

UNIVERSIDADE FEDERAL DE SÃO CARLOS

CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

**MÉTODOS PARA ESTIMATIVA DE IMAGENS
NIR A PARTIR DE IMAGENS DE CÂMERAS
RGB**

DANIEL CAIO DE LIMA

ORIENTADOR: PROF. DR. JOSÉ HIROKI SAITO

São Carlos – SP

Abril/2020

UNIVERSIDADE FEDERAL DE SÃO CARLOS

CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

**MÉTODOS PARA ESTIMATIVA DE IMAGENS
NIR A PARTIR DE IMAGENS DE CÂMERAS
RGB**

DANIEL CAIO DE LIMA

Tese apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de São Carlos, como parte dos requisitos para a obtenção do título de Doutor em Ciência da Computação, área de concentração: Processamento de Imagens e Sinais

Orientador: Prof. Dr. José Hiroki Saito

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal Superior - Brasil (CAPES) - Código de Financiamento 001

São Carlos – SP

Abril/2020



UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Ciências Exatas e de Tecnologia
Programa de Pós-Graduação em Ciência da Computação

Folha de Aprovação

Assinaturas dos membros da comissão examinadora que avaliou e aprovou a Defesa de Tese de Doutorado do candidato Daniel Caio de Lima, realizada em 17/04/2020:

Prof. Dr. José Hiroki Saito
UFSCar

Prof. Dr. Alexandre Luis Magalhães Levada
UFSCar

Prof. Dr. Celso Aparecido de Franca
UFSCar

Prof. Dr. Daniel Carlos Guimarães Pedronette
UNESP

Prof. Dr. Marcelo Andrade da Costa Vieira
EESC/USP

Certifico que a defesa realizou-se com a participação à distância do(s) membro(s) Alexandre Luis Magalhães Levada, Celso Aparecido de Franca, Daniel Carlos Guimarães Pedronette, Marcelo Andrade da Costa Vieira e, depois das arguições e deliberações realizadas, o(s) participante(s) à distância está(ao) de acordo com o conteúdo do parecer da banca examinadora redigido neste relatório de defesa.

Prof. Dr. José Hiroki Saito

AGRADECIMENTOS

Agradeço a Deus, tanto por ter sido meu apoio, ter me permitido chegar até a conclusão desta tese e realizar grandes coisas na minha vida através de teu filho, Jesus Cristo, nosso redentor. Agradeço também a intercessão da admirável e venerável Maria, Nossa Senhora, exemplo de humildade e humanidade.

Não poderia me esquecer de agradecer também aos que me acompanharam de perto, dando apoio, tendo muita paciência em vários momentos. Pessoas excepcionais que desde meus primeiros anos, me ensinaram que tudo é possível com muito esforço, trabalho e acima de tudo, fé. Meus pais, Elio e Edna. Também agradeço à minha irmã, Suzane, minha primeira e melhor amiga, que sempre me ajudou e apoiou nas minhas dificuldades, sempre me fazendo sorrir quando estava cansado ou chateado. Agradeço a Ana Carolina, que durante o desenvolvimento desta tese, passou de uma grande amiga para minha namorada e que futuramente, será minha esposa. Obrigado pela visão otimista que sempre me passou, dizendo que tudo ia dar certo. Agradeço pelo carinho, amor, orações e dedicação destas pessoas especiais colocadas por Nosso Senhor em minha vida.

Agradeço ao professor Dr. José Hiroki Saito por ter aceitado a tarefa de me orientar durante estes anos. Sempre atencioso, paciente e solícito, é um grande exemplo de profissional, assim como um grande ser humano. Aos doutores Lúcio André de Castro Jorge e Ednaldo José Ferreira, da Embrapa Instrumentação, que me proporcionaram as primeiras experiências com a pesquisa, incentivando e apoiando para que seguisse este caminho. Foram fundamentais para o desenvolvimento deste trabalho. Também gostaria de agradecer aos meus amigos, Dr. Daniel Oliva Sales, Dr. Eduardo Sacogne Fraccaroli, Gabriel Alberto Mini, Lucas dos Santos Franco e Caio Bicudo Duran, grandes amigos e companheiros de trabalho.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de financiamento 001; parcialmente com o apoio do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), projeto número 310310/2013-0 e com o apoio da Embrapa Instrumentação.

RESUMO

A Agricultura de Precisão envolve o emprego de tecnologia para gestão e tomada de decisão para que produtores rurais consigam ter melhores resultados de produção. Para surtir o efeito desejado, é necessária uma grande coleta, processamento e análise de dados das áreas cultivadas, sendo o Sensoriamento Remoto, utilizando-se de sensores especiais acoplados a *drones*, uma das principais alternativas para esta tarefa. Os índices de vegetação, obtidos por equações matemáticas, são utilizados para realçar as variações da densidade da cobertura vegetal, utilizando-se de imagens no infravermelho próximo (*Near Infrared - NIR*) e imagens no espectro visível, sendo úteis para avaliar a produtividade e biomassa de uma plantação, por exemplo. No entanto, o custo elevado dos *drones* e, principalmente, dos sensores que captam imagens NIR, podem tornar inviável ou pouco atrativo o uso desta tecnologia por pequenos produtores. O objetivo deste trabalho é propor um método para estimar imagens NIR a partir de imagens RGB obtidas por câmeras fotográficas comuns, eliminando o uso de sensores específicos e a necessidade de fazer alterações no *hardware* das câmeras, tornando assim o custo desta tecnologia mais baixo. Para atingir este objetivo, foi proposto o uso de uma arquitetura de *Deep Learning (Pix2Pix)* e um método baseado na utilização de assinaturas espectrais de referência, classificação com KNN e uma soma ponderada pelo grau de proximidade das k assinaturas de referência mais próximas, produzindo uma nova assinatura espectral. Os métodos estão descritos e avaliados no presente trabalho. Os resultados mostraram que os dois métodos investigados possuem potencial para estimar imagens NIR, apresentando alto grau de similaridade em relação a imagens NIR reais.

Palavras-chave: Sensoriamento Remoto; Estimativa de imagem NIR; *Deep Learning*; *Pix2Pix*; KNN; Assinatura Espectral

ABSTRACT

Precision Agriculture involves the use of technology for management and decision making to allow rural producers to get better production results. To gain the desired effect, it is necessary a huge data collection and processing of the cultivated areas, being Remote Sensing, using special sensors attached to drones, a powerful alternative for this task. Vegetation Indices obtained with arithmetic equations, are used to highlight vegetation cover density variations using near infrared (NIR) images and visible spectrum images, being useful to assess biomass and productivity estimation from a crop, for example. The main problem is the elevated costs of drones and principally of sensors that acquire NIR images, making the use of this technology, by small producers, unattractive or difficult. The objective of this work is to propose a method to estimate NIR images using ordinary photographic camera RGB images that could eliminate the use of specific sensors and the need to make alterations in common cameras, thus making this technology cost lower. To achieve this goal, we propose the use of a Deep Learning architecture (Pix2Pix) and a spectral signature based, KNN classification and a weighted sum by proximity degree of k nearest reference signatures method, producing a new spectral signature. These methods are described and evaluated in this document. Results showed that the two methods investigated can be used to estimate NIR images, showing high similarity to real NIR images.

Keywords: Remote Sensing; NIR image estimative; Deep Learning; Pix2Pix; KNN; Spectral signature

LISTA DE FIGURAS

2.1	Processo de desenvolvimento de tecnologias para SR.	26
2.2	Propagação da radiação eletromagnética na forma de onda, em função das oscilações ortogonais dos campos magnético (M) e elétrico (E).	27
2.3	Comparação dos comprimentos de onda, frequência e energia do espectro eletromagnético.	28
2.4	Exemplos de imagens captadas por sensor multiespectral.	32
2.5	Exemplo de imagem hiperespectral, hipercubo.	32
2.6	Exemplo de assinaturas espectrais. Esquerda: <i>Pixels</i> da imagem selecionados. Direita: Assinatura espectral para cada <i>pixel</i> selecionado. Os <i>pixels</i> e suas respectivas assinaturas espectrais possuem a mesma cor.	33
2.7	Plotagem de um espectro de referência e um espectro de teste com duas bandas. Os mesmos materiais com variações na iluminação são representados pelos vetores conectados à origem e projetados aos pontos que representam o espectro real.	34
2.8	Representação do modelo de cor RGB em forma de cubo.	39
2.9	Representação cônica do modelo de cor HSV.	40
2.10	Representação do modelo de cor CIE Lab.	42
2.11	Ilustração de um modelo de <i>Deep Learning</i>	45
2.12	Exemplo de funcionamento da <i>Pix2Pix</i>	47
2.13	Resultados de tradução imagem para imagem com a <i>Pix2Pix</i>	49
2.14	Exemplos de imagens com correspondência entre <i>pixels</i> (esquerda) e de imagens sem correspondência entre <i>pixels</i> (direita).	49

2.15	Funcionamento da <i>cycleGAN</i> . (a) Mapeamento $G : X \rightarrow Y$ e $F : Y \rightarrow X$ e seus respectivos discriminadores D_X e D_Y . D_Y força G a gerar imagens que não podem ser distinguidas das imagens do domínio Y . D_X força F a gerar imagens que não podem ser diferenciadas das imagens pertencentes ao domínio X . (b) <i>Forward cycle-consistency loss</i> : $x \rightarrow G(x) \rightarrow F(G(x)) \approx x$. (c) <i>Backward cycle-consistency loss</i> : $y \rightarrow F(y) \rightarrow G(F(y)) \approx y$	50
2.16	Exemplo de imagem RGB de um campo de arroz captada por câmera acoplada a um drone.	53
2.17	Exemplo de imagem NIR de um campo de arroz captada por câmera NIR. . . .	53
3.1	Diagrama de Blocos dos métodos propostos nesta tese. A parte 1 se refere ao método que usa <i>Deep Learning</i> . A parte número 2 se refere ao método utilizando a estimativa a partir de <i>Endmembers</i> e classificação por KNN.	56
3.2	Câmera Micasense RedEdge TM utilizada na aquisição das imagens.	58
3.3	Exemplo de imagem capturada por cada sensor.	59
3.4	Tentativa de combinar imagens Vermelho, Verde e Azul em uma matriz tridimensional.	59
3.5	Painel de calibração radiométrica da câmera. As primeiras imagens capturadas pela câmera antes do voo devem ser deste painel sob a luz solar.	60
3.6	Ortomosaicos gerados a partir do conjunto de imagens multiespectrais fornecidos pela Embrapa Instrumentação.	60
3.7	Ilustração da ferramenta para coleta de amostras de treinamento e validação desenvolvida.	61
3.8	Exemplo de arquitetura <i>U-Net</i>	62
3.9	Arquitetura da rede G . Cada coluna representa um <i>feature map</i> , com as dimensões e o número de filtros descritos acima (linhas x colunas x filtros). . . .	64
3.10	Arquitetura da rede D (<i>PatchGAN</i>).	65
3.11	Atualização de pesos da rede D	65
3.12	Atualização de pesos da rede G	66
3.13	Diagrama de funcionamento da métrica SSIM.	68

3.14	Diagrama para o sistema SSIM multi escala. L: filtragem passa-baixa. \downarrow : escala diminuída por 2.	69
3.15	Imagem RGB da plantação de cítricos no Chile.	71
3.16	Indian Pines. (a) Imagem RGB e (b) <i>Ground truth</i>	72
3.17	Salinas. (a) Imagem RGB e (b) <i>Ground truth</i>	73
3.18	Pavia Centre. (a) Imagem RGB e (b) <i>Ground truth</i>	73
3.19	Ferramenta desenvolvida para auxiliar no desenvolvimento dos experimentos.	74
3.20	Espectros coletados para a classe Cítricos (Imagem do Chile) (esq.) e e respectivo Espectro Médio (dir.).	75
3.21	Resultado da classificação com a ferramenta Weka e a "Distribuição de probabilidade" para os k vizinhos mais próximos.	77
3.22	Cáculo de estimativa do NIR utilizando grau de proximidade e <i>Endmembers</i>	78
4.1	Exemplos de imagens NIR estimadas pelo método NIRDLearn a partir de imagens RGB como entrada.	81
4.2	Melhor resultado de estimativa NIR utilizando NIRDLearn.	82
4.3	Pior resultado de estimativa NIR utilizando NIRDLearn.	83
4.4	Melhor resultado de estimativa NIR utilizando NIRDLearn utilizando a função \mathcal{L}_{Mix}	84
4.5	Pior resultado de estimativa NIR utilizando NIRDLearn utilizando a função \mathcal{L}_{Mix}	84
4.6	Melhor resultado de estimativa NIR utilizando NIRDLearn utilizando a função \mathcal{L}_{PMix}	85
4.7	Pior resultado de estimativa NIR utilizando NIRDLearn utilizando a função \mathcal{L}_{PMix}	85
4.8	Exemplos de imagens NIR estimadas por NIRDLearn a partir de imagens RGB convertidas para o modelo CIE Lab.	87
4.9	Melhor resultado de estimativa NIR utilizando NIRDLearn, a função \mathcal{L}_{L1} e imagem CIE Lab.	87
4.10	Pior resultado de estimativa NIR utilizando NIRDLearn, a função \mathcal{L}_{L1} e imagem CIE Lab.	88

4.11	Melhor resultado de estimativa NIR utilizando NIRDLearn, a função \mathcal{L}_{Mix} e imagem CIE Lab.	89
4.12	Pior resultado de estimativa NIR utilizando NIRDLearn, a função \mathcal{L}_{Mix} e imagem CIE Lab.	89
4.13	Melhor resultado de estimativa NIR utilizando NIRDLearn, a função \mathcal{L}_{PMix} e imagem CIE Lab.	90
4.14	Pior resultado de estimativa NIR utilizando NIRDLearn, a função \mathcal{L}_{PMix} e imagem CIE Lab.	91
4.15	Exemplos de imagens NIR estimadas por NIRDLearn a partir de imagens RGB convertidas para o modelo HSV.	92
4.16	Melhor resultado de estimativa NIR utilizando NIRDLearn, a função \mathcal{L}_{L1} e imagem HSV.	93
4.17	Pior resultado de estimativa NIR utilizando NIRDLearn, a função \mathcal{L}_{L1} e imagem HSV.	93
4.18	Melhor resultado de estimativa NIR utilizando NIRDLearn, a função \mathcal{L}_{Mix} e imagem HSV.	94
4.19	Pior resultado de estimativa NIR utilizando NIRDLearn, a função \mathcal{L}_{Mix} e imagem HSV.	94
4.20	Melhor resultado de estimativa NIR utilizando NIRDLearn, a função \mathcal{L}_{PMix} e imagem HSV.	95
4.21	Pior resultado de estimativa NIR utilizando NIRDLearn, a função \mathcal{L}_{PMix} e imagem HSV.	96
4.22	Exemplos de resultados ruidosos obtidos pelas funções $\mathcal{L}_{MS-SSIM}$ (a) e $\mathcal{L}_{Perceptual}$ (b).	97
4.23	RRSE entre pixels do conjunto de validação Salinas RGB e seus respectivos espectros médios.	107
4.24	RRSE entre pixels do conjunto de validação Salinas HSV e seus respectivos espectros médios.	108
4.25	RRSE entre pixels do conjunto de validação Salinas CIE Lab e seus respectivos espectros médios.	108

4.26	RRSE entre pixels do conjunto de validação de cítricos do Chile RGB e seus respectivos espectros médios.	109
4.27	RRSE entre pixels do conjunto de validação de cítricos do Chile HSV e seus respectivos espectros médios.	109
4.28	RRSE entre pixels do conjunto de validação de cítricos do Chile CIE Lab e seus respectivos espectros médios.	110
4.29	RRSE entre os espectros gerados para o conjunto de validação Salinas RGB e seus respectivos espectros médios.	110
4.30	RRSE entre os espectros gerados para o conjunto de validação Salinas HSV e seus respectivos espectros médios.	111
4.31	RRSE entre os espectros gerados para o conjunto de validação Salinas CIE Lab e seus respectivos espectros médios.	111
4.32	RRSE entre os espectros gerados para o conjunto de validação de cítricos do Chile RGB e seus respectivos espectros médios.	112
4.33	RRSE entre os espectros gerados para o conjunto de validação de cítricos do Chile HSV e seus respectivos espectros médios.	112
4.34	RRSE entre os espectros gerados para o conjunto de validação de cítricos do Chile CIE Lab e seus respectivos espectros médios.	113
4.35	Comparação entre imagens NIR originais e os resultados da estimativa de espectro para os modelos RGB, HSV e CIE Lab.	114
4.36	Resultados de estimativa de imagens NIR para os dois métodos propostos. . . .	116

LISTA DE TABELAS

3.1	Bandas espectrais capturadas pela câmera RedEdge™	58
4.1	Similaridade média entre <i>ground truth</i> e imagens estimadas para o conjunto de treinamento (Imagens RGB).	86
4.2	Similaridade média entre <i>ground truth</i> e imagens estimadas para o conjunto de validação (Imagens RGB).	86
4.3	Similaridade média entre <i>ground truth</i> e imagens estimadas para o conjunto de treinamento (Imagens CIE Lab).	90
4.4	Similaridade média entre <i>ground truth</i> e imagens estimadas para o conjunto de validação (Imagens CIE Lab).	91
4.5	Similaridade média entre <i>ground truth</i> e imagens estimadas para o conjunto de treinamento (Imagens HSV).	95
4.6	Similaridade média entre <i>ground truth</i> e imagens estimadas para o conjunto de validação (Imagens HSV).	96
4.7	Resumo dos resultados de similaridade com imagens estimadas a partir do conjunto de treinamento para os três modelos de cor	97
4.8	Resumo dos resultados de similaridade com imagens estimadas a partir do conjunto de validação para os três modelos de cor	97
4.9	Informações do <i>ground truth</i> do <i>data set</i> Indian Pines	99
4.10	Acurácia dos classificadores para o <i>data set</i> Indian Pines por algoritmo de seleção de <i>endmembers</i> (%)	100
4.11	Informações do <i>ground truth</i> do <i>data set</i> Salinas	101
4.12	Acurácia dos classificadores para o <i>data set</i> Salinas por algoritmo de seleção de <i>endmembers</i> (%)	102

4.13	Informações do <i>ground truth</i> do <i>data set</i> Pavia	103
4.14	Acurácia dos classificadores para o <i>data set</i> Pavia por algoritmo de seleção de <i>endmembers</i> (%)	104
4.15	Relação de classes e número de amostras para a imagem da plantação de cítricos do Chile.	105
4.16	Resultados do classificador KNN de acordo com o modelo de cores e os <i>data sets</i> .106	
4.17	RRSE entre espectros originais e médios X RRSE entre espectros estimados e médios	113
4.18	Resultados de similaridade entre imagens NIR reais e imagens estimadas com o método NIRKNN.	113
4.19	Média da similaridade entre imagens NIR reais e estimadas para os dois métodos.115	

GLOSSÁRIO

ATGP – *Automatic Target Generation Process*

CCA – *Canonical Correlation Analysis*

CFA – *Color Filter Array*

CIE Lab – *Modelo de cor CIE Lab*

CNN – *Convolutional Neural Network*

FIPPI – *Fast Iterative Pixel Purity Index*

GAN – *Generative Adversarial Networks*

GPS – *Global Positioning System*

GP – *Gaussian Process*

HSV – *Modelo de cor HSV*

KNN – *K Nearest Neighbor*

MLP – *Multilayer Perceptron*

MPRI – *Modified Photochemical Reflectance Index*

MS-SSIM – *Multi Scale Structural Similarity Measure*

NDVI – *Normalized Difference Vegetation Index*

NIRDLearn – *Estimador de imagens NIR baseado em Deep Learning*

NIRKNN – *Estimador de imagens NIR baseado em KNN*

NIR – *Near Infrared*

PPI – *Pixel Purity Index*

RBF – *Radial Basis Function*

RGB – *Modelo de cor RGB*

RMSE – *Root Mean Square Error*

SAM – *Spectral Angle Mapper*

SID – *Spectral Information Divergence*

SSIM – *Structural Similarity Index*

SVM – *Support Vector Machines*

VANT – *Veículo Aéreo Não Tripulado, drones*

VGGNet – *Visual Geometry Group Network*

cGAN – *Conditional Generative Adversarial Networks*

SUMÁRIO

GLOSSÁRIO

CAPÍTULO 1 – INTRODUÇÃO	18
1.1 Sensoriamento Remoto e Agricultura de Precisão	18
1.1.1 Drones na Agricultura de Precisão	21
1.2 Motivação e Justificativa	22
1.3 Objetivos	23
1.4 Objetivos Específicos	23
1.5 Principais Contribuições	24
1.6 Estrutura da tese	24
CAPÍTULO 2 – REFERENCIAL TEÓRICO E PESQUISAS BIBLIOGRÁFICAS	25
2.1 Sensoriamento Remoto	25
2.1.1 Espectro Eletromagnético	26
2.1.2 Sensoriamento Remoto e Processamento de Imagens	28
2.2 Imagens Multiespectrais e Hiperespectrais	31
2.2.1 Similaridade Espectral	33
2.2.1.1 Spectral Angle Mapper (SAM)	34
2.2.1.2 Spectral Information Divergence (SID)	35
2.2.2 Índices de Vegetação	36
2.2.2.1 <i>Normalized Difference Vegetation Index</i> - NDVI	36

2.2.2.2	<i>Modified Photochemical Reflectance Index - MPRI</i>	37
2.3	Modelos de cor	38
2.3.1	Modelo RGB	38
2.3.2	Modelo HSV	39
2.3.3	Modelo CIE Lab	41
2.4	K-Nearest-Neighbour (KNN)	43
2.5	Deep Learning	43
2.5.1	<i>Pix2Pix</i>	46
2.5.2	<i>cycleGAN</i>	48
2.6	Pesquisas Bibliográficas sobre Trabalhos relacionados - Estado da Arte	51
2.7	Considerações Finais	54
CAPÍTULO 3 – MATERIAIS E MÉTODOS		55
3.1	Descrição dos métodos propostos para a tese	55
3.2	NIRDLearn - Estimativa de imagens NIR com <i>rede Deep Learning Pix2Pix</i>	57
3.2.1	Imagens Multiespectrais utilizadas	57
3.2.2	Arquitetura da rede <i>Pix2Pix</i>	61
3.2.2.1	Gerador	61
3.2.2.2	Discriminador	64
3.2.3	Funções de custo	67
3.2.3.1	SSIM	67
3.2.3.2	MS-SSIM	69
3.2.3.3	MS-SSIM + L1 (<i>Mixed</i>)	69
3.2.3.4	Perda Perceptiva (<i>Perceptual Loss</i>)	70
3.2.3.5	<i>Perceptual</i> + L1 (P_{Mixed})	70
3.3	NIRKNN - Estimativa de imagens NIR utilizando imagens RGB e classificador KNN	71

3.3.1	Imagens Hiperespectrais utilizadas	71
3.3.2	Definição de classes referentes a cada <i>ground truth</i> e obtenção dos <i>end-members</i>	74
3.3.3	Cálculo para estimar o NIR	75
3.4	Considerações Finais	79
CAPÍTULO 4 – RESULTADOS E DISCUSSÃO		80
4.1	Estimativa de imagens NIR com o método NIRDLearn	80
4.1.1	Resultados com modelo RGB	81
4.1.2	Resultados com modelo CIE Lab	86
4.1.3	Resultados com modelo HSV	91
4.1.4	Treinamentos com função $\mathcal{L}_{MS-SSIM}$ e $\mathcal{L}_{Perceptual}$	96
4.1.5	Considerações finais sobre as redes NIRDLearn treinadas	96
4.2	Estimativa de imagens NIR com o método NIRKNN	98
4.2.1	Seleção de espectros	98
4.2.1.1	Indian Pines	99
4.2.1.2	Salinas	100
4.2.1.3	Pavia	103
4.2.1.4	Chile	104
4.2.2	Estimativa do NIR	105
4.3	Considerações finais sobre estimativa de imagens NIR utilizando NIRKNN	112
4.4	Comparação entre os dois métodos - NIRDLearn e NIRKNN	114
4.5	Considerações Finais	115
CAPÍTULO 5 – CONCLUSÕES		117
5.1	Trabalhos publicados ou submetidos	118
REFERÊNCIAS		120

Capítulo 1

INTRODUÇÃO

Este capítulo refere-se à introdução ao trabalho desenvolvido de tese e descreve o assunto abordado sobre o tema do projeto de pesquisa proposto. Além disso, apresenta as motivações e justificativas, objetivos, principais contribuições e a estrutura do trabalho.

1.1 Sensoriamento Remoto e Agricultura de Precisão

Existe uma grande preocupação com a produtividade agrícola e seus impactos na sociedade, principalmente em relação à sustentabilidade e os efeitos que o processo agrícola causa no meio ambiente. Uma das formas de amenizar essas preocupações e os impactos causados é a inserção da tecnologia, cada vez mais crescente com o projeto de equipamentos específicos e até mesmo a utilização de ferramentas de *software*, sendo esta inserção de tecnologia no ramo da agricultura chamada de agricultura de precisão. Uma definição de agricultura de precisão, dada pelo Ministério da Agricultura, Pecuária e Abastecimento (BRASIL, 2012), é uma forma de gerenciar empreendimentos agrícolas com base na variação de tempo e espaço de uma unidade produtiva com o objetivo de aumentar o retorno econômico desta unidade, assim como aumentar a sustentabilidade e minimizar os efeitos nocivos ao ambiente.

A agricultura de precisão é uma das dez maiores revoluções no ramo de agricultura (CROKSTON, 2006), envolvendo melhorias nas práticas de gestão de insumos agrícolas, como fertilizantes e sementes, alocando-os a locais corretos e no período correto, evitando prejuízos ao produtor rural (MULLA, 2013). Um exemplo de prática na agricultura de precisão é a divisão de uma extensa área cultivada em zonas de gerenciamento, nas quais esses insumos são aplicados baseados no tipo de solo, histórico de incidência de pragas e problemas relacionados à irrigação, sendo uma forma mais eficiente do que o método tradicional, no qual a tomada de providência usando os recursos é feita de maneira uniforme sobre toda a área cultivada,

evidenciando o desperdício de material. Segundo Mulla (2013), em 2012 um terço dos agricultores do meio oeste dos Estados Unidos já utilizava práticas de agricultura de precisão em suas propriedades.

Para que as práticas de agricultura de precisão tenham o efeito esperado, é necessária uma grande quantidade de coleta e análise de dados, além do gerenciamento e processamento de informações. Devido a esta necessidade de um grande volume de dados sobre as áreas cultivadas, a utilização da tecnologia para a aquisição dos dados é de primordial importância. As formas mais utilizadas para a obtenção de dados de grandes extensões são sensores acoplados a satélites, aviões, Veículos Aéreos Não Tripulados (VANTs) e veículos terrestres combinados com sistemas de posicionamento global (*Global Positioning System* - GPS), além de sensores portáteis. Estes sensores coletam informações sobre a densidade de ervas daninhas e refletância da área das folhas, status de umidade, dentre outras informações que são necessárias para um bom gerenciamento das unidades produtivas e também a uma tomada de decisão correta para o controle das pragas e deficiências existentes nas plantações. A este processo de aquisição de informações, chamamos de sensoriamento remoto.

A característica do sensoriamento remoto é a obtenção de informações de um objeto sem existir contato físico com o mesmo, e devido à possibilidade do processo ser realizado a longas distâncias, tornou-se uma abordagem cada vez mais frequente e de menor custo. Sendo assim, o sensoriamento remoto é uma excelente ferramenta para agricultura (SHIRATSUCHI et al., 2014). Suas aplicações na agricultura são baseadas na interação da radiação eletromagnética com o solo ou plantas. Em outras palavras, o sensoriamento remoto envolve medidas da radiação absorvida, transmitida ou refletida pelos objetos (MULLA, 2013), sendo que as folhas das plantas podem emitir energia por fluorescência (APOSTOL et al., 2003) ou emissão térmica (COHEN et al., 2005). Um exemplo de sensoriamento remoto por emissão térmica é a verificação do estresse hídrico em plantações, na qual a emissão de radiação em resposta a temperatura das folhas do dossel é utilizada para fazer o diagnóstico. A temperatura das folhas varia de acordo com a temperatura do ar e com a taxa de evapotranspiração.

Conforme descrito por Mulla (2013), a quantidade de radiação refletida pelas plantas é relacionada inversamente à quantidade de radiação absorvida pelos pigmentos das plantas, como a clorofila, e variam conforme o comprimento de onda da radiação incidente. Um exemplo disso é o da clorofila, que tem uma forte absorção da radiação na faixa visível do espectro eletromagnético (400 nm a 700 nm) (PINTER et al., 2003). De modo mais específico, a clorofila *a* absorve radiação nos comprimentos de onda 430 nm (correspondente ao azul) e 660 nm (correspondente ao vermelho), enquanto a clorofila *b* absorve radiação nos comprimentos 450 nm

(azul) e 650 nm (vermelho). Informações mais específicas sobre outros pigmentos das plantas, como os carotenoides, são descritos por Blackburn (2006).

Enquanto a absorção de radiação nas plantas ocorre para os comprimentos de onda na região visível do espectro eletromagnético, a refletância se dá nos comprimentos de onda que estão na região do infravermelho próximo (*Near Infrared* - NIR), faixa do espectro correspondente aos comprimentos de onda que vão dos 760nm a 1200nm, sendo que não há um consenso definido para o início e término do intervalo compreendido pelo NIR. A alta refletância nessa faixa do espectro existe devido à densidade das folhas das plantas e dos efeitos causados pelas estruturas dos dosséis (MULLA, 2013). No que se refere a energia refletida pelo solo, esta varia de acordo com a composição do mesmo, como a umidade do solo, quantidade de matéria orgânica, minerais dentre outros.

Como citado anteriormente, a aquisição de dados em Sensoriamento Remoto se dá pelo uso de sensores que captam as interações entre a radiação eletromagnética e diversos objetos alvo. As primeiras formas do uso de Sensoriamento Remoto na agricultura foram através de imagens de satélite e datam do início da década de 1970, quando o satélite Landsat1 foi lançado (1972), com sensores multiespectrais que capturavam imagens de quatro bandas (verde, vermelho e duas bandas infravermelho) com baixa resolução espacial. As imagens demoravam cerca de 18 dias para ficarem prontas para utilização. As imagens do Landsat1 foram utilizadas para classificação de panoramas agrícolas do meio oeste dos EUA em campos de plantio de soja e campos de plantio de milho (BAUER; CIPRA, 1973). As imagens geradas por satélite foram utilizadas para Agricultura de Precisão no início da década de 1990, com o uso das imagens de solo do Landsat para estimar padrões espaciais no conteúdo da matéria orgânica do solo (BHATTI; MULLA; FRAZIER, 1991), informação usada posteriormente como complemento para estimativas de padrões espaciais de quantidade de fósforo do solo e produção de grãos de trigo, feitas com medidas terrestres.

Com o passar dos anos, os satélites e sensores se modernizaram, melhorando assim a qualidade da captura, resolução espacial e o tempo de envio das imagens para a terra, porém, ainda existem limitações, pois a captura é afetada pela presença de nuvens e ainda existem fatores relacionados à calibração dos sensores.

Nos últimos anos, uma tecnologia que existia apenas para finalidades militares se popularizou e tornou-se acessível, sendo utilizada em diversos segmentos da cadeia de produção e até mesmo para fins pessoais. Os VANTs, popularmente chamados de drones, surgiram na década de 1960 para a realização do monitoramento de possíveis ameaças e evitar pôr a vida de soldados da marinha dos Estados Unidos em risco, porém, o governo estadunidense só admitiu o uso

de VANTs nos campos de batalha a partir de 1973 (PIXFORCE, 2016). Os drones são equipados com câmeras (sensores) e GPS que possibilitam a aquisição de imagens com alta resolução demarcadas com a coordenada geográfica real do momento em que sobrevoa uma determinada região.

Comparadas às imagens obtidas por satélite, as imagens capturadas por drones se destacam por obter alta resolução espacial. Cada pixel de uma imagem de satélite tem precisão de metros em sua resolução, enquanto as imagens de VANTs possuem precisão em centímetros. Uma maior resolução espacial fornece maior precisão na detecção de doenças em plantações, falhas de plantio e informações sobre desenvolvimento de plantas. Outra vantagem da utilização de drones em processos de agricultura de precisão é a periodicidade com que as imagens são obtidas. Diferente das imagens de satélite, que dependem do período de órbita do mesmo para obtenção das imagens de uma mesma localidade (cerca de 5 dias) e são afetadas pela cobertura de nuvens, drones podem obter imagens em um período de tempo mais específico, como por exemplo, antes da aplicação de determinados insumos agrícolas, causando impacto positivo na produtividade e na parte econômica (RIBEIRO, 2017).

1.1.1 Drones na Agricultura de Precisão

Segundo Jorge e Inamasu (2014), o desenvolvimento de VANTs surgiu como uma importante opção na agricultura de precisão, sendo favorecida pelo desenvolvimento tecnológico, redução do custo e do tamanho dos equipamentos necessários para o funcionamento da aeronave não tripulada.

No Brasil, as pesquisas envolvendo o uso de drones são muito recentes, portanto, estão concentradas em aplicações de vigilância policial de áreas urbanas e fronteiras, inspeção de linhas de transmissão de energia, controle de queimadas, pragas e monitoramento de práticas agrícolas (MEDEIROS, 2007).

A Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA), precisamente a EMBRAPA Instrumentação, localizada em São Carlos, é considerada pioneira nas pesquisas envolvendo o uso de VANTs na agricultura, iniciando-as em 1998 com o objetivo de substituir o uso de aviões convencionais para a realização do monitoramento de áreas agrícolas e áreas de preservação ambiental e, posteriormente, investindo no desenvolvimento de novas plataformas e aeronaves específicas para este uso (SILVA, 2015). Segundo Jorge e Inamasu (2014), um dos investimentos da EMBRAPA, em parceria com uma empresa do setor privado, foi o desenvolvimento de uma aeronave com a capacidade de voar em condições adversas relacionadas a áreas agrícolas.

Como citado anteriormente, a análise espectral da vegetação fornece um conjunto de padrões úteis de suas informações fisiológicas para identificar o seu estado de saúde por meio de Índices de Vegetação, principalmente utilizando o comprimento de onda na região do NIR, portanto, os drones utilizados para agricultura precisam de sensores baseados na realização de medidas de reflexão da radiação eletromagnética. Jorge e Inamasu (2014) fazem uma descrição dos diferentes tipos de sensores utilizados nos VANTs para fins agrícolas. O principal problema relacionado aos sensores é o custo elevado, não sendo acessíveis e interessantes para pequenos produtores rurais.

Além do monitoramento da saúde de culturas agrícolas, detecção de pragas e estimativa de produtividade, *drones* também podem ser usados para aplicação de herbicidas e pesticidas, detecção de animais doentes em atividades de pecuária, avaliar o impacto causado pela rotação de culturas, verificar invasão de ervas daninhas nas plantações, deficiência de nitrogênio em culturas, identificação de focos de incêndio, demarcação de área de plantio e monitoramento do sistema de irrigação (BEDORD, 2015) (PIXFORCE, 2016).

