... classificadores de *kernel* incluem classificadores convencionais que estimam funções de distribuição de probabilidade que usam a abordagem de 'Parzen Window' ou distribuições mistas, e classificadores que formam funções discriminantes usando funções *kernel*. Redes neurais classificadoras de kernel incluem abordagens baseadas em mapas que usam matrizes de nodos que calculam funções *kernel*, classificadores baseados no *Cerebellar Model Articulation Controller - CMAC*, e classificadores que usam o método das funções potenciais, as vezes chamadas classificadores de funções de base radial..." (Lippmann, 1989 [22]).

- Classificador de Exemplos: Esta categoria de classificadores baseia-se na distância entre os padrões de entrada e exemplos previamente treinados. A ideia é atribuir os padrões de entrada para os exemplos mais próximos calculados por alguma métrica. A métrica mais comum é a euclidiana ponderada que é usada para medir a distância entre as entradas e os centros que representam exemplos previamente selecionados durante o treino. O treino empregado geralmente é o combinado não-supervisionado/supervisionado. Alguns dos classificadores de exemplos são: o classificador de mapa-de-características, o *Learning Vector Quantizer (LVQ)*, o classificador *Restricted Coulomb Energy (RCE)*, os classificadores da Teoria da Ressonância Adaptativa, os classificadores que usam raciocínio baseado-na-memória e os classificadores que usam interpolação linear local (Todesco, 1995 [46]).

O projeto de sistemas de reconhecimento de padrões envolve três aspetos:

- 1) aquisição de dados e pré-processamento;
- 2) representação dos dados;
- 3) tomada de decisões (Campos T.E, 2001 [5]).

Geralmente o desafio encontra-se na escolha das técnicas para efetuar estes três aspetos. Um problema de reconhecimento de padrões bem definido e restrito permite uma representação compacta dos padrões e uma estratégia de decisão simples.

5.2 MLP - *Multilayer Perceptron*

O *perceptron* foi a primeira rede neuronal descrita de forma algorítmica, inventada por Rosenblatt, 1958 [32].

A MLP (do inglês *multilayer perceptron*, ou *perceptron* multicamadas) é um tipo de rede neuronal muito utilizada no reconhecimento de padrões e na aproximação de funções, sendo formada pela interligação de diversos neurónios artificiais do tipo *perceptron*, dispostos em camadas. Essa rede é capaz de resolver problemas que possuem uma superfície de decisão não linear (Mitchell, 2001 [24]), além de ser capaz de aproximar qualquer função contínua com o uso de uma camada intermédia e de aproximar qualquer função com a utilização de duas camadas intermédias (Braga; Carvalho; Ludemir 2007 [2]).

A unidade básica da MLP é o *perceptron* e é formado por diversas entradas e uma única saída. A cada entrada do *perceptron* é associado um peso sináptico, sendo a sua saída determinada em função da soma ponderada das entradas.

Diversas funções de ativação podem ser utilizadas para os *perceptrons*, mas na MLP, devido ao seu algoritmo de treino, é necessário utilizar uma função que seja diferenciável em todos os pontos, sendo muito comum o uso da função sigmóide (Russell; Norvig 2003 [35]).

A MLP é dividida em camadas formadas por um ou mais *perceptrons*, fazendo com que as saídas dos *perceptrons* de camada j possam ser utilizadas como entradas para os *perceptrons* da camada $j + 1$. Comumente, a arquitetura da rede é dividida numa camada de entrada, uma ou mais camadas intermédias e uma camada de saída, 5.1.

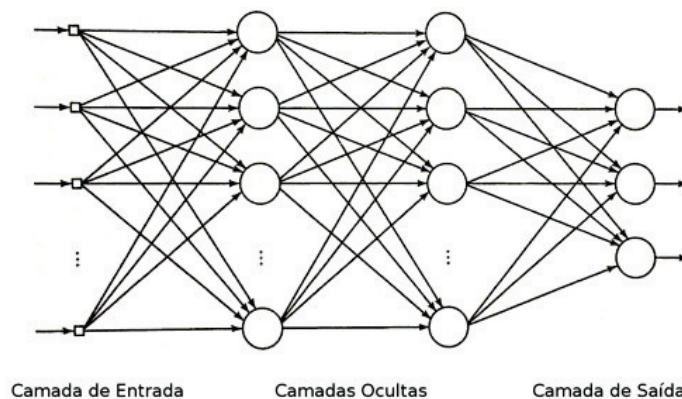


Figura 5.1: Arquitetura de um MLP

(Imagem retirada de <http://www.borges-solutions.com/rede-neural-artificial-perceptron-de-multiplas-camadas>)

A camada de entrada, diferentemente das demais, não possui pesos sinápticos, apenas recebe os atributos de entrada. A camada de saída tem o valor final computado pela rede para os atributos apresentados na camada de entrada. As camadas intermédias, também chamadas de ocultas, situam-se entre a camada de entrada e a camada de saída (Russell; Norvig, 2003 [35]). A saída da MLP é calculada pela arquitetura da rede e pelos pesos dos *perceptrons*. Para determinar os pesos dos *perceptrons* é utilizado um algoritmo conhecido como *backpropagation* (Braga; Carvalho; Ludemir, 2007 [2]). Este algoritmo é um dos métodos mais populares para treino de *perceptrons* multi-camadas. A aprendizagem *backpropagation* consiste em duas fases entre as diferentes camadas da rede: a fase *forward* e a fase *backward*. Na fase *forward*, um vetor de entrada é aplicado aos neurónios de entrada da rede, e o seu efeito é propagado

pela rede, camada a camada, até o conjunto de saída ser gerado. Na fase *forward*, todos os pesos sinápticos dos neurónios da rede são fixados. Na fase *backward*, os pesos sinápticos são ajustados de acordo com a regra de correção de erros. A resposta da rede é subtraída da resposta esperada para se obter o sinal de erro, sendo que este é propagado da frente para trás da rede contra a direção das conexões sinápticas. Sendo assim, os pesos sinápticos são ajustados com o intuito de fazer com que a resposta obtida chegue o mais próximo possível da resposta desejada (Haykin, 1999 [15]).

5.3 Aprendizagem não supervisionada *versus* Aprendizagem supervisionada

As redes neuronais artificiais são basicamente formadas por neurónios artificiais interligados paralelamente. Por essa interligação, as redes neuronais artificiais são capazes de modelar relações lineares e não lineares, podendo calcular diversas funções matemáticas complexas que são realizadas num computador. Dessa forma, o conhecimento da rede está distribuído nos diversos neurónios que a compõem.

A aquisição de conhecimento da rede neuronal ocorre por um processo de aprendizagem, denominado de treino, onde a rede neuronal tem os seus pesos ajustados (Braga; Carvalho; Ludemir 2007 [2]).

Na aprendizagem não-supervisionada, como o próprio nome sugere, não há um professor ou supervisor externo para acompanhar o processo de aprendizagem. Neste esquema de treino apenas os padrões de entrada estão disponíveis para a rede, ao contrário da aprendizagem supervisionada, cujo conjunto de treino possui pares de entrada e saída. Durante o processo de aprendizagem os padrões de entrada são apresentados continuamente à rede, e a existência de regularidades nesses dados faz com que a aprendizagem seja possível.

Aprendizagem supervisionada implica, necessariamente, a existência de um supervisor, ou professor externo, o qual é responsável por estimular as entradas da rede por meio de padrões de entrada e observar a saída calculada pela mesma, comparando-a com a saída desejada. Como a resposta da rede é função dos valores atuais do seu conjunto de pesos, estes são ajustados de forma a aproximar a saída da rede da saída desejada (Braga, 2007 [2]).

5.4 Dados não balanceados *versus* Dados balanceados

Os algoritmos de aprendizagem supervisionada são frequentemente utilizados em aplicações computacionais, tais como reconhecimento de padrões, mineração de dados, entre outras. Muitos aspetos são relevantes na avaliação do desempenho desses algoritmos de classificação. Um deles é a quantidade de amostras e a distribuição das mesmas entre as classes no conjunto de treino. Quando o número de exemplos é representativo, mas estes exemplos não estão balanceados, ou seja, há uma grande desproporção do número de dados de uma das classes de treino, surge o problema de classes não balanceadas. Nestas condições, os modelos de classificação que são otimizados em relação à exatidão global têm tendência a criar modelos triviais, que quase sempre predizem a classe maioritária.

Sendo assim, muitos sistemas de aprendizagem não estão preparados para induzir bons modelos de classificação na presença de conjuntos não balanceados. Ao treinar um modelo com esses

dados, este apresentará frequentemente uma boa exatidão para dados pertencentes à classe majoritária, mas uma exatidão inaceitável para os dados da classe minoritária. De entre as soluções propostas para o problema de classes não balanceadas, uma das técnicas mais populares é a utilização de métodos de balanceamento artificial do conjunto de treino. A maioria desses métodos baseia-se na remoção ou criação/replicação de exemplos da classe majoritária e minoritária respetivamente.

5.5 Rede neuronal e Parâmetros utilizados

Segundo Haykin (2001) [15], a aprendizagem é um processo pelo qual os parâmetros livres de uma rede neuronal são adaptados através de um processo de estimulação do ambiente no qual a rede está inserida. O tipo de aprendizagem é determinado pela maneira com a qual as mudanças nos parâmetros ocorrem. Um conjunto de regras bem definidas para a solução de um problema de aprendizagem é denominado algoritmo de aprendizagem ou algoritmo de treino. Existem alguns algoritmos de treino que, além do ajuste de pesos, provocam também mudanças na própria arquitetura da rede, como a criação ou eliminação de neurónios (Bittencourt, 2006 [4]). A maioria dos algoritmos de treino de redes neuronais é inspirada, direta ou indiretamente, na lei de Hebb (Bittencourt, 2006 [4]): A intensidade de uma ligação sináptica entre dois neurónios aumenta se ambos são excitados simultaneamente.

Em todas as experiências realizadas neste trabalho foi utilizado o algoritmo de retropropagação (*backpropagation*). A retropropagação é um algoritmo de treino supervisionado. Em problemas de reconhecimento de padrões, pretende-se uma rede neuronal para classificar as entradas num conjunto de categorias-alvo. Foram selecionadas características das imagens, utilizando a PCA, usadas como entrada para criar e treinar várias redes, e avaliar o seu desempenho com erro e matrizes de confusão. Tratou-se de redes de duas camadas *feed-forward*, com neurónios ocultos e saída sigmóide (*patternnet*), que permitem classificar vetores arbitrariamente bem, dados neurónios suficientes na sua camada escondida.

Relativamente ao treino, o total das amostras, setenta por cento foram apresentadas à rede durante a formação, e a rede foi ajustada de acordo com o seu erro; quinze por cento foram usado para medir a generalização da rede, e para parar o treino quando a generalização deixa de melhorar; os restantes quinze por cento das amostras, sem efeito sobre o formação, forneceram uma medida independente do desempenho da rede durante e após o treino.

Nos testes preliminares, foram utilizados dados não balanceados. Os resultados obtidos não refletiam a realidade e como tal, todas as experiências posteriores utilizaram dados balanceados.

Capítulo 6

Método Proposto - Pós-Processamento: Operadores Morfológicos

As operações morfológicas são operações que permitem manipular a morfologia dos objetos existentes numa imagem. A sua aplicação é particularmente relevante em imagens binárias quando se quer eliminar ruídos resultantes da binarização ou evidenciar pormenores do objeto. Neste trabalho são consideradas duas operações base, a dilatação e a erosão, e duas operações compostas, a abertura (*Open*) e o fecho (*Close*).

6.1 Erosão e Dilatação

Os operadores utilizados no processamento morfológico de imagem formam uma ampla classe de operadores não-lineares. Todos estes são construídos pela interação de duas operações básicas: a erosão e a dilatação. A erosão da imagem f para um dado pixel x é definida como o valor mínimo da imagem em uma janela definida pelo elemento estruturante B , estando a origem do elemento estruturante na posição de x [1], ou seja:

$$[\varepsilon_B(f)](x) = \min_{b \in B} f(x + b) \quad (6.1)$$

onde b é um pixel pertencente ao elemento estruturante B . Por outro lado, a dilatação é definida como:

$$[\delta_B(f)](x) = \max_{b \in B} f(x + b) \quad (6.2)$$

Aplicações alternadas da dilatação e da erosão levam à formação de duas novas operações denominadas abertura e fecho [40]. A abertura morfológica da imagem f é definida como a erosão da imagem f com elemento estruturante B , seguida de uma dilatação com o elemento estruturante transposto:

$$\gamma_B(f) = \delta_{B^T}[\varepsilon_B(f)] \quad (6.3)$$

enquanto que o fecho morfológico é definido através da sequência dilatação - erosão com o elemento estruturante transposto:

$$\phi_B(f) = \varepsilon_{B^T}[\delta_B(f)] \quad (6.4)$$

A abertura tem um efeito de alisamento do contorno, eliminando pequenas protuberâncias de uma imagem. Em contrapartida, o fecho pode preencher pequenos buracos e lacunas no contorno de uma imagem [40].

A dilatação é, então, uma operação morfológica que visa aumentar a área geométrica de um

objeto. A forma e a extensão desse aumento são controladas pelo formato do elemento estruturante utilizado. Já a erosão visa diminuir a área geométrica de um objeto, eliminando detalhes da imagem menores que o elemento estruturante. A forma e a extensão dessa diminuição são, também, controladas pelo formato do elemento estruturante utilizado.

6.1.1 Abertura e Fecho (*Open* e *Close*)

A abertura e o fecho são consideradas operações compostas, uma vez que usam a combinação das operações básicas erosão e dilatação, diferenciando-se pela sequência das operações:

- Abertura (*Open*) = Erosão + Dilatação
No Matlab®: $IM2 = imopen(IM, SE)$
Realiza a abertura morfológica em escala de cinza ou na imagem binária IM , com o elemento estruturante SE .
- Fecho (*Close*) = Dilatação + Erosão
No Matlab®: $IM2 = imclose(IM, SE)$
Executa o fecho morfológico em escala de cinza ou imagem binária IM , retornando a imagem fechada $IM2$, onde SE é o elemento estruturante.

A definição de uma abertura morfológica de uma imagem é uma erosão seguida de uma dilatação, usando o mesmo elemento estruturante para ambas as operações. A operação relacionada, o fecho morfológico de uma imagem, é o inverso, ou seja, baseia-se numa dilatação seguida de uma erosão, com o mesmo elemento estruturante.

A operação morfológica de abertura pode ser usada para remover pequenos objetos de uma imagem preservando a forma e o tamanho dos objetos maiores, enquanto a operação de fecho elimina pequenas aberturas nos objetos da imagem, preenche falhas nos contornos e junta objetos que estejam relativamente próximos.

Neste trabalho foi usada a abertura para todos os elementos naturais detetados ('água', 'casas', 'estradas') e o fecho combinado com a abertura para o caso da 'estrada'. Os elementos estruturantes utilizados foram o *disk*, *line*, *square*. No caso da 'água' foi ainda utilizado a operação *fill* para preenchimento de píxeis isolados. Os elementos estruturantes têm normalmente associados a forma e os parâmetros (*strel* (forma, parâmetros)), por exemplo:

$SE = strel('disk', R, N)$, $SE = strel('line', LEN, DEG)$ ou $SE = strel('square', W)$, onde R é o raio, LEN o comprimento e DEG graus, e W é o número de píxeis relativos à aresta do quadrado.

No caso do elemento estruturante *disk*, 6.1, usado nos elementos 'água' e 'casas', é criado um elemento plano em forma de disco de estruturação, em que R indica o raio. R deve ser um inteiro não negativo que, nestas experiências assumem valores como 1, 5, 10, 15, 20, 25 e 30. O N não foi especificado, o valor padrão assumido é quatro.

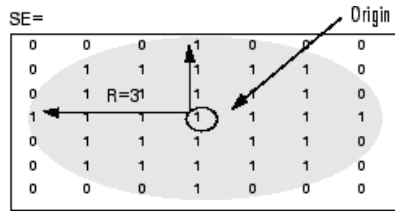


Figura 6.1: Elemento estruturante *disk*

(Imagem retirada de www.mathworks.com/help/images/ref/strel.html)

No caso do elemento estruturante *line*, é criado um elemento estruturante linear plano que é simétrico em relação ao centro da vizinhança, como pode ver-se em 6.2. DEG especifica o ângulo (em graus), da linha, medido num sentido anti-horário do eixo horizontal. LEN é aproximadamente a distância entre o centro dos membros do elemento estruturante em extremos opostos da linha. Nas experiências realizadas, foram utilizados os valores 90 e 180 para os graus e 5, 15, 30, 90, 100 e 110, para o comprimento, dependendo do elemento ser 'água' ou 'casas'.

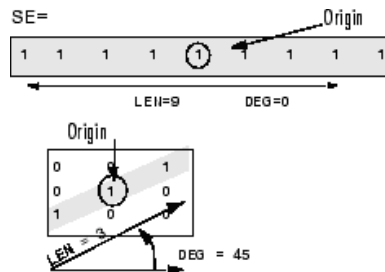


Figura 6.2: Elemento estruturante *line*

(Imagem retirada de www.mathworks.com/help/images/ref/strel.html)

Relativamente ao *square*, é criado um elemento estruturante quadrado cujo comprimento da aresta é W píxeis, como exemplificado em 6.3

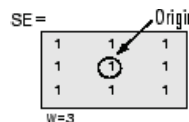


Figura 6.3: Elemento estruturante *square*

(Imagem retirada de www.mathworks.com/help/images/ref/strel.html)

Capítulo 7

Experiências e Discussão

Foram usados quatro conjuntos de imagens para o treino da rede neuronal de reconhecimento de padrões. Cada conjunto é constituído por uma imagem do *Google Earth*™, a correspondente imagem do *Google Maps*™ e a imagem onde cuja deteção de elementos foi feita manualmente recorrendo à ferramenta de edição e tratamento de imagem *Photoshop*®, usada como entrada para a rede neuronal. Todos os conjuntos contém 'estradas'; todos, excepto o conjunto número um, têm 'casas'; apenas os conjuntos números um e quatro contém 'água'.



Figura 7.1: Conjunto número 1 (Imagem *Google Earth*, Imagem *Google Maps*, Imagem 'classe')

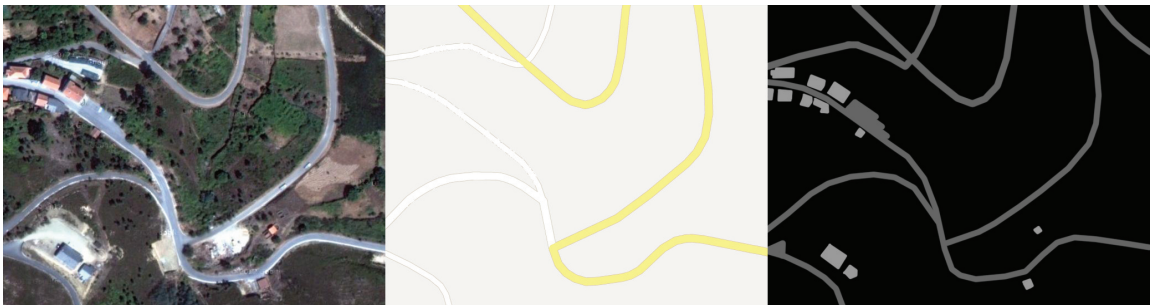


Figura 7.2: Conjunto número 2 (Imagem *Google Earth*, Imagem *Google Maps*, Imagem 'classe')

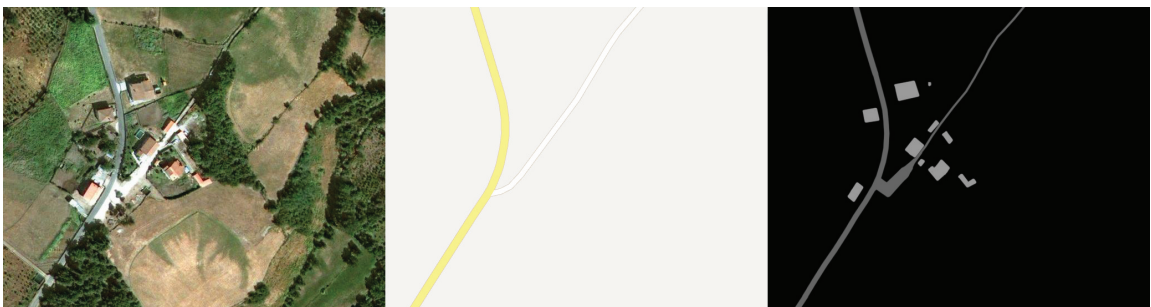


Figura 7.3: Conjunto número 3 (Imagem *Google Earth*, Imagem *Google Maps*, Imagem 'classe')

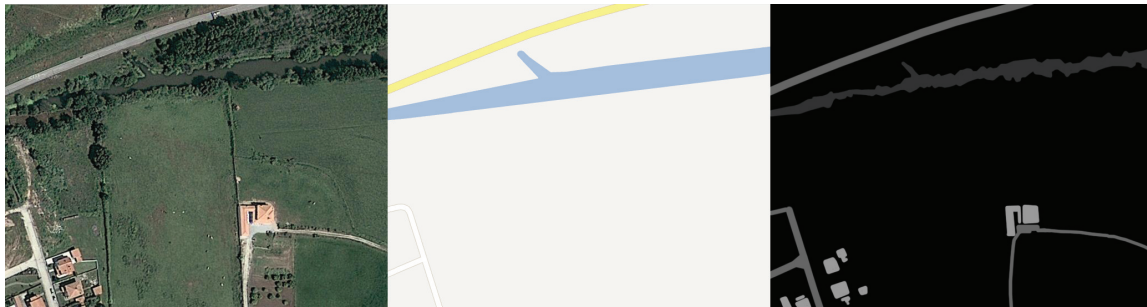


Figura 7.4: Conjunto número 4 (Imagem *Google Earth*, Imagem *Google Maps*, Imagem 'classe')

A forma apropriada da medição da taxa do erro da rede deve ser realizada submetendo dados não presentes no conjunto de treino utilizado para o cálculo dos pesos sinápticos. Medir a taxa de erro somente nos dados presentes no conjunto de treino não permite determinar a taxa de erro do algoritmo classificador proposto, já que para classificar o conjunto de dados utilizados no treino não necessita sequer de um algoritmo de aprendizagem de máquina, bastando consultar numa base de dados a classificação dada a uma instância. Logo, medir a taxa de erro no mesmo conjunto de dados utilizado na fase de treino, ou um subconjunto deste, não consiste num bom indicador da taxa de erros do classificador para dados ainda não apresentados (Witten; Frank, 2005 [48]).

Como tal, para a realização dos testes foram usados cinco novos conjuntos de imagens. Todos os conjuntos, à exceção do conjunto nove, contêm 'casas' e 'estradas'. Apenas os conjuntos cinco e nove contêm 'água'.

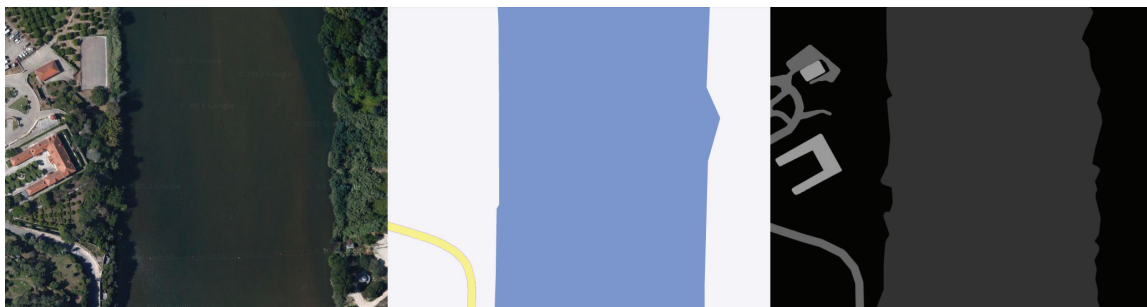


Figura 7.5: Conjunto número 5 (Imagem *Google Earth*, Imagem *Google Maps*, Imagem 'classe')

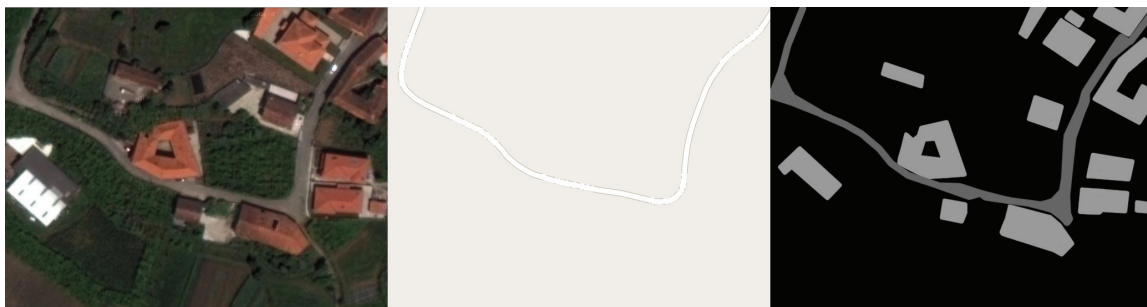


Figura 7.6: Conjunto número 6 (Imagem *Google Earth*, Imagem *Google Maps*, Imagem 'classe')

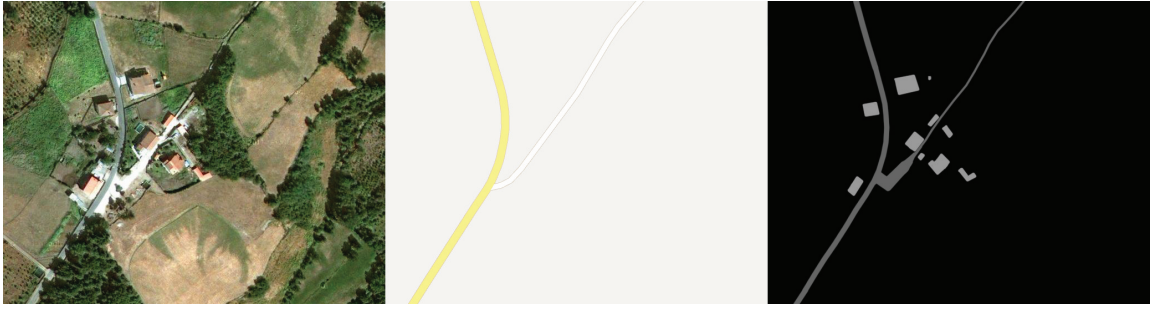


Figura 7.7: Conjunto número 7 (Imagem *Google Earth*, Imagem *Google Maps*, Imagem 'classe')

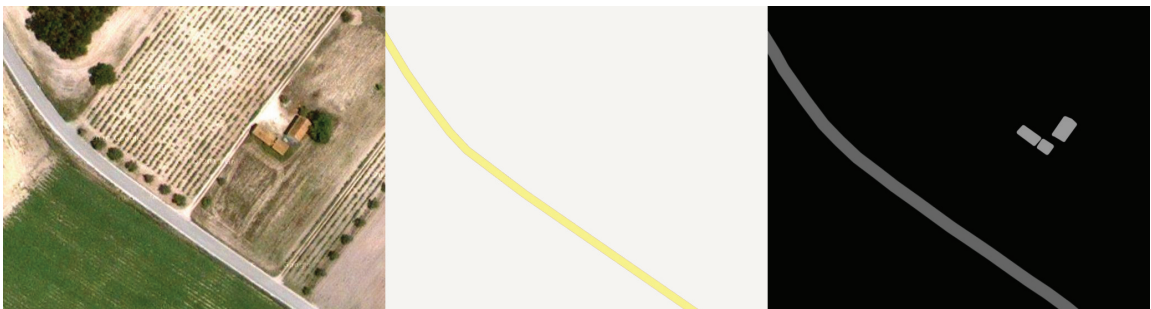


Figura 7.8: Conjunto número 8 (Imagem *Google Earth*, Imagem *Google Maps*, Imagem 'classe')

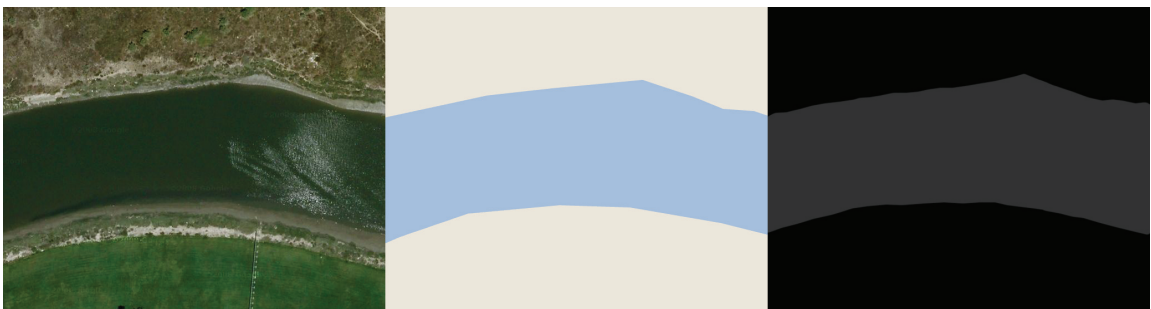


Figura 7.9: Conjunto número 9 (Imagem *Google Earth*, Imagem *Google Maps*, Imagem 'classe')

7.1 Experiências

Foram realizadas várias experiências para verificar a influência das características no resultado final, após aplicação de operadores morfológicos.

Todas as experiências foram realizadas testando cada um dos cinco conjuntos de imagens teste com cada um dos quatro conjuntos de imagens usados para treino, testes estes realizados para cada elemento natural procurado. Todos os testes foram realizados para 2, 4, 6, 8, 10, 12, 14, 16, 18 e 20 neurónios, tendo sido escolhidos para aplicação dos operadores morfológicos, os resultados com menor taxa de erro, para cada conjunto de imagens. Assim, para o conjunto número cinco foi escolhida a melhor classificação para 'água', 'estradas' e 'casas'; para os conjuntos seis, sete e oito, para 'estradas' e 'casas' e para o conjunto número nove, 'água'.

As análises foram efetuadas com base na taxa de erro resultante do teste da rede neuronal, diferença de píxeis entre a imagem 'classe' e a imagem obtida após a aplicação dos operadores morfológicos (erro total) e também nos valores relativos a falsos positivos¹ (FP) e falsos negativos² (FN). Com esses dados foram construídos gráficos para cada elemento/operador morfológico, em cada experiência, para uma melhor visualização dos resultados. São considerados verdadeiros positivos (VP) os que o sistema julgou corretamente positivos; valores positivos que o sistema julgou negativos são considerados falso negativo (FN); valores negativos que o sistema julgou como negativos são, então, verdadeiros negativos (VN); os valores negativos que o sistema julgou positivos consideram-se falso positivos (FP). Tendo em consideração a informação anterior, foi calculada a eficiência (E) de cada uma das aplicações de operadores morfológicos a cada elemento de cada experiência, usando a fórmula:

$$E = ((VP/(VP + FN)) + (VN/(VN + FP)))/2$$

Posteriormente foram selecionados em cada experiência, o elemento/operador morfológico mais eficiente e feita a respetiva matriz de confusão.

7.1.1 Características - Cor/Intensidade

Nesta experiência foram utilizadas apenas característica RGB e HSV.

O elemento 'água', no conjunto de imagens número cinco, foi melhor classificado com a rede treinada com o conjunto de imagens número quatro, com 8 neurónios; já o conjunto de imagens número nove obteve melhor classificação com a rede treinada com o conjunto de imagens número um, com 2 neurónios; relativamente ao elemento 'casas', os conjuntos número cinco e seis obtiveram menor taxa de erro quando testados com a rede treinada com o conjunto de imagens número quatro, com 14 e 18 neurónios; os conjuntos número sete e oito foram melhor classificados com a rede resultante do treino com o conjunto número dois, com 6 e 12 neurónios; o elemento 'estradas', nos conjuntos de imagens número cinco, seis e sete, obteve menor taxa de erros quando testado com a rede treinada com o conjunto de imagens número dois, com 2, 4 e 18 neurónios; para o mesmo elemento, o conjunto de imagens número oito foi melhor classificado com a rede proveniente do treino do conjunto de imagens número três, com 16 neurónios.

¹Proporção de píxeis erroneamente assinalados como sendo um elemento

²Proporção de píxeis assinalados como não elemento, quando deveria ser assinalado como elemento

7.1.1.1 Análise de erros depois da aplicação de operadores morfológicos com diferentes elementos estruturantes

Os resultados, após a aplicação dos operadores morfológicos, podem ser visualizado nos gráficos 7.10 para o elemento 'água', 7.11 para o elemento 'casas' e 7.12 para o elemento 'estradas'.