1.2 Motivação e Justificativa

Com a crescente popularização do uso de *drones* para os diversos fins e o grande interesse de mostrar os impactos positivos que a inserção desta tecnologia pode proporcionar aos produtores rurais, principalmente no gerenciamento da propriedade e na aplicação de insumos, surgiram iniciativas com o objetivo de mostrar os benefícios da aquisição das imagens das propriedades e do desenvolvimento de ferramentas de *software* que possam processar essas imagens e obter informações úteis para tomada de decisões pelos produtores rurais.

Em 2016, A Embrapa e a Qualcomm, empresa líder mundial em tecnologia móvel, fizeram uma parceria com o intuito de potencializar o uso de VANTs na agricultura e tornar essa tecnologia mais acessível para pequenos produtores. Por meio dessa parceria deverá ser construído um sistema no qual um *drone* deverá transmitir informações em pleno vôo sobre a lavoura em tempo real. Normalmente, essas informações só chegariam ao agricultor após alguns dias que o vôo com o VANT for realizado, pois as imagens capturadas teriam que passar por diversos *softwares* de processamento e posteriormente, os resultados submetidos a avaliação de um especialista (PAPP, 2016). Esse processo convencional de obtenção das informações processadas é uma das principais causas de preços elevados, restringindo o uso dessa tecnologia aos grandes produtores rurais. Segundo Papp (2016), o valor de um *drone* pode variar entre 10 mil e 300 mil reais. Além disso, os *softwares* necessários para o processamento tem alto custo e não seriam

compatíveis com o poder de processamento de computadores pessoais comuns. O valor dos VANTS descritos incluem os sensores específicos para o uso na agricultura. Sensores multiespectrais podem custar dezenas de milhares de dólares. Já os valores dos sensores hiperespectrais ultrapassam o valor de 100 mil dólares.

Outra parceria fechada foi a entre DroneDeploy e as fabricantes de sensores, Sentera e Slantrange, para o desenvolvimento de algoritmos de processamento específicos para cada sensor, um para imagens NIR e outro multiespectral (BEDORD, 2017), porém, nada que possa tornar o uso da tecnologia mais popular e atraente para pequenos produtores.

Sensores RGB que captam a reflectância na região visível do espectro eletromagnético (Azul - B, Verde - G e Vermelho - R) são muito mais acessíveis, pois câmeras fotográficas comuns podem ser utilizadas (ZHENG et al., 2018), e muitas vezes são itens que já estão acoplados ao VANT no momento da aquisição, portanto, seria interessante pesquisar sobre um método para fazer estimativas de imagens na região do NIR a partir da imagens desses sensores mais simples, tornando o valor final de um *drone* mais atraente e produzindo resultados que possam ser comparáveis aos índices de vegetação calculados com os dados obtidos por sensores mais complexos.

1.3 Objetivos

O objetivo deste trabalho é estabelecer métodos para estimar imagens na região do NIR a partir de imagens RGB obtidas por câmeras comuns acopladas aos drones, com a precisão necessária para que índices de vegetação possam ser calculados sem a necessidade da compra de sensores específicos, tornando o uso da tecnologia mais acessível para pequenos produtores rurais.

1.4 Objetivos Específicos

- Desenvolver métodos computacionais para estimar imagens NIR a partir de imagens RGB.
- Avaliar o uso de imagens multiespectrais para realização de treinamento de métodos para estimativa de imagens NIR.
- Avaliar o uso de espectros de referência (*Endmembers*) na estimativa de imagens NIR.
- Utilizar métodos de *Machine Learning* clássicos e recentes, como *Deep Learning*.

1.5 Principais Contribuições

As principais contribuições do trabalho de tese são o desenvolvimento de métodos para obtenção de imagens NIR sem a necessidade de sensores específicos ou alterações no *hardware* de câmeras comuns, estimando imagens NIR que sejam mais próximas do real; e possibilitar a diminuição do custo global do sistema de *drones* para que possam ser aplicados em processos de agricultura de precisão, assim como colaborar com cientistas que necessitem de estudos com imagens NIR, porém, não possuem acesso a sensores projetados para este fim.

1.6 Estrutura da tese

A estrutura da tese é composta, além do presente capítulo de introdução, da seguinte forma. No Capítulo 2 será apresentada toda a base teórica utilizada para fundamentar a pesquisa e os experimentos conduzidos para o presente trabalho, assim como as pesquisas bibliográficas do tema proposto. O Capítulo 3 apresenta a proposta da tese de doutorado, mostrando todas as etapas e ferramentas usadas para conduzir os experimentos em detalhes e também os cálculos utilizados para a criação dos novos espectros que darão origem à estimativa de imagem NIR. No Capítulo 4 serão exibidos os resultados dos experimentos realizados para estimativa de imagens NIR, assim como as considerações finais sobre os experimentos. Finalizando, o Capítulo 5 apresenta as conclusões deste trabalho de pesquisa.

Capítulo 2

REFERENCIAL TEÓRICO E PESQUISAS BIBLIOGRÁFICAS

Neste capítulo serão abordados os principais aspectos teóricos referentes ao desenvolvimento da tese.

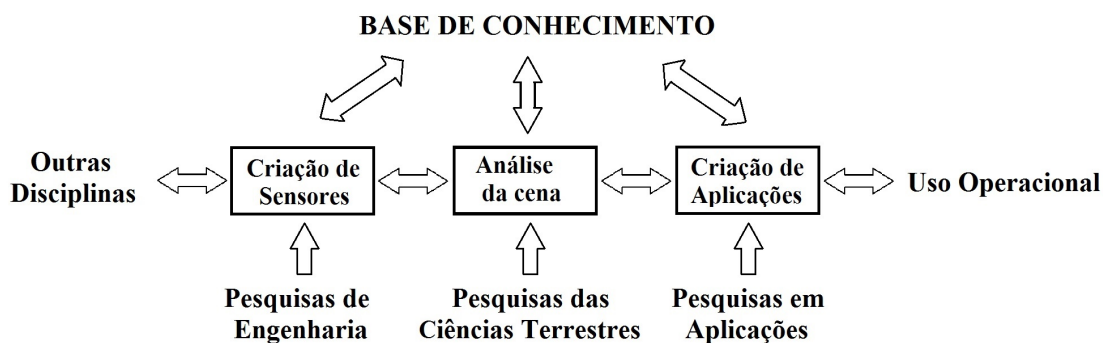
2.1 Sensoriamento Remoto

O termo Sensoriamento Remoto (SR) foi designado a partir do desenvolvimento de equipamentos especiais utilizados para captar a radiação eletromagnética refletida por objetos na superfície terrestre sem que haja contato, ou seja, para capturar a informação remotamente através de sensores. Utilizando termos mais científicos, Meneses e Almeida (2012) descrevem SR como sendo uma ciência que visa o desenvolvimento da obtenção de imagens da superfície da Terra através da detecção e medição quantitativa das respostas das interações da radiação eletromagnética com os materiais terrestres. As respostas das interações da radiação são medidas através de variações espectrais, espaciais e temporais na energia eletromagnética e gravitacional dos alvos, sendo que as medidas das variações destas energias devem ser relacionadas às informações que desejam se obter (LANDGREBE, 2005).

A definição de SR permite afirmar que os objetos imageados são registrados pelos sensores por meio de medições da radiação eletromagnética, como a luz solar incidente sobre qualquer objeto na superfície terrestre, porém, há uma grande confusão ao elevarem sensores geofísicos aerotransportados ao mesmo patamar de sensores remotos. Sensores geofísicos, como os magnetômetros, geram imagens através de medidas dos campos de força do campo magnético da superfície. Outra confusão comum é considerar como SR, imagens obtidas apenas por satélites. Fotografias aéreas eram utilizadas como formas de se observar a Terra muito tempo antes dos

primeiros satélites serem construídos. O filme fotográfico foi o primeiro material inventado pelo homem capaz de capturar a energia eletromagnética e convertê-la em uma imagem (MENESES; ALMEIDA, 2012). Outra parte importante do SR são os sistemas de processamento, que buscam dar sentido às informações obtidas utilizando o conhecimento de analistas como um complemento aos processamentos computacionais (LANDGREBE, 2005). A criação de todo o processo de SR é uma atividade multidisciplinar e está em constante desenvolvimento. São necessárias pesquisas na área das engenharias e computação e, principalmente, pesquisas relacionadas ao estudo da superfície terrestre, primordiais para a interpretação e prover sentido às informações captadas pelos sensores. A Figura 2.1 mostra um esquema para o processo de desenvolvimento de tecnologia em SR.

Figura 2.1: Processo de desenvolvimento de tecnologias para SR.



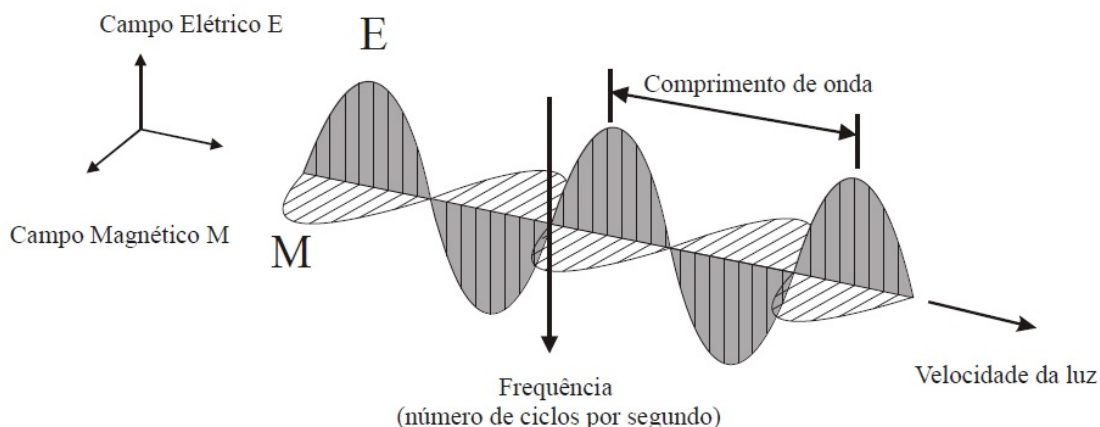
Fonte: Adaptado de Landgrebe (2005).

2.1.1 Espectro Eletromagnético

A radiação eletromagnética possui como característica a dualidade de comportamento na natureza, pois se propaga pelo espaço vazio em forma de onda e energia, assim como a luz solar, sendo assim formulada pelos modelos ondulatório e corpuscular. A coexistência da radiação na forma de onda e energia é fundamental para que seja possível explicar nas imagens tudo o que for referente à característica dos objetos (MENESES; ALMEIDA, 2012). Uma partícula carregada eletricamente gera um campo elétrico em torno de si e a partir do movimento desta, também se forma um campo magnético. Os campos elétrico e magnético da partícula possuem as mesmas amplitudes e vibram ortogonalmente entre si. A partir do momento que a partícula é acelerada, ocorre uma perturbação entre os dois campos que se propaga pelo vácuo em uma direção ortogonal aos campos. Essa perturbação entre os campos elétrico e magnético é chamada de onda eletromagnética. Sendo assim, uma onda eletromagnética é a oscilação dos campos elétrico e magnético, segundo um padrão harmônico (ondas espaçadas repetitivamente

no tempo). A Figura 2.2 é uma ilustração de como se propaga uma onda eletromagnética, onde na direção vertical observa-se a oscilação do campo elétrico E e na direção horizontal, do campo magnético M. A onda se propaga na direção das duas oscilações na velocidade da luz.

Figura 2.2: Propagação da radiação eletromagnética na forma de onda, em função das oscilações ortogonais dos campos magnético (M) e elétrico (E).



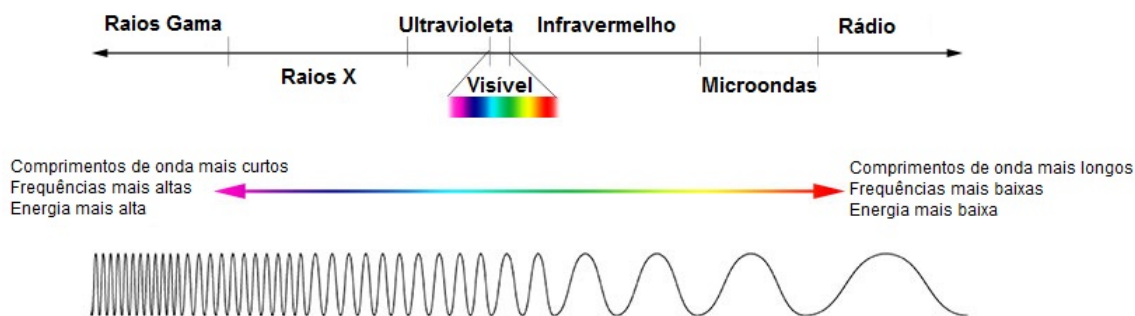
Fonte: Meneses e Almeida (2012).

Em SR costuma-se referir a radiação eletromagnética pelo seu comprimento de onda, não pela frequência. As características das imagens são explicadas levando-se em conta o tamanho das ondas e o tamanho dos objetos presentes nas cenas. Essa interação entre a onda com os objetos é chamada de interação macroscópica, que nada mais é do que a resposta da intensidade com que um objeto reflete a radiação eletromagnética em função do tamanho da onda e da textura presente na superfície do objeto. Segundo Meneses e Almeida (2012), a textura representa a rugosidade topográfica da superfície e se mostra muito eficiente para observar diferenças vistas nas imagens entre diversas formas de relevo, vegetação ou porções de água.

O espectro eletromagnético é constituído por diversas ondas eletromagnéticas, em amplas faixas de comprimentos de onda, ou variedade nas frequências de oscilação. O espectro foi dividido de maneira arbitrária em intervalos de comprimentos de onda utilizando como base os meios pelos quais essa energia eletromagnética é gerada e detectada, dando nome a cada um desses intervalos de acordo com o uso que os seres humanos encontraram para cada um deles (MENESES; ALMEIDA, 2012). A Figura 2.3 mostra a divisão e o nome de cada um dos intervalos do espectro eletromagnético, sendo possível obter informações sobre como cada uma dessas faixas do espectro podem ser aplicadas no nosso dia a dia em NASA (2013).

Em sensores remotos, as imagens não são definidas por apenas um comprimento de onda específico, mas sim em pequenos intervalos, chamados de bandas espectrais, sendo que os sensores remotos instalados em aeronaves e satélites capturam imagens apenas nas seguintes

Figura 2.3: Comparação dos comprimentos de onda, frequência e energia do espectro eletromagnético.



Fonte: Adaptado de NASA (2013).

bandas: visível, infravermelho próximo, infravermelho de ondas curtas, infravermelho de ondas médias, infravermelho termal e micro-ondas. Em grande parte das aplicações de SR na agricultura, são utilizadas bandas na região do visível e na região do infravermelho próximo.

O intervalo visível (450 - 760 nm) é assim chamado pelo fato do olho humano ser sensível a essa região do espectro, sendo muito usada em SR por conta da alta intensidade do fluxo de radiação e responsável pela interação com os minerais e pigmentos da vegetação, originando as cores. A região do infravermelho próximo, ou *Near Infrared* (NIR) (760 - 1200 nm) é o intervalo em que ocorrem interações importantes do espectro com os níveis de energia dos átomos, sendo um dos meios mais poderosos para se classificar uma vegetação saudável, uma vez que a radiação é refletida por plantas saudáveis e absorvida pela água.

2.1.2 Sensoriamento Remoto e Processamento de Imagens

Bernardi et al. (2014) descreveram aplicações de imagens multiespectrais e hiperespectrais, em conjunto com sensoriamento remoto na agricultura, realizando um levantamento de trabalhos existentes na literatura. As aplicações descritas têm finalidades para propósitos distintos, como a identificação de características para análises bioquímicas como clorofilas *a* e *b*, concentração de nitrogênio, carbono, celulose, lignina, antocianina, amido, água e sideronol - A; detecções de diferenças na refletância relacionadas à densidade da cobertura vegetal; detecções de medidas foliares e cobertura vegetal; fração da radiação fotossinteticamente ativa; medições relacionadas às características do dossel vegetal; entre outras aplicações que têm sido utilizadas como forma de apoio a operações de mapeamento de estresse vegetal, aplicação de fertilizantes e pesticidas, manejo de irrigação e monitoramento do conteúdo de nutrientes para grãos e diversas frutas.

Lawrance e Labus (2003) mostram métodos que apresentaram bons resultados em dados de imagens hiperespectrais. O objetivo do teste foi a detecção de doenças causadas pela ação de uma espécie de besouro em plantas. Para classificação foram utilizadas técnicas de análise discriminante e árvores de classificação e regressão. A aquisição das imagens foi realizada por um sensor hiperespectral acoplado a um helicóptero.

Para auxiliar na classificação de culturas vegetais, Souza (2013) utilizou informações extraídas de imagens hiperespectrais, informações de refletância bidirecional e o algoritmo de máxima verossimilhança (com medidas do campo) com o objetivo de analisar a possibilidade de melhorar a precisão de classificação de vegetação, principalmente em relação ao crescimento desses tipos de plantas. A classificação foi realizada pelo algoritmo de máxima verossimilhança, sendo as classes as culturas de beterraba, milho, trigo e grama. Os resultados apresentados mostraram que as imagens hiperespectrais são superiores às imagens multiespectrais para a resolução do problema de classificação apresentado, porém, em outras culturas, como cebola e batata, ambos os tipos de imagem não apresentaram bons resultados, sendo sugerida a exploração de outros meios de classificação para uma melhor discriminação.

A Embrapa Algodão utiliza imagens hiperespectrais. Dados de sinais infravermelhos são captados e são utilizados em processos de química analítica. A tecnologia empregada apoia análises químicas de sementes e a qualidade de fibras de algodão. Os dados são úteis para auxiliar na melhora genética da cultura de algodão. O uso de câmeras hiperespectrais permite a automatização de várias etapas de análises, contribuindo para assim agilizar e otimizar processos (SANTOS, 2014).

Em Zhang et al. (2012), imagens hiperespectrais são utilizadas para identificação de diferentes sementes de milho, utilizando características espectrais e de textura. Nesse trabalho diversas técnicas de processamento de imagens e reconhecimento de padrões foram utilizadas, dentre as quais é importante destacar a *Principal Component Analysis* (PCA), *Kernel Principal Component Analysis* (KPCA), o *Gray-Level Co-Occurrence Matrix* (GLCM), *Back Propagation Neural Network* (BPNN) e o *Least Squares-Support Vector Machine* (LS-SVM). O PCA e KPCA são utilizados para explorar as estruturas internas de dados espectrais. Já o GLCM é utilizado para extrair variáveis textuais, como contraste, homogeneidade, energia e correlação com base nos comprimentos de onda ideais. O LS-SVM e o BPNN são utilizados para estabelecer classes dos diferentes tipos de milho que serão identificadas e classificadas. Essa combinação de tecnologias possibilitou a classificação de forma rápida e precisa das variedades de sementes de milho.

Yang e Chang (2010), utilizaram o problema de infestação de bactérias em plantações de

arroz como uma oportunidade para desenvolverem uma técnica para análise das propriedades ópticas obtidas por medições de refletância espectral, sendo esta análise realizada através de um método de regressão linear, estabelecendo assim modelos espectrais. A partir dessa análise, foi possível estabelecer um grau de gravidade da infestação de bactérias de acordo com variações de cor e aparência da folha. Outra aplicação da junção de técnicas de sensoriamento remoto com imagens hiperespectrais, é a estimativa de produtividade agrícola, possibilitando a obtenção de informações através da condutividade elétrica do solo, informações estas que permitem determinar características como textura e salinidade. A condutividade elétrica do solo pode ser estimada a partir de bandas do espectro vermelho, infravermelho, azul e verde. As características obtidas nestas estimativas podem auxiliar a determinar técnicas de gestão que devem ser aplicadas no campo para as melhoras do rendimento e da qualidade da produção (PINTER et al., 2003).

Em seu trabalho, Reid, Ramos e Sukkarieh (2011) realizaram a classificação de ambientes naturais para estudos de geoestatística extraíndo cores e textura das imagens para gerar vetores de características, sendo a classificação realizada por *Gaussian Processes* (GP). Os resultados da classificação mostraram uma precisão de até 88% entre quatro classes distintas de árvores.

Hung, Bryson e Sukkarieh (2011) utilizaram imagens obtidas por um VANT para classificar a copa de árvores utilizando informações de cor e textura, formando um vetor de características composto por três canais de cores e por trinta canais de textura. Para a classificação de cada vetor de característica foi proposta a utilização de SVM, obtendo uma precisão de 95% na segmentação, sendo que, em comparação à classificação manual da copa das árvores, o método automático atingiu uma precisão de 80%.

Fröhlich et al. (2013) descreveram um método para classificar vários tipos de cobertura terrestre (solo descoberto, água, árvores, gramados, edifícios e impenetrável) em imagens obtidas por satélite utilizando informações contextuais, sendo mais específico, a técnica *Iterative Context Forests* (FRÖHLICH; RODNER; DENZLER, 2012), uma técnica desenvolvida para tarefas de detecção de fachadas em imagens, porém adaptada para conseguir atingir os objetivos propostos. O método utilizado para atribuição dos rótulos, *Iterative Context Forests* (ICF), tem o mesmo princípio da técnica *Semantic Texton Forests* (SHOTTON et al., 2006), sendo que a diferença básica entre os dois métodos é o modo no qual as características para classificação são calculadas. Outra técnica utilizada como base neste trabalho foi *Random Decision Forests* (BREIMAN, 2001), que previne o *overfitting* das árvores de decisão.

Souza et al. (2015) descreveram um método para detecção do fungo *Ceratocystis Fimbriata* em áreas de cultivo de eucalipto por meio de imagens RGB aéreas obtidas por um VANT. Neste trabalho, foi proposta a separação de três classes a serem extraídas da imagem: solo, árvores

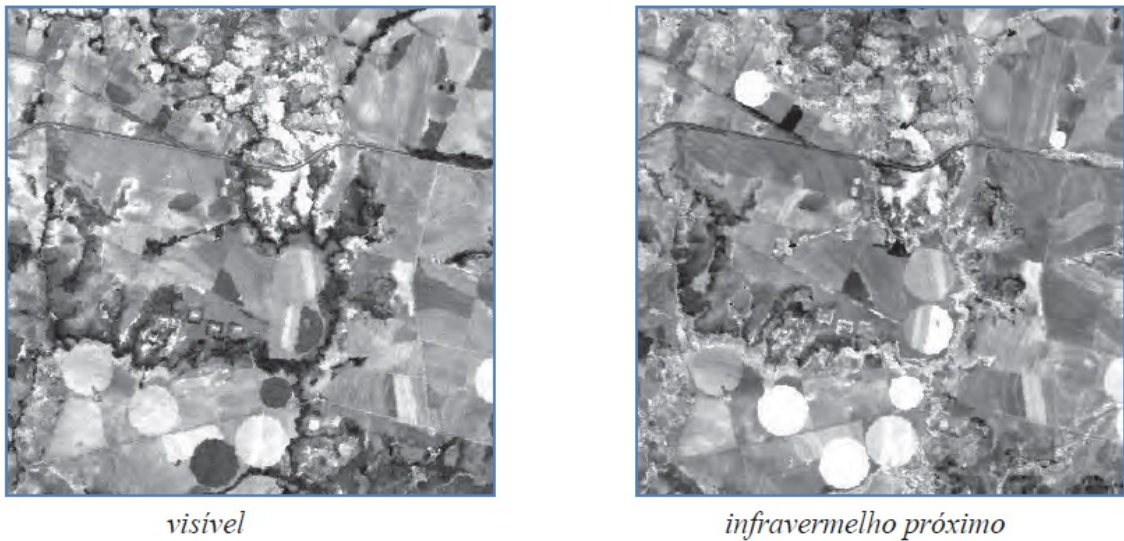
saudáveis e árvores doentes; e quatro métodos de aprendizado de máquina foram testados para realizar a classificação, sendo estes o *K-Nearest Neighbor* (KNN), *Random Forests* (RF), Redes Neurais Artificiais (*Multilayer Perceptron* - MLP) e GP. As características utilizadas para a rotulação dos pixels foram os canais do modelo de cor CIELab (convertido do modelo RGB) e a textura por *Local Binary Pattern* (LBP) (OJALA; PIETIKAINEN; MAENPAA, 2002), extraída da imagem em níveis de cinza. Os resultados mostraram que o classificador que obteve o melhor desempenho para a detecção das árvores doentes foi o GP.

2.2 Imagens Multiespectrais e Hiperespectrais

Segundo Meneses e Almeida (2012), o SR tem as bases no desenvolvimento de câmeras fotográficas que poderiam ser acopladas a aviões para obter fotografias da superfície terrestre, porém, as aplicações desse tipo de imageamento eram limitadas. A partir do surgimento de novas tecnologias para a substituição do filme fotográfico convencional e dos primeiros satélites, novos sensores com maior resolução espectral começaram a ser projetados de modo a obter informações terrestres a partir do espaço. Ao invés do uso de filmes fotográficos que geram imagens analógicas, as imagens passaram a ser digitais e geradas pelos sensores eletrônicos.

Existem dois tipos de sensores remotos utilizados em processos de agricultura de precisão, os sensores multiespectrais e os hiperespectrais. Os sensores multiespectrais dividem o espectro em poucas bandas espectrais (JORGE; INAMASU, 2014), separadas por segmentos espectrais nos quais nenhuma medida é realizada (SHIPPERT, 2004), possibilitando análises de refletância espectral dos alvos de estudo de modo razoável (MENESES; ALMEIDA, 2012). Geralmente, possuem bandas na região visível do espectro eletromagnético e uma banda na região do NIR. A Figura 2.4 mostra duas imagens monocromáticas captadas por sensor multiespectral, sendo uma imagem captada na banda visível do espectro e outra na banda do NIR.

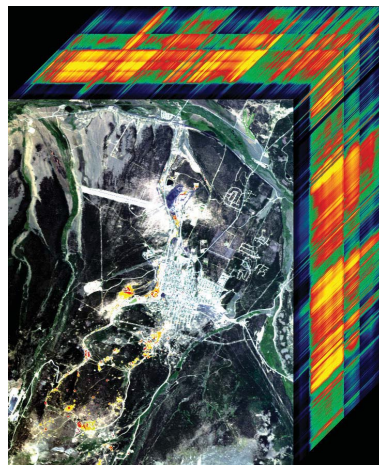
Imagens hiperespectrais fornecem uma grande quantidade de informação espectral que permite identificar e fazer distinção entre materiais com espectros semelhantes (SHIPPERT, 2004). A ideia básica do imageamento hiperespectral tem fundamento no fato de que, para qualquer material existente, a quantidade de radiação que é refletida, absorvida ou emitida, varia de acordo com os comprimentos de onda. Sensores hiperespectrais medem a energia emitida dos materiais, para cada pixel, em um grande número de bandas espectrais adjacentes (MANOLAKIS; MARDEN; SHAW, 2003), provendo assim um potencial para extração de informações mais precisas e detalhadas do que outras formas de dados coletados remotamente (SHIPPERT, 2004). Jorge e Inamasu (2014) dizem que o grande número de bandas espectrais captadas por

Figura 2.4: Exemplos de imagens captadas por sensor multiespectral.

Fonte: Meneses e Almeida (2012).

estes sensores permite detalhar precisamente, através de uma curva, o comportamento espectral de uma área monitorada, assim dando o potencial de discriminação e identificação dos componentes da área, confirmando o que foi dito por Shippert (2004).

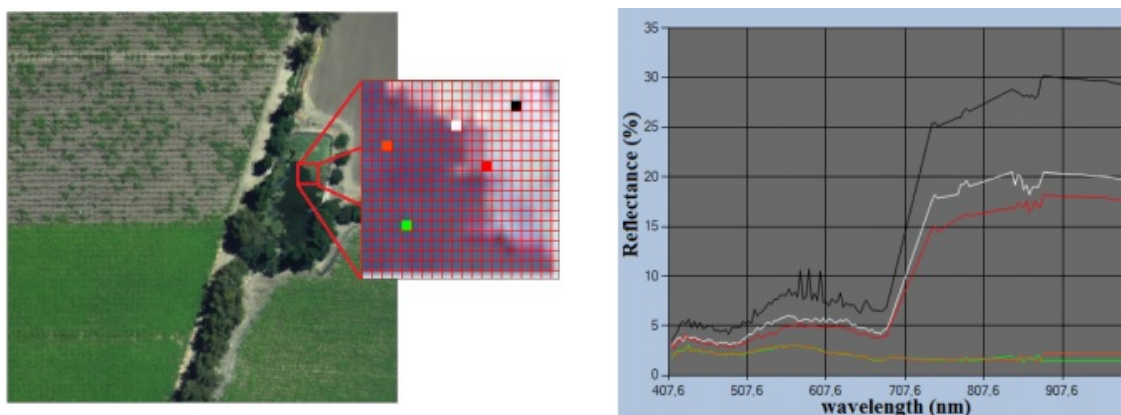
Imagens hiperespectrais também são conhecidas como hipercubo (Figura 2.5) (JORGE; INAMASU, 2014), por causa de ter grande dimensão espectral, além das duas dimensões espaciais (altura e largura) (MENESES; ALMEIDA, 2012). Cada *pixel* da imagem carrega uma ampla quantidade de informação distribuída ao longo do espectro, sendo capaz de reproduzir uma curva ou assinatura espectral de um alvo de observação. A Figura 2.6 mostra exemplos de assinatura espectral para cada *pixel* de uma dada imagem.

Figura 2.5: Exemplo de imagem hiperespectral, hipercubo.

Fonte: Shippert (2004).

Para cada *pixel* dessa imagem, as características espectrais formam vetores nos quais cada

Figura 2.6: Exemplo de assinaturas espectrais. Esquerda: *Pixels* da imagem selecionados. Direita: Assinatura espectral para cada *pixel* selecionado. Os *pixels* e suas respectivas assinaturas espectrais possuem a mesma cor.



Fonte: Elaborada pelo autor.

elemento corresponde a uma das bandas contíguas, abrangendo do espectro visível ao infravermelho. As características espectrais de cada *pixel* são determinadas pelos materiais presentes em sua posição geográfica e pela região que o cerca, além de ruídos causados por variações de iluminação, ângulo de visão e por fatores ambientais, como aplicação de aerossóis e umidade (SOLTANI-FARANI; RABIEE; HOSSEINI, 2015).

A resolução espacial em imagens hiperespectrais de satélite faz com que cada *pixel* presente nela represente áreas em metros quadrados reais. Em uma área de um metro quadrado podem existir vários objetos ou materiais (concreto, grama ou solo exposto, por exemplo), cada um com uma assinatura espectral distinta. O fato de que em um *pixel* possa existir uma mistura de espectros de diferentes materiais puros faz com que, para que a classificação do alvo seja feita com maior acurácia, seja necessário determinar a quantidade de materiais puros que o compõe e porção que eles representam no *pixel*. Cada espectro que caracteriza um material puro é chamado de *endmember* (PLAZA et al., 2012).

2.2.1 Similaridade Espectral

Como visto na subseção anterior, *pixels* de imagens hiperespectrais são compostos por uma mistura de espectros de diferentes materiais que estão no mesmo espaço físico (CHANG, 2000). Uma das formas para se classificar objetos presentes dentro de uma imagem utilizando o espectro é verificando a proximidade de um espectro presente no *pixel* em relação a um espectro de referência para determinado material, ou seja, um *endmember*. Dois métodos bem conhecidos da literatura usados para verificar a proximidade entre os espectros, ou seja, a similaridade espectral, são os *Spectral Angle Mapper* (SAM) (KRUSE et al., 1993) e o *Spectral Information*

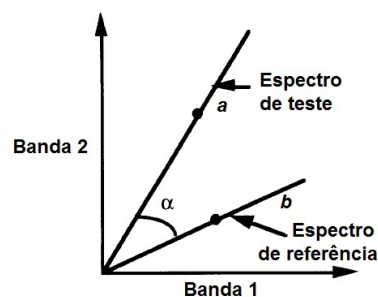
Divergence (SID) (CHANG, 2000).

2.2.1.1 Spectral Angle Mapper (SAM)

O SAM é um classificador espectral que verifica a similaridade entre um espectro de um *pixel* da imagem analisada com um espectro de referência (classe), como dito anteriormente, contido dentro de uma biblioteca espectral que pode ser obtida da própria imagem. O modo como o classificador verifica a similaridade entre os espectros de referência e os da imagem é considerando-os como vetores em um espaço n -dimensional, em que n é o número de bandas do espectro (KRUSE et al., 1993)(MENESES; ALMEIDA, 2012). Quanto menor o ângulo entre eles, mais similares os espectros são, portanto, os *pixels* que possuem menor ângulo em relação a um espectro de referência, são classificados como pertencentes a mesma classe do espectro de referência.

Kruse et al. (1993) dá o que seria uma explicação simplificada para o método ao considerar-se dois espectros, sendo um deles o de referência, de apenas duas bandas representadas em um espaço bi-dimensional como dois pontos (Figura 2.7). As retas que conectam cada um dos pontos e a origem contêm todas as posições de intensidade de iluminação para o material.

Figura 2.7: Plotagem de um espectro de referência e um espectro de teste com duas bandas. Os mesmos materiais com variações na iluminação são representados pelos vetores conectados à origem e projetados aos pontos que representam o espectro real.



Fonte: Adaptado de Kruse et al. (1993).

Ao se observar a Figura 2.7, é possível notar que o ângulo entre os vetores (espectros de teste e de referência) é o mesmo, independentemente de seus comprimentos. O algoritmo generaliza a interpretação desse espaço para n -dimensional, sendo que a similaridade entre os dois espectros é calculada como sendo o arco-cosseno do produto escalar entre eles, como pode ser visto na Equação 2.1, onde n é o número de bandas.

$$\alpha = \cos^{-1} \left(\frac{\sum_{i=1}^n a_i b_i}{\left(\sum_{i=1}^n a_i^2\right)^{\frac{1}{2}} \left(\sum_{i=1}^n b_i^2\right)^{\frac{1}{2}}} \right) \quad (2.1)$$

2.2.1.2 Spectral Information Divergence (SID)

Por causa de fatores atmosféricos, a informação contida em cada *pixel* varia durante a aquisição das imagens. A *Spectral Information Measure* (SIM), uma medida espectral baseada na teoria da informação, modela esta variação de informação entre bandas como resultado da incerteza causada pela aleatoriedade, considerando cada *pixel* da imagem como sendo uma variável aleatória e define a distribuição de probabilidade ideal fazendo a normalização do histograma espectral para unidade, além de gerar estatísticas de alto nível dos *pixels* baseados na assinatura espectral (CHANG, 2000).