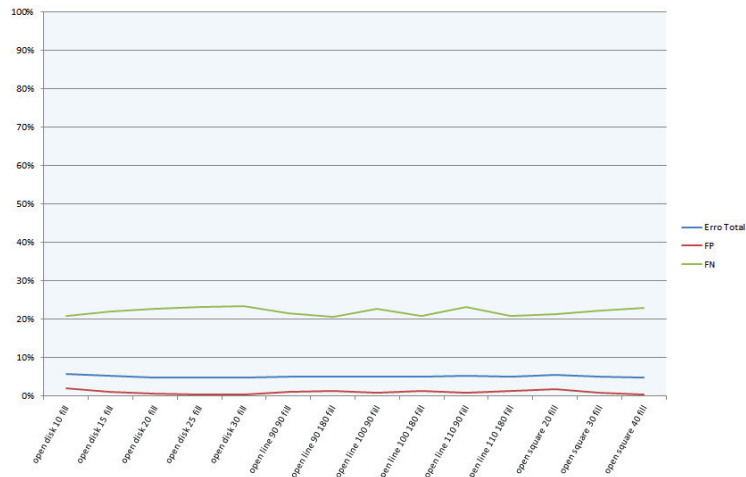


Figura 7.10: Gráfico - 'Água': Taxa de erro / Operadores morfológicos

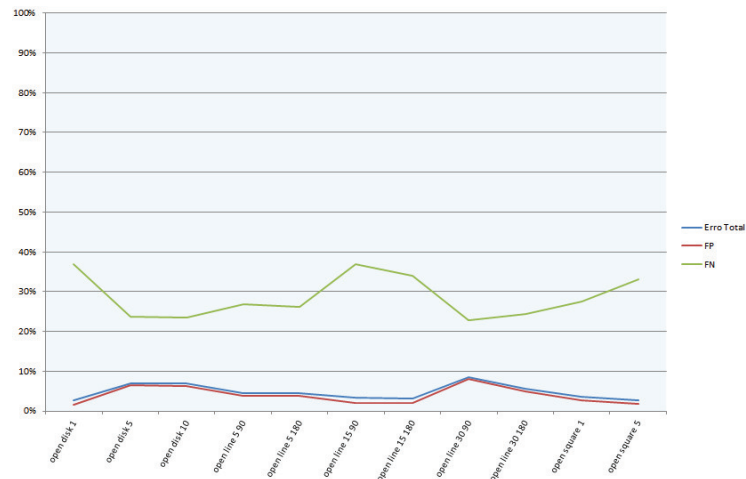


Figura 7.11: Gráfico - 'Casas': Taxa de erro / Operadores morfológicos

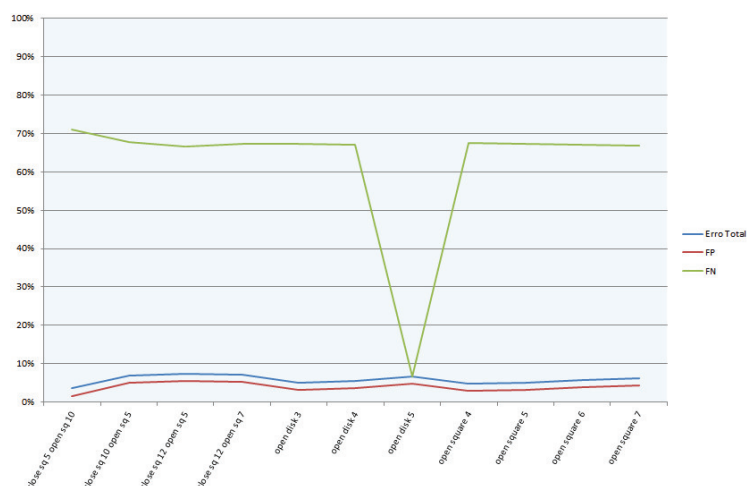


Figura 7.12: Gráfico - 'Estradas': Taxa de erro / Operadores morfológicos

7.1.1.2 Análise da Matriz Confusão do resultado com maior eficiência

Nas tabelas 7.1, 7.2 e 7.3 encontram-se as matrizes confusão dos melhores resultados para cada elemento, obtidos nesta experiência, tendo em consideração a eficiência.

Tabela 7.1: Matriz Confusão do melhor resultado para o elemento 'água'

	Positivos	Negativos
Positivos	19,17	13,41
Negativos	20,72	80,83

Tabela 7.2: Matriz Confusão do melhor resultado para o elemento 'casas'

	Positivos	Negativos
Positivos	3,36	27,83
Negativos	27,56	96,64

Tabela 7.3: Matriz Confusão do melhor resultado para o elemento 'estradas'

	Positivos	Negativos
Positivos	2,89	4,79
Negativos	66,64	97,11

Nas matrizes de confusão dos elementos 'água' e 'casas', os valores de falsos negativos situam-se entre os 20 e os 28%, ou seja, essa percentagem de píxeis 'elemento' não foram assinalados como tal. Para o elemento 'estradas', a percentagem de falsos negativos é significativamente mais elevada, indicando que apenas 40% dos píxeis 'estradas' foram corretamente detetados. Em contrapartida, para o mesmo elemento, a taxa de falsos positivos é baixa. Para os dois outros elementos, a percentagem de píxeis assinalados como elemento, não o sendo, situa-se entre os 20 e os 30%.

7.1.1.3 Análise das imagens

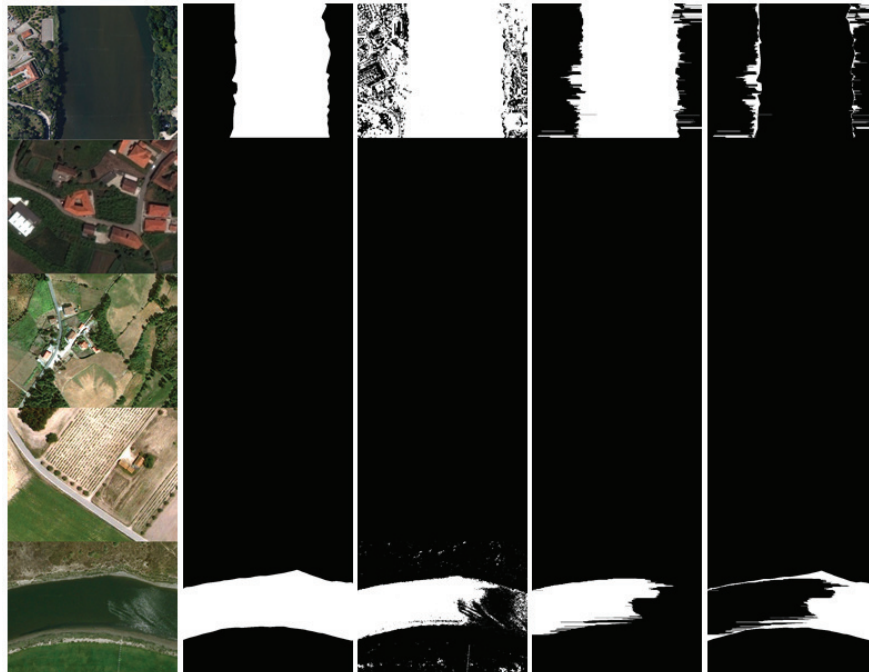


Figura 7.13: Imagens originais; Imagens 'classe'; Melhor classificação rede; Melhor classificação operadores pós-processamento; Imagens XOR

Neste caso, o elemento 'água' do conjunto de imagens número cinco apresenta-se mais completo do que a do conjunto nove, podendo tal facto, estar relacionado com o brilho da água, nesse conjunto de imagens. No entanto, em ambos os casos, embora não na totalidade, o elemento foi detetado, usando apenas características relativas à cor.

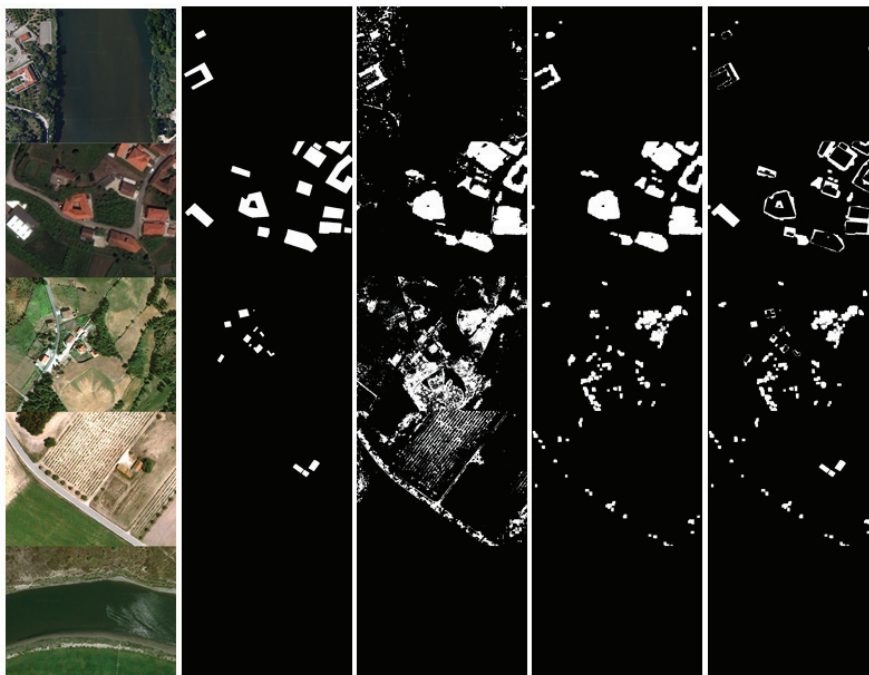


Figura 7.14: Imagens originais; Imagens 'classe'; Melhor classificação rede; Melhor classificação operadores pós-processamento; Imagens XOR

Para o elemento 'casas', os conjuntos de imagens números cinco e seis apresentam resultados razoáveis, após aplicação dos operadores morfológicos. Já nos conjuntos de imagens números sete e oito, não foi possível identificar corretamente o elemento 'casas', facto que pode estar relacionado com a dimensão reduzida e cor mais suave dos telhados.

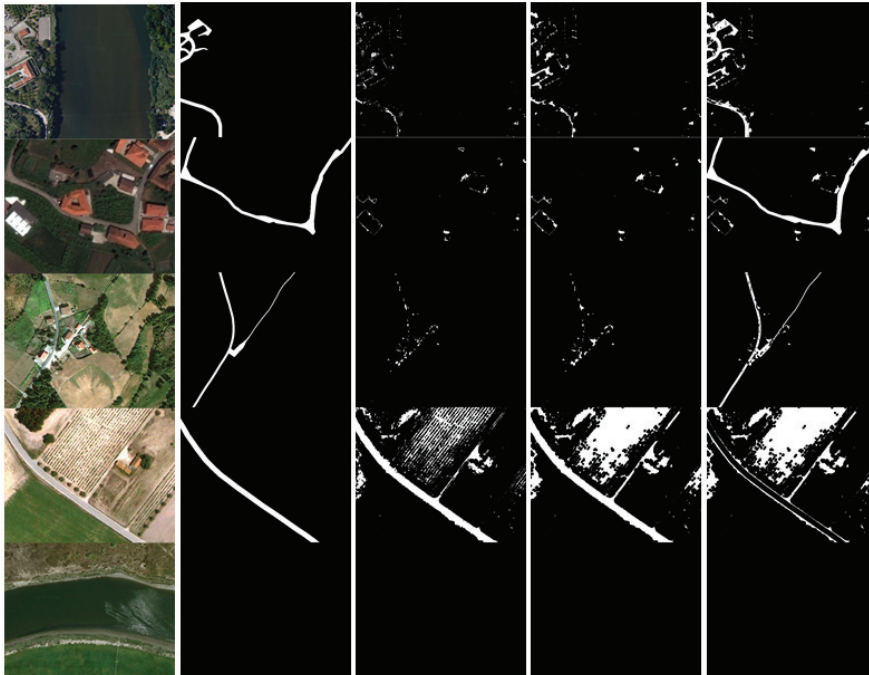


Figura 7.15: Imagens originais; Imagens 'classe'; Melhor classificação rede; Melhor classificação operadores pós-processamento; Imagens XOR

Recorrendo apenas a características da cor, os resultados para o elemento 'estrada' não se mostram satisfatórios. Apenas no conjunto de imagens número oito foi detetado o elemento, mas tendo sido também marcado como 'estrada' uma parte do terreno agrícola. Nos conjuntos de imagens números cinco e seis, os resultados são mesmo insatisfatórios.

7.1.2 Características - Matriz de Co-Ocorrência e Função Entropia

Nesta experiência foram utilizadas características extraídas através da matriz de co-ocorrência de tons cinza e da função entropia.

O elemento 'água' obteve a melhor classificação com a rede resultante do treino do conjunto de imagens número um, com 2 e 12 neurónios; para o elemento 'casas', o conjunto de imagens cinco obteve menor taxa de erros quando testada com a rede resultante do treino do conjunto de imagens número dois, com 4 neurónios; o conjunto de imagens número seis foi melhor classificado com a rede treinada com o conjunto de imagens número quatro, com 2 neurónios; os conjuntos de imagens números sete e oito obtiveram melhor classificação quando testados com as redes treinadas com o conjunto de imagens número três, com 2 e 10 neurónios; o elemento 'estradas', no conjunto de imagens número cinco, apresentou menor taxa de erro quando foi usada a rede treinada com o conjunto de imagens número dois, com 12 neurónios, e o conjunto número seis, com a rede treinada com o conjunto de imagens número um, com 10 neurónios; o conjunto número sete apresentou melhores resultados quando testado com com a rede resultante do treino com o conjunto de imagens número três, com 8 neurónios e, finalmente, o conjunto número oito foi melhor classificado com a rede treinada com o conjunto número

quatro, com 4 neurónios.

7.1.2.1 Análise de erros depois da aplicação de operadores morfológicos com diferentes elementos estruturantes

Após aplicação dos operadores morfológicos, os resultados constam dos gráficos 7.16, 7.17 e 7.18.

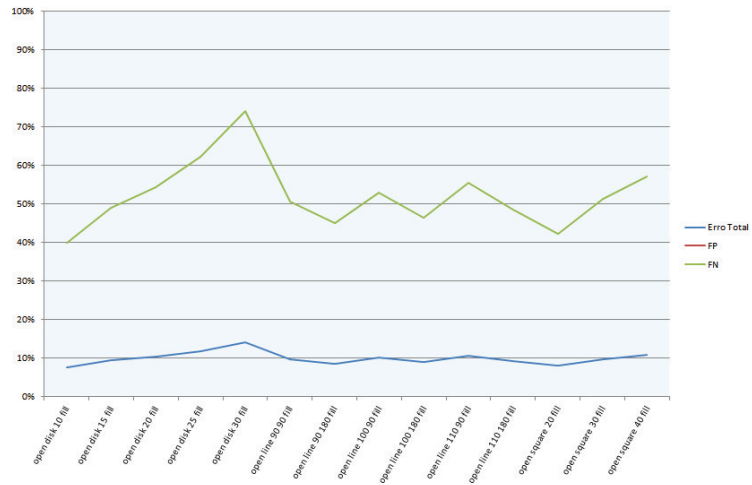


Figura 7.16: Gráfico - 'Água': Taxa de erro / Operadores morfológicos

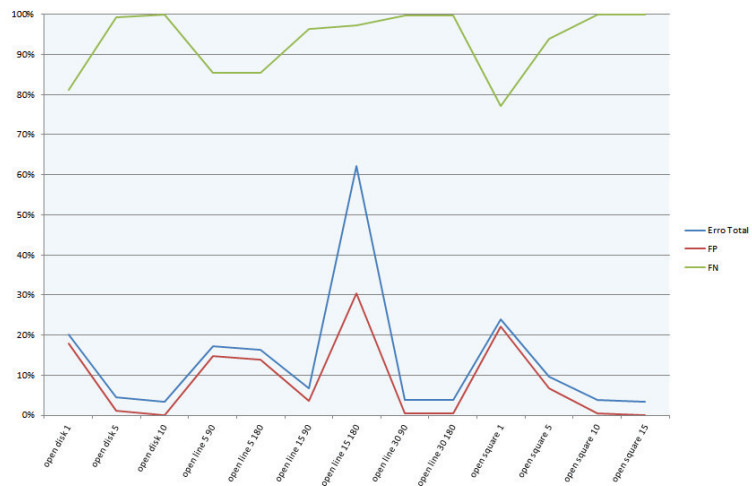


Figura 7.17: Gráfico - 'Casas': Taxa de erro / Operadores morfológicos

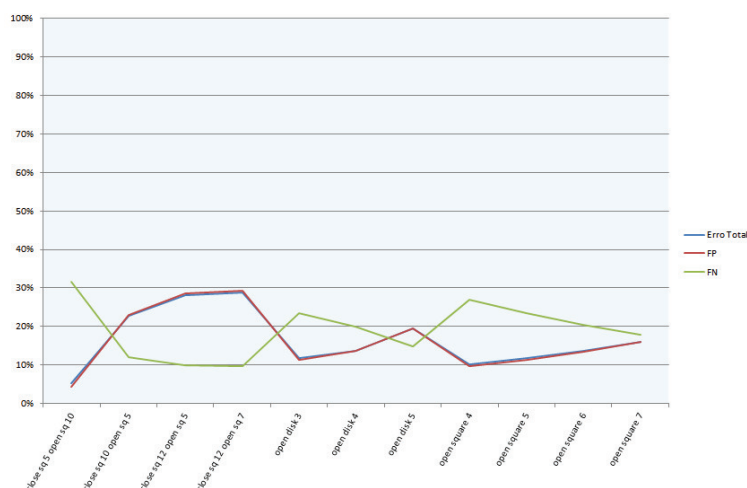


Figura 7.18: Gráfico - 'Estradas': Taxa de erro / Operadores morfológicos

7.1.2.2 Análise da Matriz Confusão do resultado com maior eficiência

Nas tabelas 7.4, 7.5 e 7.6 encontram-se as matrizes confusão dos melhores resultados para cada elemento, obtidos nesta experiência, tendo em conta a eficiência.

Tabela 7.4: Matriz Confusão do melhor resultado para o elemento 'água'

	Positivos	Negativos
Positivos	19,17	0,00
Negativos	39,84	80,83

Tabela 7.5: Matriz Confusão do melhor resultado para o elemento 'casas'

	Positivos	Negativos
hline Positivos	3,36	22,13
Negativos	77,08	96,64

Tabela 7.6: Matriz Confusão do melhor resultado para o elemento 'estradas'

	Positivos	Negativos
Positivos	2,89	4,41
Negativos	31,54	97,11

Da análise das matrizes de confusão é possível concluir que, para o elemento 'casas', o valor de falsos negativos é bastante acentuado (quase 80%), sendo também significativo no caso dos outros elementos, ou seja, existe um grande número de píxeis que deveriam ter sido assinalados como elemento e não o foram. A taxa de falsos positivos é baixa, embora ligeiramente mais notória no caso do elemento 'casas'.