A SID é uma técnica de classificação por similaridade espectral entre *pixels* baseada no SIM, sendo que cada *pixel* é classificado por meio da avaliação da diferença entre os comportamentos probabilísticos entre as assinaturas espectrais do *pixel* da imagem e do espectro de referência. Assim como no SAM, quanto menor o valor retornado pelo cálculo de SID, mais similares são os espectros. Considerando-se duas assinaturas espectrais r_i e r_j , o cálculo de SID pode ser definido pela equação 2.2.

$$SID(r_i, r_j) = D(r_i \parallel r_j) + D(r_j \parallel r_i) \quad (2.2)$$

onde:

$$D(r_i \parallel r_j) = \sum_{l=1}^L p_l \log_2 \left(\frac{p_l}{q_l} \right) \quad (2.3)$$

Na equação 2.3, $D(r_i \parallel r_j)$ é a medida de informação de Kullback-Leibler, também conhecida como divergência dirigida ou entropia cruzada; $p = (p_1, \dots, p_L)^T$ e $q = (q_1, \dots, q_L)^T$ são respectivamente os vetores de probabilidade do *pixel* a ser classificado e a assinatura espectral de referência. O vetor de probabilidade é resultante de um vetor do *pixel* hiperespectral $x = (x_1, \dots, x_L)^T$, em que cada componente x_l é um valor de reflectância ou radiância no comprimento de onda l pertencente a um dos L comprimentos de onda captados na imagem hiperespectral. A probabilidade para um *pixel* x em um comprimento de onda j é dada pela equação 2.4.

$$p_j = \frac{x_j}{\sum_{l=1}^L x_l} \quad (2.4)$$

2.2.2 Índices de Vegetação

No Capítulo 1 foi descrito que a aplicação de sensoriamento remoto na agricultura se dá a partir da interação da radiação eletromagnética com o solo ou com as plantas, existindo regiões do espectro eletromagnético em que ocorrem maior reflexão e maior absorção de luz por parte da vegetação.

Os índices de vegetação foram criados com base nessas respostas espectrais das plantas em relação a radiação (SHIRATSUCHI et al., 2014), sendo chamados de transformações espectrais desenvolvidas para aprimorar a contribuição das propriedades da vegetação (HUETE et al., 2002). São resultados de operações matemáticas entre imagens, transformando os dados a partir da combinação *pixel a pixel* entre duas ou mais imagens monocromáticas de bandas diferentes a partir de uma regra matemática pré-definida.

Segundo Meneses e Almeida (2012), os índices de vegetação são operações de divisão entre bandas que tem como objetivo realçar as variações de densidade da cobertura vegetal. Na vegetação, a principal banda de absorção é a que se encontra no comprimento de onda de 650 nm, região do espectro visível onde se localiza a luz vermelha; e a região do espectro em que ocorre o maior nível de reflectância se encontra na região do NIR, entre os comprimentos de onda de 760 nm a 1200 nm (JACKSON; HUETE, 1991)(MENESES; ALMEIDA, 2012).

Existem diversos índices espectrais utilizados na agricultura e na avaliação de processos específicos, como processos fotoquímicos relacionados à fotossíntese ou estimativa do conteúdo da pigmentação de folhas. Dois índices de vegetação serão descritos a seguir: *Normalized Difference Vegetation Index* e *Modified Photochemical Reflectance Index*.

2.2.2.1 *Normalized Difference Vegetation Index* - NDVI

O índice NDVI tende a ser linearmente proporcional à biomassa da vegetação, sendo o seu uso mais apropriado quando há a necessidade de fazer comparações de uma mesma área ao decorrer do tempo porque sofre menos influência das mudanças nas condições atmosféricas (MENESES; ALMEIDA, 2012), o que o torna bastante utilizado no monitoramento de culturas agrícolas (SHIRATSUCHI et al., 2014). Além disso, é um índice de vegetação amplamente utilizado, sendo que as principais aplicações do mesmo são em estudos de medida de área foliar,

avaliação do percentual de cobertura vegetal e de estimativas da radiação fotossinteticamente ativa (usadas em modelos de estudo de fotossíntese e sequestro de carbono).

O cálculo do NDVI utiliza as mesmas bandas do espectro eletromagnético que índice *Simple Ratio*, calculado pela razão entre NIR e vermelho; porém, é realizada a divisão da subtração da imagem da banda NIR e da imagem da banda vermelha pela soma entre as imagens dessas duas bandas (JACKSON; HUETE, 1991) (MENESES; ALMEIDA, 2012), conforme a equação 2.5, onde $Imagem_{NIR}$ e $Imagem_{Red}$ representam respectivamente matrizes para a banda NIR e banda vermelha. A operação é realizada *pixel a pixel*.

$$NDVI = \frac{Imagem_{NIR} - Imagem_{Red}}{Imagem_{NIR} + Imagem_{Red}} \quad (2.5)$$

Os valores obtidos a partir do NDVI são valores modularizados do SR, variam de -1 a 1 e são melhores indicativos para um índice de vegetação, pois em áreas de vegetação densa o valor retornado é próximo de 1, em locais com ausência de vegetação os valores são negativos e locais com solo exposto tem valores próximos de zero.

2.2.2.2 *Modified Photochemical Reflectance Index - MPRI*

Assim como o NDVI, o MPRI, proposto por Yang, Willis e Mueller (2008), também é baseado em uma diferença normalizada entre duas imagens monocromáticas geradas por duas bandas distintas, porém, as duas bandas utilizadas para o cálculo estão na região visível do espectro eletromagnético, mais especificamente, nas regiões verde e vermelha do espectro. As duas bandas foram selecionadas pois a relação entre elas é similar à relação das bandas na região do vermelho e NIR (LINHARES; ROCHA; AMARAL, 2013).

Apesar do intuito da criação deste índice de vegetação ter sido o de analisar a eficiência da utilização de luz e refletir o estresse hídrico em culturas agrícolas, muitos especialistas o utilizam como uma alternativa ao NDVI, devido ao acesso limitado a sensores multiespectrais ou hiperespectrais. Linhares, Rocha e Amaral (2013) realizaram um estudo para avaliar a correlação entre estes dois índices de vegetação, assim como validar o MPRI como um índice vegetativo. Ao final do estudo, concluíram que existe alta correlação entre os dois índices vegetativos (acima de 80%), assim como comportamentos semelhantes quando os alvos de análise são solo e vegetação, porém, quando há presença de corpos d'água na imagem, os resultados se tornam conflitantes, pois a água tem comportamento distinto para os dois índices, sendo esta uma grande desvantagem em relação ao NDVI. A equação 2.6 mostra como este índice é calculado.

$$MPRI = \frac{Imagem_{Green} - Imagem_{Red}}{Imagem_{Green} + Imagem_{Red}} \quad (2.6)$$

2.3 Modelos de cor

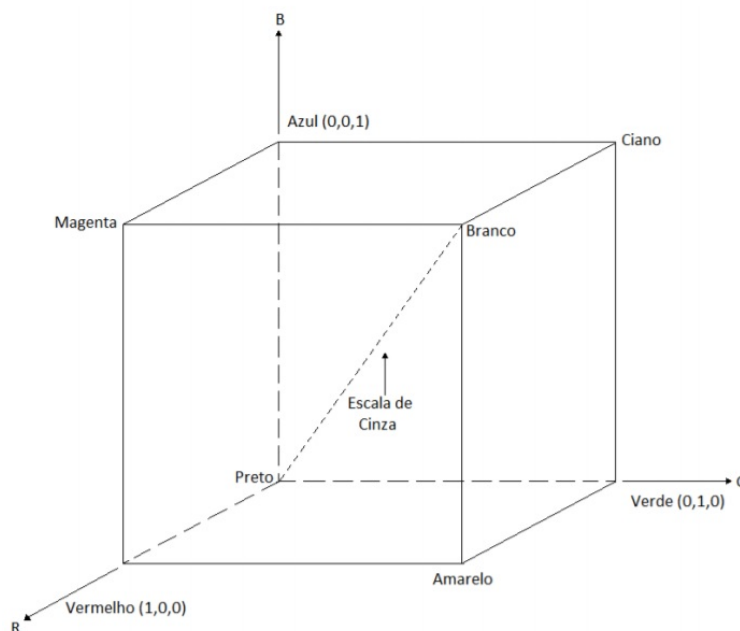
Como o objetivo do trabalho de pesquisa é realizar a estimativa de imagens NIR a partir de imagens de câmeras comuns (padrão RGB), é necessário revisar alguns modelos de cores e suas particularidades, de modo a verificar quais obterão os melhores resultados para inferir a nova imagem.

2.3.1 Modelo RGB

A superfície da retina do olho humano é composta de receptores discretos de luz, chamados de cones e bastonetes, sendo que os cones são sensíveis à cor e os bastonetes sensíveis à luminosidade. Os cones da retina podem ser divididos em três categorias de sensoriamento, de acordo com o tipo de cor captada: Vermelho, Verde ou Azul (GONZALEZ; WOODS, 2010) (TKALCIC; TASIC, 2003).

O modelo de cores RGB (*Red*, *Green* e *Blue*) trabalha com as cores representadas pelas componentes Vermelho (R), Verde (G) e Azul (B) (cores primárias), assim como os cones da retina. É chamado de modelo aditivo, pois todas as outras cores conhecidas são obtidas a partir de combinações entre as componentes R, G e B. Um exemplo disso é a combinação para obtenção da cor amarela, que é composta pela combinação de valores máximos para as componentes R e G e pelo valor nulo da componente B (R = 255, G = 255 e B = 0). Se todas as componentes tiverem seu valor máximo (255), a combinação entre elas resultará na cor branca. Segundo Chaves-González et al. (2009), é o modelo de cor mais conhecido e é o utilizado por dispositivos eletrônicos de visualização, portanto, as componentes R, G e B podem ser obtidas diretamente das imagens digitais. Sendo, assim, é o modelo de cores mais básico e suas componentes podem ser usadas para obter componentes de outros modelos de cor (TKALCIC; TASIC, 2003). As principais desvantagens desse modelo é não ser tolerante a variações na iluminação e ser dependente das funções de sensibilidade do dispositivo de captura da imagem.

Na Figura 2.8 é mostrada a conhecida representação deste modelo na forma de um cubo, em que as cores primárias e secundárias (ciano, magenta e amarelo) estão localizadas. No ponto de origem do cubo está a cor preta e no ponto superior à frente do preto está a cor branca. Na diagonal formada pelas posições das cores preto e branco estão os valores dos níveis de cinza.

Figura 2.8: Representação do modelo de cor RGB em forma de cubo.

Fonte: Adaptado de Gonzalez e Woods (2010).

2.3.2 Modelo HSV

Diferente do modelo de cor RGB, em que as cores são compostas a partir da combinação das três cores primárias, o modelo HSV, também conhecido como HSL e HSI (CHAVES-GONZÁLEZ et al., 2009), é baseado na forma em que o seres humanos descrevem as cores (TKALCIC; TASIC, 2003), baseados nas componentes matiz (*Hue - H*), saturação (*Saturation - S*) e valor (*Value - V*).

A componente matiz é a própria cor (azul, roxo, amarelo), diferenciando uma cor de outra. Os valores que esta componente pode ter estão dentro do intervalo de 0 a 360. A saturação é a componente que descreve a intensidade ou a pureza da cor definida pela matiz, assumindo valores de 0 a 1. Quanto menor o valor da componente saturação, mais fraca e apagada será a cor, apresentando um aspecto desbotado. Quanto maior o valor da componente S, mais viva é a cor. Por fim, a componente valor é relacionada à intensidade da luminosidade ou brilho da cor, tendo valores que vão de 0 a 1, assim como a saturação (CHAVES-GONZÁLEZ et al., 2009).

A representação desse modelo de cores é dada por um cone (Figura 2.9), sendo que os valores da componente matiz são os ângulos ao redor do eixo vertical e este representa os valores da componente valor. Os valores da componente saturação são representados pela distância radial em relação ao eixo vertical do cone.

Os valores das três componentes deste modelo de cor podem ser obtidas a partir das três

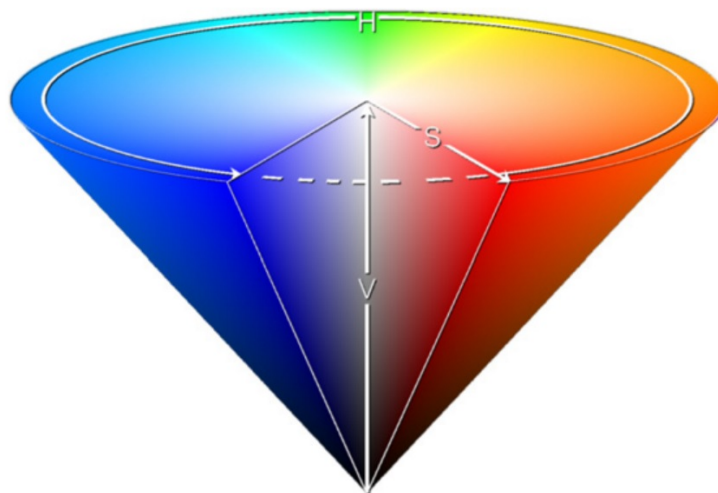
componentes do modelo RGB utilizando as seguintes equações, onde MAX e MIN representam os valores máximo e mínimo, respectivamente, entre as componentes R, G e B:

$$H = \begin{cases} \textit{indefinido} & , \text{ se } MAX = MIN \\ 60 \left(\frac{G-B}{MAX-MIN} \right) & , \text{ se } MAX = R \text{ e } G \geq B \\ 60 \left(\frac{G-B}{MAX-MIN} \right) + 360 & , \text{ se } MAX = R \text{ e } G < B \\ 60 \left(\frac{G-B}{MAX-MIN} \right) + 120 & , \text{ se } MAX = G \\ 60 \left(\frac{G-B}{MAX-MIN} \right) + 240 & , \text{ se } MAX = B \end{cases} \quad (2.7)$$

$$S = \begin{cases} 0 & , \text{ se } MAX = 0 \\ 1 - \left(\frac{MIN}{MAX} \right) & , \text{ caso contrário} \end{cases} \quad (2.8)$$

$$V = MAX \quad (2.9)$$

Figura 2.9: Representação cônica do modelo de cor HSV.



Fonte: Chaves-González et al. (2009).

Segundo Tkalcic e Tasic (2003), é um modelo de cor dependente do dispositivo, uma vez que as componentes dos canais são obtidas por transformações lineares dos canais RGB e não incluem informação de cromaticidade e existe uma correlação ruim entre a iluminação calculada e a iluminação percebida.

2.3.3 Modelo CIE Lab

A CIE (*Commission Internationale de l'Éclairage*), organização internacional de padronização de cor e luz, desenvolveu em 1931 um novo modelo de cores a partir de um experimento em que um observador deveria formar uma cor ao observar uma projeção de três fontes de luz em uma tela, cada uma delas sendo uma das cores primárias. Conforme a intensidade das fontes de luz variava, novas cores eram formadas, porém, algumas cores não puderam ser formadas utilizando esse método, sendo necessário utilizar intensidades negativas nas fontes luminosas para obtê-las. Sendo assim, foi definido o modelo de cor XYZ, de modo a criar sistemas que utilizem apenas combinações de valores positivos das cores primárias para representar qualquer cor. Neste caso, as componentes primárias são X, Y e Z (PRATT, 2006).

Em 1976, a CIE propôs um novo modelo de cor para criar um modelo de cor dito como perceptualmente uniforme, uma vez que o XYZ não é (BILLMEYER; SALTZMAN, 1981). A importância do modelo de cor ser perceptualmente uniforme é a garantia de que uma mudança nos valores de suas componentes possa refletir proporcionalmente no modo em que os seres humanos percebem essas cores resultantes das alterações. O modelo CIE Lab é composto por três componentes: L, a^* e b^* . A componente L representa o brilho associado a cor, assim como a componente V do modelo HSV. As duas outras componentes, a^* e b^* , representam escalas de matizes, variando de valores negativos a positivos. A componente a^* , em termos de cor, varia da cor verde (valores negativos) a vermelho (valores positivos), e a componente b^* varia da cor azul (valores negativos) a amarelo (valores positivos), conforme pode ser visto na Figura 2.10.

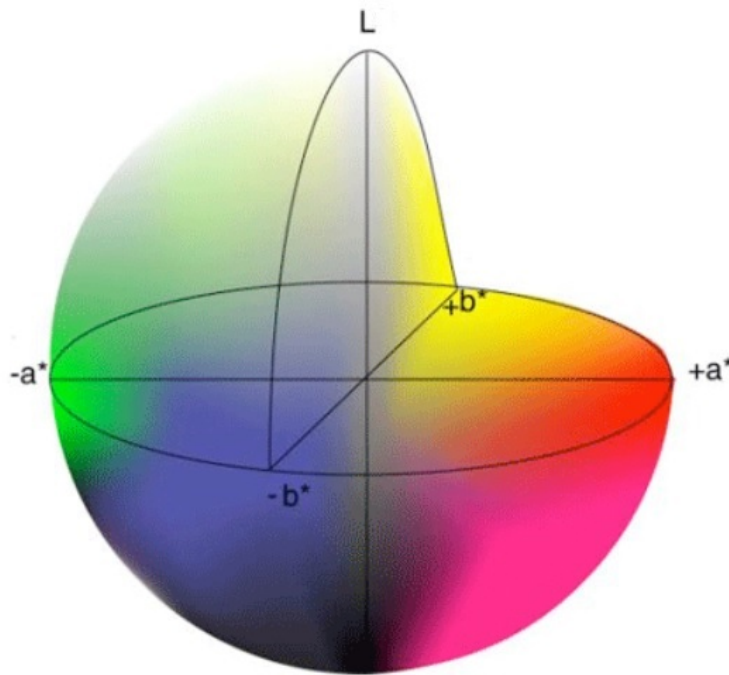
Para o cálculo das componentes do modelo CIE Lab, é necessário que as componentes da imagem estejam no modelo XYZ, ou seja, não é possível as componentes deste modelo serem obtidas diretamente de uma imagem RGB. As componentes do modelo XYZ são obtidas através das equações 2.10 a 2.12:

$$X = 0,49018626R + 0,30987954G + 0,19993420B \quad (2.10)$$

$$Y = 0,17701522R + 0,81232418G + 0,01066060B \quad (2.11)$$

$$Z = 0,01007720G + 0,98992280B \quad (2.12)$$

sendo que R, G, B, X, Y e $Z \in [0, 1]$.

Figura 2.10: Representação do modelo de cor CIE Lab.

Fonte: <https://www.flexoglobal.com/flexomag/08-September/flexomag-ploumidis.htm>.

Após a transformação das componentes do modelo RGB para o modelo XYZ, é possível calcular as componentes para o modelo CIE Lab utilizando as equações 2.13 a 2.15:

$$L = 116f\left(\frac{Y}{Y_n}\right) - 16 \quad (2.13)$$

$$a^* = 500 \left[f\left(\frac{X}{X_n}\right) - f\left(\frac{Y}{Y_n}\right) \right] \quad (2.14)$$

$$b^* = 200 \left[f\left(\frac{Y}{Y_n}\right) - f\left(\frac{Z}{Z_n}\right) \right] \quad (2.15)$$

sendo que:

$$f(k) = \begin{cases} k^{\frac{1}{3}} & , \text{ se } k > \left(\frac{6}{29}\right)^3 \\ \frac{1}{3} \left(\frac{29}{6}\right)^2 k + \frac{4}{29} & , \text{ caso contrário.} \end{cases} \quad (2.16)$$

Nas equações anteriores, X_n , Y_n e Z_n são valores CIE XYZ para o iluminante referência, sendo que foi utilizado para as conversões dos modelos de cor o padrão D65 (padrão de iluminante que simula a cor do dia artificialmente), em que $X_n = 95,047$, $Y_n = 100,000$ e $Z_n =$

108,883.

2.4 K-Nearest-Neighbour (KNN)

O classificador de k vizinhos mais próximos (*k-nearest-neighbour*) é um dos classificadores mais populares e mais simples existentes na literatura (WEBB; COPSEY, 2011). O procedimento básico para a classificação consiste em determinar k vizinhos mais próximos dentro de um conjunto de treinamento para uma amostra não classificada, podendo utilizar uma métrica de proximidade como medida de distância. Quanto menor a distância entre a amostra de teste x e um dado x_i do conjunto de treinamento, mais próximos eles são. Devem ser fornecidos para o algoritmo o número de vizinhos desejados (k) e a métrica de proximidade como parâmetros e o conjunto de treinamento.

O valor de k deve ser escolhido de modo que este seja grande o bastante, para diminuir a sensibilidade a ruído, e pequeno o suficiente para que o número de vizinhos não ultrapasse o domínio de outras classes (WEBB; COPSEY, 2011). A métrica de proximidade mais comum utilizada nos classificadores de vizinhos mais próximos é a Distância Euclidiana, representada na equação 2.17.

$$d(x_i, x) = |x_i - x| \quad (2.17)$$

Na equação 2.17, $d(x_i, x)$ é a distância entre a amostra de teste x e a amostra x_i do conjunto de treinamento ($i = 1, \dots, n$). A distância é calculada para todo o conjunto de treinamento e após isso, os k vizinhos x_j com as menores distâncias são encontrados, sendo ($j = 1, \dots, k$).

A partir dos k vizinhos x_j mais próximos de x , a amostra é classificada de acordo com a maioria dos rótulos das classes destes vizinhos (votos). Por exemplo, se $k = 5$, 3 dos vizinhos possuem rótulo A e dois possuem rótulo B , então, a amostra de teste será classificada como pertencente à classe A , pois obteve maior número de votos dessa classe.

2.5 Deep Learning

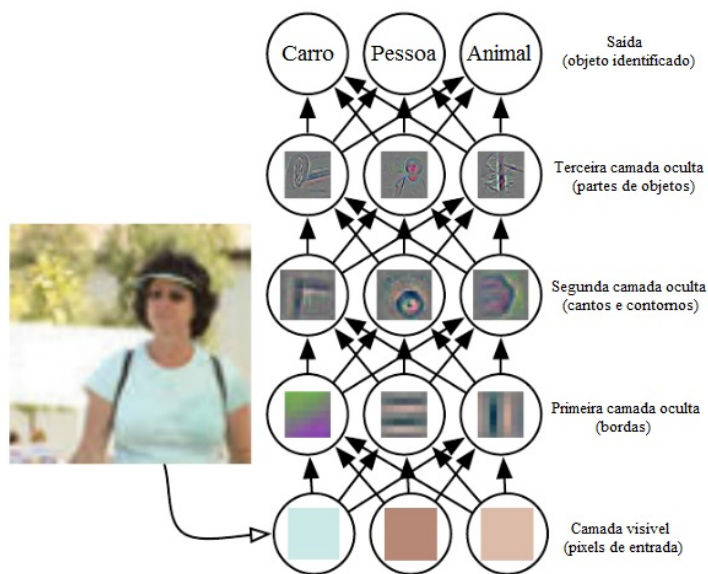
Um obstáculo bastante comum em tarefas de reconhecimento de padrões é a alta dimensionalidade dos dados, pois a complexidade de aprendizado aumenta exponencialmente quando a dimensão dos dados aumenta linearmente. Este fenômeno foi chamado por Richard Bellman de Maldição da Dimensionalidade. A forma mais comum de lidar com este fenômeno é

o pré-processamento dos dados para diminuir a quantidade de dados que serão efetivamente processados, chamado de seleção de características (AREL; ROSE; KARNOWSKI, 2010). Durante muito tempo, a construção de sistemas de aprendizado de máquina ou de reconhecimento de padrões necessitavam de métodos cuidadosos e grande domínio para se desenvolver algoritmos de seleção de características que pudessem transformar os dados brutos em dados que sejam adequados para os classificadores (LECUN; BENGIO; HINTON, 2015). Se características ruins ou incompletas são extraídas, o processo de classificação tem desempenho limitado (AREL; ROSE; KARNOWSKI, 2010).

Descobertas da neurociência sobre o modo como a informação é representada no cérebro inspiraram novas ideias para o desenvolvimento de sistemas que representam informações. O neocórtex, parte do cérebro associada a habilidades cognitivas, não realiza o pré-processamento dos sinais sensoriais e os permite ser propagados por uma hierarquia complexa de estruturas que, com o tempo, aprendem a representar observação com base na regularidade com as quais são exibidas. Esta descoberta foi o que motivou o surgimento de *Deep Learning* (Aprendizagem Profunda) (AREL; ROSE; KARNOWSKI, 2010).

Segundo LeCun, Bengio e Hinton (2015), *Deep Learning* permite modelos computacionais compostos de múltiplas camadas de processamento aprenderem a representação com vários níveis de abstração. Os métodos de aprendizagem profunda consistem na combinação de módulos simples, mas não lineares, que transformam a representação de um nível (começando pelos dados brutos) ao outro nível mais alto, sendo este um pouco mais abstrato. Com essa composição de muitas dessas transformações, funções muito complexas podem ser aprendidas. Em tarefas de classificação, as camadas de representação superiores amplificam aspectos dos dados de entrada que são importantes para a discriminação e suprime variações irrelevantes, ou seja, cada uma das camadas aprende uma característica, inicialmente simples, sendo que esta é transmitida para outra camada aprender outra característica, sendo esta mais abstrata e complexa do que a da camada anterior, eliminando assim a necessidade de especialistas especificarem formalmente todo o conhecimento necessário para o aprendizado (GOODFELLOW; BENGIO; COURVILLE, 2016). Na Figura 2.11 é mostrado um exemplo de sistema de *deep learning* capaz de inferir a representação de uma fotografia de uma pessoa em função da combinação de conceitos mais simples. Dada uma imagem representada pelos seus *pixels*, a primeira camada pode facilmente identificar bordas utilizando brilho e vizinhança entre *pixels*; a partir das bordas, a segunda camada pode aprender cantos e contornos; e a partir destas características aprendidas, a terceira camada oculta pode detectar partes complexas de objetos, que poderão ser identificados na camada de saída.

Figura 2.11: Ilustração de um modelo de *Deep Learning*.



Fonte: Adaptado de Goodfellow, Bengio e Courville (2016).

Dentre os tipos de redes neurais para aprendizagem profunda existentes, se destacam as *Convolutional Neural Networks* (CNNs). As CNNs são redes neurais convolucionais desenvolvidas especificamente para serem utilizadas com dados bidimensionais, como imagens e vídeos (AREL; ROSE; KARNOWSKI, 2010), tendo muito sucesso em aplicações práticas (GOODFELLOW; BENGIO; COURVILLE, 2016). A arquitetura típica de uma CNN é uma estrutura composta de várias etapas. Os primeiros poucos estágios iniciais são compostos de dois tipos de camadas: camadas convolucionais e camadas de agrupamento (*pooling*). O papel das camadas convolucionais é detectar conjunções de características locais de camadas anteriores e o papel das camadas de agrupamento é fundir características semanticamente similares em uma única (LECUN; BENGIO; HINTON, 2015).

Novas arquiteturas de redes *Deep Learning* foram propostas para geração de imagens a partir de uma ou mais imagens de referência. A este processo dá-se o nome de *image-to-image translation* (tradução imagem para imagem). Para cada tarefa de mapeamento de uma imagem para outra, existe um algoritmo específico de aprendizado profundo para que o resultado desejado seja obtido, como métodos específicos para extração de bordas ou para mapear imagens em rótulos semânticos, por exemplo. Isola et al. (2017) propôs um método, utilizando redes de arquitetura do tipo GAN (*Generative Adversarial Networks*), para realização de diferentes tarefas de tradução imagem para imagem, sendo este método popularmente conhecido como *Pix2Pix*.

2.5.1 Pix2Pix

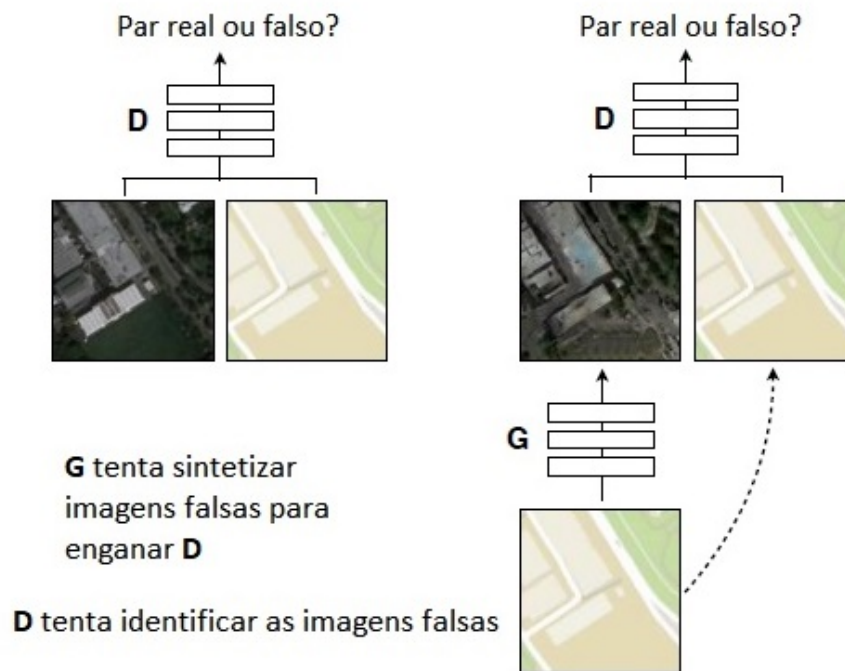
Goodfellow et al. (2014) propuseram um modelo de redes neurais composto por duas redes *perceptron* multicamadas que competem entre si, na qual uma rede cria novas amostras a partir de um ruído aleatório (modelo gerador - *Generative*) e a outra rede que aprende se esta amostra é real ou falsificada (modelo discriminador - *Discriminative*), sendo este modelo chamado de *Generative Adversarial Networks* (GAN).

O funcionamento da GAN é análogo ao problema de falsificação de produtos ou documentos, no qual um grupo de falsificadores criam mercadorias falsificadas que possam ser comercializadas sem que ninguém note que elas não são autênticas (modelo gerador), enquanto investigadores policiais tentam detectar se produtos vendidos são autênticos ou falsificados (modelo discriminador). Em ambos os modelos, existe um esforço enorme para que os dois tenham sucesso, porém, o modelo gerador tende a melhorar seu método de criação para que as amostras produzidas sejam cada vez mais difíceis de serem diferenciadas de uma amostra real.

A partir do princípio da GAN, novos modelos inspirados em redes que competem entre si surgiram, dentre eles, o *Conditional Generative Adversarial Networks* (cGAN) (MIRZA; OSINDERO, 2014). O cGAN condiciona informações adicionais (rótulos de classes, por exemplo) às redes geradora e discriminadora, tornando-se este modelo adequado para o mapeamento de uma imagem para outra (ISOLA et al., 2017). Diferente dos outros modelos de GAN, nesta arquitetura, o objetivo do modelo gerador não é só enganar o discriminador, mas também produzir imagens mais semelhantes às imagens de referência (*ground truth*).

Isola et al. (2017) utilizaram o conceito das cGAN para criar um modelo de propósito geral para traduções de imagem para imagem, chamado *Pix2Pix*. Diferente do modelo de cGAN de Mirza e Osindero (2014), no qual duas redes *perceptron* multicamadas era utilizadas como modelos geradores e discriminadores, o *Pix2Pix* utiliza arquiteturas de aprendizado profundo, sendo mais específico, arquitetura baseada na rede convolucional *U-Net* (RONNEBERGER; FISCHER; BROX, 2015) para o módulo gerador e para o módulo discriminador, um classificador convolucional *PatchGAN*, chamado também como Discriminador Markoviano. A figura 2.12 representa o funcionamento da rede *Pix2Pix* em um contexto de geração de imagens aéreas realistas a partir de mapas. Na figura da esquerda, o discriminador D aprende a classificar entre imagem falsa e real, recebendo como entrada uma dupla (mapa, foto). O gerador G aprende a enganar o discriminador, gerando uma imagem falsa a partir do mapa de entrada. Na figura da direita, a imagem de entrada é transformada numa foto pelo gerador G, na parte inferior; e o discriminador compara uma foto gerada com o mapa de entrada.

Figura 2.12: Exemplo de funcionamento da Pix2Pix.



Fonte: Adaptado de Isola et al. (2017).

O processo de classificação do *PatchGAN* em imagens reais ou falsas é feito por meio de *patches* (partes) da imagem. Cada *patch* de tamanho $N \times N$ da imagem é analisado, por meio de convoluções, e classificado como real ou falsificado, calculando a média para obtenção do resultado definitivo de classificação do discriminador. Segundo Isola et al. (2017), o tamanho N do *patch* pode ser muito inferior ao tamanho da imagem e ainda gerar resultados de alta qualidade. O *PatchGAN* também pode ser considerado como uma função de custo para textura ou estilo (ISOLA et al., 2017).

Para a atualização de pesos das redes, o modelo utiliza uma função objetivo dupla, misturando a função de erro *Conditional GAN* (equação 2.18) com a função de erro L1 (soma das diferenças entre *pixels* da imagem gerada e do *ground truth*) (equação 2.19). Os pesos da rede Discriminadora são atualizados diretamente após a verificação da autenticidade da imagem. Os pesos da rede Geradora são atualizados a partir da atualização dos pesos da rede Discriminadora. Nas equações 2.18 e 2.19, x corresponde à imagem *ground truth*, y é imagem gerada e z é uma imagem de ruído aleatório de entrada. G é a função geradora que tem como entrada as imagens x e z e gera a imagem de saída y . D é a função discriminante, que é treinada para detectar imagens falsas.

$$\mathcal{L}_{cGAN}(G, D) = E_{x,y} [\log D(x, y)] + E_{x,z} [\log(1 - D(x, G(x, z)))] \quad (2.18)$$

$$\mathcal{L}_{L1}(G) = E_{x,y,z} [\|y - G(x, z)\|_1] \quad (2.19)$$

Na equação 2.18, o gerador G tenta minimizar o objetivo em relação ao seu oponente, o discriminador D , que tenta maximizar seu objetivo, ambos utilizando a imagem *ground truth* x . Na equação 2.19, o objetivo é verificar o quanto G produz imagens y mais parecidas com o *ground truth* x , utilizando como entrada uma imagem de ruído aleatório z . A combinação das duas funções objetivo G^* é mostrada na equação 2.20, onde λ é um valor de ponderação para $\mathcal{L}_{L1}(G)$.

$$G^* = \arg \min_G \max_D \mathcal{L}_{cGAN}(G, D) + \lambda \mathcal{L}_{L1}(G) \quad (2.20)$$

A figura 2.13 mostra resultados de geração de imagens para diferentes tipos de imagem de entrada. O exemplo do canto superior esquerdo, denominado "Rótulos para cena de rua", mostra uma imagem de saída próxima de uma imagem real, a partir de uma imagem de entrada sem detalhes dos objetos em cena. No segundo exemplo, na parte central superior, é vista uma imagem de uma fachada de edifício, sendo a imagem de entrada sem detalhes dos elementos do edifício e a imagem de saída, próxima do real. O exemplo do canto superior direito refere-se a geração de colorações na imagem de saída, em relação à imagem de entrada em branco e preto. No canto inferior esquerdo, tem-se o exemplo de uma imagem aérea de entrada e uma imagem de saída mostrando um mapa de ruas correspondente. Na parte inferior do centro, tem-se o exemplo de uma imagem de entrada de uma paisagem diurna, transformada para uma imagem de saída noturna. Finalmente, no canto inferior direito, tem-se uma imagem de bordas de uma bolsa como entrada e uma imagem de saída de uma bolsa com detalhes de coloração e textura.