7.1.2.3 Análise das Imagens

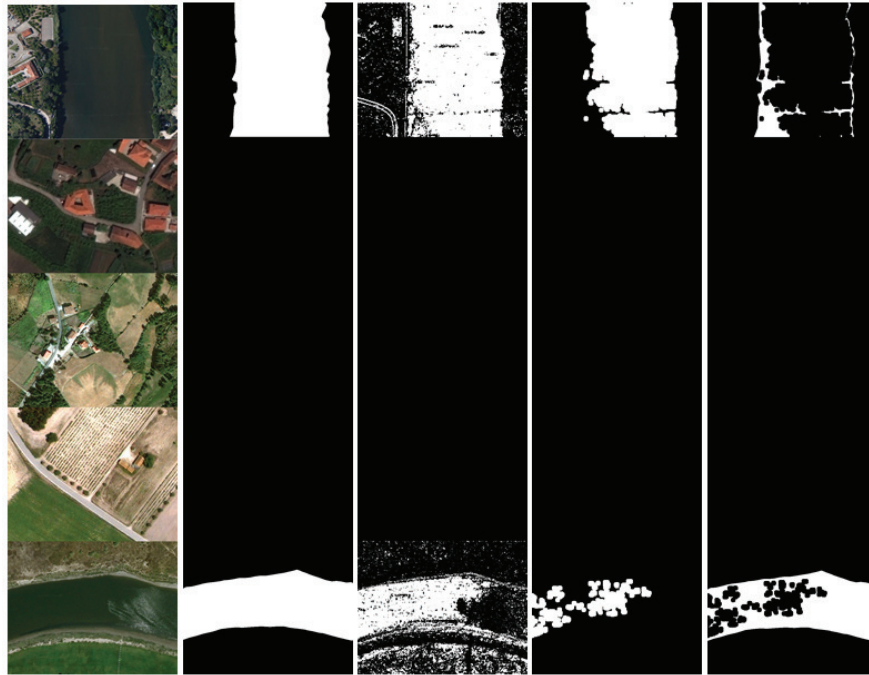


Figura 7.19: Imagens originais; Imagens 'classe'; Melhor classificação rede; Melhor classificação operadores pós-processamento; Imagens XOR

O elemento 'água' foi novamente melhor classificado para o conjunto de imagens número cinco do que para o número nove. Possivelmente pelo motivo descrito em 7.13.

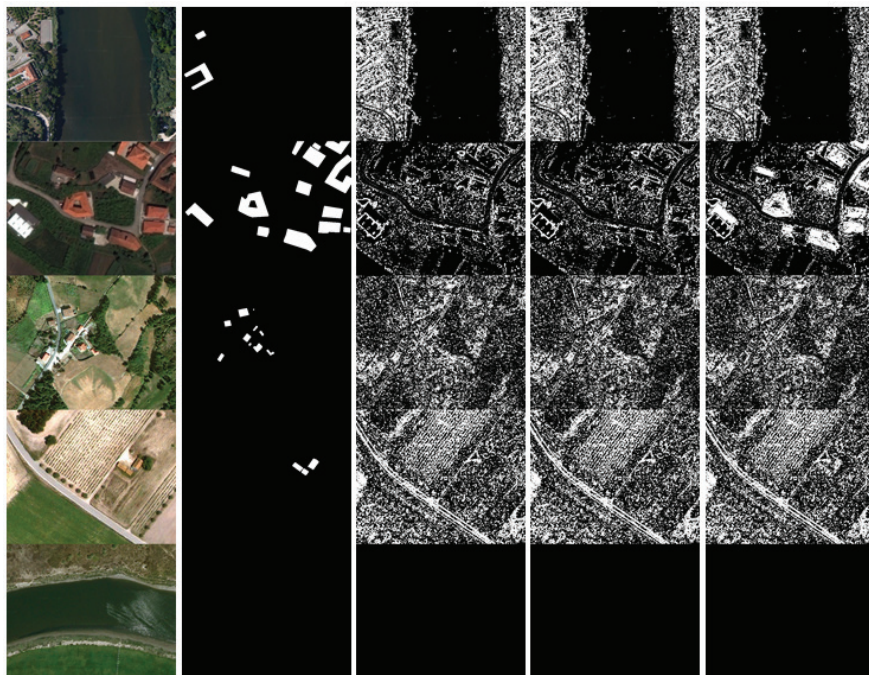


Figura 7.20: Imagens originais; Imagens 'classe'; Melhor classificação rede; Melhor classificação operadores pós-processamento; Imagens XOR

As características extraídas com recurso à Matriz de Co-ocorrência de tons cinza em conjunto

com a função entropia, não permitem, no caso do elemento 'casas', resultados aceitáveis. Em nenhum dos conjuntos de imagens o elemento em causa foi corretamente detetado.

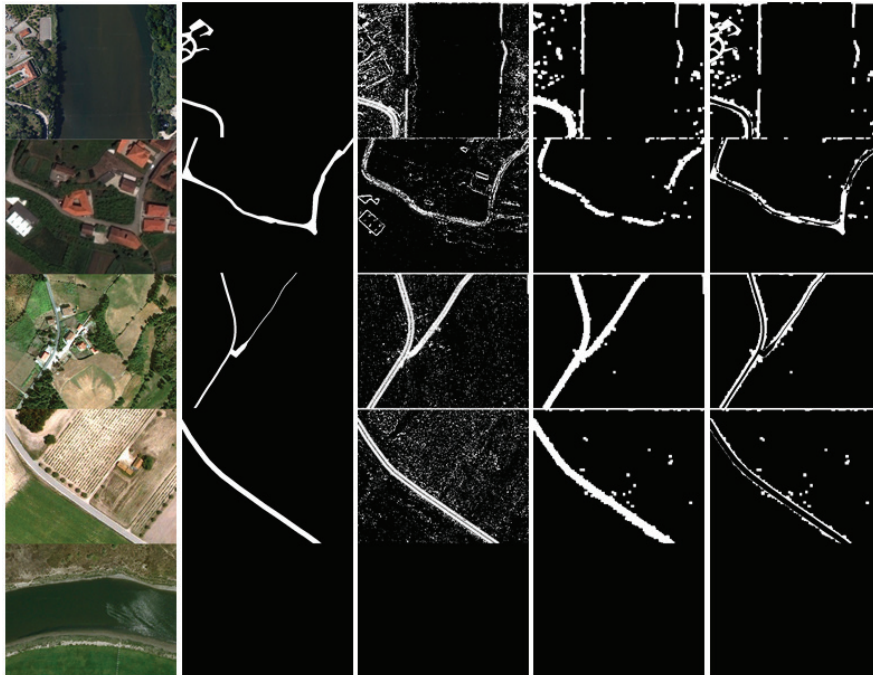


Figura 7.21: Imagens originais; Imagens 'classe'; Melhor classificação rede; Melhor classificação operadores pós-processamento; Imagens XOR

Embora nos conjuntos de imagens números cinco e seis o elemento 'estradas' não tenha sido totalmente detetado, é bem visível nos restantes conjuntos onde ele está presente, embora a área detetada esteja ligeiramente exagerada.

7.1.3 Características - Filtros Gabor

Nesta experiência foram utilizadas apenas as características extraídas através dos filtros de Gabor.

Para a classificação do elemento 'água', os resultados com um menor erro foram os obtidos com a rede neuronal treinada com o conjunto de imagens número um, com 10 neurónios; no que respeita ao elemento 'casas', os melhores resultados foram conseguidos com as redes treinadas com a imagem quatro, com 8 e 18 neurónios, com excepção do conjunto de imagens número sete, que obteve menores erros com a rede treinada com o conjunto de imagens número três, de vinte neurónios; para o elemento 'estradas', os conjuntos de imagens números cinco e seis obtiveram melhores resultados com as redes treinadas com o conjunto de imagens número um, com 16 e 18 neurónios, e os conjuntos de imagens sete e oito, com as redes treinadas com a imagem um, com 2 e 12 neurónios.

7.1.3.1 Análise de erros depois da aplicação de operadores morfológicos com diferentes elementos estruturantes

Nos gráficos seguintes constam os erros, falsos positivos e falsos negativos, após aplicação dos operadores morfológicos.

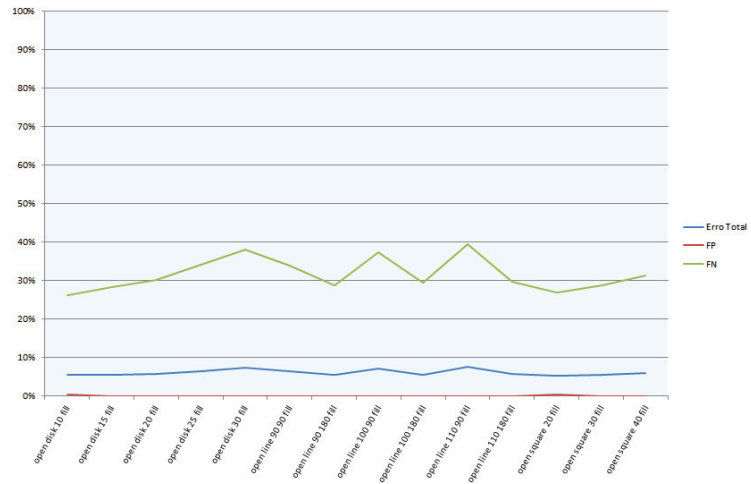


Figura 7.22: Gráfico - 'Água': Taxa de erro / Operadores morfológicos

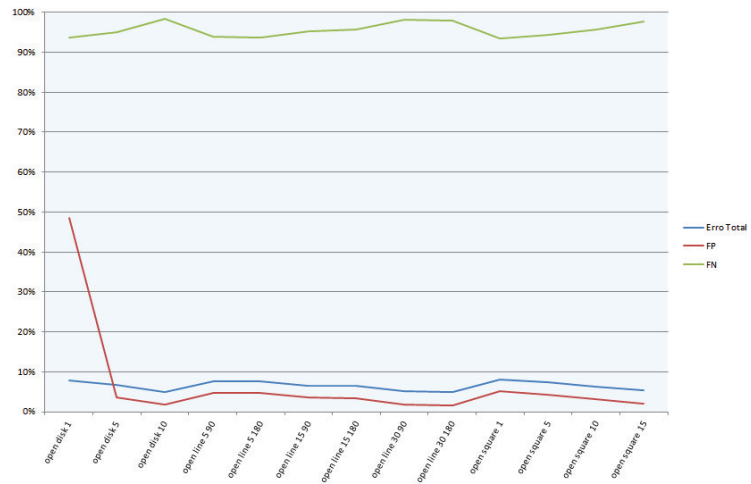


Figura 7.23: Gráfico - 'Casas': Taxa de erro / Operadores morfológicos

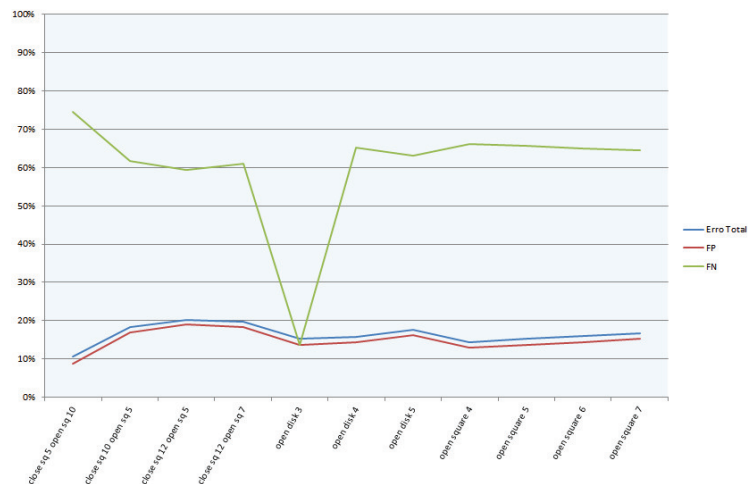


Figura 7.24: Gráfico - 'Estradas': Taxa de erro / Operadores morfológicos

7.1.3.2 Análise da Matriz Confusão do resultado com maior eficiência

Nas tabelas 7.7, 7.8 e 7.9 encontram-se as matrizes confusão dos melhores resultados para cada elemento, obtidos nesta experiência, tendo em conta a eficiência.

Tabela 7.7: Matriz Confusão do melhor resultado para o elemento 'água'

	Positivos	Negativos
Positivos	19,17	0,06
Negativos	26,30	80,83

Tabela 7.8: Matriz Confusão do melhor resultado para o elemento 'casas'

	Positivos	Negativos
Positivos	3,36	4,69
Negativos	93,32	96,64

Tabela 7.9: Matriz Confusão do melhor resultado para o elemento 'estradas'

	Positivos	Negativos
Positivos	2,89	13,70
Negativos	13,70	97,11

As matrizes de confusão relativas à experiência com características extraídas através dos filtros de *Gabor* mostram que existe uma taxa bastante elevada de falsos negativos no caso do elemento 'casas', ou seja, de todos os píxeis 'casas', apenas menos de 10% foi corretamente assinalado. Relativamente aos restantes elementos, a taxa de falsos negativos é bastante menor. A taxa de falsos positivos mostra-se reduzida para todos os elementos, especialmente para o elemento 'água', ou seja, de todos os píxeis assinalados como elemento, apenas uma pequena percentagem foi incorretamente identificada.

7.1.3.3 Análise das Imagens

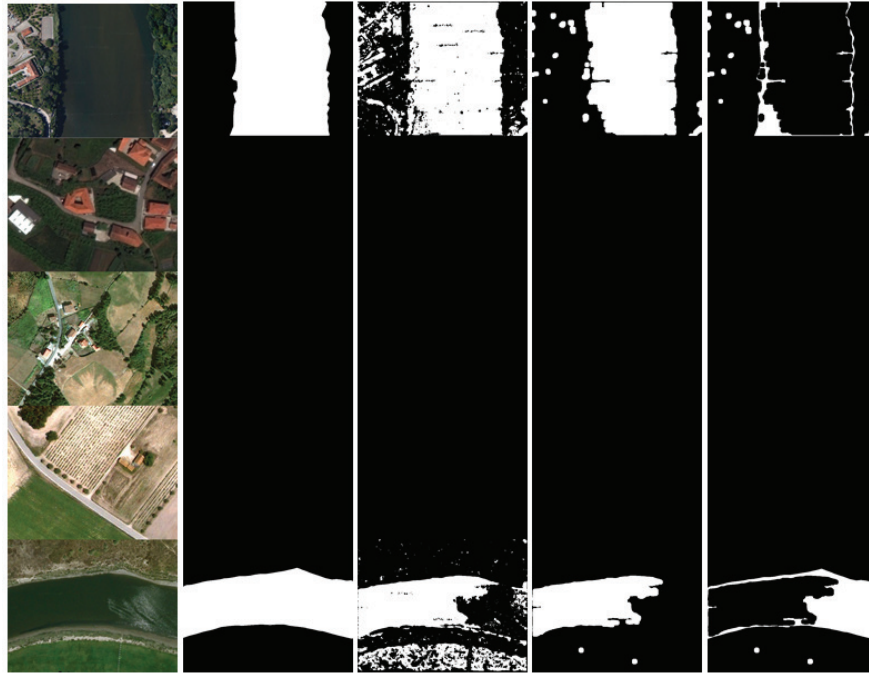


Figura 7.25: Imagens originais; Imagens 'classe'; Melhor classificação rede; Melhor classificação operadores pós-processamento; Imagens XOR

No caso do elemento 'água', não houve uma detecção perfeita, mas grande parte do elemento está identificado. A detecção no conjunto de imagens número cinco é, novamente, mais satisfatória do que no conjunto de imagens número nove, havendo uma significativa diferença entre a imagem da classificação da rede e a imagem após aplicação dos operadores morfológicos.

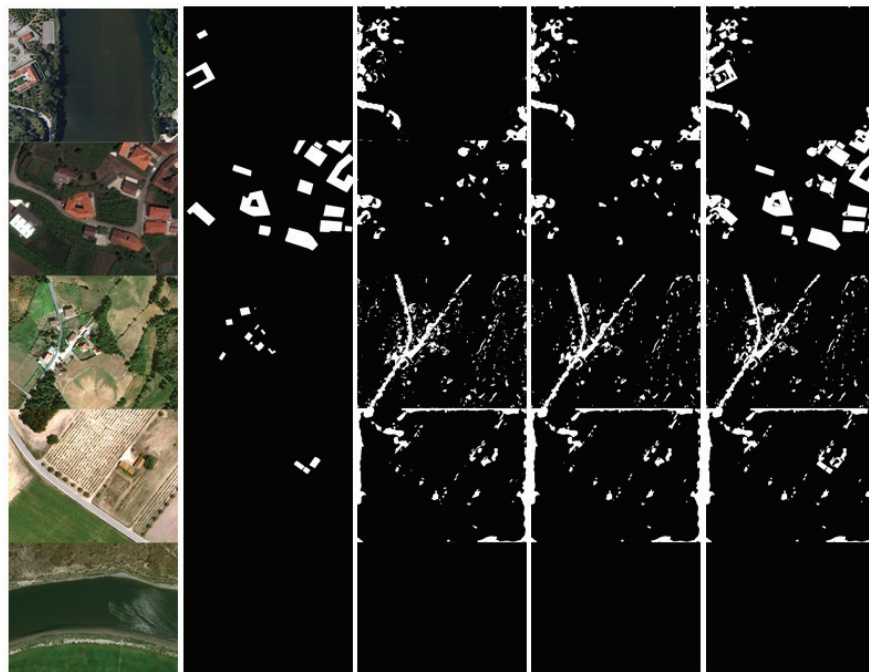


Figura 7.26: Imagens originais; Imagens 'classe'; Melhor classificação rede; Melhor classificação operadores pós-processamento; Imagens XOR

Nesta experiência, apenas no conjunto de imagens número seis, apresenta o elemento 'casas' de forma perceptível. Nos restantes conjuntos de imagens existe um grande número de píxeis erradamente assinalados como elemento.

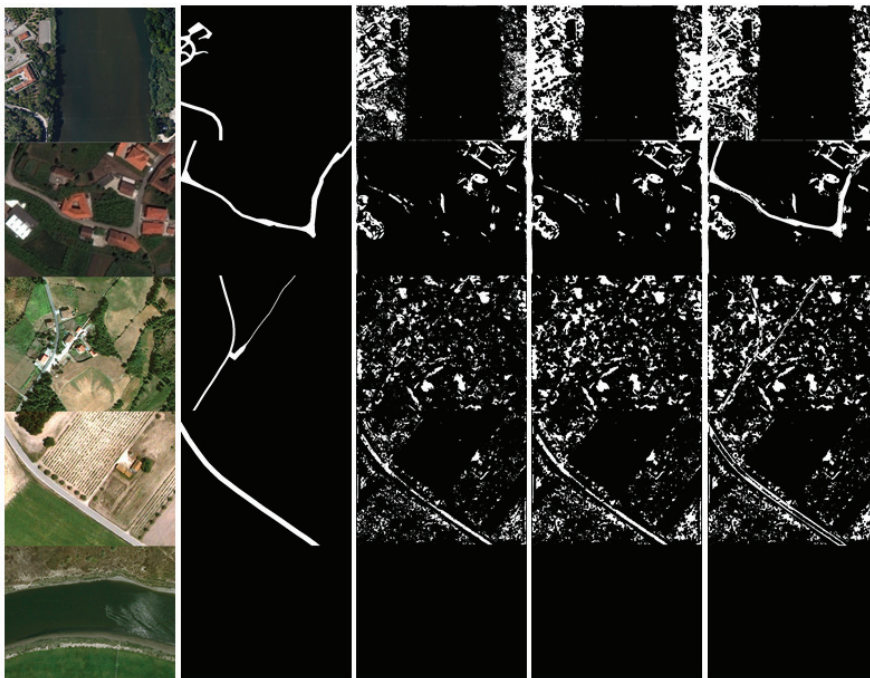


Figura 7.27: Imagens originais; Imagens 'classe'; Melhor classificação rede; Melhor classificação operadores pós-processamento; Imagens XOR

Embora no conjunto de imagens número oito seja perceptível a representação do elemento 'estradas', nenhum dos conjuntos apresenta uma classificação aceitável.

7.1.4 Características - Filtros de Gabor, Matriz de Co-Ocorrência e Função Entropia

Esta experiência foi realizada com as características obtidas da aplicação do filtros de Gabor, da matriz de co-ocorrência de tons cinza e da função Entropia.

A melhor classificação para o elemento 'água' foi obtida com a rede resultante do treino com o conjunto de imagens número um, com 10 e 12 neurónios; para o elemento 'casas', os melhores resultados foram conseguidos com as redes treinadas com o conjunto de imagens número três, com 10, 12 e 16 neurónios; relativamente ao elemento 'estradas', os menores erros foram obtidos com a rede treinada com o conjunto de imagens número um, com 6, 16 e 20 neurónios, à excepção do conjunto de imagens número cinco, que foi melhor classificado com a rede treinada com o conjunto número dois, com 10 neurónios.

7.1.4.1 Análise de erros depois da aplicação de operadores morfológicos com diferentes elementos estruturantes

Os resultados relativos ao erro, falsos positivos e falsos negativos após aplicação dos operadores morfológicos, podem ser visualizados em 7.28 para o elemento 'água', em 7.29 para o elemento 'casas' e em 7.30 para o elemento 'estradas'.

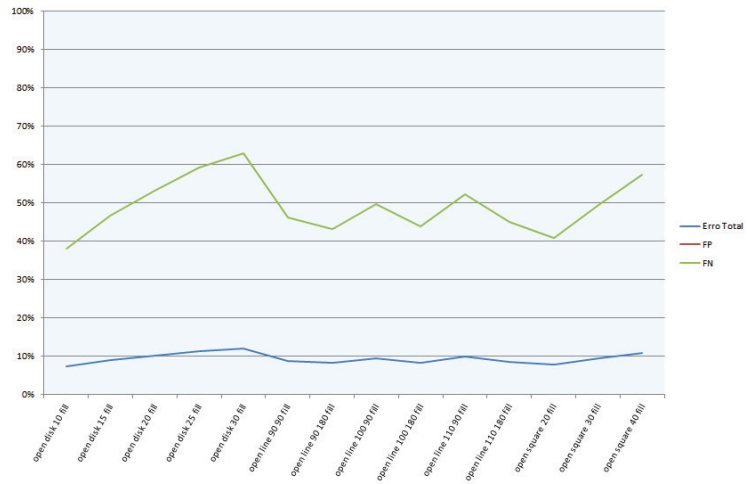


Figura 7.28: Gráfico - 'Água': Taxa de erro / Operadores morfológicos

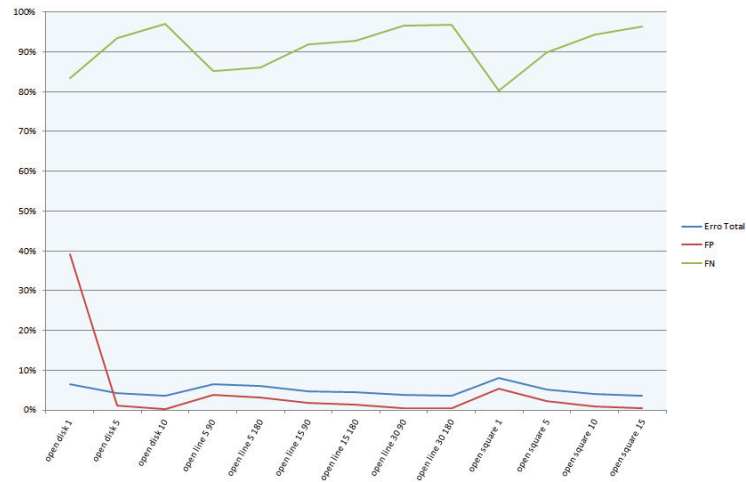


Figura 7.29: Gráfico - 'Casas': Taxa de erro / Operadores morfológicos

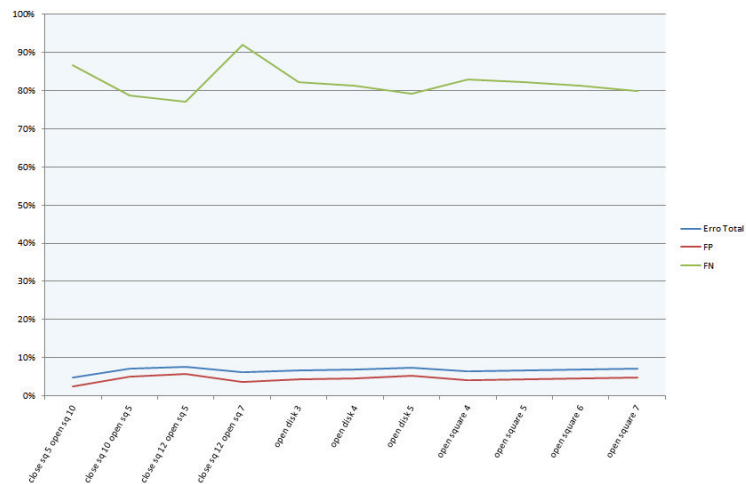


Figura 7.30: Gráfico - 'Estradas': Taxa de erro / Operadores morfológicos

7.1.4.2 Análise da Matriz Confusão do resultado com maior eficiência

Nas tabelas 7.10, 7.11 e 7.12 encontram-se as matrizes confusão dos melhores resultados para cada elemento, obtidos nesta experiência, tendo em conta o cálculo da eficiência.

Tabela 7.10: Matriz Confusão do melhor resultado para o elemento 'água'

	Positivos	Negativos
Positivos	19,17	0,00
Negativos	57,35	80,83

Tabela 7.11: Matriz Confusão do melhor resultado para o elemento 'casas'

	Positivos	Negativos
Positivos	3,36	0,32
Negativos	89,13	96,64

Tabela 7.12: Matriz Confusão do melhor resultado para o elemento 'estradas'

	Positivos	Negativos
Positivos	2,89	5,65
Negativos	77,03	97,11

Da análise das matrizes de confusão para esta experiência podemos concluir que o valor de falsos negativos é muito elevado para o elemento 'casas', ou seja, dos píxeis relativos a esse elemento, a maioria deles não foi assinalada como tal. O valor de falsos positivos é bastante baixo para o elemento 'água', ou seja, pouco elementos 'não água' foram erradamente assinalados como sendo 'água'.

7.1.4.3 Análise das Imagens

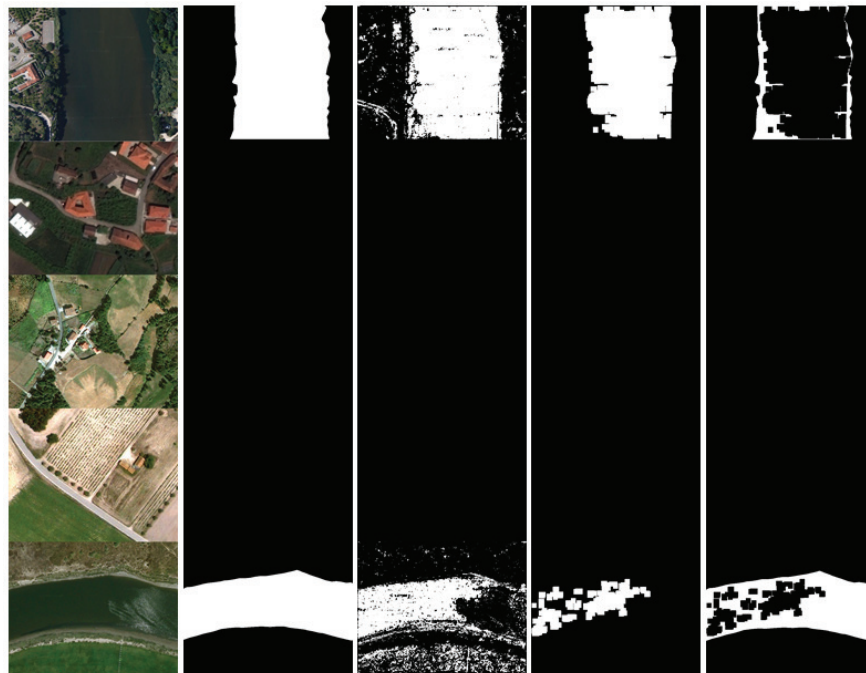


Figura 7.31: Imagens originais; Imagens 'classe'; Melhor classificação rede; Melhor classificação operadores pós-processamento; Imagens XOR

O elemento 'água' apresenta agora os resultados menos satisfatórios, tendo em consideração as experiências anteriores, principalmente para conjunto de imagens número nove, depois da aplicação dos operadores morfológicos.

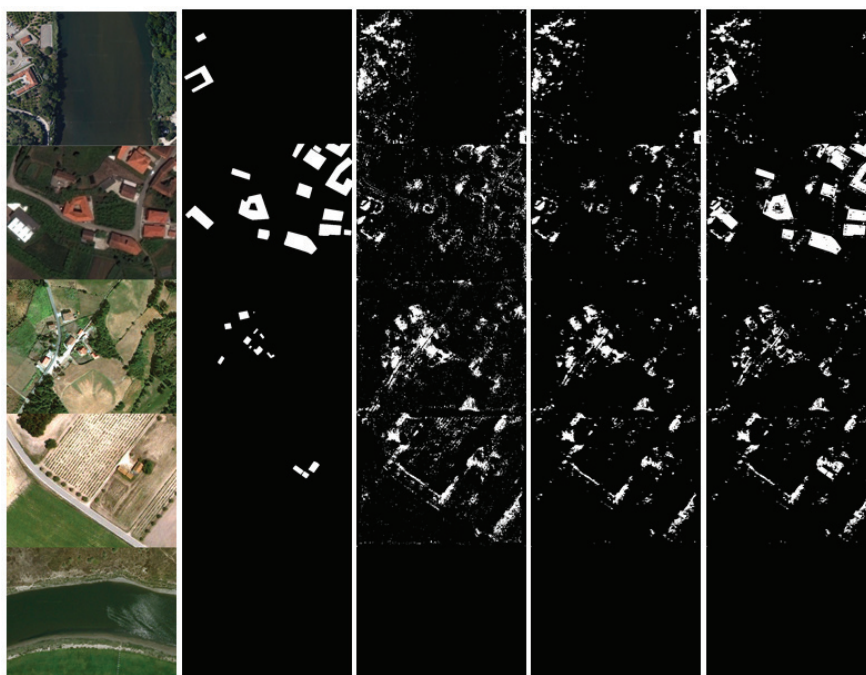


Figura 7.32: Imagens originais; Imagens 'classe'; Melhor classificação rede; Melhor classificação operadores pós-processamento; Imagens XOR

Esta experiência não apresenta resultados aceitáveis no que diz respeito ao elemento 'casas', pois em nenhum dos conjuntos de imagens ele foi corretamente detetado.

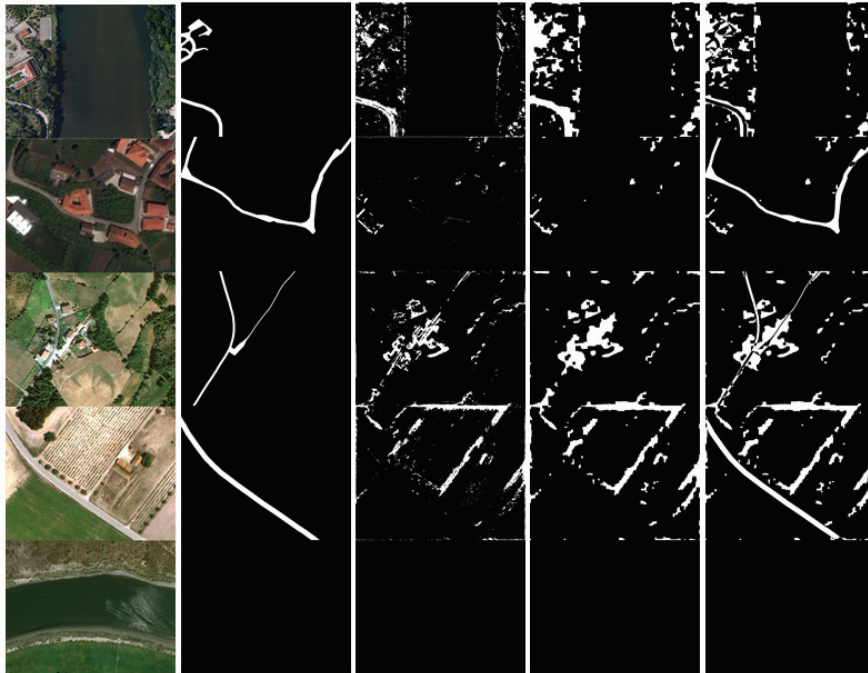


Figura 7.33: Imagens originais; Imagens 'classe'; Melhor classificação rede; Melhor classificação operadores pós-processamento; Imagens XOR

Também no caso do elemento 'estradas' não houve resultados satisfatórios; embora no conjunto de imagens número cinco o elemento tenha sido identificado, píxeis 'não-estradas' foram marcados como 'estradas'.

7.1.5 Características - Cor/Intensidade e Filtros de Gabor

Utilizando características obtidas a partir da cor, intensidade e dos filtros de Gabor, foi possível constatar que o elemento 'água' obteve melhores resultados quando testado com a rede proveniente do treino do conjunto de imagens número quatro, com 2 e 8 neurónios; as melhores classificações para o elemento 'casas' dos conjuntos de imagens número cinco e oito foram obtidas com as redes treinadas com o conjunto de imagens número dois, com 12 e 18 neurónios; já para o conjunto número seis, a melhor classificação resultou do teste com a rede treinada com o conjunto de imagens número quatro, com 16 neurónios e para o conjunto número sete, foi a rede proveniente do treino com o conjunto número três, com 20 neurónios; o elemento 'estradas' obteve o menor erro quando testado com as redes treinadas com o conjunto de imagens número dois, com 4, 14 e 16 neurónios.