2.5.2 cycleGAN

As redes *Pix2Pix* tem um bom desempenho com pares de imagens alinhados, ou seja, os *pixels* da imagem e de seu respectivo *ground truth* devem ter a mesma correspondência. Em algumas tarefas de mapeamento de imagem para imagem, não é possível a obtenção de dados de treinamento que possuem essa correspondência entre *pixels* da imagem e do *ground truth* (Figura 2.14). De modo a lidar com o problema da falta de imagens com alinhamento entre *pixels*, foi proposta uma nova abordagem, baseada na *Pix2Pix*, chamada de *cycleGAN* (ZHU et

Figura 2.13: Resultados de tradução imagem para imagem com a *Pix2Pix*.

Fonte: Adaptado de Isola et al. (2017).

al., 2017), técnica bastante utilizada para transferência de estilos (transformação de fotografias em pinturas com estilo de algum artista renomado, por exemplo).

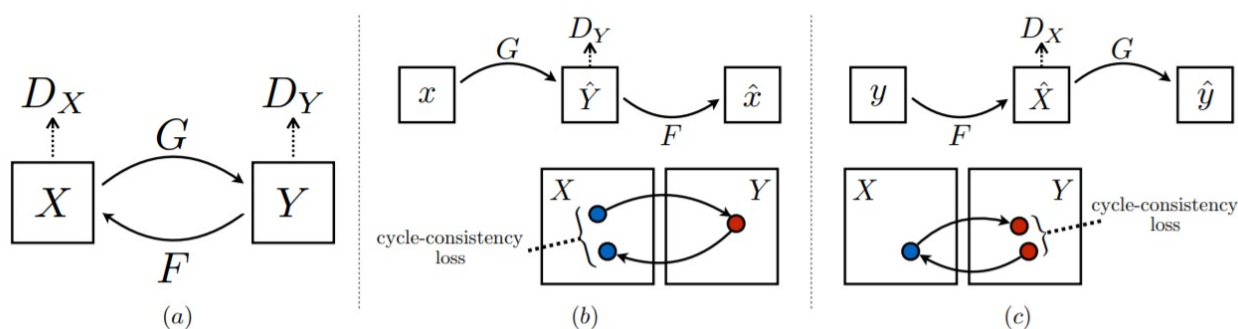
Figura 2.14: Exemplos de imagens com correspondência entre *pixels* (esquerda) e de imagens sem correspondência entre *pixels* (direita).

Fonte: Adaptado de Zhu et al. (2017).

A *cycleGAN* possui duas redes geradoras e duas redes discriminadoras, sendo assim composta por duas GANs treinadas simultaneamente. Cada GAN (composta pelo par gerador e discriminador) tem o funcionamento semelhante ao da *Pix2Pix*, porém, as duas são complementares. Pode ser descrito um exemplo ilustrativo para o funcionamento do treinamento da rede no contexto de uma aplicação que tem como objetivos transformar imagens capturadas no verão para que fiquem com aparência de inverno: A GAN 1 transforma uma imagem de verão real para imagem "inverno", sendo esta imagem a entrada da GAN 2, que a transforma em uma imagem "verão" e esta saída é comparada com a entrada da GAN 1. Neste caso, a rede GAN 1 é

denominada geradora. Depois, a GAN 2 transforma uma imagem de inverno real em "verão" e esta é transformada em imagem "inverno" pela GAN 1, sendo esta imagem comparada com a imagem de entrada da GAN 2, que neste caso é denominada rede geradora. De modo mais simplificado, uma rede geradora é denotada G e tem como objetivo converter imagens do domínio X para o domínio Y ($G : X \rightarrow Y$). A outra rede geradora, chamada de F , faz o processo inverso, convertendo imagens do domínio Y para o domínio X ($F : Y \rightarrow X$). Cada rede geradora possui uma rede discriminadora correspondente, chamadas D_X e D_Y , respectivamente, responsáveis por distinguir as imagens reais (x e y) das imagens sintéticas ($G(x)$ e $F(y)$), conforme pode ser visto na Figura 2.15.

Figura 2.15: Funcionamento da cycleGAN. (a) Mapeamento $G : X \rightarrow Y$ e $F : Y \rightarrow X$ e seus respectivos discriminadores D_X e D_Y . D_Y força G a gerar imagens que não podem ser distinguidas das imagens do domínio Y . D_X força F a gerar imagens que não podem ser diferenciadas das imagens pertencentes ao domínio X . (b) *Forward cycle-consistency loss*: $x \rightarrow G(x) \rightarrow F(G(x)) \approx x$. (c) *Backward cycle-consistency loss*: $y \rightarrow F(y) \rightarrow G(F(y)) \approx y$.



Fonte: Zhu et al. (2017).

Assim como na *Pix2Pix*, as arquiteturas escolhidas para as redes geradoras e discriminadoras são, respectivamente, a *U-Net* e *PatchGAN*. Como pode ser visto na Figura 2.15, a função de erro utilizada para medir a diferença entre a imagem do domínio de origem com a imagem produzida pelo segundo gerador (x e $F(G(x))$, por exemplo) é chamada de *Cycle Consistency Loss* (Erro de Consistência do Ciclo). Consistência de Ciclo é um conceito de medida de perda utilizado em tradução de máquina. Ao se traduzir uma sentença do idioma A para o idioma B e, posteriormente, traduzir de volta do idioma B para o idioma A . A sentença gerada ao final do processo deve ser idêntica à inicial, no caso ideal (ZHU et al., 2017).

Apesar de preencher a lacuna relacionada a pares de imagens em que os *pixels* não possuem correspondência, o desempenho apresentado pela *cycleGAN* com imagens que possuem *pixels* alinhados se mostram inferiores ao da *Pix2Pix* (ZHU et al., 2017) (KAJI; KIDA, 2019).

2.6 Pesquisas Bibliográficas sobre Trabalhos relacionados - Estado da Arte

Existem poucos trabalhos na literatura relacionados à estimativa de imagens NIR a partir de imagens de câmeras RGB, sendo que o método mais comum envolve alterações no *hardware* das câmeras, como os trabalhos de Lu et al. (2009) e Rabatel, Gorretta e Labbé (2011).

Câmeras fotográficas comuns são inerentemente sensíveis à banda NIR do espectro luminoso, porém possuem um filtro bloqueador da banda NIR na frente do sensor para impedir que o NIR contamine a imagem visível. Em seu trabalho, Lu et al. (2009) descreve um método para obtenção de imagens no espectro visível e NIR simultaneamente usando um único sensor. O trabalho consistiu em desenvolver um novo modelo de câmera, sem o filtro bloqueador de NIR, um novo algoritmo de "demosaicamento" (*demosaiicing*, em inglês) e com um novo *Color Filter Array* (CFA).

Um CFA é um arranjo composto de um mosaico de um material colorido translúcido que normalmente só permite a passagem de uma porção do espectro, sendo porcentagens de vermelho (R), verde (G) e azul (B) para cada *pixel*. O novo CFA combina as porções do espectro visível (R, G e B) em uma única imagem, mais uma imagem de intensidades NIR como a leitura resultante do sensor, que, a partir do algoritmo de "demosaicamento", produz imagens no espectro visível e NIR de alta resolução. Utilizando o *Minimum Mean Square Error* (MMSE), foi definido o método de construção da matriz de demosaicamento otimizado, sendo esta utilizada para desenvolver o novo CFA.

O objetivo do trabalho realizado por Rabatel, Gorretta e Labbé (2011) foi o de extrair uma imagem do índice de vegetação NDVI a partir de uma câmera fotográfica comum, obtendo imagens NIR de alta resolução. Para que o objetivo fosse atingido, o primeiro passo foi de capturar imagens nos comprimentos de onda na região do NIR. A alternativa encontrada para a captura desses comprimentos de onda foi a retirada de um dos filtros que bloqueiam o NIR e a adição de um filtro passa-baixa para bloquear um dos canais RGB (no caso, o canal B), permitindo obter apenas o NIR.

O filtro passa-baixa considerado ideal seria aquele que fizesse com que a sensibilidade espectral do NIR (obtida de imagem hiperespectral) se tornasse mais similar à sua projeção ortogonal. O método para análise da similaridade espectral entre a sensibilidade espectral do NIR e a sua projeção ortogonal foi o SAM.

Os resultados dos cálculos do NDVI obtidos foram comparados com valores de referência, mostrando que foi possível obter bandas com qualidade para o cálculo deste índice com alta

resolução espacial e com custo menor do que os sensores multiespectrais, porém, a remoção dos filtros que bloqueiam o NIR pode ser uma tarefa delicada dependendo do modelo da câmera, correndo o risco de danificar o equipamento.

Por meio de *software*, Reiter et al. (2006) desenvolveram um método 3D de reconstrução facial e previsão de texturas faciais no espectro NIR utilizando regressão múltipla baseado na ideia de que o relacionamento da profundidade e da aparência das faces podem ser modelados com um pequeno número de fatores pares, como, por exemplo, características lineares correlacionadas no espaço de profundidade das imagens e imagens coloridas. A estimativa da imagem de texturas faciais no NIR foi resultado de uma aplicação criada para demonstrar a aplicação do método proposto, usando Análise de Correlação Canônica (*Canonical Correlation Analysis* - CCA) como método de regressão para gerar a imagem NIR a partir de uma imagem colorida, obtida por uma câmera comum.

A CCA é uma ferramenta adequada para relacionar dois conjuntos de medidas, assim, os arranjos de imagens RGB e os mapas de profundidade foram considerados como os dois vetores aleatórios, cada um com sua dimensão, sendo chamados de variações canônicas. Para estimar as imagens NIR, o CCA foi utilizado com imagens NIR "sem forma" e imagens RGB obtidos a partir do uso de *Active Appearance Models* (COOTES; EDWARDS; TAYLOR, 2001). Os resultados da estimativa da imagem NIR com CCA foram levemente melhores do que os resultados obtidos com um método de regressão múltipla comum.

Outra proposta para se estimar uma imagem NIR foi feita por Arai et al. (2016), utilizando informações extraídas de imagens no espectro visível ao invés de realizar alterações no hardware de sensores. O objetivo foi estimar uma imagem NIR para avaliar danos causados por insetos em plantações de arroz da cidade de Saga (Japão), utilizando o índice de vegetação NDVI. Uma informação de grande importância é a existência de alta correlação entre a reflectância do NIR e a reflectância do canal verde.

Imagens RGB de campos de arroz, como as da Figura 2.16 foram capturadas por um drone em dias e horários diferentes, assim como as imagens obtidas por uma câmera NIR, porém, as imagens foram captadas em direções diferentes, como pode ser visto na Figura 2.17. As imagens foram divididas em dois conjuntos: áreas normais e áreas danificadas.

Foram analisadas as reflectâncias espectrais para os dois conjuntos definidos e concluiu-se que em regiões danificadas, os valores de reflectância para os comprimentos de onda NIR são inferiores aos valores dos comprimentos de onda NIR das regiões consideradas normais. Utilizando os valores das intensidades dos canais R,G,B e NIR, aparentemente valores médios para as imagens de cada canal, foi feita uma análise regressiva com esses dados, cujo resultado

Figura 2.16: Exemplo de imagem RGB de um campo de arroz captada por câmera acoplada a um drone.



Fonte: Arai et al. (2016).

Figura 2.17: Exemplo de imagem NIR de um campo de arroz captada por câmera NIR.



Fonte: Arai et al. (2016).

mostrou que existe correlação entre a banda G e a NIR, com coeficiente de determinação maior que 0,94. Sendo assim, concluiu-se que é possível estimar o NIR utilizando o canal G da imagem RGB. O artigo não apresenta exemplos de imagens estimadas e não há correspondência entre os *pixels* das imagens RGB e NIR utilizadas nos experimentos, além de serem capturadas em dias diferentes, com diferentes condições climáticas e de iluminação, não possibilitando realizar comparações com os métodos descritos nesta tese.

No trabalho de Han, Kim e Song (2017), foi proposto um método baseado em CNN para melhorar a resolução de imagens NIR utilizando, simultaneamente, imagens NIR de baixa resolução e imagens de alta resolução no espectro visível capturadas sob baixas condições de iluminação. O método foi proposto pois imagens NIR capturadas em condições precárias de iluminação possuem baixa resolução. Devido a limitações no sensor, são imagens mais claras e possuem as bordas borradas. Já as imagens no espectro visível são mais escuras e ruidosas, porém, possuem resolução espacial maior e melhor definição. O algoritmo proposto consiste de três etapas. A primeira é a extração de imagens de alta frequência a partir das imagens de entrada (NIR e visível). A segunda etapa é relacionada à inferência com o uso de uma CNN,

composta de camadas de concatenação e convolução, onde as imagens obtidas na etapa anterior são combinadas. A última etapa consiste na geração da imagem NIR de alta resolução, adicionando a imagem de saída da rede à imagem NIR de baixa resolução. Todas as imagens utilizadas para o experimento (utilizadas previamente em outros trabalhos da literatura) foram capturadas por câmeras multiespectrais, sendo que os pares visível/NIR possuem a mesma resolução espacial. As imagens NIR de baixa resolução são obtidas a partir de transformações nas imagens NIR de alta resolução e as imagens no espectro visível utilizadas sem nenhuma transformação. Os resultados obtidos mostraram que o método proposto superaram os resultados do estado da arte.

Um trabalho interessante a ser citado faz o caminho inverso, no qual o objetivo é colorir imagens NIR utilizando técnicas de Deep Learning, mais especificamente, uma CNN (LIMMER; LENSCH, 2016). Feixes de luz infravermelho podem ser usados para iluminar cenários com condições de iluminação baixa, porém, os sensores que captam esses feixes de luz produzem imagens monocromáticas, tornando difícil a interpretação de usuários de sistemas avançados de assistência ao condutor, uma vez que é difícil em situações de ausência de cor. No trabalho foi proposta uma CNN de múltipla escala para estimar os valores de baixa frequência dos canais R, G e B a partir da imagem NIR monocromática. Foi realizado um pré processamento em duas etapas: a primeira era obter uma pirâmide de imagens com n níveis, sendo que em cada nível a escala da imagem é reduzida por um fator de 0.5. A segunda etapa era normalizar todos os níveis da pirâmide de imagens utilizando como unidade a média da variância, produzindo uma imagem de textura aprimorada. Após a inferência das cores da imagem pela CNN, utilizando a pirâmide como parâmetro de entrada, a imagem resultante foi submetida a um filtro bilateral para a remoção de ruídos e, posteriormente, os detalhes da imagem original foram transferidos para a imagem colorida utilizando a imagem de textura aprimorada obtida no pré processamento. Segundo os autores, o método falhou apenas em colorir objetos artificiais corretamente, porque a aparência deles não se correlaciona com uma cor específica.

2.7 Considerações Finais

Neste capítulo foi descrita a teoria básica que será utilizada para desenvolver todos os experimentos do presente trabalho, assim como uma pesquisa bibliográfica sobre os trabalhos relacionados a esta temática. No capítulo seguinte serão descritos os métodos utilizados para cumprir os objetivos propostos, mostrando as etapas de processamento e a descrição detalhada dos experimentos desenvolvidos.

Capítulo 3

MATERIAIS E MÉTODOS

Neste capítulo serão mostradas as etapas necessárias para a realização dos experimentos, bem como os materiais e métodos utilizados, de modo a atingir os objetivos propostos no trabalho de pesquisa de doutorado.

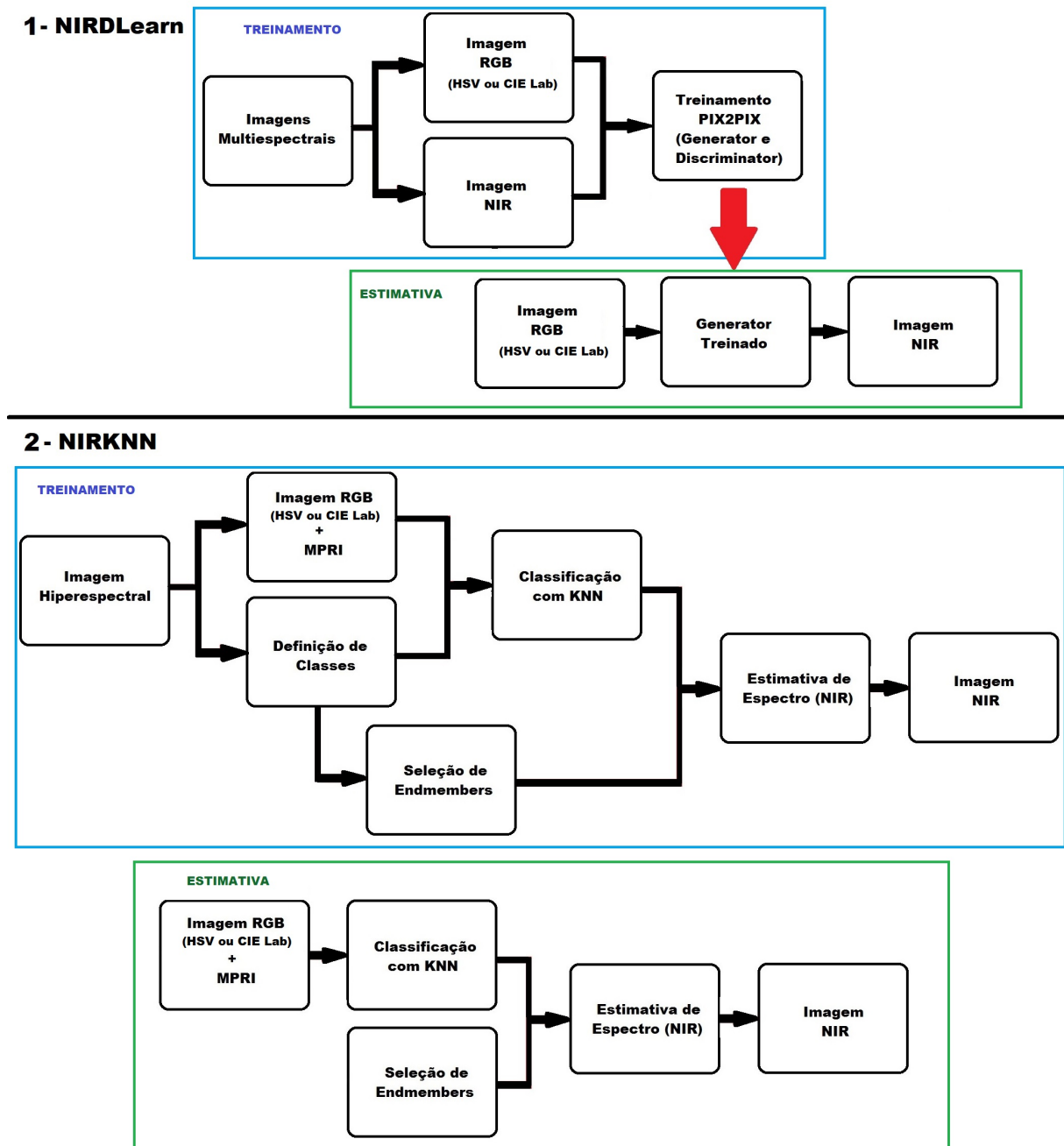
3.1 Descrição dos métodos propostos para a tese

Nesta tese são propostos dois diferentes métodos para estimar imagens NIR a partir de imagens RGB: um método envolvendo *Deep Learning (Pix2Pix)* com imagens multiespectrais, denotado NIRDLearn; e outro método utilizando o classificador KNN e uma equação para estimativa da imagem NIR utilizando espectros de referência extraídos de imagens hiperespectrais, denotado NIRKNN. A Figura 3.1 mostra um diagrama de blocos que exemplifica como estão divididas as etapas que constituem o método proposto.

O método NIRDLearn, referente à parte 1 da Figura 3.1, utiliza como entrada para treinamento das redes de geração e discriminação da *Pix2Pix*, imagens multiespectrais, de onde são extraídas as imagens RGB e NIR. Utilizando-se da rede de geração treinada, cujo treinamento é mostrado pelos blocos superiores da figura, é então possível obter uma imagem NIR estimada a partir de uma entrada de imagem RGB, conforme mostrado pelos blocos inferiores.

O método NIRKNN, referente à parte 2 da Figura 3.1, refere-se ao uso do classificador KNN. Inicialmente, uma imagem hiperespectral é utilizada como entrada para a obtenção das imagens RGB e o índice MPRI. Essas imagens hiperespectrais também possibilitam a definição de classes de assinaturas de *pixels* para a classificação pelo KNN, conforme mostram os blocos superiores do diagrama. Com a classificação feita pelo KNN e os *endmembers* selecionados, é possível realizar a estimativa do espectro NIR e a obtenção da imagem NIR. Essa descrição

Figura 3.1: Diagrama de Blocos dos métodos propostos nesta tese. A parte 1 se refere ao método que usa *Deep Learning*. A parte número 2 se refere ao método utilizando a estimativa a partir de *Endmembers* e classificação por KNN.



Fonte: Elaborada pelo autor.

refere-se à etapa de treinamento do método NIRKNN, portanto, foi necessário o uso da imagem hiperespectral como entrada. Após o treinamento, as imagens NIR são estimadas apenas com as entradas da imagem RGB e os *endmembers* selecionados.

Os métodos propostos estão descritos nas seções 3.2 e 3.3. Na seção 3.2 será descrito o método NIRDLearn para estimativa de imagens NIR utilizando a arquitetura de aprendizado

profundo *Pix2Pix*, sendo descrita na subseção o conjunto de imagens multiespectrais e o pré processamento das mesmas antes do treinamento da rede. Na subseção 3.2.2 serão descritos os detalhes dos módulos Gerador e Discriminador da rede *Pix2Pix* e na subseção 3.2.3, serão mostradas as funções de custo alternativas à função de custo utilizada no trabalho de Isola et al. (2017). Na seção 3.3 será descrito o método NIRKNN de estimativa de imagens NIR a partir de dados hiperespectrais. Na subseção 3.3.1 são descritas as imagens hiperespectrais utilizadas e métodos utilizados para manipulação das mesmas. Na subseção 3.3.2 são apresentados os métodos utilizados para coletar espectros (assinatura espectral de cada *pixel*), definição de classes e seleção de *endmembers* que serão utilizados para atingir os objetivos propostos. Na Subseção 3.3.3 serão mostradas as etapas que envolvem a estimativa da imagem NIR, como a classificação das informações do espectro visível com o classificador KNN nas classes que estão na imagem e, por fim, o cálculo utilizado para gerar a imagem NIR a partir dos *endmembers*.

Os métodos propostos neste trabalho foram desenvolvidos na linguagem de programação Python, utilizando os *frameworks* OpenCV¹ (conhecida biblioteca que possui vários algoritmos de visão computacional), KERAS², TensorFlow³ (bibliotecas para desenvolvimento de redes neurais e *Deep Learning*), Spectral Python (SPy)⁴ e PySptools⁵ (bibliotecas que possuem implementações de métodos conhecidos para processamento de dados hiperespectrais), além da ferramenta de mineração de dados Weka (FRANK; HALL; WITTEN, 2016).

3.2 NIRDLearn - Estimativa de imagens NIR com *rede Deep Learning Pix2Pix*

Nesta seção serão descritos os materiais e métodos utilizados nos experimentos de estimativa de imagens NIR com o modelo de rede *Deep Learning Pix2Pix* adaptado à aplicação desejada, NIRDLearn.

3.2.1 Imagens Multiespectrais utilizadas

Para a realização de experimentos relativos ao NIRDLearn foi necessário o uso de dados de *Ground Truth* para avaliar o desempenho da classificação, bem como o próprio aprendizado. O Laboratório de Imagens da Embrapa Instrumentação forneceu conjuntos de imagens aéreas

¹<http://opencv.org/>

²<http://keras.io>

³<http://www.tensorflow.org>

⁴<http://www.spectralpython.net>

⁵<https://pysptools.sourceforge.io>

de regiões agrícolas para que os experimentos de estimativa de imagens NIR fossem feitos. As imagens foram capturadas por uma câmera Micasense RedEdgeTM (Figura 3.2) acoplada em um *drone*. As imagens foram capturadas entre janeiro e julho de 2018, abrangendo plantações de milho, soja, cana de açúcar e algodão em diferentes condições atmosféricas.

Figura 3.2: Câmera Micasense RedEdgeTM utilizada na aquisição das imagens.



Fonte: Elaborada pelo autor.

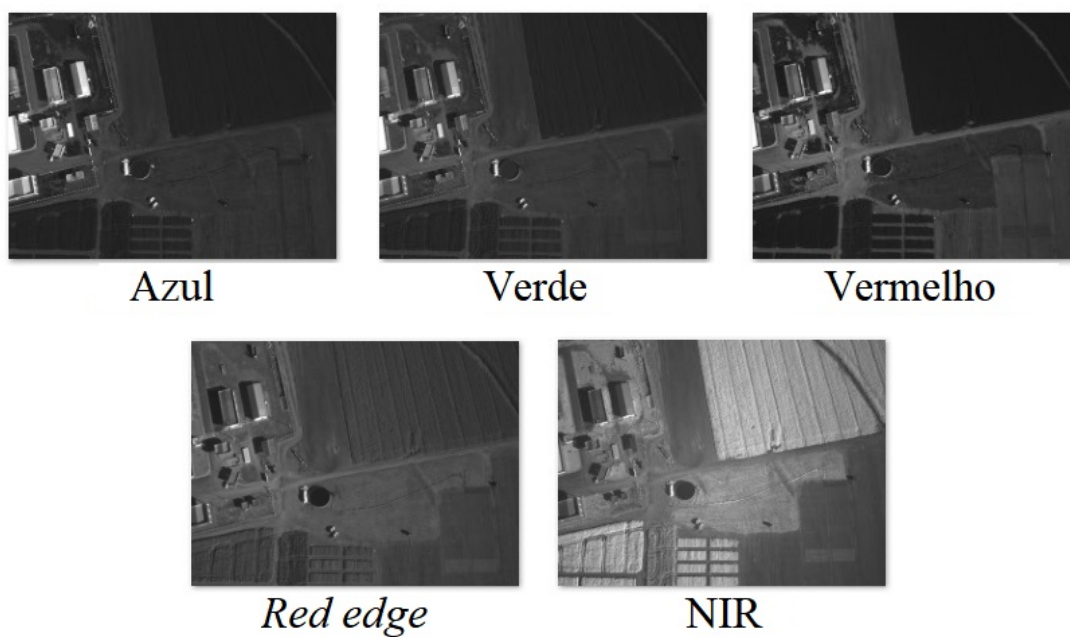
Como pode ser observado na Figura 3.2, a câmera utilizada possui cinco entradas de sensores, que capturam cada uma das três bandas espectrais visíveis e mais duas bandas não visíveis: Azul, Verde, Vermelho, *Red edge* e NIR. Na Tabela 3.1 são mostradas as informações sobre as regiões do comprimento de onda capturadas para cada banda.

Tabela 3.1: Bandas espectrais capturadas pela câmera RedEdgeTM.

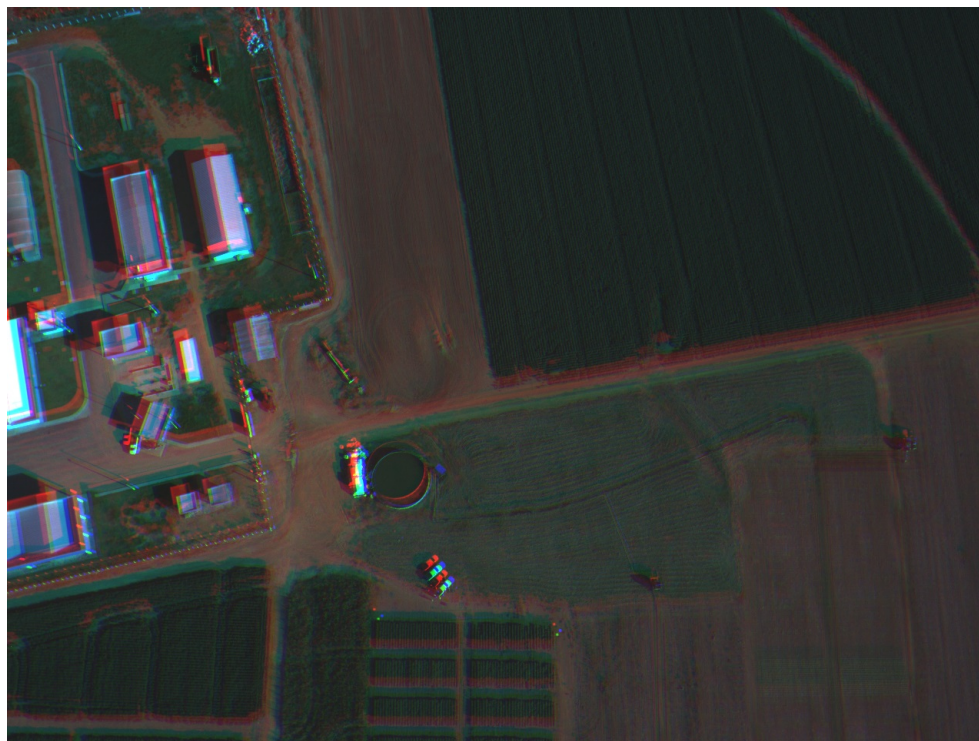
Banda	Comprimento de onda central (nm)	Largura à meia altura (nm)
Azul	475	20
Verde	560	20
Vermelho	668	10
<i>Red edge</i>	717	10
NIR	840	40

Na Figura 3.3 são mostradas imagens obtidas por cada sensor, onde é possível notar que, assim como existe uma distância entre os sensores na câmera e possíveis atrasos entre os mesmos no momento da captura, as imagens estão deslocadas entre si, não sendo possível uma simples combinação entre as imagens Vermelho, Verde e Azul em uma matriz tridimensional para gerar uma imagem RGB, conforme pode ser visto na Figura 3.4.

As imagens precisam passar por processos de correção radiométrica, utilizando imagens capturadas de um painel de calibração que acompanha a câmera (Figura 3.5); e um procedimento para alinhar as imagens. Foi utilizado o *software* Agisoft Metashape para a realização

Figura 3.3: Exemplo de imagem capturada por cada sensor.

Fonte: Elaborada pelo autor.

Figura 3.4: Tentativa de combinar imagens Vermelho, Verde e Azul em uma matriz tridimensional.

Fonte: Elaborada pelo autor.

dessas tarefas e para a geração de ortomosaicos, os quais foram utilizados para extração de amostras de treinamento para a rede *Pix2Pix*. Um ortomosaico refere-se a um mosaico com-

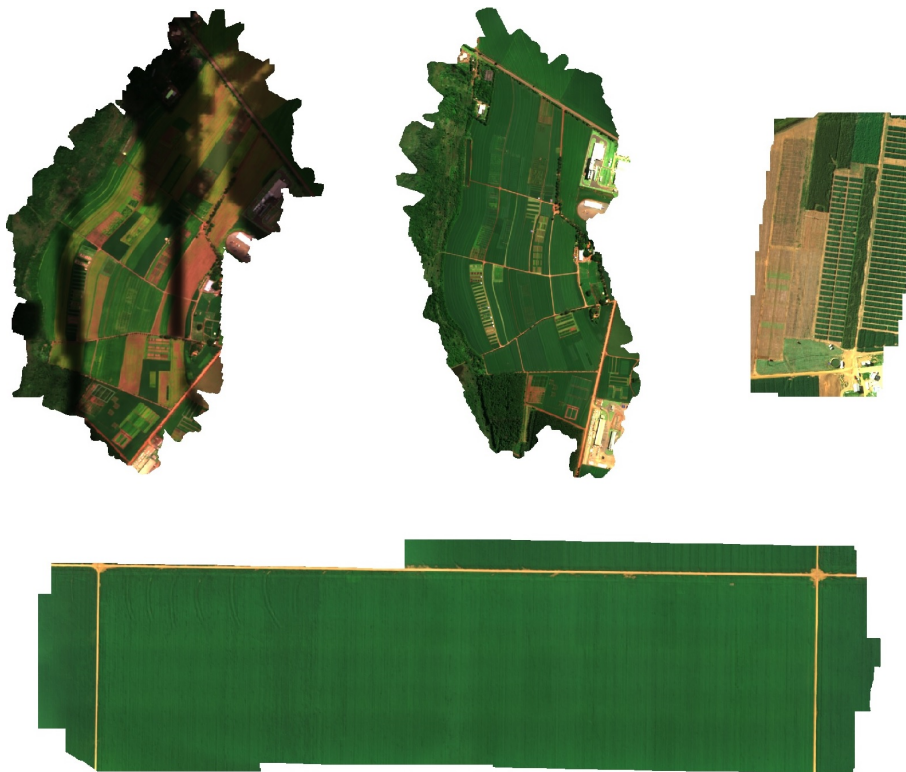
posto de várias fotos, cujas distorções foram corrigidas. Os ortomosaicos gerados possuem resolução de 8 cm quadrados por *pixel* e podem ser visualizados na Figura 3.6.

Figura 3.5: Painel de calibração radiométrica da câmera. As primeiras imagens capturadas pela câmera antes do vôo devem ser deste painel sob a luz solar.



Fonte: Elaborada pelo autor.

Figura 3.6: Ortomosaicos gerados a partir do conjunto de imagens multiespectrais fornecidos pela Embrapa Instrumentação.

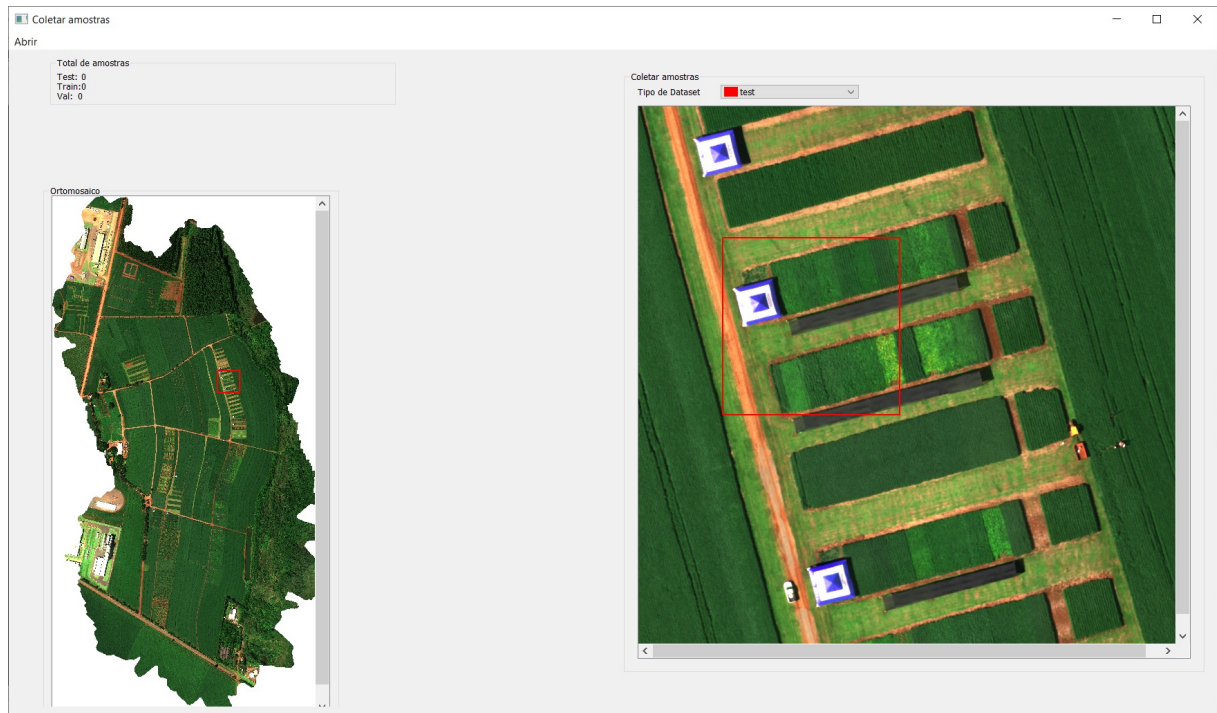


Fonte: Elaborada pelo autor.