7.1.5.1 Análise de erros depois da aplicação de operadores morfológicos com diferentes elementos estruturantes

Aplicados os operadores morfológicos com os diferentes elementos estruturantes, os resultados foram os constantes dos seguintes gráficos:

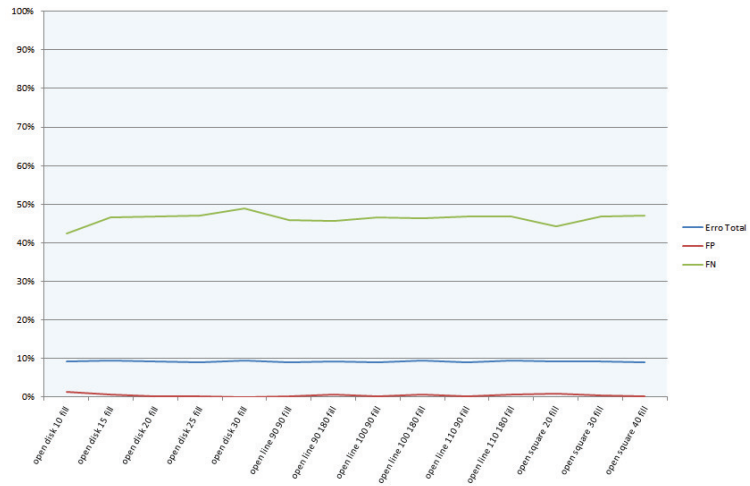


Figura 7.34: Gráfico - 'Água': Taxa de erro / Operadores morfológicos

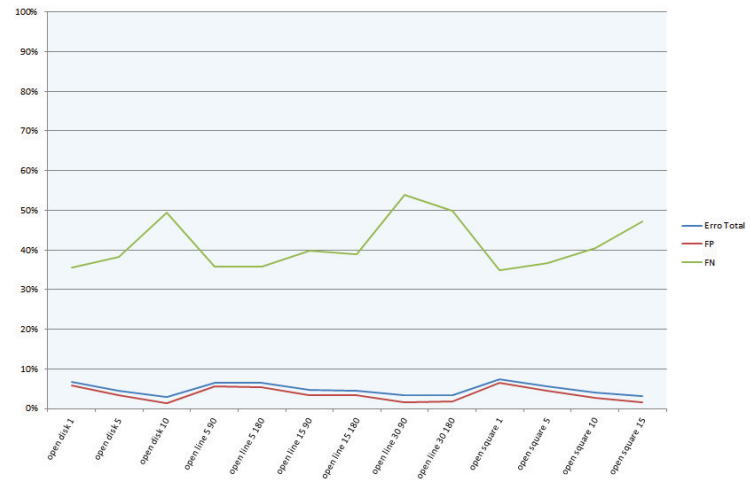


Figura 7.35: Gráfico - 'Casas': Taxa de erro / Operadores morfológicos

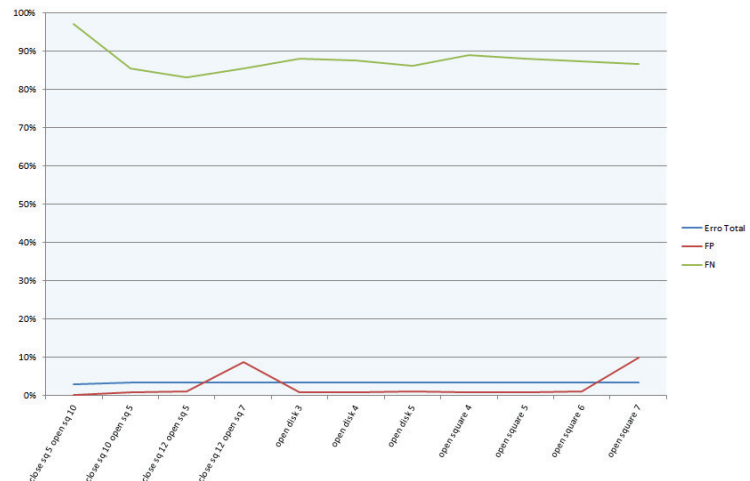


Figura 7.36: Gráfico - 'Estradas': Taxa de erro / Operadores morfológicos

7.1.5.2 Análise da Matriz Confusão do resultado com maior eficiência

Nas tabelas 7.13, 7.14 e 7.15 encontram-se as matrizes confusão dos melhores resultados para cada elemento, obtidos nesta experiência, tendo sido escolhidos os que apresentaram a maior eficiência.

Tabela 7.13: Matriz Confusão do melhor resultado para o elemento 'água'

	Positivos	Negativos
Positivos	19,17	1,30
Negativos	42,40	80,83

Tabela 7.14: Matriz Confusão do melhor resultado para o elemento 'casas'

	Positivos	Negativos
Positivos	3,36	5,82
Negativos	35,61	96,64

Tabela 7.15: Matriz Confusão do melhor resultado para o elemento 'estradas'

	Positivos	Negativos
Positivos	2,89	1,06
Negativos	83,16	97,11

Analisando as matrizes de confusão, podemos concluir que a taxa de falsos negativos é bastante alta no caso do elemento 'estradas', ou seja, dos píxeis 'estradas', uma grande percentagem não foi corretamente detetada; para os elementos 'água' e 'casas', os valores são significativamente mais baixos. Em todos os casos, a taxa de falsos positivos é baixa, embora ligeiramente superior no caso do elemento 'casas'.

7.1.5.3 Análise das Imagens

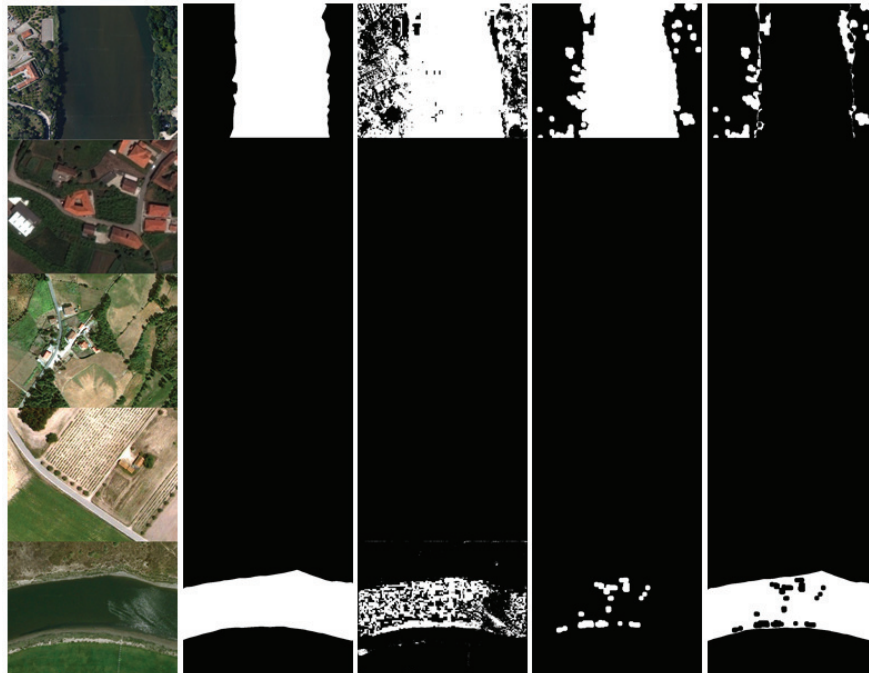


Figura 7.37: Imagens originais; Imagens 'classe'; Melhor classificação rede; Melhor classificação operadores pós-processamento; Imagens XOR

A imagem resultado desta experiência, relativamente ao elemento 'água' mostra um excesso de deteção no conjunto de imagens número cinco, ao contrário do que acontece com o conjunto número nove, onde a deteção é fraca.

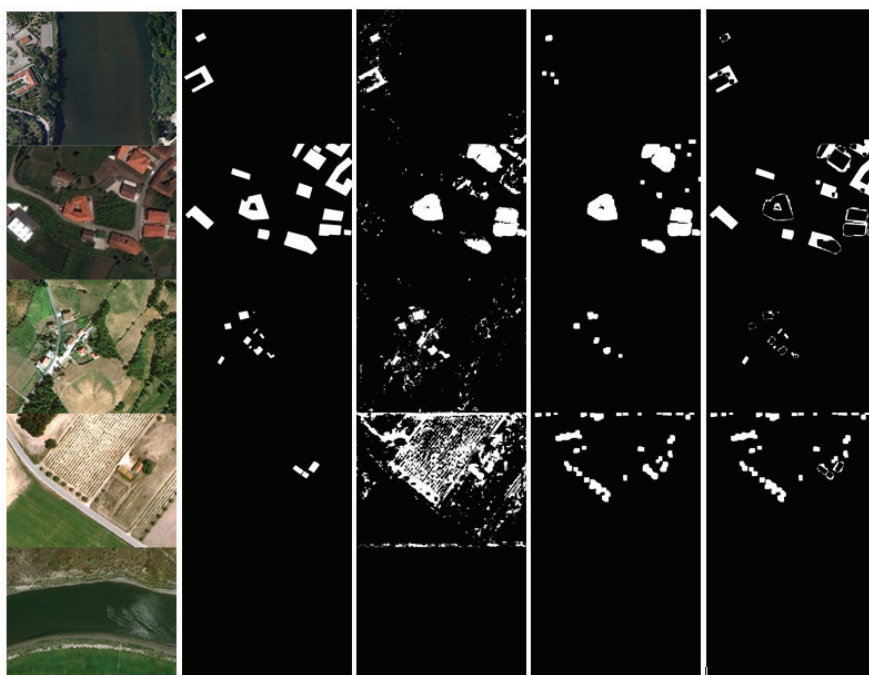


Figura 7.38: Imagens originais; Imagens 'classe'; Melhor classificação rede; Melhor classificação operadores pós-processamento; Imagens XOR

No caso do elemento 'casas', ainda que os operadores morfológicos utilizados na experiência não tenham permitido uma boa limpeza, ele é perceptível em todas as imagens onde deveria ser encontrado.

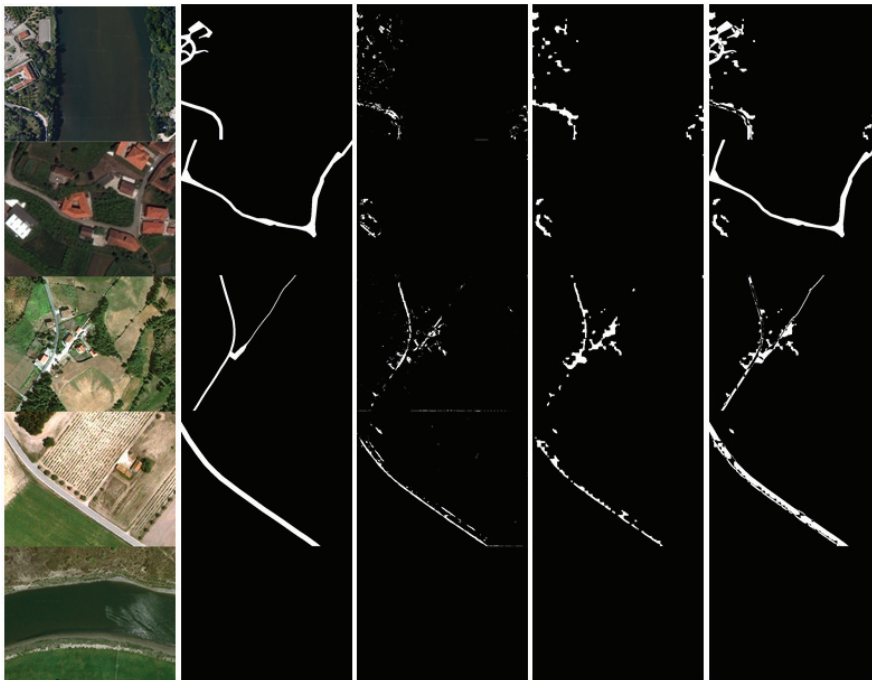


Figura 7.39: Imagens originais; Imagens 'classe'; Melhor classificação rede; Melhor classificação operadores pós-processamento; Imagens XOR

O elemento 'estradas' foi detetado, com bastante dificuldade, nos conjuntos de imagens números cinco, seis e oito. No conjunto sete não foi detetado.

7.1.6 Características - Cor, Matriz de Co-Ocorrência e Função Entropia

Nesta experiência foram utilizadas características relativas à cor, à matriz de co-ocorrência de tons cinza e à função entropia.

As melhores classificações para o elemento 'água' foram obtidas com a aplicação da rede treinada com o conjunto de imagens número quatro, com 4 e 16 neurónios; para o elemento 'casas', o conjunto de imagens número cinco e seis foram melhor classificados com a rede resultante do treino com o conjunto de imagens dois, com 6 e 12 neurónios; os conjuntos de imagens números sete e oito obtiveram melhores resultados com o rede treinada com o conjunto de imagens número três, com 4 e 18 neurónios; o elemento 'estradas' obteve erros menores quanto testado com as redes resultantes do treino com o conjunto de imagens número dois, com 8, 10 e 14 neurónios.

7.1.6.1 Análise de erros depois da aplicação de operadores morfológicos com diferentes elementos estruturantes

Os resultados após a aplicação dos operadores morfológicos podem ser visualizados nos seguintes gráficos:

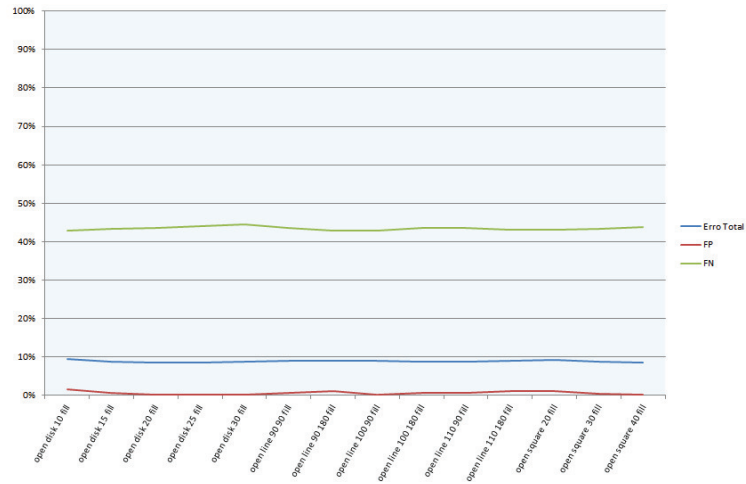


Figura 7.40: Gráfico - 'Água': Taxa de erro / Operadores morfológicos

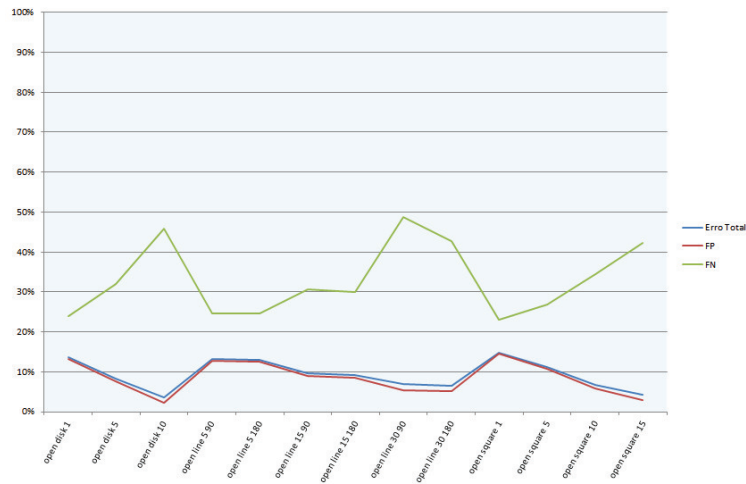


Figura 7.41: Gráfico - 'Casas': Taxa de erro / Operadores morfológicos

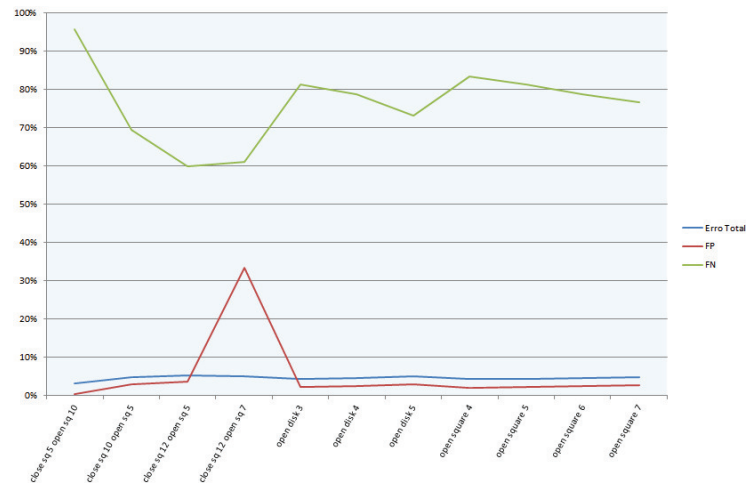


Figura 7.42: Gráfico - 'Estradas': Taxa de erro / Operadores morfológicos

7.1.6.2 Análise da Matriz Confusão do resultado com maior eficiência

Nas tabelas 7.16, 7.17 e 7.18 encontram-se as matrizes confusão dos melhores resultados para cada elemento, obtidos nesta experiência, tendo em conta a eficiência.

Tabela 7.16: Matriz Confusão do melhor resultado para o elemento 'água'

	Positivos	Negativos
Positivos	19,17	1,11
Negativos	42,91	80,83

Tabela 7.17: Matriz Confusão do melhor resultado para o elemento 'casas'

	Positivos	Negativos
Positivos	3,36	5,78
Negativos	34,51	96,64

Tabela 7.18: Matriz Confusão do melhor resultado para o elemento 'estradas'

	Positivos	Negativos
Positivos	2,89	3,61
Negativos	60,02	97,11

Da análise das matrizes de confusão, é possível verificar que, no caso do elemento 'água' existe uma taxa de falsos negativos de cerca de 43%, ou seja, quase metade dos pixels 'água' foram assinalados como sendo 'não-água', já a taxa de pixels 'não-água' assinalados como 'água' é bastante baixa. Para o elemento 'casas', o valor de falsos negativos é significativo. Quanto ao elemento 'estradas', a taxa de falsos negativos é de cerca de 60%, ou seja, menos de metade dos pixels 'estradas' foram corretamente assinalados.