A partir dos ortomosaicos da Figura 3.6 e do auxílio de uma ferramenta desenvolvida em

Python (Figura 3.7), foram extraídos 1520 pares de imagens RGB-NIR de 256×256 pixels, separados aleatoriamente em dois grupos, treinamento (1080) e validação do método (440).

Figura 3.7: Ilustração da ferramenta para coleta de amostras de treinamento e validação desenvolvida.



Fonte: Elaborada pelo autor.

Na Figura 3.7, parte esquerda, é mostrada uma miniatura do ortomosaico, onde é possível selecionar regiões de 800×800 pixels, exibidas em tamanho ampliado na parte direita da figura. Nesta parte, são selecionadas as amostras de 256×256 pixels (quadrado vermelho) que serão utilizadas para o treinamento ou validação do método.

3.2.2 Arquitetura da rede Pix2Pix

Na seção 2.5.1 foi descrito, de maneira geral, o conceito e funcionamento da rede Pix2Pix. Nesta seção serão descritas de forma mais detalhada as arquiteturas das redes componentes da Pix2Pix, os módulos Gerador (G) e Discriminador (D).

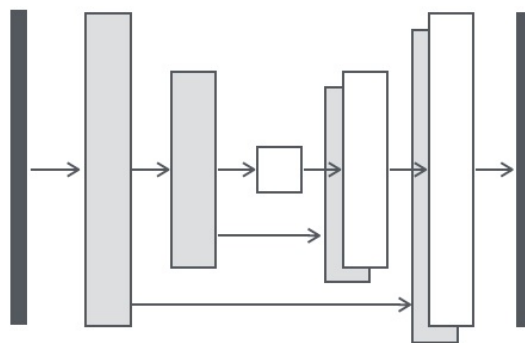
3.2.2.1 Gerador

Como descrito anteriormente, para G foi utilizada uma rede de arquitetura do tipo U-Net (RONNEBERGER; FISCHER; BROX, 2015). A U-Net é uma arquitetura de redes convolucionais mais comumente utilizada em tarefas de segmentação de imagens. Segundo Isola et

al. (2017), existe muita informação de baixo nível que pode ser compartilhada entre a entrada e a saída de uma rede para geração de imagens a partir de uma referência, sendo importante propagá-la através da rede. No caso da estimativa de imagens NIR, as imagens RGB de entrada e as imagens de saída da rede compartilham a mesma informação de textura e localização de bordas.

Na rede em questão, a imagem de entrada é transmitida pela rede através de uma série de camadas, sendo que em cada camada, a imagem é compactada por meio de convoluções em vetores de menor dimensão, até que seja atingida uma camada que contém o menor número de nós, chamada *Bottleneck layer*. A partir desta camada, o processo é revertido por meio de deconvoluções para que a saída da rede tenha as mesmas dimensões da imagem de entrada. Uma particularidade dessa arquitetura é a existência de conexões adicionais entre camadas diferentes da rede, sendo que normalmente a saída de uma camada é a entrada da outra subsequente. Essas conexões são chamadas de *skip connections*, pois "pulam" camadas que estão em sequência para se conectar às camadas mais distantes. Sendo assim, existem *skip connections* entre camadas i (compactam informação) e camadas $n - i$ (restauram informação), onde n é o número total de camadas da rede. A Figura 3.8 mostra um exemplo das conexões entre as camadas da rede.

Figura 3.8: Exemplo de arquitetura *U-Net*.



Fonte: Isola et al. (2017).

Cada camada da rede é composta por operações de Convolução, *Batch Normalization* e de retificação linear, mais especificamente, *Rectified Linear Unit* (ReLU). A função das convoluções foi descrita na Seção 2.5. Proposta por Ioffe e Szegedy (2015), *Batch Normalization*, ou normalização em lotes, é uma técnica utilizada para acelerar o treinamento de redes de aprendizado profundo, melhorar o desempenho e a estabilidade das redes. Esta melhora ocorre a partir do ajuste das médias e variâncias das camadas de entrada. Ao fazer isto, é reduzido deslocamento covariado interno (*internal covariate shift*), um problema decorrente da mudança de parâmetros que pode ocorrer entre as saídas e entradas (entradas podem ser "desnormalizadas")

nas camadas ocultas da rede durante o treinamento.

A ReLU é uma função de ativação não linear, porém, muito próxima a linear. É a função mais amplamente recomendada em projetos de redes neurais. Segundo Goodfellow, Bengio e Courville (2016), esta função preserva muitas propriedades que tornam modelos lineares fáceis de otimizar. A função ReLU retorna 0 se o valor do argumento de ativação x (somatório da multiplicação das entradas da camada por seus respectivos pesos) for menor ou igual a 0, caso contrário, a saída da função é o próprio valor do argumento. Apesar da vantagem, transformar valores menores do que 0 em 0 pode "matar" neurônios da rede durante o treinamento, fazendo com que os pesos desses neurônios não sejam atualizados. A representação matemática da função ReLU pode ser vista na equação 3.1.

$$R(x) = \max(0, x) \quad (3.1)$$

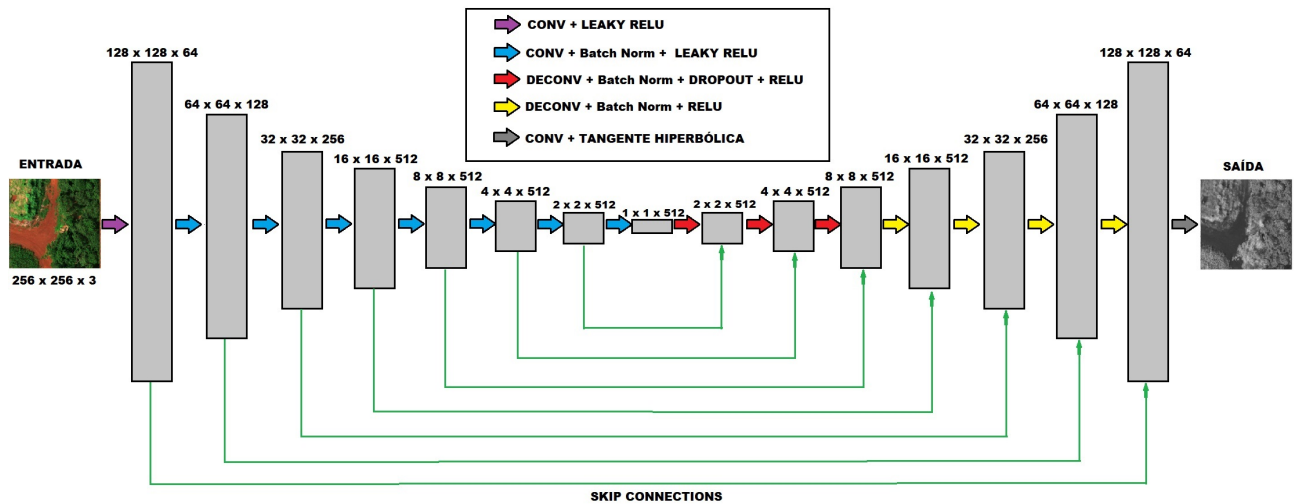
Enquanto a função ReLU foi utilizada para algumas camadas da rede, uma outra função foi utilizada para outras camadas. Essa função é uma versão aprimorada da ReLU, chamada *Leaky ReLU* (MAAS; HANNUN; NG, 2013), que ao invés de retornar 0 para os argumentos menores ou iguais a zero, é inserida uma pequena inclinação α ($\alpha = 0,2$, na *Pix2Pix*) na parte negativa do domínio, evitando assim o problema apresentado pela função ReLU. A função *Leaky ReLU* é representada pela equação 3.2.

$$R_{Leaky}(x) = \begin{cases} x, & \text{se } x > 0 \\ \alpha x, & \text{se } x \leq 0 \end{cases} \quad (3.2)$$

O modelo da *U-Net* proposto por Isola et al. (2017) é composto por 16 camadas, sendo 8 camadas de convolução e 8 camadas de deconvolução. A cada camada da rede, foram utilizados filtros convolucionais de 4×4 com movimento de 2 saltos (*strides*). Para cada convolução, o *feature map* (obtido a partir da imagem de entrada após a primeira camada convolucional) é redimensionado para um tamanho 2 vezes menor do que o da camada anterior. Para cada deconvolução, o *feature map* é redimensionado para um tamanho 2 vezes maior que o da camada anterior. Nas três primeiras camadas deconvolucionais, são adicionadas técnicas de *Dropout* de 50%, ou seja, 50% dos neurônios destas camadas são aleatoriamente desativados a cada mini-lote (*mini batch*) de treinamento. O objetivo do *dropout* é tornar a rede mais robusta, pois é como se várias redes diferentes fossem treinadas, compartilhando os mesmos recursos e se adaptando de forma diferente para realizar classificações corretas (HINTON et al., 2012). No final da última camada de deconvolução, é realizada uma convolução para mapear o *feature*

map para o número de canais da imagem de saída, seguindo uma função de ativação tangente hiperbólica. O modelo final da rede G é mostrado na Figura 3.9.

Figura 3.9: Arquitetura da rede G. Cada coluna representa um *feature map*, com as dimensões e o número de filtros descritos acima (linhas x colunas x filtros).



Fonte: Elaborada pelo autor.

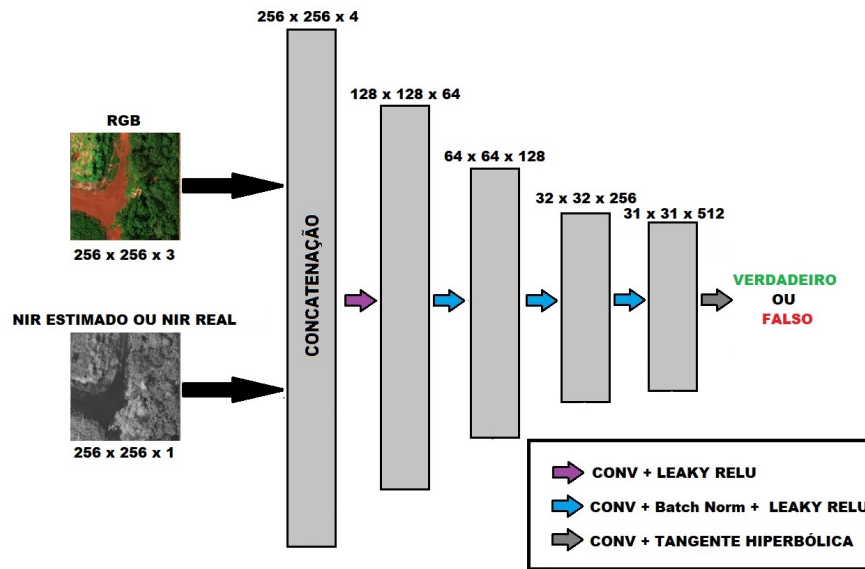
3.2.2.2 Discriminador

Para a rede D é utilizado um classificador convolucional chamado de *PatchGAN*, sendo o funcionamento deste descrito na seção 2.5.1. O modelo é composto por 4 camadas convolucionais, cada uma delas seguindo as operações de Convolução, *Batch Normalization* e ativação *Leaky ReLU* (equação 3.2, com $\alpha = 0,2$), com exceção da primeira camada convolucional, assim como feito para a porção convolucional da rede G. Após a última camada, uma convolução é aplicada para que a rede tenha uma saída unidimensional e posteriormente, esta saída é enviada para uma função sigmoideal. Para os experimentos de estimativa de imagens NIR foram utilizados *patches* de 64×64 *pixels*. O modelo da *PatchGAN* é mostrado na Figura 3.10.

Como pode ser observado na Figura 3.10, a rede D tem dois pares de imagens como parâmetros de entrada, que são concatenadas. A rede D olha para o par RGB/NIR real (R) e para o par RGB/NIR estimado (E), fazendo comparações entre eles a fim de julgar se o par E é verdadeiro ou falso. Após a etapa de julgamento, são calculados os erros em relação à informação de referência para cada par (R - Real, E - Estimado). Os pesos são então ajustados de acordo com o erro de classificação dos pares R e E. A Figura 3.11 mostra o exemplo do ajuste dos pesos da rede D.

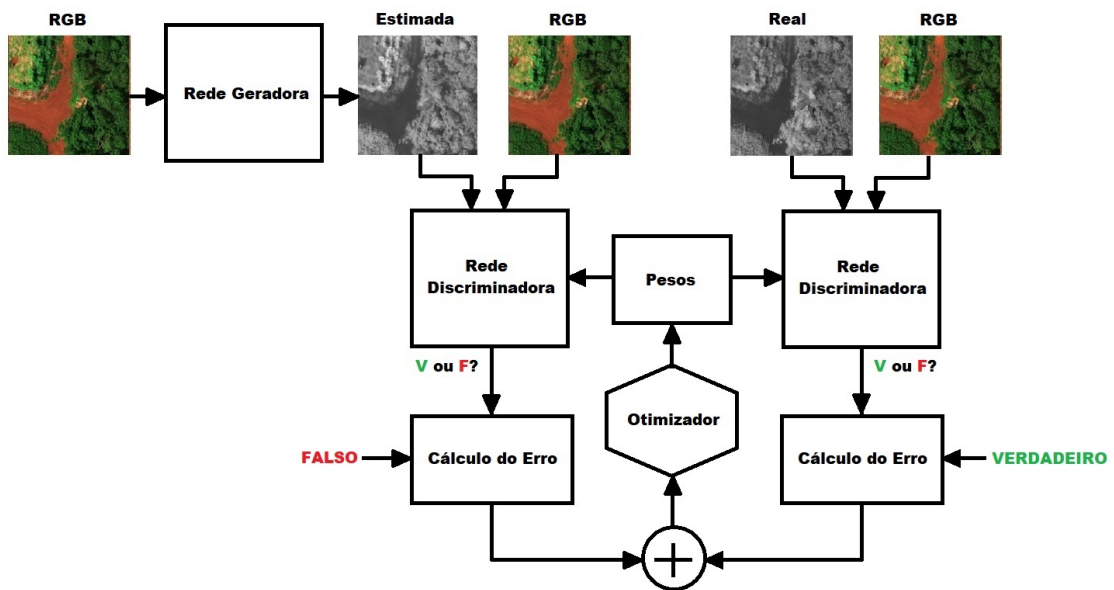
Como dito na Seção 2.5.1, os pesos da rede G são atualizados a partir da atualização da rede D. Depois de gerada uma imagem sintética, esta é enviada à rede D junto com a imagem RGB.

Figura 3.10: Arquitetura da rede D (PatchGAN).



Fonte: Elaborada pelo autor.

Figura 3.11: Atualização de pesos da rede D.

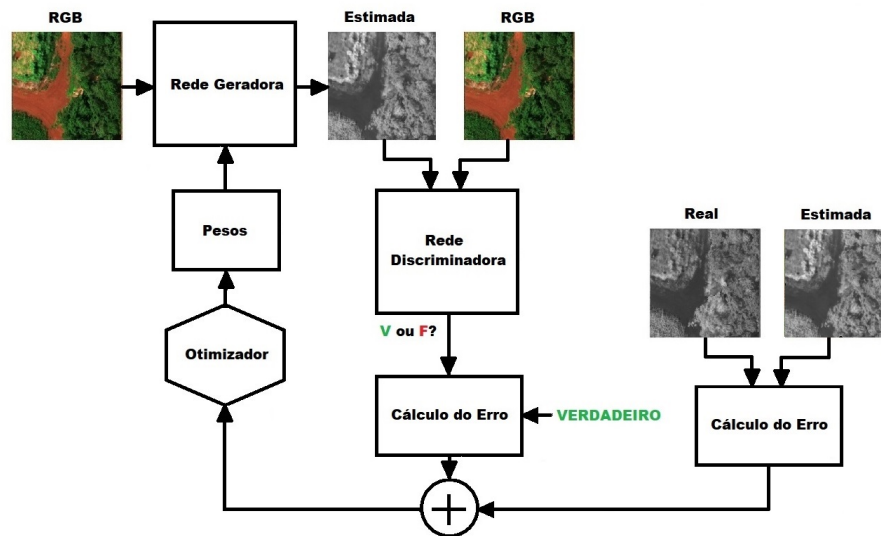


Fonte: Elaborada pelo autor.

É calculado o erro deste par em relação à referência, aferindo como a imagem sintética pode ser classificada (Verdadeiro ou Falso), assim como pode ser visto na metade esquerda da Figura 3.11. Na metade direita da Figura 3.11, a imagem real e a imagem RGB são utilizadas como entrada da rede D para calcular o erro entre o par de imagens que deve resultar na aferição da imagem real como verdadeira. Os dois erros, tanto do par de imagens (estimado, RGB), como do par (real, RGB), são adicionados e otimizados para a atualização dos pesos da rede D, como mostram os blocos centrais da Figura 3.11.

É também calculado o erro entre a imagem real e a imagem sintética como mostrado na parte direita da Figura 3.12, para assim realizar a atualização dos pesos da rede G, juntamente com o resultado do cálculo de erro entre a imagem estimada e RGB, mostrado na parte central da figura. A atualização dos pesos da rede G passa por um otimizador em função dos dois erros somados, como mostrado nos blocos da parte esquerda da figura.

Figura 3.12: Atualização de pesos da rede G.



Fonte: Elaborada pelo autor.

Tanto nas Figuras 3.11 e 3.12 é possível notar a presença de um módulo otimizador. Os otimizadores são os responsáveis por maximizar ou minimizar as funções de custo (*loss*) (GOODFELLOW; BENGIO; COURVILLE, 2016), neste caso a função objetivo (equação 2.20), para realizar a atualização dos pesos da rede neural. Para os dois módulos, G e D, foi utilizado o otimizador *Adam* (KINGMA; BA, 2014). Este método calcula taxas de aprendizado adaptáveis para diferentes parâmetros a partir das estimativas do primeiro e segundo momento dos gradientes. Os parâmetros de configuração do otimizador são os recomendados como padrão por Kingma e Ba (2014), sendo estes: taxa de aprendizado = 10^{-3} ; $\beta_1 = 0,9$; $\beta_2 = 0,999$ e $\epsilon = 10^{-8}$. Os parâmetros β_1 e β_2 são taxas de decaimento exponencial para as estimativas do primeiro e do segundo momento, respectivamente. O parâmetro ϵ é um valor numérico utilizado para evitar divisões por 0.

Nos experimentos conduzidos para estimar as imagens NIR, foram utilizadas as imagens no padrão RGB e convertidas para os modelos de cor HSV e CIE Lab. Também foram testadas outras funções de custo, de modo a verificar a influência desta no resultado final da estimativa.

3.2.3 Funções de custo

Para que exista aprendizado, é necessário que os erros sejam avaliados e mensurados. No trabalho de Isola et al. (2017), a função utilizada para verificar o erro entre as imagens *ground truths* e as imagens geradas pela rede foi a L1 (equação 2.19). Apesar de produzir resultados com borramento em tarefas de geração de imagens, a função de custo L1 captura com precisão as informações de baixa frequência (ISOLA et al., 2017), deixando o papel de lidar com altas frequências para a rede D.

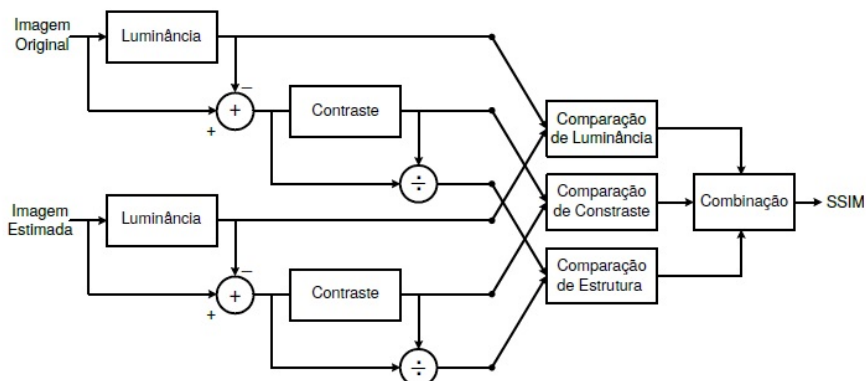
Zhao et al. (2016) fizeram um estudo sobre funções de custo para redes neurais utilizadas para restauração de imagens, verificando o efeito de diferentes métricas para mensurar o custo, dentre elas, a L1. O objetivo era estabelecer alternativas para a métrica L2, conhecida como também como método dos mínimos quadrados (métrica padrão até então), pois esta apresentava resultados com baixa qualidade. Nos testes realizados, a métrica L1 apresentou melhores resultados que a L2, porém, também foram avaliadas duas métricas que apresentaram resultados com qualidade superior aos que utilizaram a métrica L1, as métricas *Structural Similarity* (SSIM) e *Multiscale Structural Similarity* (MS-SSIM). Também foi proposta uma nova métrica, uma combinação entre a L1 e a MS-SSIM, que apresentou desempenho superior a todas as métricas de similaridade consideradas para o estudo.

3.2.3.1 SSIM

Os *pixels* de uma imagem possuem fortes dependências, principalmente quando estão espacialmente próximos. Estas dependências carregam informações importantes sobre a estrutura dos objetos presentes na imagem (WANG et al., 2004). A métrica SSIM avalia as imagens levando em consideração o fato do sistema visual humano ser sensível a mudanças nas estruturas locais, normalizadas pela luminância e pelo contraste. O SSIM é utilizado para medir a similaridade entre imagens, comparando as estruturas da imagem de referência e a imagem modificada (no caso desta tese, a imagem gerada). A Figura 3.13 mostra um diagrama do funcionamento do SSIM, em que dadas as imagens original e estimada à esquerda, os valores de luminância, contraste e estrutura são obtidos, comparados e combinados, para a obtenção da métrica SSIM à direita.

As expressões utilizadas para calcular as comparações de luminância (l), contraste (c) e estrutura (s) são apresentadas nas equações 3.3, 3.4 e 3.5,

$$l(x, y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}, \quad (3.3)$$

Figura 3.13: Diagrama de funcionamento da métrica SSIM.

Fonte: Adaptado de Wang et al. (2004).

$$c(x,y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}, \quad (3.4)$$

$$s(x,y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3}, \quad (3.5)$$

em que x e y são respectivamente as imagens original e estimada, μ_x e μ_y são a média de x e y . As variáveis σ_x e σ_y representam as variâncias de x e y , respectivamente, σ_{xy} representa a covariância entre as duas imagens e C_1 , C_2 e C_3 são pequenas constantes calculadas com base no intervalo dos valores dos *pixels* em imagens, que para 8 bits é 255. O método para calcular o índice SSIM com base na luminância, contraste e estrutura é mostrado na equação 3.6,

$$SSIM(x,y) = [l(x,y)]^\alpha \cdot [c(x,y)]^\beta \cdot [s(x,y)]^\gamma, \quad (3.6)$$

em que α , β e γ são parâmetros que definem a ordem de importância das três componentes. Assumindo-se que $\alpha = \beta = \gamma = 1$ e $C_3 = C_2/2$, tem-se a seguinte forma:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (3.7)$$

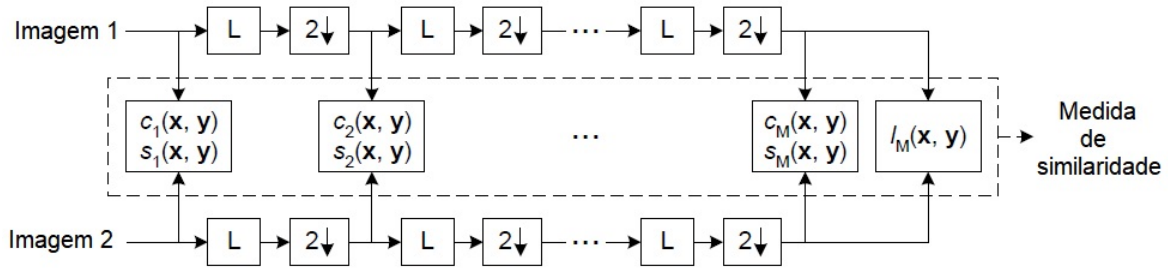
O valor do cálculo varia entre -1 e 1, porém, são mais comuns valores entre 0 e 1. Quanto mais próximo de 1, mais similares são as imagens comparadas. A função de erro utilizando a métrica SSIM é dada pela equação 3.8 (ZHAO et al., 2016).

$$\mathcal{L}_{SSIM}(x,y) = 1 - SSIM(x,y) \quad (3.8)$$

3.2.3.2 MS-SSIM

Segundo Wang, Simoncelli e Bovik (2003), a perceptibilidade dos detalhes em uma imagem depende da densidade de amostragem dos sinais da imagem, da distância do plano da imagem ao observador e da capacidade de percepção do sistema visual do observador, pois a avaliação subjetiva desta imagem varia de acordo com a variação destes fatores. Sendo assim, foi desenvolvido um método para analisar a similaridade estrutural em múltiplas escalas das imagens. O funcionamento do método é mostrado no diagrama da Figura 3.14.

Figura 3.14: Diagrama para o sistema SSIM multi escala. L: filtragem passa-baixa. ↓: escala diminuída por 2.



Fonte: Adaptado de Wang, Simoncelli e Bovik (2003).

Como pode ser observado na Figura 3.14, as imagens são submetidas a M filtragens passa-baixa e reduções na escala, realizando comparações entre os contrastes (equação 3.4) e as estruturas (equação 3.5) em cada escala, sendo que a luminância (equação 3.3) é comparada apenas na menor escala. O cálculo para o MS-SSIM é mostrado na equação 3.9

$$MSSSIM(x, y) = [l_M(x, y)]^{\alpha_M} \prod_{j=1}^M [c_j(x, y)]^{\beta_j} [s_j(x, y)]^{\gamma_j} \quad (3.9)$$

O valor do cálculo do MS-SSIM varia entre -1 e 1, porém, são mais comuns valores entre 0 e 1, assim como no SSIM. A função de erro utilizando a métrica SSIM para múltiplas escalas é dada pela equação 3.10 (ZHAO et al., 2016).

$$\mathcal{L}_{MSSSIM}(x, y) = 1 - MSSSIM(x, y) \quad (3.10)$$

3.2.3.3 MS-SSIM + L1 (Mixed)

Segundo Zhao et al. (2016), MS-SSIM e SSIM não são particularmente sensíveis a uniformidades, podendo levar a mudança no brilho ou cores, tornando-as opaca. Apesar disso, a

MS-SSIM preserva melhor o contraste em regiões de alta frequência do que as outras funções de custo utilizadas no estudo, porém, a função L1 preserva melhor as cores e a luminância. Sendo assim, foi proposta uma nova função de custo de modo a combinar as vantagens da MS-SSIM e da L1, a qual foi chamada de *Mix* pelos autores (equação 3.11),

$$\mathcal{L}_{Mix}(x,y) = \alpha \cdot \mathcal{L}_{MSSSIM} + (1 - \alpha) \cdot \mathcal{L}_{L_1}, \quad (3.11)$$

onde α é um valor de ponderação para as duas funções de *loss*, definindo empiricamente como 0,84; justificando a escolha deste valor como sendo uma maneira de fazer com que a contribuição das duas funções (\mathcal{L}_{MSSSIM} e \mathcal{L}_{L_1}) seja equilibrada (ZHAO et al., 2016).

3.2.3.4 Perda Perceptiva (*Perceptual Loss*)

Em seu trabalho, Johnson, Alahi e Fei-Fei (2016) discorreram sobre o problema existente nas tarefas de transformação de imagens de entrada para outras imagens, enfatizando que uma das formas de se atingir esse objetivo é a partir do uso de redes neurais convolucionais utilizando funções de custo que verificam a diferença entre imagens de saída e *ground truths pixel a pixel* (informações de baixo nível), porém, estas funções não eram capazes de capturar diferenças "perceptivas" (*perceptual*, do inglês) entre estas imagens, uma vez que, ao se considerar imagens idênticas que estão deslocadas a um *pixel* de distância, seriam imagens completamente diferentes ao serem analisadas utilizando a abordagem *pixel a pixel*.

Sendo assim, propuseram utilizar funções de custo baseadas nas diferenças entre representações de características de alto nível da imagem, sendo estas extraídas por redes neurais convolucionais pré treinadas. No caso do trabalho em questão, foi utilizada a décima sexta camada de uma rede convolucional VGGnet (SIMONYAN; ZISSERMAN, 2014) pré treinada com o *dataset* ImageNet (RUSSAKOVSKY et al., 2015).

3.2.3.5 *Perceptual + L1* (P_{Mixed})

De modo análogo ao feito por Zhao et al. (2016), foi proposta para a tese uma combinação da vantagem da *Perceptual Loss* ($\mathcal{L}_{Perceptual}$) e da informação obtida pela função L1, que preserva características importantes ao analisar as diferenças utilizando informações de baixo nível das imagens de saída da rede e as imagens *ground truth*. Substituindo-se a função \mathcal{L}_{MSSSIM} pela função $\mathcal{L}_{Perceptual}$ na Equação 3.11, tem-se a *Perceptual Mixed Loss*, $\mathcal{L}_{PMix}(x,y)$ (Equação 3.12),

$$\mathcal{L}_{PMix}(x,y) = \alpha \cdot \mathcal{L}_{Perceptual} + (1 - \alpha) \cdot \mathcal{L}_{L_1}, \quad (3.12)$$

onde α é um valor de ponderação para as duas funções de perda (*loss*), definido empiricamente como 0,84, assim como na Equação 3.11. De acordo com testes empíricos, alterações do α para valores muito inferiores contribuem negativamente para a qualidade dos resultados obtidos. Já para valores maiores, não há ganho significativo para a imagem de saída da rede.

3.3 NIRKNN - Estimativa de imagens NIR utilizando imagens RGB e classificador KNN

Nesta seção será descrito o método utilizando o classificador KNN, proposto para a estimativa de imagens NIR, NIRKNN, sendo inicialmente apresentadas as imagens hiperespectrais utilizadas para a extração das classes de *pixels*, necessárias para o treinamento do sistema.

3.3.1 Imagens Hiperespectrais utilizadas

Devido a ausência de uma vasta biblioteca gratuita de imagens hiperespectrais, foram utilizadas algumas imagens bastante conhecidas na literatura e uma imagem hiperespectral de uma plantação de cítricos no Chile, capturada no dia 17 de janeiro de 2011 por um sensor HySpex VNIR-1600 acoplado a um avião. Esta imagem capturada pelo sensor acoplado ao avião possui 160 bandas que abrangem do espectro visível ao NIR. A Figura 3.15 mostra a imagem RGB obtida a partir da imagem hiperespectral da plantação de citrus, utilizando as bandas 55 (611 nm), 41 (560,2 nm) e 12 (453,8 nm) para criar, respectivamente, os canais R, G e B da imagem.

Figura 3.15: Imagem RGB da plantação de cítricos no Chile.



Fonte: Elaborada pelo autor.

As imagens hiperespectrais utilizadas em diversos trabalhos da literatura⁶ são as seguintes:

⁶http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_remote_sensing_scenes

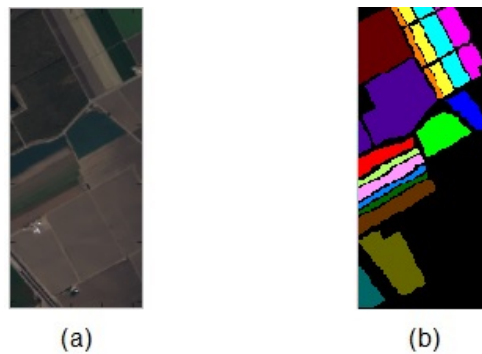
- **Indian Pines:** Esta imagem foi coletada por um sensor hiperespectral AVIRIS (*Airborne Visible/Infrared Imaging Spectrometer*) no dia 12 de junho de 1992 sobre a fazenda de agronomia da Universidade de Purdue (noroeste do estado de Indiana, Estados Unidos) e seus arredores. A área em que a universidade se localiza inclui grandes porções das bacias hidrográficas Indian Creek e Pine Creek, sendo assim, o *data set* ficou conhecido como Indian Pine (BAUMGARDNER; BIEHL; LANDGREBE, 2015). A imagem possui dimensão de 145 x 145 *pixels*, sendo que dois terços da área coberta correspondem a culturas agrícolas e a dimensão espectral de 224 bandas que estão na faixa de comprimento de onda que vai de 400 nm a 2500 nm, porém, bandas relacionadas à absorção de água foram removidas, restando 220 bandas espectrais. Também é disponibilizada uma imagem de *ground truth*, que possui 16 classes relacionadas aos objetos presentes na imagem (vegetação e construções). Um exemplo de imagem do *data set* Indian Pines pode ser visto na Figura 3.16.

Figura 3.16: Indian Pines. (a) Imagem RGB e (b) *Ground truth*.

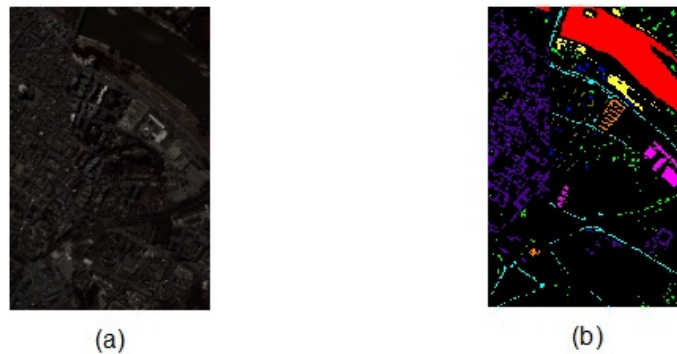


Fonte: Elaborada pelo autor.

- **Salinas:** Outra imagem coletada por um sensor AVIRIS, porém, na região de Salinas Valley, na Califórnia (Estados Unidos), possuindo uma resolução espectral maior do que a imagem anterior e contém informações de vegetação, solo e campos de vinhedo, divididos em 16 classes no *ground truth*. A imagem possui dimensão de 512 x 215 *pixels* e foi capturada com 224 bandas, mas assim como no *data set* Indian Pines, foram removidas as bandas que possuem relação com absorção de água. A Figura 3.17 representa uma imagem do *data set* Salinas e seu respectivo *ground truth*.
- **Pavia Centre:** Esta imagem hiperespectral foi capturada na cidade de Pavia, norte da Itália, utilizando um sensor ROSIS (*Reflective Optics System Imaging Spectrometer*). A imagem possui dimensão de 715 x 1096 *pixels* e 102 bandas espectrais, representando 9 classes, como água, asfalto, solo, pastagens e construções. A Figura 3.18 mostra uma imagem RGB composta com bandas do *data set* da Pavia Centre e seu respectivo *ground truth*.