7.1.6.3 Análise das Imagens

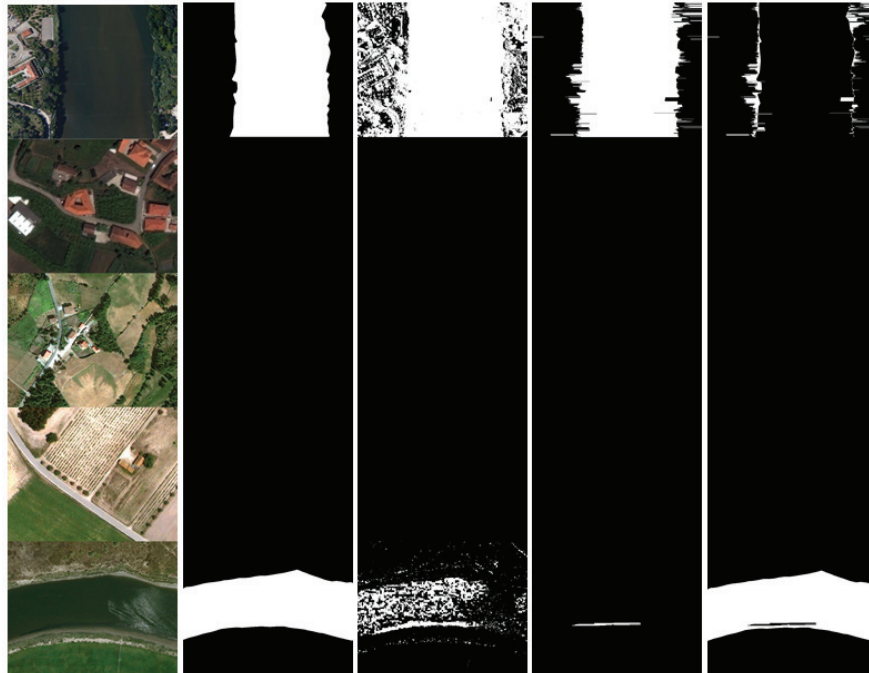


Figura 7.43: Imagens originais; Imagens 'classe'; Melhor classificação rede; Melhor classificação operadores pós-processamento; Imagens XOR

Nesta experiência é visível a falha na detecção do elemento 'água' no conjunto de imagens número nove, mas o mesmo não acontece com o conjunto de imagens número cinco.

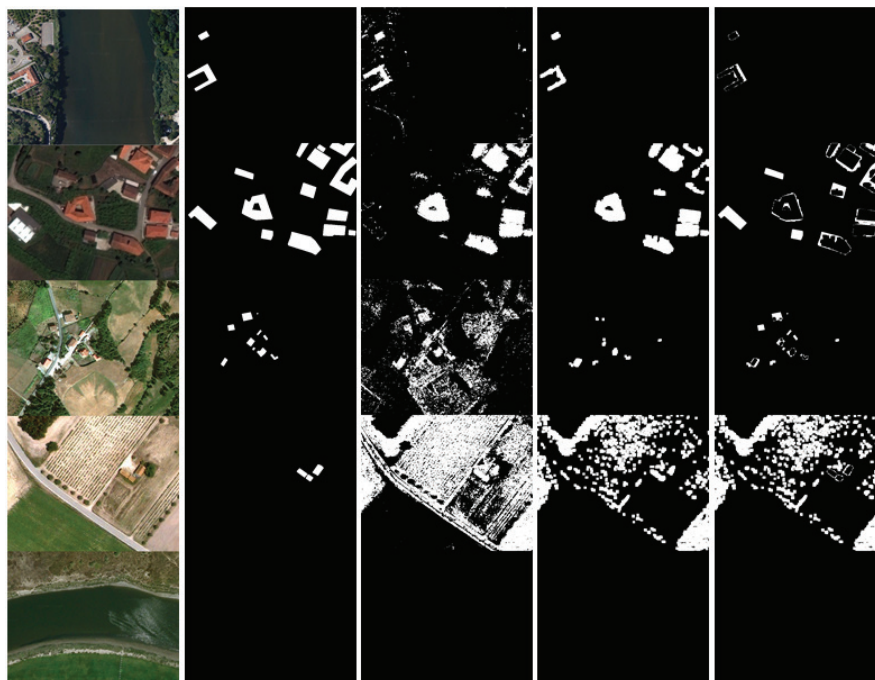


Figura 7.44: Imagens originais; Imagens 'classe'; Melhor classificação rede; Melhor classificação operadores pós-processamento; Imagens XOR

Nos conjuntos de imagens números cinco e seis, o elemento 'casas' foi satisfatoriamente dete-

tado, embora o mesmo não tenha acontecido nos conjuntos sete e oito.

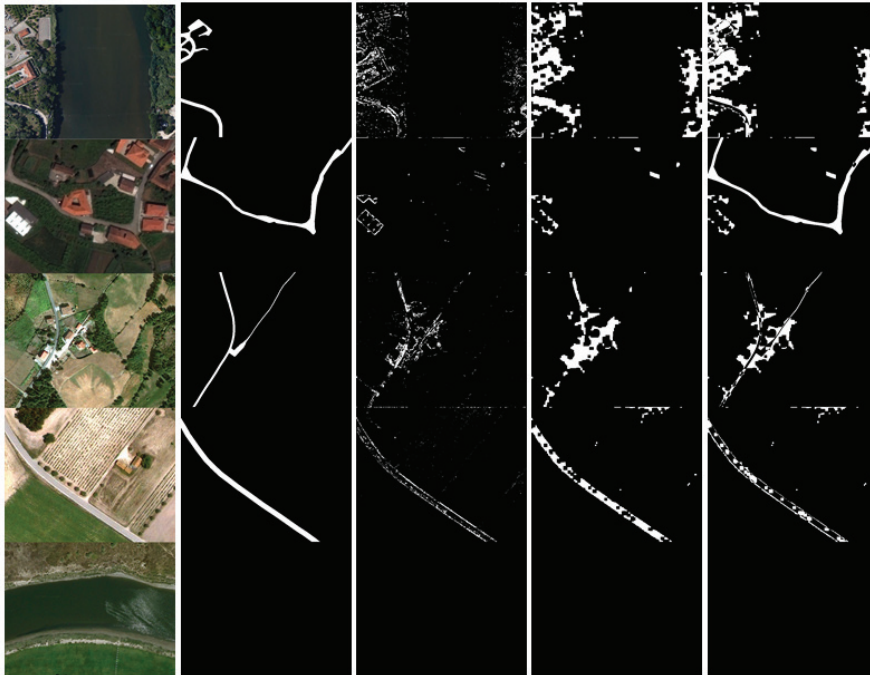


Figura 7.45: Imagens originais; Imagens 'classe'; Melhor classificação rede; Melhor classificação operadores pós-processamento; Imagens XOR

No caso do elemento 'estradas', nesta experiência, os resultados são apenas aceitáveis no caso do conjunto de imagens número oito, sendo que nos restantes conjuntos o elemento foi incorreta ou insatisfatoriamente assinalado. Neste caso, a aplicação dos operadores morfológicos causou um aumento de casos de falsos positivos.

7.1.7 Características - Cor/Intensidade, Filtros de Gabor, Matriz de Co-Ocorrência e Função Entropia

Esta experiência foi realizada utilizando características provenientes da cor e intensidade, da aplicação dos filtros de Gabor, matriz de co-ocorrência de tons cinza e função entropia.

Depois de testados os conjuntos de imagens, verificou-se que o elemento 'água' obteve melhores resultados com as redes treinadas com o conjunto de imagens número quatro, com 16 e 4 neurónios; para o elemento 'casas', as melhores classificações foram obtidas com as redes treinadas com o conjunto de imagens número três, com 10, 12 e 16 neurónios; relativamente ao elemento 'estradas' os menores valores de erro foram conseguidos com as redes treinadas com o conjunto quatro, com 12 e 14 neurónios para todos os conjuntos de teste, à excepção do conjunto número sete, que obteve melhores resultados com a rede treinada com o conjunto três, com 14 neurónios.

7.1.7.1 Análise de erros depois da aplicação de operadores morfológicos com diferentes elementos estruturantes

Após a aplicação dos operadores morfológicos, foram obtidos os resultados assinalados nos gráficos 7.46, 7.47 e 7.48:

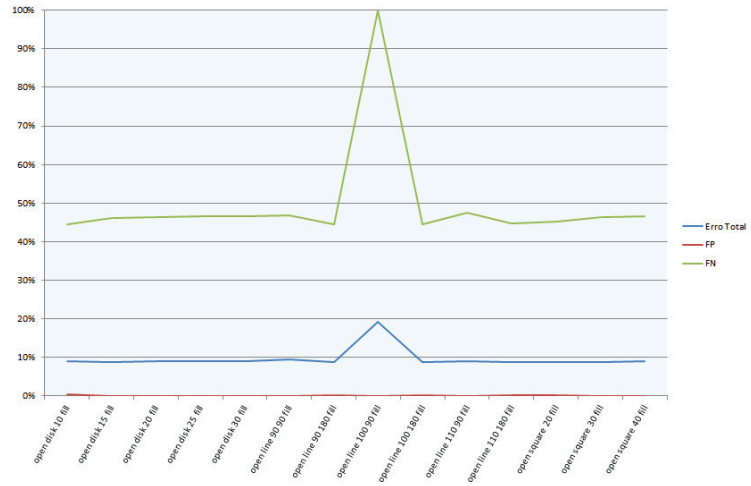


Figura 7.46: Gráfico - 'Água': Taxa de erro / Operadores morfológicos

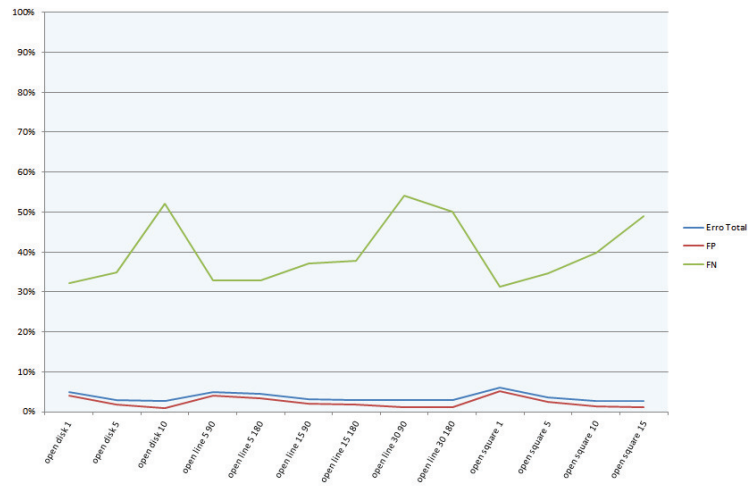


Figura 7.47: Gráfico - 'Casas': Taxa de erro / Operadores morfológicos

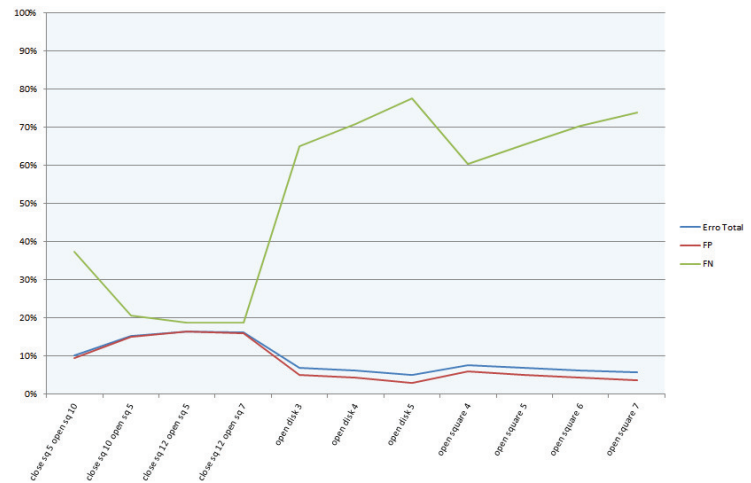


Figura 7.48: Gráfico - 'Estradas': Taxa de erro / Operadores morfológicos

7.1.7.2 Análise da Matriz Confusão do resultado com maior eficiência

Nas tabelas 7.19, 7.20 e 7.21 encontram-se as matrizes confusão dos melhores resultados para cada elemento, obtidos nesta experiência, tendo em conta a eficiência.

Tabela 7.19: Matriz Confusão do melhor resultado para o elemento 'água'

	Positivos	Negativos
Positivos	19,17	0,28
Negativos	44,45	80,83

Tabela 7.20: Matriz Confusão do melhor resultado para o elemento 'casas'

	Positivos	Negativos
Positivos	3,36	1,40
Negativos	39,84	96,64

Tabela 7.21: Matriz Confusão do melhor resultado para o elemento 'estradas'

	Positivos	Negativos
Positivos	2,89	16,04
Negativos	18,72	97,11

Nesta experiência, que incluiu a extração de todas as características abordadas, os valores de falsos negativos são mais baixos para o elemento 'estradas', aumentando para o elemento 'casas' e atingindo os 44% para o elemento 'água', ou seja, quase metade dos píxeis 'água' foram assinalados como sendo 'não-água'. A taxa de falsos positivos é mais elevada para o elemento 'estradas'.

7.1.7.3 Análise das Imagens

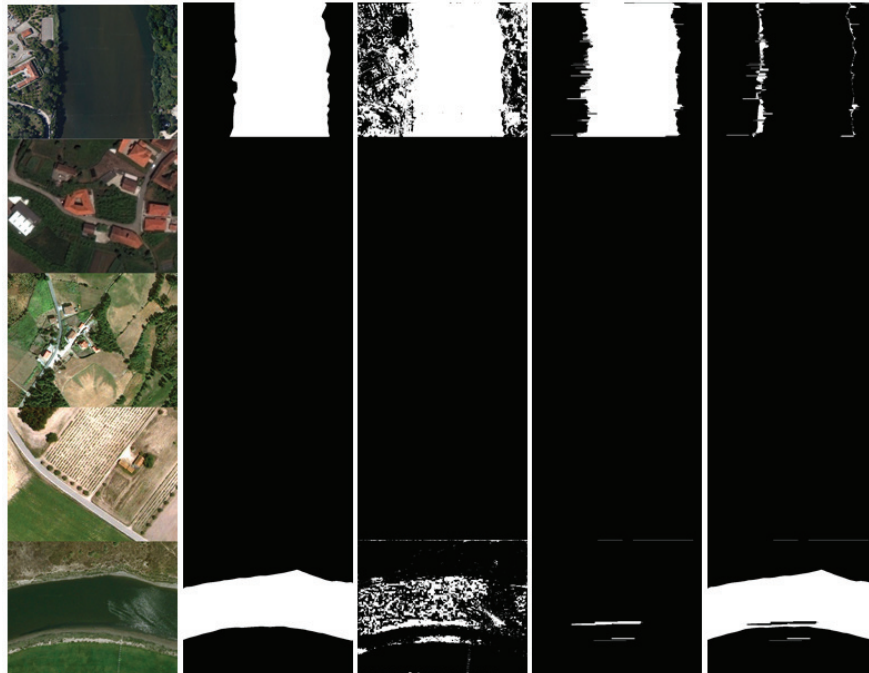


Figura 7.49: Imagens originais; Imagens 'classe'; Melhor classificação rede; Melhor classificação operadores pós-processamento; Imagens XOR

O elemento água foi detetado no conjunto de imagens número cinco, mas no conjunto de imagens número nove é insatisfatório.

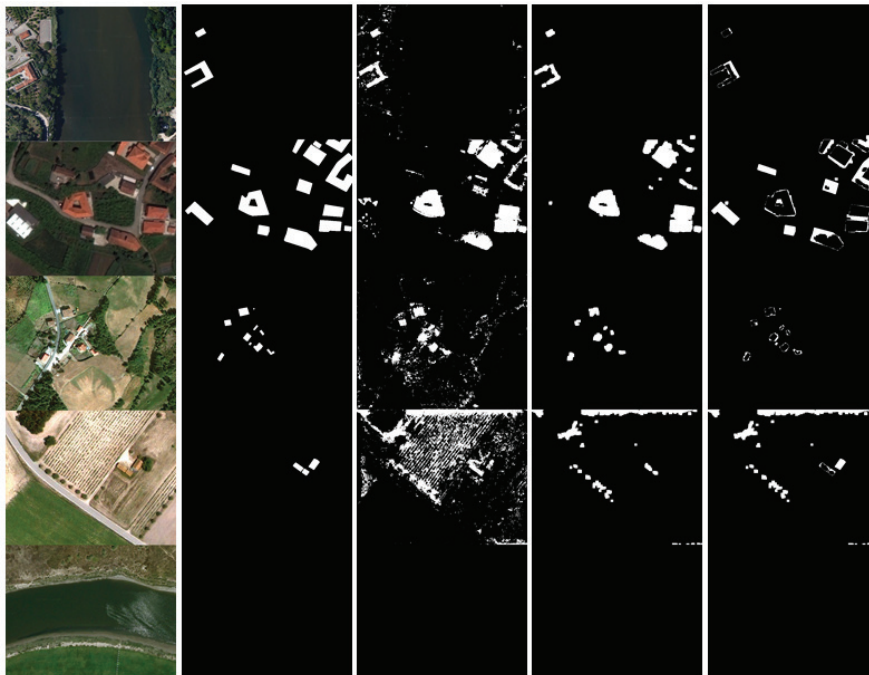


Figura 7.50: Imagens originais; Imagens 'classe'; Melhor classificação rede; Melhor classificação operadores pós-processamento; Imagens XOR

O elemento 'casas' foi satisfatoriamente detetado em todas as imagens onde ele existe, embora

os operadores morfológicos utilizados nestas experiências não tenham permitido remover todo o ruído existente.