Figura 3.17: Salinas. (a) Imagem RGB e (b) Ground truth.

Fonte: Elaborada pelo autor.

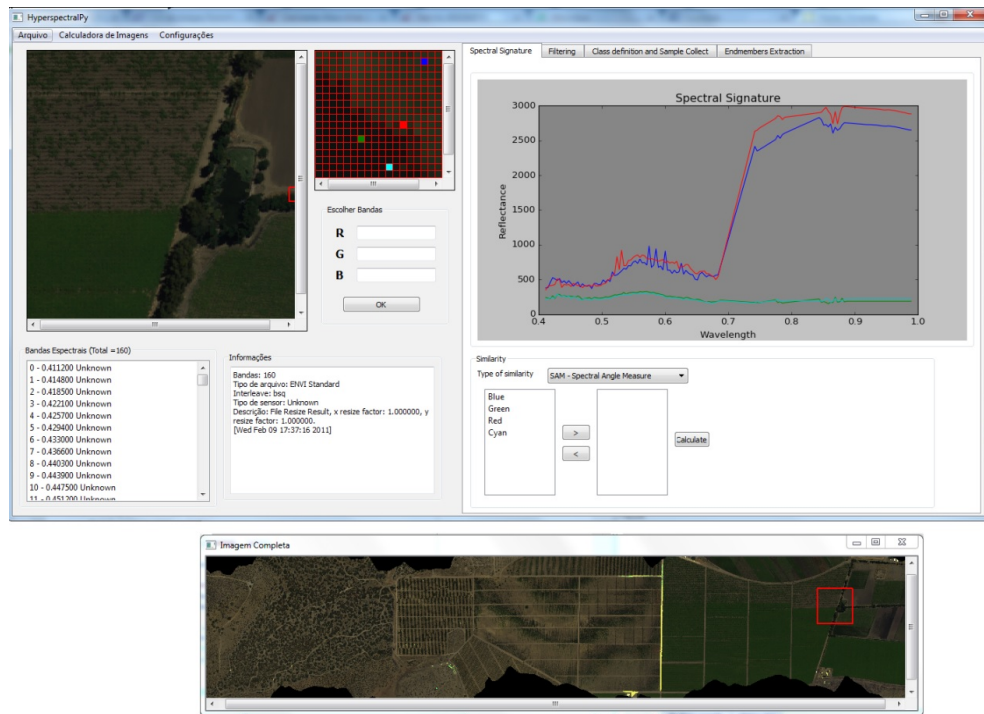
Figura 3.18: Pavia Centre. (a) Imagem RGB e (b) Ground truth.

Fonte: Elaborada pelo autor.

Foi desenvolvida uma ferramenta para visualização e manipulação dos dados das imagens hiperespectrais (Figura 3.19), assim como definição de classes de pertinência para cada *pixel* (assinatura espectral), coleta de assinaturas espectrais, seleção de *endmembers* e classificação utilizando as medidas de similaridade citados na seção 2.2.1.

Na parte inferior da Figura 3.19, é mostrada uma representação RGB da imagem hiperespectral, na qual se pode selecionar regiões de 400×400 *pixels*, as quais serão exibidas em tamanho real no canto superior direito do painel. Essa imagem será usada para selecionar áreas de 20×20 *pixels*, sendo ampliadas na imagem quadriculada. Cada quadrado delimita um *pixel*, que, ao ser selecionado (pintado na área quadriculada), exibe sua respectiva assinatura espectral, no lado direito do painel. Os campos descritivos do painel contém informações sobre as bandas (parte inferior esquerda), metadados da imagem (parte inferior central) e seleção de assinaturas espectrais para verificação de similaridade (parte inferior direita).

Figura 3.19: Ferramenta desenvolvida para auxiliar no desenvolvimento dos experimentos.



Fonte: Elaborada pelo autor.

3.3.2 Definição de classes referentes a cada *ground truth* e obtenção dos *endmembers*

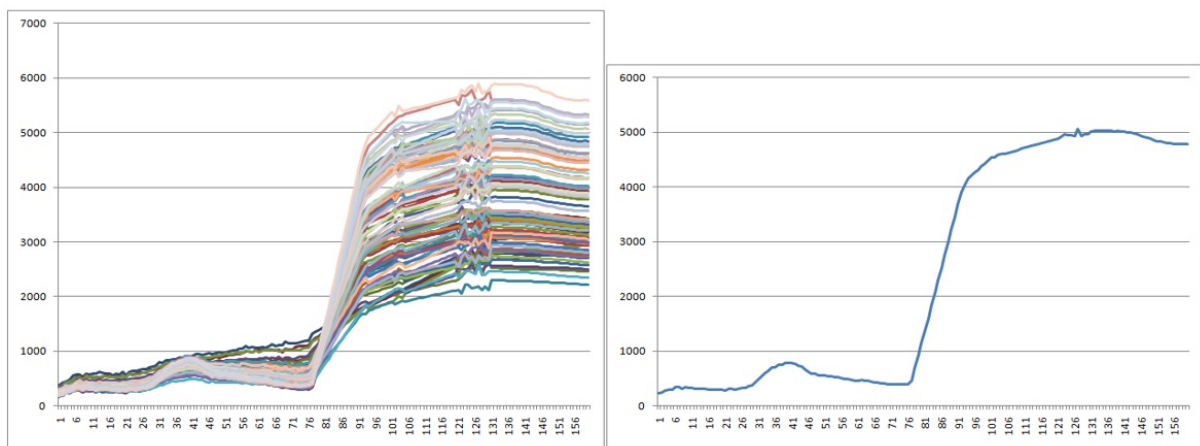
De modo a investigar o comportamento espectral dos *pixels* das imagens hiperespectrais, *pixels* de diferentes regiões foram coletados com o auxílio da ferramenta mostrada na Figura 3.19, sendo que cada um deles foi atribuído às classes descritas em seus respectivos *ground truths*.

Os *pixels* selecionados formam um arquivo referente às classes a qual pertence, sendo que para cada imagem de teste escolhida, existe um conjunto de arquivos de classes específicos. Cada classe possui 400 *pixels*, ou seja, 400 amostras que a representam, com exceção da imagem Indian Pines, em que foram escolhidos apenas 15 *pixels* por classe, pois no *ground truth* existem poucos *pixels* para representar algumas classes. Cada amostra pode ser vista na sua representação espectral conforme a Figura 3.20 (esquerda), onde são vistas várias assinaturas espectrais, ou simplesmente espectros, pertencentes à classe.

Depois da etapa de coleta e definição dos arquivos contendo os espectros de cada classe, foi necessário estabelecer qual espectro seria o mais representativo (com menos mistura de materiais) para cada classe, o chamado *endmember*. Para encontrar o *endmember* de cada classe, foram utilizados alguns métodos clássicos da literatura para a extração de espectros puros, sendo

eles: *Automatic Target Generation Process - ATGP* (REN; CHANG, 2003), *Pixel Purity Index - PPI* (SOLUTIONS, 1999), *N-FINDR* (WINTER, 1999) e *Fast Iterative Pixel Purity Index - FIPPI* (CHANG; PLAZA, 2006), além da média dos valores de reflectância dos 400 *pixels* de cada classe, formando um **espectro médio** (Figura 3.20, direita). O cálculo dos espectros médios foi considerado devido aos estudos iniciais realizados e que mostraram que o espectro médio preserva o formato característico da curva que representa a assinatura espectral e suaviza a interferência de *outliers* se estes existirem nas amostras coletadas.

Figura 3.20: Espectros coletados para a classe Cítricos (Imagem do Chile) (esq.) e e respectivo Espectro Médio (dir.).



Fonte: Elaborada pelo autor.

Após os *endmembers* serem extraídos dos arquivos de classes para cada imagem, cada um deles foi utilizado como espectro de referência nas tarefas de classificação utilizando os métodos de similaridade espectral SAM e SID, ambos citados na subseção 2.2.1. Saber qual é o espectro que melhor representa cada uma das classes é de fundamental importância para a próxima etapa do método proposto: NIRKNN.

3.3.3 Cálculo para estimar o NIR

Nesta seção é descrito o procedimento utilizado no NIRKNN, para a obtenção de imagem NIR.

Depois de encontrados os *endmembers* de cada classe existente nas imagens, foi necessário desenvolver uma técnica para transformar uma imagem RGB em uma imagem monocromática do NIR. Primeiramente, foram escolhidas 3 bandas de cada imagem, na região do espectro visível referentes à banda azul, verde e vermelha, para gerar uma imagem RGB normalizando os valores de reflectância dessas bandas dentro do intervalo de 0 a 255. A partir desta imagem, para cada *pixel* os valores de R, G e B são extraídos, assim com os valores de R e G são utilizados

para calcular o índice MPRI, sendo cada uma dessas informações associadas a uma das classes da imagem, definindo assim um conjunto de treinamento. Além disso, os valores de R, G e B também foram utilizados na obtenção dos valores dos parâmetros para os modelos de cor HSV e CIE Lab e associados às classes, de modo a verificar se esses modelos apresentam melhor resultado de classificação pelo algoritmo KNN. Também foi gerado um conjunto de dados para teste, ou seja, dados extraídos das imagens que não possuem rótulos associados. Todas essas informações serviram como entrada para o método NIRKNN. Essa parte refere-se ao bloco de "Imagem RGB + MPRI" mostrado no diagrama da Figura 3.1.

O próximo passo do método NIRKNN é a classificação das amostras com o algoritmo KNN, conforme mostrado na Figura 3.1. A classificação é realizada para cada *pixel* da imagem RGB, incluindo como atributos, além dos valores R, G e B, o índice MPRI. Também foram realizados experimentos utilizando outros modelos de cores: HSV e CIE Lab.

Como citado no início deste capítulo, foi utilizada a ferramenta WEKA para a realização da tarefa de classificação das amostras. O número de k vizinhos escolhido foi 5, após testes com outros valores; a métrica de distâncias foi a Distância Euclidiana e a contribuição de cada um dos k vizinhos mais próximos (amostras do conjunto de treinamento) foi ponderada pelo inverso da distância ($1/d(a,b)$), onde a seria a amostra (*pixel*) a ser classificado e b uma amostra do conjunto de treinamento. Portanto, foi utilizado o algoritmo KNN ponderado.

A ferramenta WEKA foi escolhida por ter um recurso interessante: para cada amostra classificada, é exibido o grau de proximidade que a amostra possui para cada uma das k classes mais próximas para o respectivo problema de classificação, sendo esta classificada de acordo com o maior grau de pertinência. A informação de quão próximos são cada um dos k vizinhos mais próximos desempenha um papel fundamental para o propósito deste trabalho de pesquisa. Não foi encontrada documentação sobre os métodos que a ferramenta WEKA utiliza para o cálculo do grau de proximidade que os k vizinhos possuem em relação à amostra a ser classificada, denominado pela ferramenta como *Probability Distribution* (Figura 3.21), porém, como é uma ferramenta *open source*, foi possível verificar e analisar a implementação do algoritmo que realiza o cálculo.

Inicialmente, o algoritmo cria um vetor *distribution* com dimensão igual ao número de classes do problema de classificação. Cada elemento de *distribution* tem um valor inicial, chamado de "correção do classificador", definido pela equação 3.13.

$$correcao = \frac{1}{MAX(1, Num_{Instancias})} \quad (3.13)$$

Figura 3.21: Resultado da classificação com a ferramenta Weka e a "Distribuição de probabilidade" para os k vizinhos mais próximos.

```

=== Predictions on test split ===
inst#, actual, predicted, error, probability distribution
1 1:ET1 1:ET1 *0.757 0.243 0 0 0 0 0 0 0 0 0 0 0 0 0 0
2 1:ET1 1:ET1 *1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
3 1:ET1 1:ET1 *1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
4 1:ET1 1:ET1 *1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
5 1:ET1 1:ET1 *1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
6 1:ET1 1:ET1 *1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
7 1:ET1 1:ET1 *1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
8 1:ET1 14:Solon1 + 0.18 0 0 0 0 0 0 0 0 0 0 0 0 0 0.82 0 0
9 1:ET1 1:ET1 *1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
10 1:ET1 1:ET1 *1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
11 1:ET1 14:Solon1 + 0 0.411 0 0 0 0 0 0 0 0 0 0 0 0.589 0 0 0

```

Fonte: Elaborada pelo autor.

Para cada um dos k vizinhos mais próximos, o algoritmo calcula um peso $W_i, 1 \leq i \leq k$, utilizando a distância de cada um deles em relação à amostra a ser classificada e o número de atributos utilizados como entrada para a classificação da mesma, como pode ser visto na equação 3.14.

$$W_i = \frac{1}{\left(\sqrt{\frac{distance_i^2}{Num_{Atributos}}} + 0,001 \right) distance_i}, 1 \leq i \leq k \quad (3.14)$$

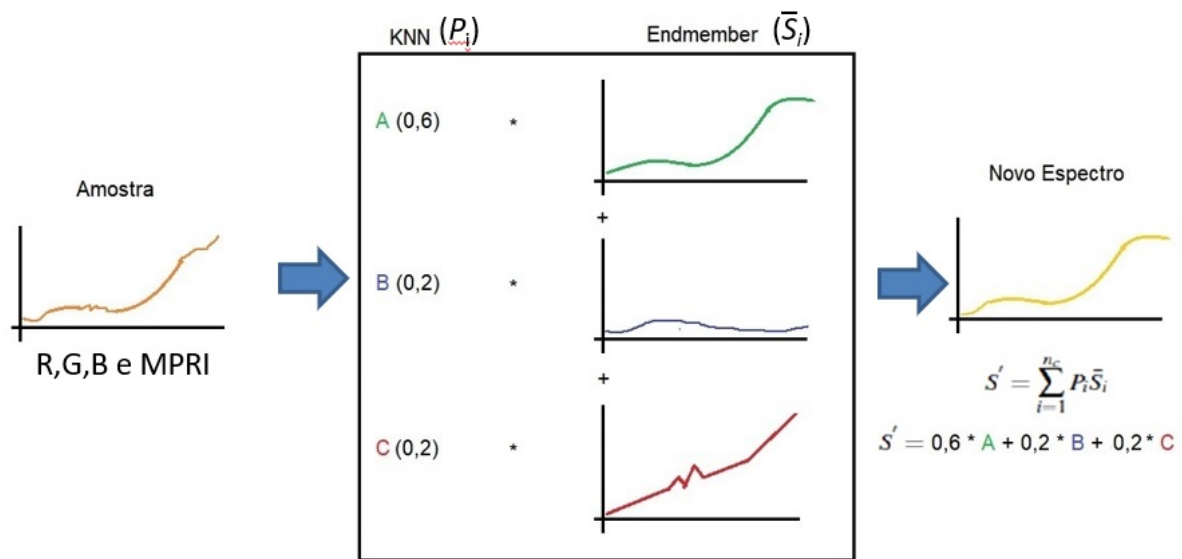
O vetor *distribution* é atualizado apenas nas posições referentes à classe de cada um dos k vizinhos mais próximos, conforme a equação 3.15 e após a atualização, o algoritmo faz uma normalização do vetor (equação 3.16), sendo este o vetor do grau de proximidade P_i . Esse vetor normalizado estende-se para todas as classes existentes na imagem RGB.

$$distribution_i = correcao + W_i \quad (3.15)$$

$$P_i = \frac{distribution_i}{\sum_{j=1}^k W_j}, 1 \leq i \leq Num_{Classes} \quad (3.16)$$

A partir das classificações feitas pelo KNN, para cada *pixel* de imagem; e a definição dos *endmembers* para cada classe, é possível fazer a estimativa da imagem NIR, que é a próxima etapa mostrada no diagrama da Figura 3.1. A Figura 3.22 mostra o exemplo de como é feito o cálculo de um novo espectro S' utilizando o grau de proximidade P_i e os *Endmembers* de cada classe.

Conforme a Figura 3.22, a partir de uma amostra, esta é classificada de acordo com os valores extraídos para a entrada do KNN (R,G,B e MPRI). Tendo posse dos *endmembers* encontrados e do grau de proximidade das amostras classificadas a seus k vizinhos, a estimativa

Figura 3.22: Cálculo de estimativa do NIR utilizando grau de proximidade e *Endmembers*.

Fonte: Elaborada pelo autor.

de um novo espectro pode ser realizada, com a mesma quantidade de bandas que os espectros do conjunto de treinamento possui. Os *endmembers* que representam todas as classes do *ground truth* são somados utilizando seus respectivos graus de proximidade como fator de ponderação, resultando em um novo espectro S' , conforme a equação 3.17.

$$S' = \sum_{i=1}^{n_c} P_i \bar{S}_i \quad (3.17)$$

Na equação 3.17, n_c é o número de classes do *ground truth* da imagem original, P_i é o grau de proximidade dos valores RGB, HSV ou CIE Lab do *pixel* em relação à classe i e \bar{S}_i representa o *endmember* que representa a classe i .

Uma vez obtidos os resultados dos novos espectros S' para cada *pixel* da imagem, os valores de *pixels* correspondentes ao NIR podem ser extraídos e obtida a imagem NIR estimada, representada pelo bloco mais à direita da Figura 3.1.

Nota-se que na descrição apresentada, foram utilizadas imagens hiperespectrais como entrada, porém, como os *endmembers* já foram obtidos, as imagens NIR podem ser estimadas diretamente das imagens RGB, que é o objetivo desta tese.

3.4 Considerações Finais

Neste capítulo foram descritos os métodos utilizados para o desenvolvimento do trabalho, assim como uma descrição das imagens utilizadas nos experimentos realizados. O próximo capítulo apresenta os experimentos realizados e os resultados obtidos pelos dois métodos propostos para estimar imagens NIR, fazendo uma análise sobre cada um deles, comparando os resultados obtidos por eles utilizando o mesmo conjunto de imagens.

Capítulo 4

RESULTADOS E DISCUSSÃO

Neste capítulo serão mostrados os resultados obtidos para os experimentos realizados com o objetivo de estimar imagens NIR a partir de imagens RGB obtidas diretamente das imagens digitais, assim como os resultados das etapas intermediárias que envolvem o processo.

A Seção 4.1 refere-se aos resultados obtidos com o método NIRDLearn e a Seção 4.2, com o método NIRKNN.

4.1 Estimativa de imagens NIR com o método NIRDLearn

Foram realizados experimentos com o método NIRDLearn, alterando o modelo de cor das imagens de entrada (RGB, HSV e CIE Lab) e também a função de *loss* (\mathcal{L}_{L1} , \mathcal{L}_{Mix} e \mathcal{L}_{PMix}) na função objetivo (equação 2.20), resultando no total de nove configurações de treinamento distintas.

Substituindo a função \mathcal{L}_{L1} por \mathcal{L}_{Mix} (eq. 3.11) na equação 2.20, tem-se a equação 4.1:

$$G^* = \arg \min_G \max_D \mathcal{L}_{cGAN}(G, D) + \lambda \mathcal{L}_{Mix}(G) \quad (4.1)$$

Análogo ao feito para equação anterior, ao se substituir a função \mathcal{L}_{L1} pela função \mathcal{L}_{PMix} (eq. 3.12) na equação 2.20, é obtida a seguinte equação 4.2:

$$G^* = \arg \min_G \max_D \mathcal{L}_{cGAN}(G, D) + \lambda \mathcal{L}_{PMix}(G) \quad (4.2)$$

Para todas as configurações de treinamento, foram utilizadas 400 épocas de treinamento, valor utilizado por Isola et al. (2017) em alguns *datasets*. O valor de λ utilizado nas equações

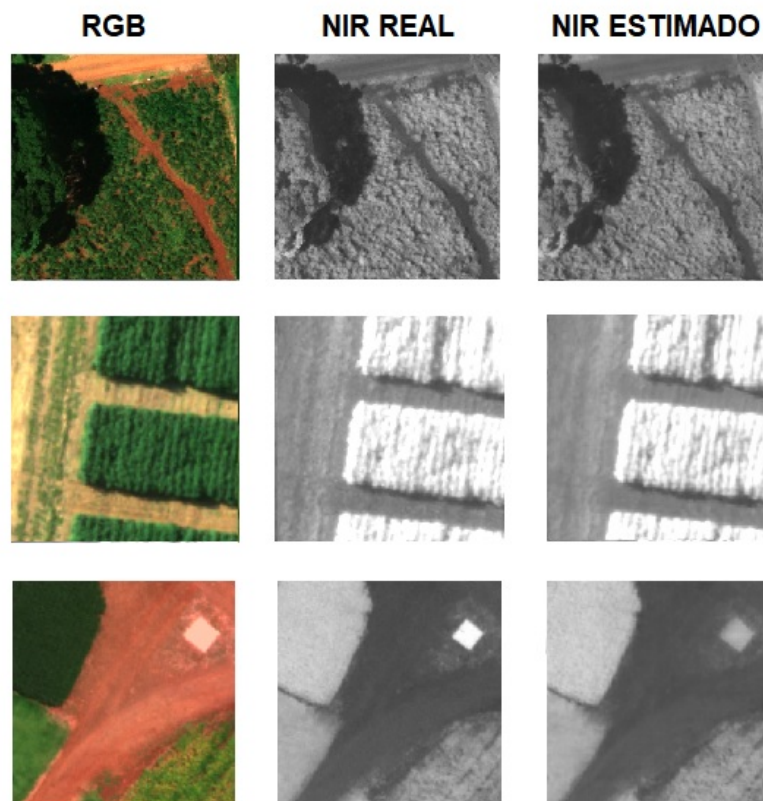
2.20, 4.1 e 4.2 foi o mesmo utilizado por Isola et al. (2017) ($\lambda = 100$), assim como os parâmetros para o otimizador *Adam* também permanecem os recomendados por Kingma e Ba (2014) (taxa de aprendizado = 10^{-3} ; $\beta_1 = 0,9$; $\beta_2 = 0,999$ e $\varepsilon = 10^{-8}$).

Para verificar a similaridade entre as imagens geradas e o *ground truth*, foram utilizadas duas métricas, as quais são as mais utilizadas para verificar a similaridade entre imagens: a MS-SSIM (equação 3.9) e a Raíz do Erro quadrático médio (RMSE). As imagens NIR reais e estimadas tiveram ajuste de brilho e contraste (com os mesmos valores) para melhor visualização destas no texto.

4.1.1 Resultados com modelo RGB

A Figura 4.1 mostra três exemplos de resultado de estimativa utilizando o método NIRDLearn com entrada de imagens RGB.

Figura 4.1: Exemplos de imagens NIR estimadas pelo método NIRDLearn a partir de imagens RGB como entrada.

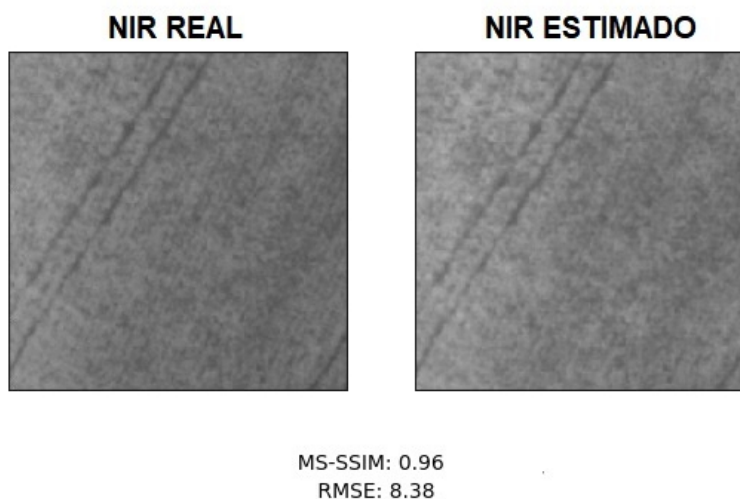


Fonte: Elaborada pelo autor.

A primeira configuração de rede treinada e testada foi aquela que atende ao padrão utilizado pela rede *Pix2Pix*, no qual a imagem de entrada pertence ao modelo de cores RGB e a função

objetivo é a mesma proposta por Isola et al. (2017). Após o treinamento com os 1080 pares de amostra RGB-NIR, as imagens RGB do par foram apresentadas à rede G , sendo utilizadas para verificar o nível de similaridade produzida entre as imagens NIR reais e suas respectivas imagens estimadas durante o treinamento. O valor médio para a métrica MS-SSIM foi de 0,93; mostrando que as imagens estimadas são muito similares às imagens NIR reais. No caso do RMSE, o valor foi de 17,25 (escala de 0 a 255); representando a diferença média entre os valores dos *pixels* das imagens comparadas. Para as 440 amostras de validação, a média do MS-SSIM foi de 0,89; ainda apresentando grau elevado de similaridade entre imagens reais e estimadas. O RMSE para o conjunto de validação foi de 24,48. Na Figura 4.2 é possível observar a imagem estimada pela rede com maior similaridade com uma imagem NIR real pertencente ao conjunto de validação. Na Figura 4.3 tem-se a imagem estimada com a menor similaridade em comparação à imagem NIR real do conjunto de validação.

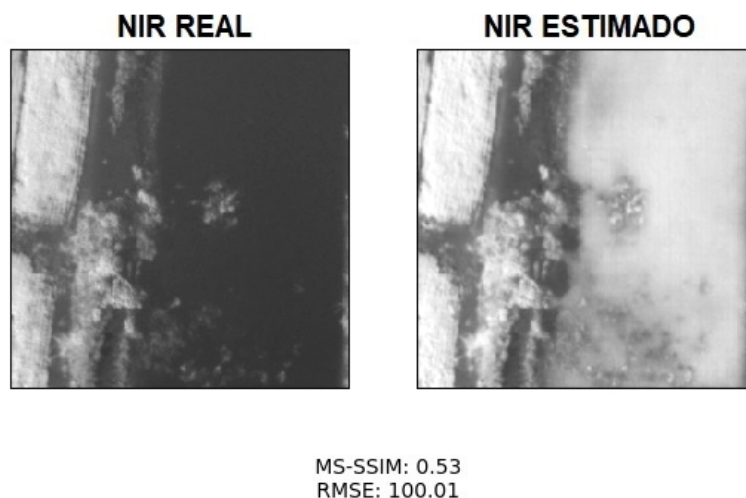
Figura 4.2: Melhor resultado de estimativa NIR utilizando NIRDLearn.



Fonte: Elaborada pelo autor.

A imagem NIR real presente na Figura 4.3 contém um rio, representado pela parte mais escura da imagem, sendo a parte mais clara, a vegetação. Observando a imagem estimada, é possível notar que a área da vegetação apresenta valores mais altos para os *pixels* e a região do rio contém valores completamente diferentes em comparação à imagem real, evidenciado pelo valor elevado do RMSE (100,01).

A segunda configuração treinada e testada foi a que utiliza a equação 4.1 como função objetivo. O resultado da estimativa de NIR utilizando as imagens RGB do conjunto de treinamento também evidenciou o grande nível de similaridade entre as imagens reais e estimadas, tendo valores médios de 0,938 para o MS-SSIM e 17,258 para o RMSE. Em comparação ao resultado

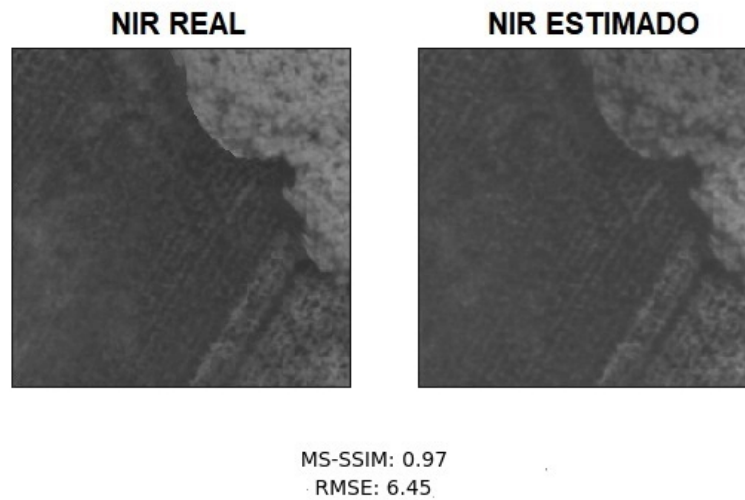
Figura 4.3: Pior resultado de estimativa NIR utilizando NIRDLearn.

Fonte: Elaborada pelo autor.

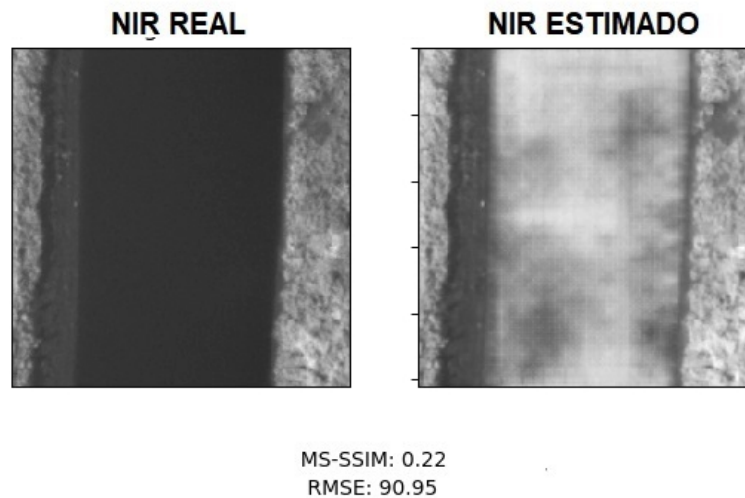
anterior, é possível notar que a substituição da função \mathcal{L}_{L1} pela função \mathcal{L}_{Mix} gerou um pequeno aumento na similaridade estrutural entre as imagens estimadas e reais, porém, um pequeno aumento na diferença média entre os valores dos *pixels*. A estimativa de NIR utilizando as imagens RGB do conjunto de validação resultou em valores de similaridade 0,8885 (MS-SSIM) e 24,463 (RMSE), valor inferior de MS-SSIM e uma diferença menor entre os *pixels* das imagens em comparação ao do treinamento anterior. Na Figura 4.4 tem-se o resultado de estimativa de imagem NIR com maior similaridade em relação à imagem NIR real. É possível verificar que a imagem estimada possui regiões mais escuras em locais em que deveriam ter valores mais altos para os *pixels*, mais especificamente, na parte superior direita da imagem.

Ao observar a imagem estimada com menores valores para a métrica MS-SSIM e maior valor para o RMSE, é observado que a imagem NIR real compreende uma região com presença de rio, assim como observado no experimento que utiliza a função \mathcal{L}_{L1} . Outro fato notado foi que, neste experimento, as áreas que contém vegetação também possuem valores de *pixels* muito mais altos na imagem estimada do que na imagem NIR real.

Por fim, foi realizado o último treinamento da rede utilizando imagens RGB como entrada, porém, utilizando a função \mathcal{L}_{PMix} . Os resultados mostraram uma melhora pequena em relação à configuração da rede que utiliza \mathcal{L}_{L1} , porém, inferior ao resultado obtido pelo experimento anterior, comparando-se a similaridade estrutural. Para o índice MS-SSIM, o valor médio do conjunto de treinamento foi de 0,931; enquanto para o RMSE, o valor foi 17,779; o maior valor de diferença média entre os valores de *pixels* para as três redes treinadas com imagens RGB. As estimativas com as imagens de validação mostraram que a média do MS-SSIM foi de 0,892;

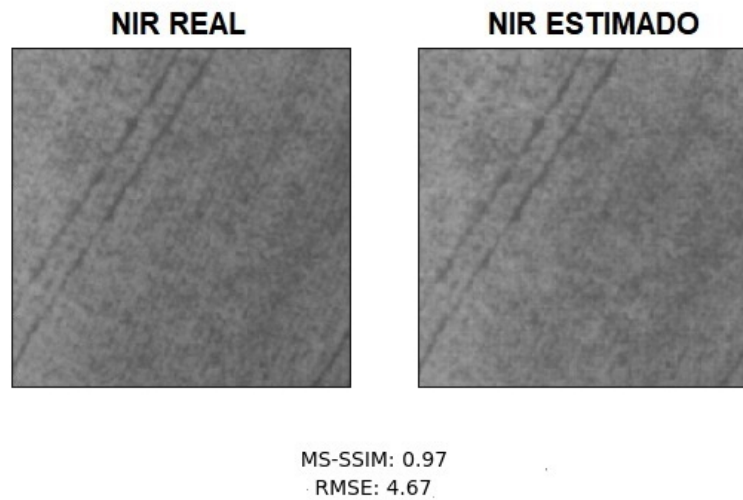
Figura 4.4: Melhor resultado de estimativa NIR utilizando NIRDLearn utilizando a função \mathcal{L}_{Mix} .

Fonte: Elaborada pelo autor.

Figura 4.5: Pior resultado de estimativa NIR utilizando NIRDLearn utilizando a função \mathcal{L}_{Mix} .

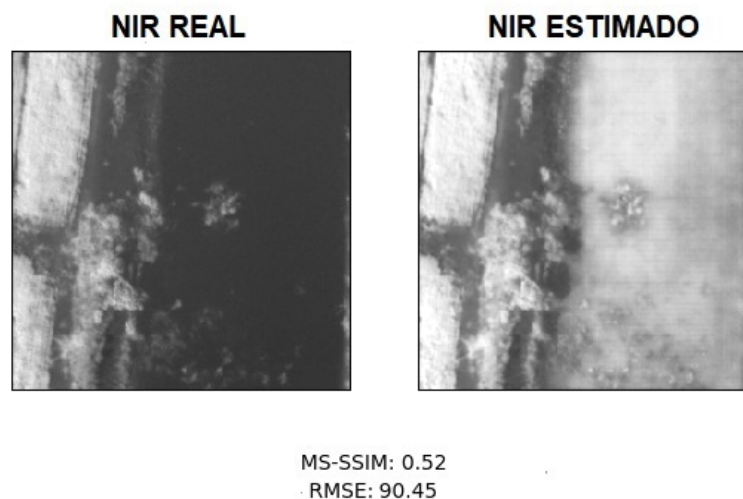
Fonte: Elaborada pelo autor.

o maior valor de similaridade obtido para o conjunto de validação até então, porém, o valor do RMSE (24,5275) mostrou que a função \mathcal{L}_{PMix} produz *pixels* com maiores diferenças de valores entre imagens NIR estimadas e reais para entradas da rede com dados RGB. Os pares de imagens com maior e menor similaridade são os mesmos que os estimados pela rede que utiliza a função \mathcal{L}_{L1} , porém, os valores para o MS-SSIM e RMSE são diferentes. Para ambas as imagens, o resultado de similaridade foi superior com o uso da função \mathcal{L}_{PMix} . A Figura 4.6 mostra a imagem RGB do conjunto de validação que gerou uma imagem NIR com maior similaridade estrutural e menor erro médio entre *pixels*.

Figura 4.6: Melhor resultado de estimativa NIR utilizando NIRDLearn utilizando a função \mathcal{L}_{PMix} .

Fonte: Elaborada pelo autor.

Na Figura 4.7 é apresentada a imagem estimada com menor similaridade estrutural e erro entre valores de *pixels*. Por se tratar da mesma imagem da Figura 4.3, a região da imagem estimada em que se encontra o rio permanece com valores muito maiores do que os da imagem real, apresentando tons de cinza mais claros, assim como acontece na vegetação.

Figura 4.7: Pior resultado de estimativa NIR utilizando NIRDLearn utilizando a função \mathcal{L}_{PMix} .

Fonte: Elaborada pelo autor.

As tabelas 4.1 e 4.2 resumem os resultados de similaridade para os conjuntos de treinamento e validação, respectivamente.

Tabela 4.1: Similaridade média entre *ground truth* e imagens estimadas para o conjunto de treinamento (Imagens RGB).

Loss	MS-SSIM	RMSE
\mathcal{L}_{L1}	0,930	17,250
\mathcal{L}_{Mix}	0,938	17,258
\mathcal{L}_{PMix}	0,931	17,779

Tabela 4.2: Similaridade média entre *ground truth* e imagens estimadas para o conjunto de validação (Imagens RGB).

Loss	MS-SSIM	RMSE
\mathcal{L}_{L1}	0,890	24,480
\mathcal{L}_{Mix}	0,888	24,463
\mathcal{L}_{PMix}	0,892	24,527

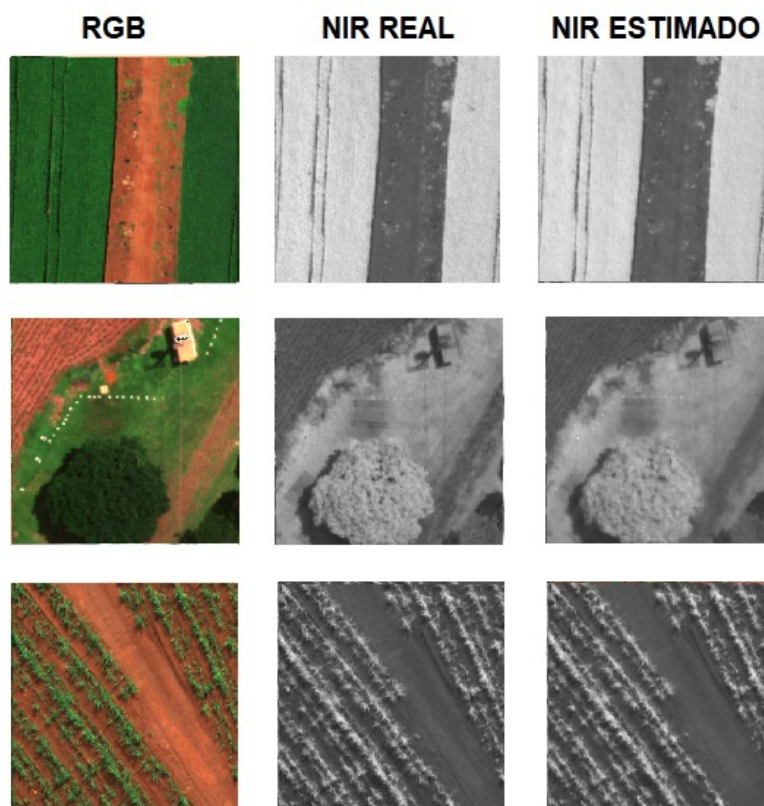
4.1.2 Resultados com modelo CIE Lab

Foram realizados os testes utilizando as três funções de perda (*loss*) propostas, porém, convertendo as imagens de entrada para o modelo de cor CIE Lab. A Figura 4.8 mostra três exemplos de resultado de estimativa utilizando NIRDLearn com imagens transformadas para este modelo de cor. Para efeitos de melhor visualização, as imagens de entrada estão no modelo de cor RGB.

Assim como no experimento que utiliza imagens RGB como entrada da rede, a primeira rede treinada utilizava a função de erro \mathcal{L}_{L1} . A alteração do modelo de cor nas imagens de entrada resultou em maiores valores de similaridade estrutural e menor erro entre os *pixels* das imagens reais e as estimadas. O valor médio do MS-SSIM para as imagens estimadas a partir do conjunto de treinamento foi de 0,941; maior do que o atingido com imagens RGB (0,93). Com relação ao valor do RMSE, o erro entre os *pixels* foi de 6,47; muito inferior ao dos retornados pelos testes anteriores. Os resultados de similaridade avaliados após estimativa com as imagens do conjunto de validação foram de 0,895 e 16,5895 para a MS-SSIM e a RMSE, respectivamente. Os resultados para o conjunto de validação mostraram que a mudança do modelo de cor influenciou positivamente na estimativa das imagens NIR, uma vez que as diferenças *pixel a pixel* entre as imagens se tornaram menores. A imagem estimada a partir do conjunto de validação que apresentou maior similaridade com a imagem real foi a mesma das Figuras 4.2 e 4.6, porém, o valor do RMSE foi o menor até então, sendo possível verificar na Figura 4.9.

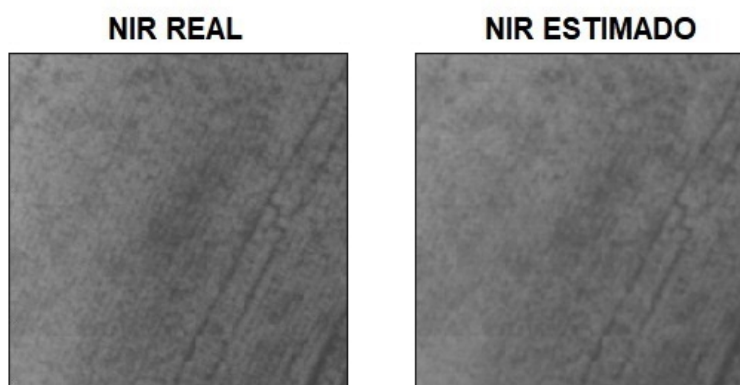
Na Figura 4.10 é possível verificar que, novamente, a imagem que possui menor similaridade contém uma região com água, apresentando as mesmas características encontradas até

Figura 4.8: Exemplos de imagens NIR estimadas por NIRDLearn a partir de imagens RGB convertidas para o modelo CIE Lab.



Fonte: Elaborada pelo autor.

Figura 4.9: Melhor resultado de estimativa NIR utilizando NIRDLearn, a função \mathcal{L}_{L1} e imagem CIE Lab.

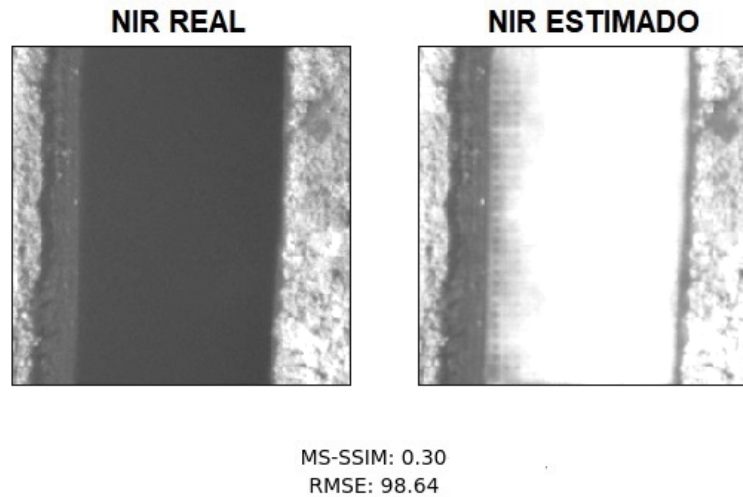


MS-SSIM: 0.97
RMSE: 3.71

Fonte: Elaborada pelo autor.

então, áreas esbranquiçadas onde existe água e vegetação com valores mais altos para os níveis de cinza.

Figura 4.10: Pior resultado de estimativa NIR utilizando NIRDLearn, a função \mathcal{L}_{L1} e imagem CIE Lab.



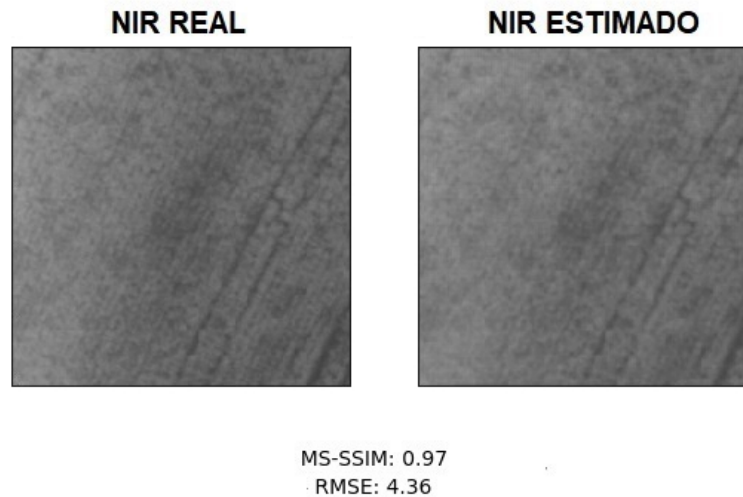
Fonte: Elaborada pelo autor.

Ao se substituir a função \mathcal{L}_{L1} pela função \mathcal{L}_{Mix} , a média da similaridade estrutural se manteve com o mesmo valor (MS-SSIM = 0,941) para as imagens do conjunto de treinamento, apresentando um leve aumento na média da diferença entre *pixels* das imagens NIR reais e os *pixels* das imagens estimadas. O valor do RMSE foi de 6,52; representando um aumento de 0,05 no erro avaliado. Curiosamente, para as imagens do conjunto de validação, o valor do MS-SSIM (0,897) apresentou valor um pouco superior comparado ao valor resultante do uso da função \mathcal{L}_{L1} (0,895). Em relação ao RMSE, o valor para as imagens do conjunto de validação foi de 16,2755; valor inferior ao da rede treinada anteriormente. A imagem estimada a partir do conjunto de validação que obteve maior índice de similaridade para esta rede (Figura 4.11 foi a mesma do experimento anterior, a diferença se dá no valor do RMSE, agora 4,36, um aumento de 0,65 para a imagem.

Observa-se na Figura 4.12 que a imagem estimada a partir do conjunto de validação com menor similaridade foi a mesma da rede treinada anteriormente, porém, o valor da diferença entre os *pixels* reais e estimados é 95,69; 2,95 menor do que o da rede anterior. Não houve alterações significativas em relação aos *pixels* referentes à água.

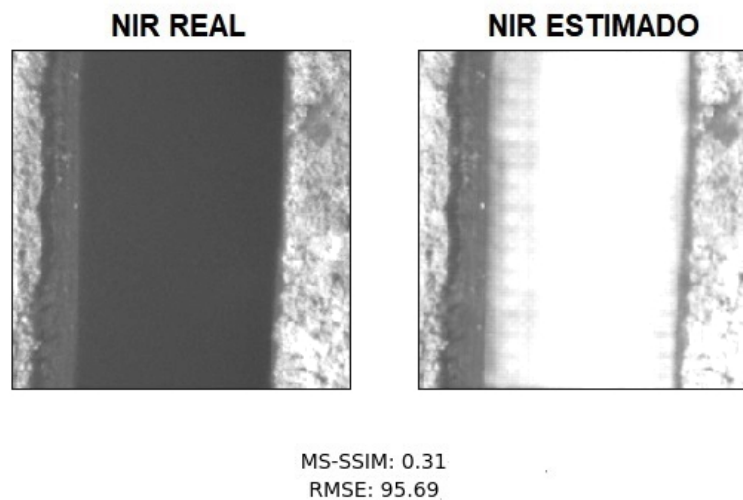
A rede treinada com a função \mathcal{L}_{PMix} foi a que produziu melhores resultados de similaridade utilizando as imagens de treinamento, tanto para a análise estrutural, quanto para a análise *pixel a pixel*. O valor da média do MS-SSIM para as 1080 imagens do conjunto de validação

Figura 4.11: Melhor resultado de estimativa NIR utilizando NIRDLearn, a função \mathcal{L}_{Mix} e imagem CIE Lab.



Fonte: Elaborada pelo autor.

Figura 4.12: Pior resultado de estimativa NIR utilizando NIRDLearn, a função \mathcal{L}_{Mix} e imagem CIE Lab.

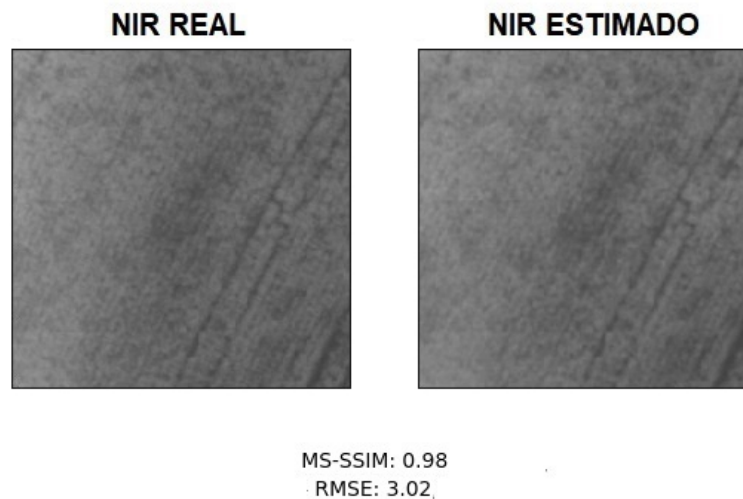


Fonte: Elaborada pelo autor.

foi de 0,95; o melhor resultado de similaridade estrutural até então, considerando os experimentos realizados com imagens RGB. O valor de RMSE também foi o menor dentre todos os experimentos realizados, sendo 5,948 o seu valor. Apesar dos melhores resultados ao se considerar as imagens estimadas a partir do conjunto de treinamento, os resultados de estimativa com as imagens de validação apresentaram os menores valores de similaridade dentre todos os experimentos utilizando o modelo de cor CIE Lab, com 0,891 para o MS-SSIM e 16,624.

As imagens estimadas a partir das imagens de validação que geraram maior e menor similaridade em relação a imagem NIR real podem ser vistas, respectivamente, nas figuras 4.13 e 4.14. Os valores de MS-SSIM e RMSE para a imagem com maior nível de similaridade foram respectivamente de 0,98 e 3,02; indicando melhora em relação aos outros dois resultados anteriores. Para a imagem com menor similaridade, o valor do MS-SSIM foi de 0,43 e o de RMSE foi de 56,24; os melhores resultados para estas imagens. Nesta mesma imagem, é possível notar que a imagem estimada possui *pixels* mais escuros na região em que se localiza o rio do que as outras imagens estimadas. Apesar dos piores resultados ao se avaliar a média das medidas de similaridade para o conjunto de validação, apresentou os melhores resultados para os dois extremos (imagem estimada mais similar e imagem estimada menos similar).

Figura 4.13: Melhor resultado de estimativa NIR utilizando NIRDLearn, a função \mathcal{L}_{PMix} e imagem CIE Lab.



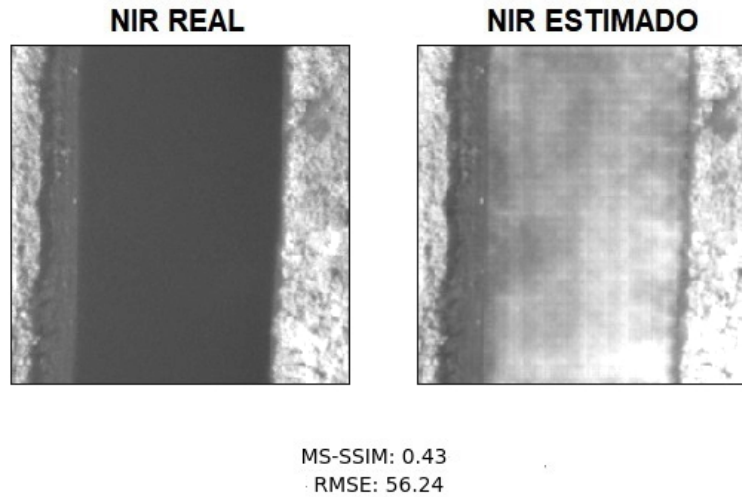
Fonte: Elaborada pelo autor.

Os resultados de similaridade para as redes treinadas estão resumidos nas tabelas a seguir. Na tabela 4.3 estão os resultados das medidas de similaridade para as imagens do conjunto de treinamento, enquanto na tabela 4.4 se encontram os valores para as imagens do conjunto de validação.

Tabela 4.3: Similaridade média entre *ground truth* e imagens estimadas para o conjunto de treinamento (Imagens CIE Lab).

Loss	MS-SSIM	RMSE
\mathcal{L}_{L1}	0,941	6,470
\mathcal{L}_{Mix}	0,941	6,520
\mathcal{L}_{PMix}	0,950	5,948

Figura 4.14: Pior resultado de estimativa NIR utilizando NIRDLearn, a função \mathcal{L}_{PMix} e imagem CIE Lab.



Fonte: Elaborada pelo autor.

Tabela 4.4: Similaridade média entre *ground truth* e imagens estimadas para o conjunto de validação (Imagens CIE Lab).

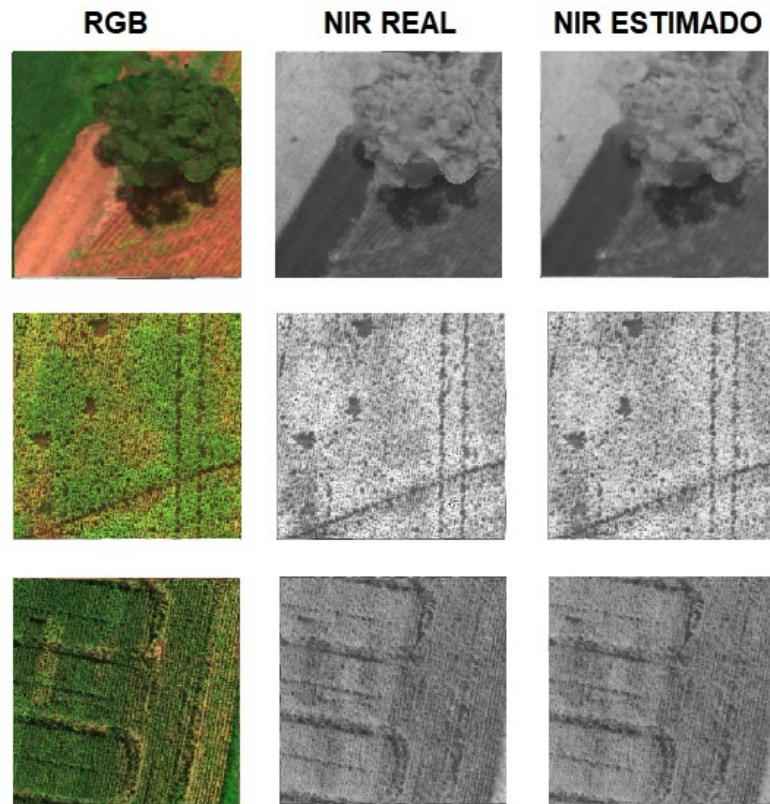
Loss	MS-SSIM	RMSE
\mathcal{L}_{L1}	0,895	16,589
\mathcal{L}_{Mix}	0,897	16,275
\mathcal{L}_{PMix}	0,891	16,624

4.1.3 Resultados com modelo HSV

O último dos modelos de cor a ser utilizado para as imagens de entrada das três diferentes NIRDLearn foi o modelo HSV. Na Figura 4.15 podem ser vistos três resultados da estimativa de imagens NIR a partir de imagens HSV.

O resultado da estimativa de imagens NIR a partir de imagens HSV do conjunto de treinamento, utilizando a função \mathcal{L}_{L1} , mostrou que as imagens estimadas possuem alta similaridade estrutural e pouca diferença entre *pixels* reais e estimados, considerando-se a média. O valor da média do MS-SSIM foi de 0,94 e o da média do RMSE foi de 6,586, valores indicativos de resultados de estimativa maiores do que os gerados por imagens RGB, principalmente considerando-se o valor do RMSE, porém, os resultados estão próximos aos resultados obtidos ao se utilizar o modelo de cor CIE Lab. Para as imagens do conjunto de validação, as médias do MS-SSIM e do RMSE foram, respectivamente, 0,894 e 16,414, valores menor para MS-SSIM relação ao obtido com imagens CIE Lab, porém, com RMSE menor comparado ao obtido a partir das imagens CIE Lab. Isto mostra que a rede gerou menor diferença entre valor de *pixel*

Figura 4.15: Exemplos de imagens NIR estimadas por NIRDLearn a partir de imagens RGB convertidas para o modelo HSV.



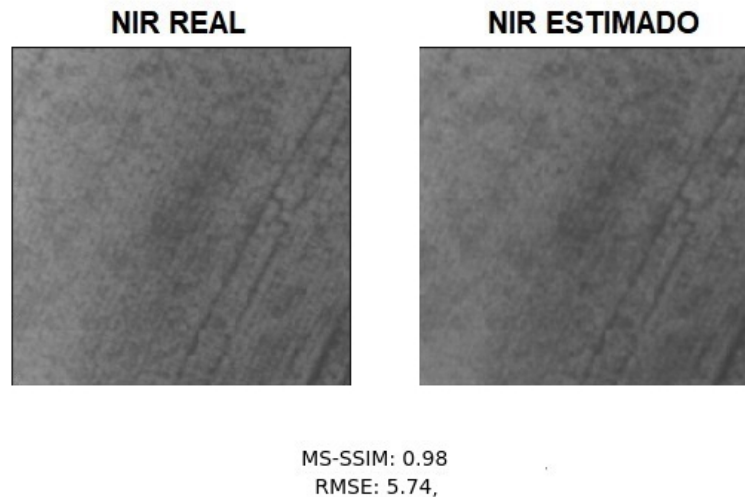
Fonte: Elaborada pelo autor.

real e estimado. Na Figura 4.16 é mostrada a imagem estimada com maior similaridade em relação a imagem NIR real, sendo possível notar que é a mesma imagem mostrada nas figuras dos experimentos anteriores, com exceção à Figura 4.4.

Novamente, a imagem estimada a partir das imagens do conjunto de validação é a mesma dos experimentos anteriores. A utilização do modelo HSV não resultou em melhoras significativas com o uso da função \mathcal{L}_{L1} , conforme a Figura 4.17.

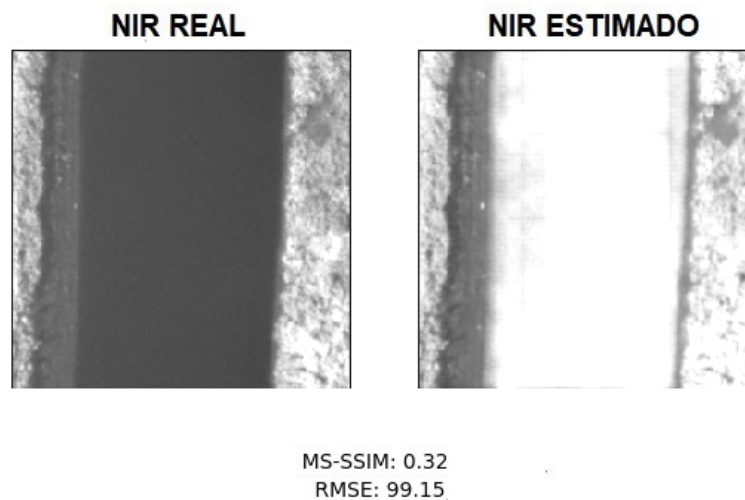
A rede treinada com imagens HSV e a função \mathcal{L}_{Mix} obteve o mesmo índice de similaridade estrutural para as imagens estimadas a partir do conjunto de treinamento em relação ao modelo de cor utilizado na subseção anterior (MS-SSIM = 0,941), porém, com maior diferença entre *pixels* de imagens estimadas e *pixels* NIR reais (RMSE = 6,62). Em relação aos valores das imagens estimadas a partir do conjunto de validação, o resultado do índice MS-SSIM foi de 0,895, sendo este inferior ao obtido na seção anterior para a mesma função de *loss*, porém, o valor do RMSE diminuiu em relação aos resultados obtidos anteriormente com imagens CIE Lab, caindo de 16,2755 para 16,1255. A Figura 4.18 mostra a imagem estimada a partir do conjunto de validação com maior similaridade em relação a imagem NIR real. É possível notar

Figura 4.16: Melhor resultado de estimativa NIR utilizando NIRDLearn, a função \mathcal{L}_{L1} e imagem HSV.



Fonte: Elaborada pelo autor.

Figura 4.17: Pior resultado de estimativa NIR utilizando NIRDLearn, a função \mathcal{L}_{L1} e imagem HSV.

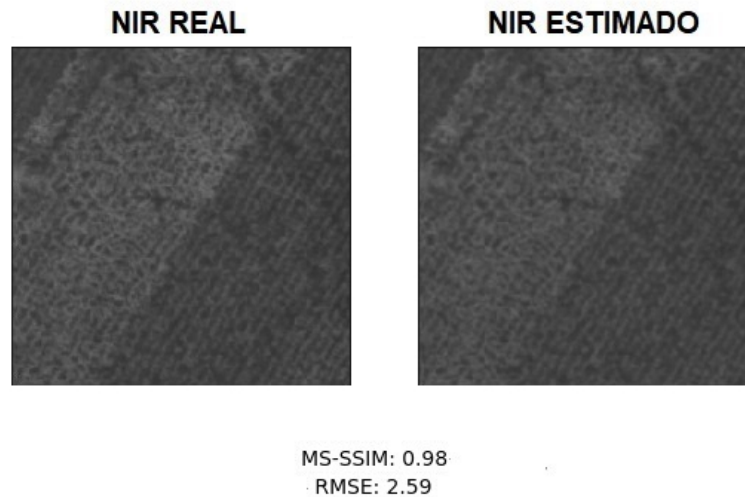


Fonte: Elaborada pelo autor.

que a imagem foi alterada em relação aos experimentos anteriores e que o valor do RMSE para esta imagem foi o menor obtido até então (2,59). Na Figura 4.19 é mostrada a imagem gerada que possui menor similaridade em relação a imagem NIR real, sendo possível notar que é a mesma imagem que apresentou índice de similaridade inferior para as outras redes treinadas anteriormente. O valor do MS-SSIM foi menor em relação ao gerado pela rede treinada com imagens HSV e com a função \mathcal{L}_{L1} , representando assim uma similaridade estrutural inferior. Apesar disso, o valor do RMSE foi inferior, mostrando que os valores *pixels* estimados e reais

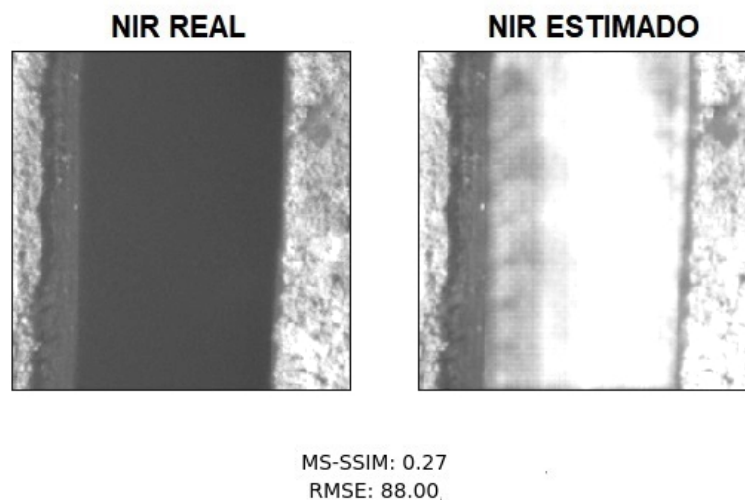
estão menos distantes para este resultado de estimativa.

Figura 4.18: Melhor resultado de estimativa NIR utilizando NIRDLearn, a função \mathcal{L}_{Mix} e imagem HSV.



Fonte: Elaborada pelo autor.

Figura 4.19: Pior resultado de estimativa NIR utilizando NIRDLearn, a função \mathcal{L}_{Mix} e imagem HSV.

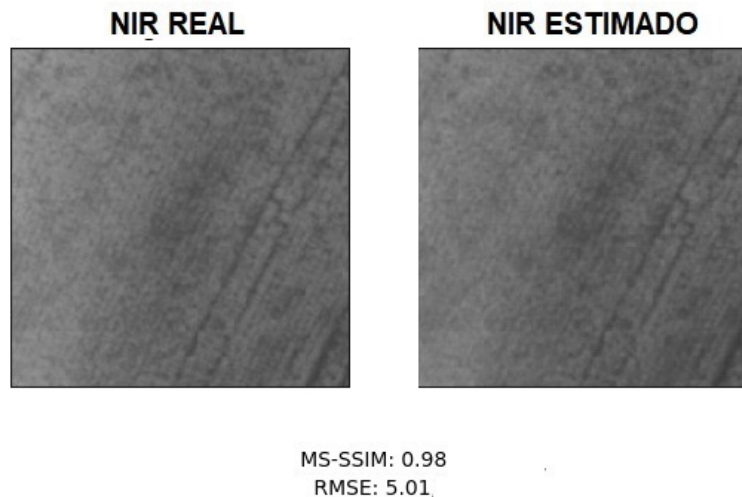


Fonte: Elaborada pelo autor.

Por fim, a última configuração de rede treinada foi a que utiliza imagens HSV como parâmetro de entrada e a função \mathcal{L}_{PMix} para cálculo de erro e ajuste de pesos. O valor da média do MS-SSIM para as imagens estimadas a partir do conjunto de treinamento foi 0,95, o mesmo índice de similaridade estrutural obtido utilizando a mesma função e imagens CIE Lab, sendo que a média do valor do RMSE foi de 5,978, valor levemente maior ao obtido com a mesma função na

subseção anterior. A média dos valores de MS-SSIM e RMSE para imagens estimadas a partir do conjunto de validação foram de 0,888 e 16,3575, respectivamente. A similaridade estrutural foi a mais baixa dentre todos os resultados de estimativa utilizando o conjunto de validação para todas as redes treinadas. O valor do RMSE foi menor do que o obtido para o conjunto da validação, utilizando a mesma função de *loss* para a subseção anterior. Para esta rede, a imagem estimada com maior similaridade em relação à imagem NIR real foi a imagem mais recorrente para os melhores resultados de estimativa, sendo possível observar o valor do MS-SSIM e do RMSE na Figura 4.20. A imagem estimada com menor similaridade em relação a imagem NIR (4.21) foi a imagem mais frequente em todas as redes treinadas, apresentando um dos menores valores para o índice MS-SSIM, sendo o mesmo valor visto na Figura 4.19. O valor do RMSE para esta imagem estimada foi inferior ao da Figura 4.19.

Figura 4.20: Melhor resultado de estimativa NIR utilizando NIRDLearn, a função \mathcal{L}_{PMix} e imagem HSV.



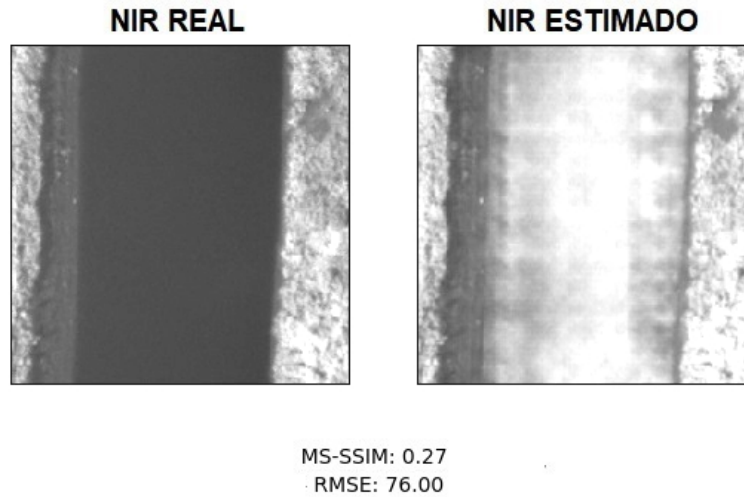
Fonte: Elaborada pelo autor.

Os resultados da avaliação de similaridade para as imagens estimadas utilizando as imagens HSV nas diferentes redes treinadas podem ser vistos nas Tabelas 4.5 e 4.6.

Tabela 4.5: Similaridade média entre *ground truth* e imagens estimadas para o conjunto de treinamento (Imagens HSV).

Loss	MS-SSIM	RMSE
\mathcal{L}_{L1}	0,940	6,586
\mathcal{L}_{Mix}	0,941	6,620
\mathcal{L}_{PMix}	0,950	5,978

Figura 4.21: Pior resultado de estimativa NIR utilizando NIRDLearn, a função \mathcal{L}_{PMix} e imagem HSV.



Fonte: Elaborada pelo autor.

Tabela 4.6: Similaridade média entre *ground truth* e imagens estimadas para o conjunto de validação (Imagens HSV).

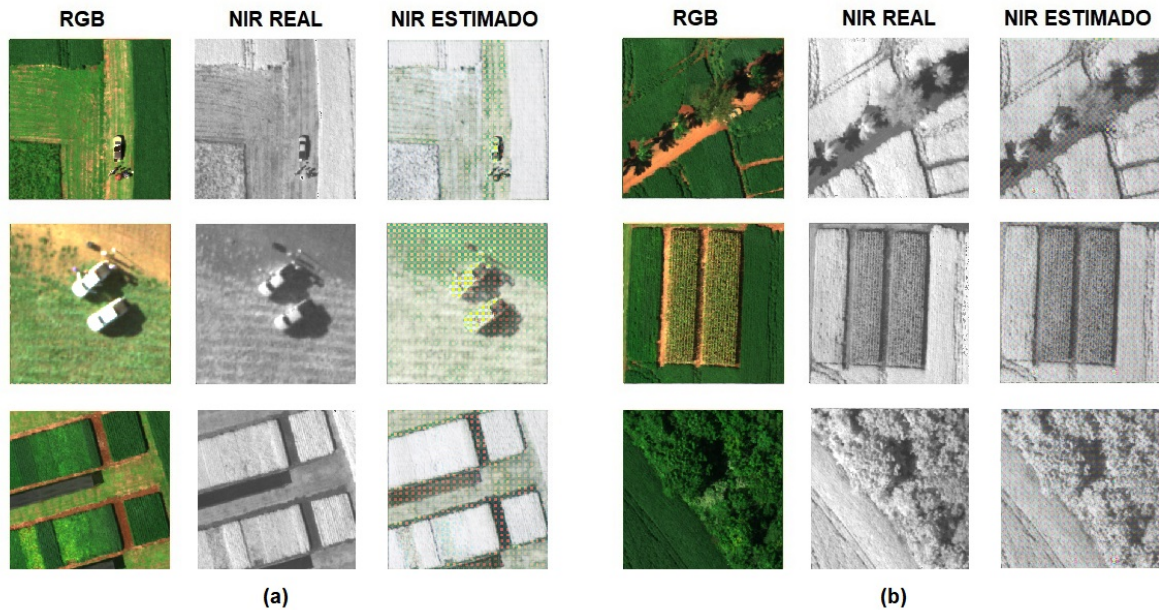
Loss	MS-SSIM	RMSE
\mathcal{L}_{L1}	0,894	16,414
\mathcal{L}_{Mix}	0,895	16,125
\mathcal{L}_{PMix}	0,888	16,357

4.1.4 Treinamentos com função $\mathcal{L}_{MS-SSIM}$ e $\mathcal{L}_{Perceptual}$

Além das redes treinadas anteriormente, foram treinadas NIRDLearn substituindo a função \mathcal{L}_{L1} na equação 2.20 pelas funções $\mathcal{L}_{MS-SSIM}$ (3.10) e $\mathcal{L}_{Perceptual}$. Os resultados de estimativa geraram *pixels* ruidosos até mesmo quando verificadas as imagens geradas a partir do conjunto de treinamento, afetando a qualidade da estimativa do NIR. Foram realizados os treinamentos com essas duas funções utilizando imagens RGB e CIE Lab. Para os dois modelos de cor, os resultados apresentaram defeitos, portanto, pode-se concluir que a presença de ruído produzido por estas duas funções de *loss* independe do modelo de cor utilizado nas imagens de entrada da rede. A Figura 4.22 mostra exemplos de resultados gerados por estas redes.