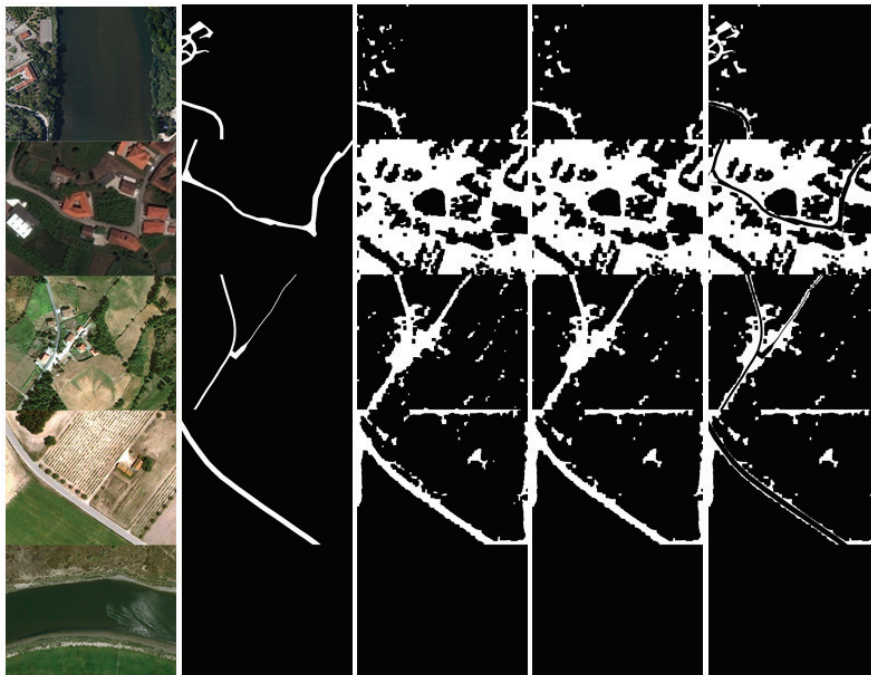


Figura 7.51: Imagens originais; Imagens 'classe'; Melhor classificação rede; Melhor classificação operadores pós-processamento; Imagens XOR

Embora não seja das melhores classificações, o elemento 'estradas' foi detetado nos conjuntos de imagens números cinco, sete e oito.

7.1.7.4 Operadores Morfológicos - Elementos estruturantes

Vários foram os elementos estruturantes associados aos operadores morfológicos utilizados nestas experiências. Para cada elemento, foram escolhidos vários daqueles que depois de uma bateria de testes mostraram ser os mais apropriados. Mas houve os que se destacassem e fossem mais vezes escolhidos pela maior eficiência.

São eles:

Para o elemento 'casas':

- open(disk, 10);
- open(line, 90, 180);
- open(square, 20);

Para o elemento 'casas':

- open(square, 10);
- open(line, 5, 90);
- open(line, 5, 180);
- open(square, 1);

Para o elemento 'estradas':

- close(square, 12) open(square, 5);
- close(square, 12) open(square, 7);
- open(disk, 3);
- open(disk, 5).

O elemento estruturante, associado à abertura, que se mostrou mais eficaz para o elemento 'água' foi o *disk* 10; No caso do elemento 'casas', o que melhor resultados apresentou foi o *square* 10, associado à abertura. Para o elemento 'estradas', os menores erros foram conseguidos com uma combinação de fecho com elemento estruturante *square* 12, seguida de abertura *square* 5.

7.2 Discussão

Os resultados obtidos nas experiências indicam que o elemento 'água' foi sempre detetado, embora nunca na sua totalidade, principalmente no conjunto de imagens número nove. Esta falha pode dever-se ao brilho existente na imagem, uma vez que é nessa região que se verificam os maiores problemas de classificação, em todas as experiências.

O elemento 'casas' foi mais facilmente detetado nas experiências que envolveram a extração de características relativas à cor e intensidade. Os piores resultados foram os provenientes da experiência onde foram usadas características extraídas através da matriz de co-ocorrência de tons cinza e da função entropia.

Já a melhor classificação para o elemento 'estradas' foi obtido na experiência que utilizou características extraídas da matriz de co-ocorrência de tons cinza e da função entropia. Em todas as restantes experiências, os resultados mostraram alguns problemas na sua deteção.

Existem outros aspetos, além das características, que podem ter influenciado os resultados, tais como o facto de a seleção das melhores imagens, logo após o teste na rede neuronal, ter sido feito apenas com base no erro geral e não ter sido levado em consideração a taxa de falsos negativos e falsos positivos; o facto de todas as redes, treino e teste, terem utilizado neurónios no intervalo [2-20], de dois em dois; a escolha dos elementos estruturantes dos operadores morfológicos que foram aplicados a todas as imagens em simultâneo: o operador morfológico/elemento estruturante que pode melhorar o elemento 'água' da imagem do conjunto cinco, pode não ser útil para melhorar o elemento da imagem do conjunto nove.

Nas experiências realizadas, não foi possível chegar a conclusões sobre o número ideal de neurónios a utilizar para obter melhores resultados, visto que os melhores resultados não apresentaram nenhuma regra ou padrão quanto a esse aspeto.

Capítulo 8

Conclusões e Trabalho Futuro

8.1 Conclusões

Os resultados obtidos neste trabalho, apesar de poderem ter sido feitas muitas outras experiências, permitiram chegar a várias conclusões relativamente aos métodos para extração de características, à utilização de operadores morfológicos e aos diferentes elementos estruturantes utilizados nas operações morfológicas.

Os melhores resultados para o elemento 'água' foram obtidos com as características extraídas da cor e intensidade e utilizando os filtros de Gabor; já para o elemento 'casas', os melhores resultados foram obtidos com as características relativas à cor e intensidade; o elemento 'estradas' foi mais facilmente detetado quando foram utilizadas características extraídas da matriz de co-ocorrência de tons cinza e função entropia.

Os resultados podem ter sido negativamente influenciados pela heterogeneidade dos elementos dos vários conjuntos de imagens. O elemento 'água' do conjunto de imagens número cinco tem características diferentes do mesmo elemento no conjunto de imagens número nove, sendo que neste último caso, é notória a influência do brilho presente no elemento em causa, em todas as experiências realizadas. O elemento 'casas' do conjunto de imagens número seis tem propriedades que o tornam mais facilmente identificável, quer pela cor, quer pelo tamanho dos telhados; no conjunto de imagens sete e oito, os telhados, além de terem tamanhos mais reduzidos, apresentam uma cor mais suavizada relativamente ao conjunto anterior. O elemento 'estradas' também se apresenta bastante heterogénio de conjunto para conjunto de imagens, e no conjunto número oito, as características do campo agrícola envolvente pode suscitar alguma confusão.

Na maioria das experiências realizadas, o valor de falsos negativos mostrou-se elevado, ao contrário do valor relativo a falsos positivos. Os operadores morfológicos e elementos estruturantes aplicados mostram que diminuir os casos de falsos negativos provoca, na maioria dos casos, um aumento dos falsos positivos e *vice-versa*.

O elemento estruturante, associado à abertura, que se mostrou mais eficaz para o elemento 'água', foi o *disk* 10; já para o elemento 'casas', o que melhor resultados apresentou foi o *square* 1, associado à abertura. Para o elemento 'estradas', os menores erros foram conseguidos com uma combinação de fecho com elemento estruturante *square* 12, seguida de uma abertura *square* 5.

8.2 Trabalho Futuro

Muitas experiências mais poderiam ter sido realizadas, tais como experiências relativas ao valor da variância do PCA, experiências para reconhecimento dos três elementos usando uma única rede para o treino e uma única para o teste, em vez de um rede por cada elemento, outras experiências com maior variação do número de neurónios e até com muitos mais elementos estruturantes para os operadores morfológicos, no sentido de encontrar um que fosse adequado

Google Earth™ Search Engine - Implementação e Validação do Método de Reconhecimento

às características específicas de todas as imagens com um determinado elemento. Dada a escassez de tempo, todas essas experiências são remetidas para trabalho futuro. Nesse sentido podem ainda ser adicionados alguns aspectos que tornem o sistema automático, mais poderoso e capaz de obter bons resultados em imagens com um grau de homogeneidade elevado. Um dos aspectos seria acrescentar um conjunto de características mais específicas a cada um dos elementos a detectar, pois em nenhuma das experiências foram obtidos resultados satisfatórios para todos os elementos em simultâneo.

Bibliografia

- [1] Infantosi A.F.C. Azevedo C.M. Pereira W.C.A. Alvarenga, A.V. Application of morphological operators on the segmentation and contour detection of ultrasound breast images. *Revista Brasileira de Engenharia Biomédica*, 19(2):91-101, 2003. 51
- [2] Teresa Bernarda Ludemir António de Pádoa Braga, André Ponce de Leon F. de Carvalho. *Redes Neurais Artificiais - Teoria e Aplicações*, 2/E. LTC, 2007. 47, 48
- [3] Dev Ashish. Land-use classification of aerial images using artificial neural networks, 2002. 5, 30
- [4] G. Bittencourt. *Inteligência Artificial Ferramentas e Teorias*. UFSC, 2006. 45, 49
- [5] Teofilo E. Campos. Técnicas de seleção de características com aplicações em reconhecimento de faces, 2001. 46
- [6] Yohay Carmel and Ronen Kadmon. Computerized classification of mediterranean vegetation using panchromatic aerial photographs. *Journal of Vegetation Science*, 9(3):445-454, 1998. 11, 30
- [7] Russell G. Congalton. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, 37(1):35 - 46, 1991. 14
- [8] John G. Daugman. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *J. Opt. Soc. Am. A*, 2(7):1160-1169, July 1985. 38
- [9] R. O. Duda and P. E. Hart. *Pattern Classification and Scene Analysis*. John Willey & Sons, New York, 1973. 45
- [10] Vagner Morais e Cláudio Vieira. *Matlab 7& 6 - Curso Completo*, 3/E. FCA, Lisboa, Portugal, 2006. 33
- [11] D. Gabor. Theory of communication. *Journal of the Institute of Electrical Engineers*, page 429-457, 1946. 37
- [12] D. Gabor. Theory of communication. part 1: The analysis of information. *Electrical Engineers - Part III: Radio and Communication Engineering*, *Journal of the Institution of*, 93(26):429-441, 1946. 38
- [13] Kendra Cherry About.com Guide. Young-helmholtz trichromatic theory of color vision, 2013. 34
- [14] R.M. Haralick, K. Shanmugam, and Its'Hak Dinstein. Textural features for image classification. *Systems, Man and Cybernetics, IEEE Transactions on*, SMC-3(6):610-621, 1973. 8, 9, 30, 35
- [15] Simon Haykin. *Neural Networks: a Comprehensive Foundation*, 2/E. Prentice Hall, 1999. 45, 48, 49
- [16] A. Hinz. The z/i digital aerial camera system. In *Proceedings of the 47th Photogrammetric Week*, pages 109-115. Wichmann Verlag, Heidelberg, 1999. 18

- [17] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 20(11):1254-1259, 1998. 19
- [18] P. Kahn, L. Kitchen, and E. M. Riseman. A fast line finder for vision-guided robot navigation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 12(11):1098-1102, November 1990. 28
- [19] M. Kendall. *Multivariate Analysis*. MacMillan Publishing Co. INC., 1980. 41
- [20] Zhu Le-qing and Zhang Zhen. Auto-classification of insect images based on color histogram and glcm. In *Fuzzy Systems and Knowledge Discovery (FSKD), 2010 Seventh International Conference on*, volume 6, pages 2589-2593, 2010. 20, 35
- [21] Y. LeCun, L. Bottou, G. Orr, and K. Muller. Efficient backprop. In G. Orr and Muller K., editors, *Neural Networks: Tricks of the trade*. Springer, 1998. 41
- [22] Richard P. Lippmann. Review of neural networks for speech recognition. *Journal Neural Computation archive*, pages 1-38, 1989. 46
- [23] Juho Lumme. Supervised classification using aerial photographs and lidar data. 2006. 14, 30
- [24] Thomas M. Mitchell. *Machine Learning*. McGraw-Hill, Inc., New York, NY, USA, 1 edition, 2001. 47
- [25] Nick Kingsbury (ngk10@cam.ac.uk). Entropy, 2003. 37
- [26] Aude Oliva and Antonio Torralba. Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*, 42(3):145-175, 2001. 25
- [27] K. Pearson. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 2(6):559-572, 1901. 41
- [28] Zhi qiang Lv and Qi gang Zhou. Monitoring of urban sprawl using remote sensing technology in the rapid urbanized region, china. In *Geoscience and Remote Sensing (IITA-GRS), 2010 Second IITA International Conference on*, volume 1, pages 426-429, 2010. 1
- [29] P. Quezel, R. Morandini, and R. Tomaselli. *Mediterranean forests and maquis: ecology, conservation, and management*. Number vol. 22 in MAB technical notes. Unesco, 1977. 11
- [30] Vladimir Risojevi, Snjeana Momi, and Zdenka Babi. Gabor descriptors for aerial image classification. In Andrej Dobnikar, U. Lotri, and Branko, editors, *Adaptive and Natural Computing Algorithms*, volume 6594 of *Lecture Notes in Computer Science*, pages 51-60. Springer Berlin Heidelberg, 2011. 25
- [31] V. Risojevic and Z. Babic. Orientation difference descriptor for aerial image classification. In *Systems, Signals and Image Processing (IWSSIP), 2012 19th International Conference on*, pages 150-153, 2012. 24, 31
- [32] Frank Rosenblatt. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Reviews*, 65(6):386-408, November 1958. 47
- [33] A. Rosenfeld and M. Thurston. Visual texture analysis, 2. *Computer Sci. Cent., Univ. Maryland, College Park*, pages 70-129, 1970. 9

- [34] Viñas O. Domingo A. Marco V. Ruiz, A. Tree species classification from aerial images and lidar in agricultural areas. 2009. 15, 31
- [35] Stuart J. Russell and Peter Norvig. *Artificial Intelligence A Modern Approach, 2/E*. Prentice Hall, 2003. 47
- [36] V.N. Sedykh. Using aerial photography and satellite imagery to monitor forest cover in western siberia. In MichaelJ. Apps, DavidT. Price, and Joe Wisniewski, editors, *Boreal Forests and Global Change*, pages 499-507. Springer Netherlands, 1995. 1
- [37] Young-Woo Seo, Nathan Ratliff, and Christopher Urmson. Self-supervised aerial image analysis for extracting parking lot structure. In *Proceedings of International Joint Conference on Artificial Intelligence (IJCAI-09)*. AAAI Press, July 2009. 27, 31
- [38] Short D. Short, D. *Studying long term community dynamics using image processing. In: Tenhunen, J.D. (ed.) Plant response to stress in Mediterranean climates*. Springer-Verlag, Berlin, 1987. 30
- [39] Alvy Ray Smith. Rgb to hsv transforms. aug 1976. 35
- [40] Pierre Soille. Erosion und dilatation. In *Morphologische Bildverarbeitung*, pages 51-93. Springer Berlin Heidelberg, 1998. 51
- [41] Rohlf F.J. Sokal, R.R. *Biometry*. Freeman and Co., 1981. 14
- [42] Guindon B. Goodenough Teillet, P.M. On the slope-aspect correction of multispectral scanner data. *Canadian Journal of Remote Sensing*, (8):84-106, 1982. 12
- [43] PrintWiki the Free Encyclopedia of Print. Maxwell's triangle, 2006. 34
- [44] Inc. The MathWorks. Mathworks - pca, 1994-2013. 42
- [45] Sergios Theodoridis and Konstantinos Koutroumbas. *Pattern Recognition, Third Edition*. Academic Press, Inc., Orlando, FL, USA, 2006. 45
- [46] José L. Todesco. Reconhecimento de padrões usando uma rede neuronal artificial com uma função de base radial: uma aplicação na classificação de cromossomos humanos, 1995. 46
- [47] E. B. Troy, E. S. Deutsch, and A. Rosenfeld. Gray-level manipulation experiments for texture analysis. *Systems, Man and Cybernetics, IEEE Transactions on*, SMC-3(1):91-98, 1973. 9
- [48] I. H. Witten and E. Frank. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2 edition, 2005. 56
- [49] Sheng Xu, Tao Fang, Hong Huo, and Deren Li. A novel method of aerial image classification based on attention-based local descriptors. *Procedia Earth and Planetary Science*, 1(1):1133 - 1139, 2009. special issue title: Proceedings of the International Conference on Mining Science and Technology. 18, 31
- [50] Sheng Xu, Tao Fang, Deren Li, and Shiwei Wang. Object classification of aerial images with bag-of-visual words. *Geoscience and Remote Sensing Letters, IEEE*, 7(2):366-370, 2010. 20, 31

- [51] Haipeng Yu, Jun Cao, Wei Luo, and Yixing Liu. Image retrieval of wood species by color, texture, and spatial information. In *Information and Automation, 2009. ICIA '09. International Conference on*, pages 1116-1119, 2009. 35