4.1.5 Considerações finais sobre as redes NIRDLearn treinadas

Como pode ser observado pelos resultados gerados pelas redes treinadas, a melhoria do resultado das estimativas de imagens NIR foi mais influenciada pela alteração do modelo de cor da imagem de entrada para modelos de cor que consideram componentes como cor e iluminação,

Figura 4.22: Exemplos de resultados ruidosos obtidos pelas funções $\mathcal{L}_{MS-SSIM}$ (a) e $\mathcal{L}_{Perceptual}$ (b).

Fonte: Elaborada pelo autor.

principalmente considerando-se o valor do RMSE. A Tabela 4.7 resume os resultados de similaridade estrutural e entre níveis de cinza para imagens estimadas a partir do conjunto de treinamento, destacando os melhores indicadores de similaridade. Na Tabela 4.8 é exibido o resumo dos resultados para imagens estimadas a partir do conjunto de validação.

Tabela 4.7: Resumo dos resultados de similaridade com imagens estimadas a partir do conjunto de treinamento para os três modelos de cor

Modelo de Cor	MS-SSIM			RMSE		
	\mathcal{L}_{L1}	\mathcal{L}_{Mix}	\mathcal{L}_{PMix}	\mathcal{L}_{L1}	\mathcal{L}_{Mix}	\mathcal{L}_{PMix}
RGB	0,930	0,938	0,931	17,250	17,258	17,779
CIE Lab	0,941	0,941	0,950	6,470	6,520	5,948
HSV	0,940	0,941	0,950	6,586	6,620	5,978

Tabela 4.8: Resumo dos resultados de similaridade com imagens estimadas a partir do conjunto de validação para os três modelos de cor

Modelo de Cor	MS-SSIM			RMSE		
	\mathcal{L}_{L1}	\mathcal{L}_{Mix}	\mathcal{L}_{PMix}	\mathcal{L}_{L1}	\mathcal{L}_{Mix}	\mathcal{L}_{PMix}
RGB	0,890	0,888	0,892	24,483	24,463	24,527
CIE Lab	0,895	0,897	0,891	16,589	16,2755	16,624
HSV	0,894	0,895	0,888	16,414	16,125	16,357

Os resultados com o conjunto de validação mostraram que os melhores resultados para os modelos de cor CIE Lab e HSV foram obtidos pelas redes treinadas usando a função \mathcal{L}_{Mix} , gerando valores de média altos para a similaridade estrutural e menores valores para a média

da diferença entre *pixels* estimados e reais. Em relação aos modelos de cor, o modelo RGB é o menos indicado para realizar a estimativa utilizando NIRDLearn. Os resultados de similaridade obtidos com os modelos CIE Lab e HSV apresentam valores muito próximos (diferença de 0,02 para MS-SSIM e de 0,15 para RMSE), portanto, ambos são indicados para uso com esta arquitetura de rede. A limitação se deu para estimativas de NIR em regiões que possuem porções de água, uma vez que para todas as configurações de rede treinadas, as imagens estimadas apresentaram baixa similaridade em relação às imagens NIR reais.

4.2 Estimativa de imagens NIR com o método NIRKNN

A partir desta seção serão descritas as etapas intermediárias e seus respectivos resultados para estimativa de imagens NIR utilizando NIRKNN. Serão descritos os métodos utilizados para selecionar os espectros de referência e para estimar imagens NIR, passando pelos resultados de classificação com KNN.

4.2.1 Seleção de espectros

Os primeiros experimentos foram realizados de modo a escolher os espectros de referência, *endmembers*, que melhor representam os materiais/objetos que estão presentes nas posições dos *pixels* das imagens hiperespectrais utilizados para a realização dos testes, uma vez que esses representantes das classes foram utilizados em função da equação 3.17 para a geração de um novo espectro.

A partir dos espectros selecionados de cada *pixel* classificado conforme descrito na Seção 3.3.2, foram avaliados quatro métodos clássicos da literatura, também já citados na Seção 3.3.2, além do método do espectro médio, para extração dos espectros mais representativos utilizando a acurácia da classificação com os métodos SAM e SID para verificar qual o método de extração de características apresenta o melhor desempenho em tarefas de classificação. Nesta subseção será exibida a acurácia atingida pelos classificadores para cada conjunto de características selecionadas por cada um dos algoritmos. A classificação foi realizada para cada *pixel* da imagem, verificando-se a sua similaridade espectral para cada espectro de referência. O *pixel* da imagem foi classificado de acordo com o espectro de referência mais similar.

4.2.1.1 Indian Pines

A Tabela 4.9 mostra as 16 classes em que os *pixels* da imagem Indian Pines foram divididos e a quantidade de amostras (espectros) para cada uma delas.

Tabela 4.9: Informações do *ground truth* do *data set* Indian Pines

	Classe	Amostras
1	Alfalfa	46
2	Corn-notill	1428
3	Corn-mintill	830
4	Corn	237
5	Grass-pasture	483
6	Grass-trees	730
7	Grass-pasture-mowed	28
8	Hay-windrowed	478
9	Oats	20
10	Soybean-notill	972
11	Soybean-mintill	2455
12	Soybean-clean	593
13	Wheat	205
14	Woods	1265
15	Buildings-Grass-Trees-Drives	386
16	Stone-Steel-Towers	93

O primeiro método testado para extração dos melhores espectros que representam cada uma das classes foi o método ATGP. Cada um dos espectros selecionados foi utilizado como espectro de referência para classificar as imagens utilizando os algoritmos SAM e SID, sendo que o valor 0,3 foi escolhido como *threshold* para o classificador, pois foi o valor que apresentou melhor resultado de segmentação. O classificador SAM apresentou acurácia de 25,66%. A acurácia com o classificador SID foi de 24,28%.

Em sequência, o método de seleção de espectros avaliado foi o PPI. Este algoritmo apresentou espectros semelhantes entre si para representar cada uma das classes da imagem, assim como o algoritmo ATGP, porém, espectros com assinatura espectrais muito diferentes foram escolhidos para representar algumas classes em relação ao método utilizado anteriormente. A classificação com os novos espectros selecionados apresentou resultado inferior em relação àquela com os espectros selecionados pelo método ATGP para o classificador SAM. A acurácia com o classificador SAM foi de 22,64% e para o classificador SID, a acurácia foi de 25,29%.

Com o método de extração N-FINDR verificou-se o mesmo padrão que foi encontrado nos espectros extraídos pelos métodos utilizados anteriormente, com assinaturas espectrais semelhantes para diferentes classes, evidenciando assim que existem materiais com características

muito parecidas no conjunto de dados. A acurácia do classificador SAM utilizando os espectros selecionados por este método foi de 23,98%. Já com o classificador SID, o resultado da classificação foi superior ao do outro classificador, apresentando acurácia de 32,31%.

O próximo método testado para a extração dos espectros que melhor caracterizam cada classe do *ground truth* foi o FIPPI, porém, assim como nos espectros selecionados pelos métodos anteriormente testados, os espectros selecionados para cada classe são semelhantes entre si. Os resultados da classificação com os novos *endmembers* utilizando os dois classificadores foi superior aos resultados com os espectros escolhidos pelos métodos anteriormente testados, apresentando acurácia de 31,02% para a classificação com SAM e 33,90% para a classificação com o SID.

O último método para encontrar espectros de referência que foi avaliado com este *data set* foi o cálculo de espectros médios para cada uma das classes. Após o cálculo destes espectros, foi possível verificar que as assinaturas espectrais também são muito semelhantes entre si. Os espectros médios foram usados como espectros de referência para os classificadores SAM e SID, apresentando melhores resultados que os obtidos utilizando os espectros selecionados pelo algoritmo FIPPI. A acurácia com o classificador SAM foi de 33,33% e com o classificador SID, a acurácia foi de 40,08%.

A partir dos resultados obtidos utilizando os métodos de seleção de espectros que melhor representam cada uma das classes, é possível concluir que os objetos que representam as classes são muito similares entre si, tendo assinaturas espectrais parecidas, refletindo na baixa acurácia de classificação para todos os métodos de extração de *endmembers*. Também é possível observar que para esta imagem hiperespectral, o método que forneceu os espectros mais representativos para os classificadores foi o cálculo de espectros médios, conforme pode ser visto na Tabela 4.10.

Tabela 4.10: Acurácia dos classificadores para o *data set* Indian Pines por algoritmo de seleção de *endmembers* (%)

	ATGP	PPI	N-FINDR	FIPPI	Espectros Médios
SAM	25,66	22,64	23,98	31,02	33,33
SID	24,28	25,29	32,31	33,90	40,08

4.2.1.2 Salinas

A Tabela 4.11 mostra a divisão da imagem hiperespectral Salinas em 16 classes e a quantidade de amostras (espectros) para cada uma delas.

Tabela 4.11: Informações do *ground truth* do *data set* Salinas

	Classe	Amostras
1	Brocoli_green_weeds_1	2009
2	Brocoli_green_weeds_2	3726
3	Fallow	1976
4	Fallow_rough_plow	1394
5	Fallow_smooth	2678
6	Stubble	3959
7	Celery	3579
8	Grapes_untrained	11271
9	Soil_vinyard_develop	6203
10	Corn_senesced_green_weeds	3278
11	Lettuce_romaine_4wk	1068
12	Lettuce_romaine_5wk	1927
13	Lettuce_romaine_6wk	916
14	Lettuce_romaine_7wk	1070
15	Vinyard_untrained	7268
16	Vinyard_vertical_trellis	1807

Assim como feito no *data set* Indian Pines, o primeiro método testado para extração dos melhores espectros representantes das classes pertencentes ao *ground truth* foi o método ATGP. Para realizar os testes de classificação, o valor 0,3 foi escolhido como valor de *threshold* para os classificadores SAM e SID, assim como feito nos experimentos com a imagem Indian Pines.

Ao observar os espectros selecionados pelo algoritmo ATGP, verifica-se que existem assinaturas espectrais com curvas muito semelhantes para representar cada uma das classes, porém, os valores nas primeiras bandas são bem distintos entre algumas classes, sendo possível presumir que haverá resultados de classificação superiores em relação aos resultados obtidos com a imagem anterior. A classificação com o algoritmo SAM apresentou acurácia de 29,59%. Já a acurácia com o classificador SID foi de 34,61%.

Dando sequência, o método PPI apresentou assinaturas espectrais com características diferentes daquelas extraídas pelo algoritmo ATGP, fazendo com que espectros com assinaturas espectrais muito diferentes fossem escolhidos para representar as mesmas classes em relação ao método utilizado anteriormente. A classificação com as assinaturas espectrais extraídas do conjunto de dados apresentou resultados superiores em relação às extraídas pelo método ATGP para ambos os classificadores. A acurácia com o classificador SAM foi de 44,45% e para o classificador SID, a acurácia foi de 42,15%.

Com o método de extração N-FINDR verificou-se o mesmo padrão que foi encontrado nos espectros extraídos pelos métodos utilizados anteriormente, sendo que as principais diferenças

entre os espectros se dão nas primeiras bandas da imagem e cada um dos algoritmos avaliados com esta imagem seleciona algumas assinaturas com características distintas para representar a mesma classe. A acurácia do classificador SAM utilizando os espectros selecionados por este método foi de 52,10%. O classificador SID apresentou acurácia de 48,89%, apresentando resultados de classificação que superam os resultados do algoritmo de extração PPI.

Os espectros extraídos pelo algoritmo FIPPI, assim como nos espectros selecionados pelos métodos anteriormente testados, apresentam distinção entre si nas primeiras bandas da imagem, sendo que os espectros selecionados por este método também possuem assinaturas com pouca semelhança às assinaturas selecionadas pelos outros métodos para representar as mesmas classes. Os resultados da classificação com estes *endmembers* utilizando os dois classificadores foram superiores aos resultados com os espectros escolhidos pelos métodos anteriormente testados, apresentando acurácia de 54,72% para a classificação com SAM e 53,43% para a classificação com o SID.

O cálculo de espectros médios para cada uma das classes também resultou em assinaturas espectrais com características distintas das assinaturas extraídas pelos métodos da literatura. Após o cálculo dos espectros médios, foi possível verificar que as assinaturas espectrais possuem valores que se diferenciam entre as primeiras bandas da imagem hiperespectral. Os espectros médios foram usados como espectros de referência para os classificadores SAM e SID, apresentando melhores resultados que os obtidos pelos métodos existentes na literatura. A acurácia com o classificador SAM foi de 62,14%. Com o classificador SID, a acurácia foi de 59,49%.

A partir dos resultados obtidos utilizando os métodos de seleção de espectros que melhor representam cada uma das classes, é possível concluir que as primeiras 100 bandas da imagem seriam suficientes para fazer a distinção entre os materiais, resultando em uma precisão maior para os resultados de classificação. Também é possível observar que para esta imagem hiperespectral, assim como para a imagem Indian Pines, o método que resultou nos espectros mais representativos para os classificadores foi o cálculo de espectros médios, conforme pode ser visto na Tabela 4.12.

Tabela 4.12: Acurácia dos classificadores para o *data set* Salinas por algoritmo de seleção de *endmembers* (%)

	ATGP	PPI	N-FINDR	FIPPI	Espectros Médios
SAM	29,59	44,45	52,10	54,72	62,14
SID	34,61	42,15	48,89	53,43	59,49

4.2.1.3 Pavia

A Tabela 4.13 possui informações sobre as classes presentes no *data set* Pavia, assim como a quantidade de amostras (*pixels*) existentes para cada uma delas.

Tabela 4.13: Informações do *ground truth* do *data set* Pavia

	Classe	Amostras
1	Water	824
2	Trees	820
3	Asphalt	816
4	Self-Blocking Bricks	808
5	Bitumen	808
6	Tiles	1260
7	Shadows	476
8	Meadows	824
9	Bare Soil	820

Seguindo os mesmos procedimentos realizados para os *data sets* anteriores, o primeiro método de seleção de espectros avaliado foi o ATGP. Cada um dos espectros selecionados foi utilizado como espectro de referência para classificar as imagens utilizando os algoritmos SAM e SID. Foi escolhido o valor 0,3 como valor de *threshold* para os classificadores.

Ao observar os espectros selecionados pelo algoritmo, verifica-se que existem espectros que são bem distintos entre si, mostrando que os materiais presentes na imagem possuem naturezas distintas, o que reflete no resultado da classificação para todos os outros métodos, sendo, possivelmente, resultados melhores do que os obtidos com as outras imagens hiperespectrais analisadas anteriormente. A classificação com o SAM apresentou acurácia de 48,10%. A acurácia com o classificador SID foi de 69,14%.

O algoritmo PPI apresentou espectros com assinaturas diferentes daqueles selecionados pelo ATGP para algumas das classes da imagem. A classificação com estes novos espectros selecionados apresentou resultado muito superior em relação ao método ATGP para o classificador SAM. A acurácia com o classificador SAM foi de 81,79% e para o classificador SID, a acurácia foi de 58,96%, sendo esta inferior à acurácia com os espectros do ATGP para o mesmo classificador.

Nas assinaturas espectrais retornadas pelo N-FINDR verificou-se que novamente, as curvas formadas por cada espectro tem padrões diferentes das formadas pelos espectros extraídos por outros algoritmos ATGP e PPI. A acurácia do classificador SAM utilizando os espectros selecionados pelo N-FINDR foi de 79,34%, inferior ao resultado com os espectros extraídos pelo PPI. Com o classificador SID, o resultado da classificação apresentou acurácia de 73,06%,

sendo superior ao resultado do método PPI para este mesmo classificador.

O algoritmo FIPPI também apresentou espectros com características distintas para representar as mesmas classes em relação aos demais métodos avaliados. O resultado da classificação com estes *endmembers* utilizando o classificador SAM foi superior aos resultados com os espectros escolhidos pelos métodos testados até então, apresentando acurácia de 82,94%. Já para o classificador SID, apresentou o pior desempenho, com acurácia de 57,61% para a classificação.

Os espectros médios, novamente, apresentaram os melhores resultados para a classificação da imagem. A acurácia com o classificador SAM foi de 88,71% e com o classificador SID, a acurácia foi de 84,27%.

O resumo dos resultados obtidos pode ser visto na Tabela 4.14 e, a partir desta, pode se concluir que mais uma vez, o uso dos espectros médios como espectros de referência para os classificadores se mostrou mais preciso que os outros métodos avaliados.

Tabela 4.14: Acurácia dos classificadores para o *data set* Pavia por algoritmo de seleção de *endmembers* (%)

	ATGP	PPI	N-FINDR	FIPPI	Espectros Médios
SAM	48,10	81,79	79,34	82,94	88,71
SID	69,14	58,96	73,06	57,61	84,27

4.2.1.4 Chile

Os resultados de classificação com a imagem de plantação de cítricos do Chile não apresentaram bons resultados para nenhum dos métodos de extração de *endmembers*, uma vez que os *pixels* de cada classe foram coletados usando pequenos *frames* da imagem, e dentro dessa pequena imagem, existem muitos *pixels* que representam materiais diferentes, refletindo assim na seleção incorreta dos espectros. O *ground truth* também não foi definido adequadamente, pois essa imagem tem uma resolução maior do que as outras da literatura. Por ter resolução maior, em cada *pixel* da imagem existe mistura de materiais menor do que nos *pixels* das imagens de resolução menor. Por exemplo, mesmo que existam partes com solo exposto nas plantação da imagem Indian Pines, os espectros do solo e das plantas se fundem, ficando praticamente uniformes. Nesta imagem de cítricos do Chile, as partes com solo descoberto estão bem evidentes e as assinaturas espectrais do solo e da plantação são bem diferentes entre si. A acurácia para as classificações com os espectros escolhidos ficou abaixo dos 10% para todos os métodos de extração. A imagem *ground truth* precisaria ser refeita, classificando individualmente cada um dos *pixels* que a compõem.

4.2.2 Estimativa do NIR

Como pode ser observado nos resultados dos experimentos anteriores, o espectro médio apresentou os melhores resultados de classificação para as imagens utilizadas neste trabalho de pesquisa, portanto, espectros médios de cada classe que existam dentro das imagens serão utilizados para estimar novos espectros utilizando a equação 3.17. Para cada imagem foi criado um banco de espectros, sendo que para as imagens Indian Pines, Salinas e Pavia serão utilizados o mesmo conjunto de espectros usados anteriormente para a verificação do melhor método de extração de *endmembers*. Para a imagem de cítricos do Chile, a imagem toda (Figura 3.15) foi utilizada para coletar individualmente 12.848 *pixels* e estes foram divididos em 16 categorias, conforme a Tabela 4.15.

Tabela 4.15: Relação de classes e número de amostras para a imagem da plantação de cítricos do Chile.

Classe	Número de amostras (Pixels)
Estrada de Terra 1	415
Estrada de Terra 2	640
Lago 1	672
Lago 2	663
Plantação 1	861
Plantação 2	846
Plantação 3	660
Plantação 4	1.045
Plantação 5	976
Solo 1	1.024
Solo 2	900
Solo 3	820
Solo 4	877
Solo Nativo 1	772
Solo Nativo 2	647
Vegetação Nativa	1.030
Total	12.848

Foram realizados experimentos com as quatro imagens hiperespectrais, tendo como ponto de partida imagens RGB geradas a partir de 3 bandas de cada imagem. Para as três imagens da literatura, foram utilizadas as bandas 30, 20 e 10 para gerar os canais R,G e B, respectivamente. As bandas utilizadas para gerar a imagem RGB da plantação de cítricos já foram descritas na seção 3.3.1. Tendo posse da imagem de três bandas, foi calculado o índice de vegetação MPRI para cada um dos *pixels* da imagem e os valores RGB foram convertidos para os outros dois modelos de cor propostos, HSV e CIE Lab. Serão exibidos os resultados da classificação com o KNN para cada modelo de cor, sendo que o melhor resultado de classificação será utilizado

para a realização da estimativa do espectro novo. O espectro novo também será criado para a imagem de cítricos do Chile de modo a validar o método, uma vez que não há informações definidas por especialistas sobre esta imagem.

Conjuntos de teste foram escolhidos para avaliar qual modelo de cor apresentava melhores resultados de classificação com o algoritmo KNN. Para a imagem Indian Pines, foram selecionados 5 *pixels* que não estão no conjunto de treinamento para cada classe. Para as imagens Salinas e Pavia, foram selecionados 50 *pixels* para cada classe, sendo estes diferentes daqueles que estão nos respectivos conjuntos de treinamento. Já para a imagem de cítricos do Chile, foram coletadas 4190 amostras individuais para cada classe, sendo um conjunto de teste desbalanceado.

Para a imagem Indian Pines, os resultados de classificação com o algoritmo KNN apresentaram acurácia muito baixa para os três modelos de cor analisados, com 56,61% de amostras classificadas corretamente para o modelo RGB; 53,18% para o modelo de cor HSV e 55,21% para o modelo CIE Lab. Já para a imagem Salinas, o classificador obteve ótimos resultados, todos eles com valor de acurácia bem próximos. A acurácia para este *data set* com os modelos RGB e HSV foi de 93,76%, já para o modelo CIE Lab, a acurácia apresentou leve aumento, indo para 93,88%. Com a imagem Pavia, os resultados de classificação foram medianos, apresentando de 69,39%, 71,32% e 71,12% para os modelos de cor RGB, HSV e CIE Lab, respectivamente. Os resultados de classificação também foram medianos para o conjunto de dados da imagem do Chile. Com o RGB, a acurácia foi de 64,46%. Já com os outros dois modelos de cor, a acurácia para o HSV foi de 64,84% e a do CIE Lab foi de 64,86%. A Tabela 4.16 resume os resultados de classificação com o KNN.

Tabela 4.16: Resultados do classificador KNN de acordo com o modelo de cores e os *data sets*.

Imagens	RGB	HSV	CIE Lab
Indian Pines	56,61%	53,18%	55,21%
Salinas	93,76%	93,76%	93,88%
Pavia	69,39%	71,33%	71,12%
Chile	64,46%	64,84%	62,86%

Conforme citado anteriormente, para avaliação dos resultados da estimativa do novo espectro, e por consequência, a imagem NIR, seria escolhida a imagem que obtivesse melhores resultados de classificação com o KNN e a imagem de cítricos do Chile. Para a imagem com melhores resultados de classificação como pode ser visto na Tabela 4.16, a imagem Salinas foi escolhida. Para as duas imagens, utilizando as distribuições de probabilidade das classificações para os três modelos de cor e os espectros médios das classes, os novos espectros foram calculados utilizando a equação 3.17. Os espectros dos *pixels* do conjunto de validação foram

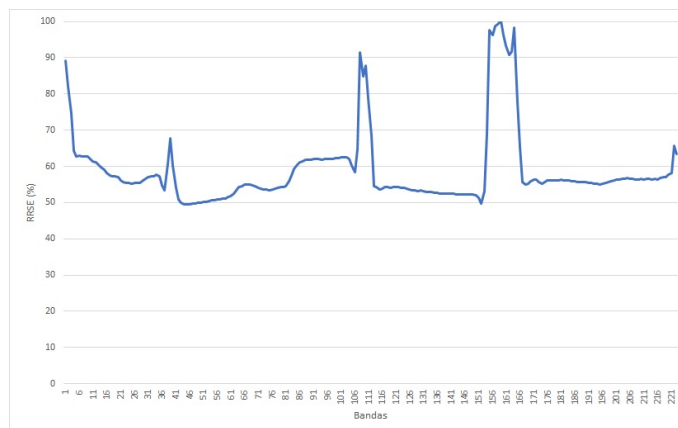
confrontados com os espectros médios das classes as quais foram classificados pelo KNN. Foi calculado o erro entre os espectros, utilizando o método *Root Relative Squared Error* (RRSE), definido pela equação 4.3.

$$RRSE(i) = \sqrt{\frac{\sum_{j=1}^n (\bar{S}_{i,j} - W_{i,j})^2}{\sum_{j=1}^n (\bar{W}_{i,j} - W_{i,j})^2}} \quad (4.3)$$

Na equação anterior, n se refere ao número de amostras do conjunto de validação, $\bar{S}_{i,j}$ representa o espectro médio da classe com o qual a amostra j foi classificada, referente ao comprimento de onda i , $W_{i,j}$ é o valor da reflectância do *pixel* de validação j no comprimento de onda i e $\bar{W}_{i,j}$ é a média dos valores de reflectância dos *pixels* de validação para o comprimento de onda i .

O RRSE entre as bandas dos espectros do conjunto de teste e as bandas dos espectros médios, sendo estes referentes aos rótulos das amostras classificadas pelo KNN, foram calculados, sendo que o RRSE médio foi de 58,88% para o conjunto de teste extraído da imagem Salinas classificado com RGB, 60,56% para o classificado com HSV e 59,87% para o conjunto classificado com CIE Lab, mostrando que há pouca variação nos resultados de classificação para cada amostra usando os três modelos de cor propostos neste trabalho. As Figuras 4.23, 4.24 e 4.25 mostram o RRSE para cada banda em forma de gráfico para os conjuntos classificados com RGB, HSV e CIE Lab, respectivamente, sendo possível notar que as curvas de erro são muito semelhantes.

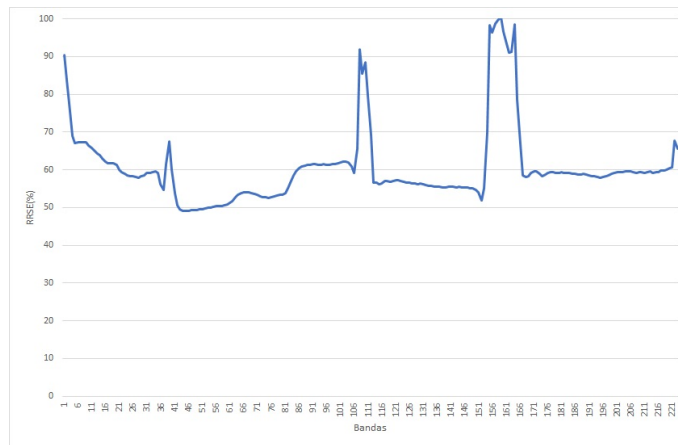
Figura 4.23: RRSE entre pixels do conjunto de validação Salinas RGB e seus respectivos espectros médios.



Fonte: Elaborada pelo autor.

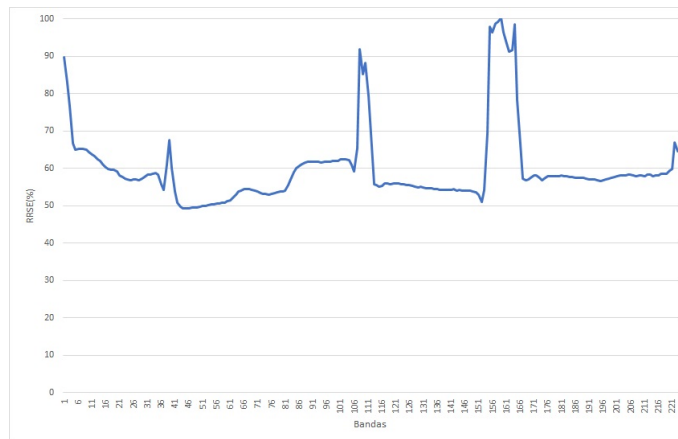
O RRSE entre as bandas dos espectros do conjunto de teste e as bandas dos espectros médios relativos às amostras classificadas foram calculados, sendo que o RRSE médio foi de 52,73% para o conjunto de teste extraído da imagem de cítricos Chile classificado com RGB,

Figura 4.24: RRSE entre pixels do conjunto de validação Salinas HSV e seus respectivos espectros médios.



Fonte: Elaborada pelo autor.

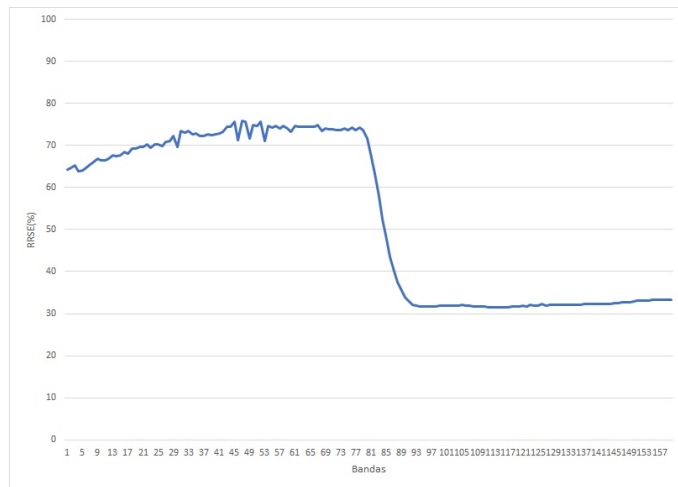
Figura 4.25: RRSE entre pixels do conjunto de validação Salinas CIE Lab e seus respectivos espectros médios.



Fonte: Elaborada pelo autor.

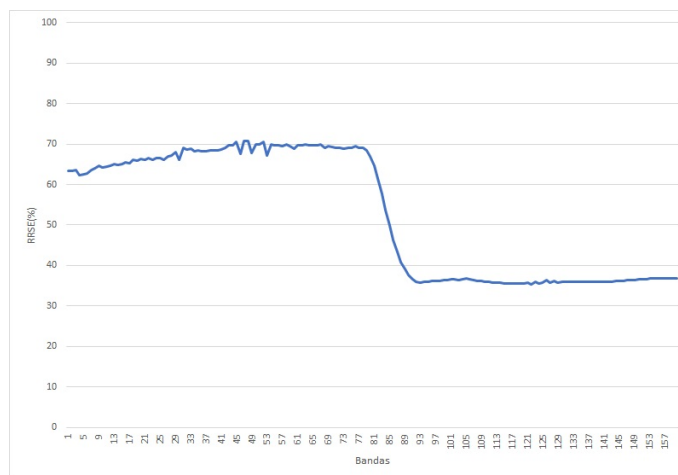
52,86% para o classificado com HSV e 54,22% para o conjunto classificado com CIE Lab. Conforme o resultado do cálculo do erro entre os espectros dos *pixels* e os espectros médios relativos às classes as quais foram classificados, pode se concluir que existe pouca variação nos rótulos atribuídos para a mesma amostra nos processos de classificação com os dados RGB, HSV e CIE Lab. As Figuras 4.26, 4.27 e 4.28 mostram o RRSE para cada banda em forma de gráfico para os conjuntos classificados com RGB, HSV e CIE Lab. A observação destes gráficos evidencia a existência de pouca variação entre os resultados de classificação para cada amostra do conjunto de teste para os modelos de cores propostos. As curvas formadas tem aspecto similar, sendo possível concluir que os espectros médios referentes ao resultado de classificação para cada amostra eram iguais para os três modelos de cor na maioria das vezes.

Figura 4.26: RRSE entre pixels do conjunto de validação de cítricos do Chile RGB e seus respectivos espectros médios.



Fonte: Elaborada pelo autor.

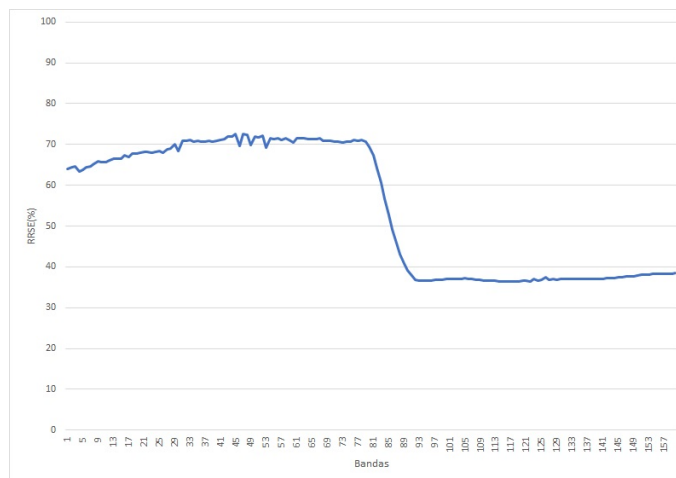
Figura 4.27: RRSE entre pixels do conjunto de validação de cítricos do Chile HSV e seus respectivos espectros médios.



Fonte: Elaborada pelo autor.

O RRSE calculado com os novos espectros gerados para cada pixel mostrou grande queda nos erros médios para os conjuntos das duas imagens e para todos os modelos de cor, sendo que para os conjuntos de teste extraídos da imagem Salinas, o erro médio ficou em 8,12% para os três modelos de cor. Para os conjuntos de teste da imagem de cítricos do Chile, o erro médio ficou em 8,48% para o RGB, 7,77% para o HSV e 7,41% para o CIE Lab. Os espectros novos, gerados com o uso da distribuição de probabilidade dos k vizinhos mais próximos e os espectros médios, se aproximam mais dos espectros de referência (espectros médios) para cada amostra classificada. As Figuras 4.29, 4.30 e 4.31 mostram os gráficos do RRSE para os modelos de cor RGB, HSV e CIE Lab, respectivamente, para os conjuntos da imagem Salinas. As Figuras 4.32, 4.33 e 4.34 mostram os gráficos do RRSE para os modelos de cor RGB, HSV e CIE Lab

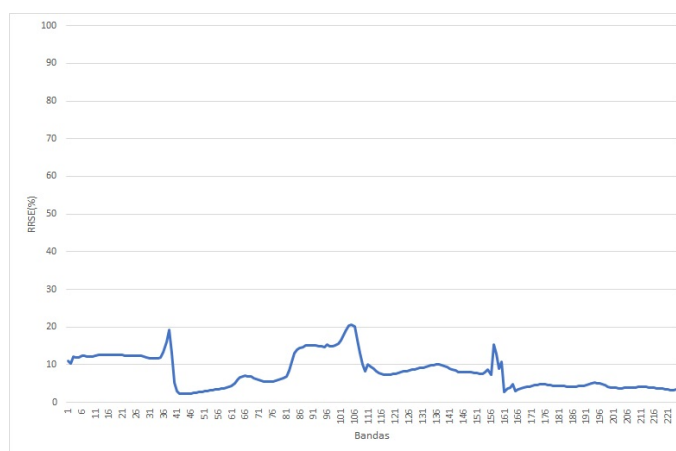
Figura 4.28: RRSE entre pixels do conjunto de validação de cítricos do Chile CIE Lab e seus respectivos espectros médios.



Fonte: Elaborada pelo autor.

para a imagem de cítricos Chile.

Figura 4.29: RRSE entre os espectros gerados para o conjunto de validação Salinas RGB e seus respectivos espectros médios.

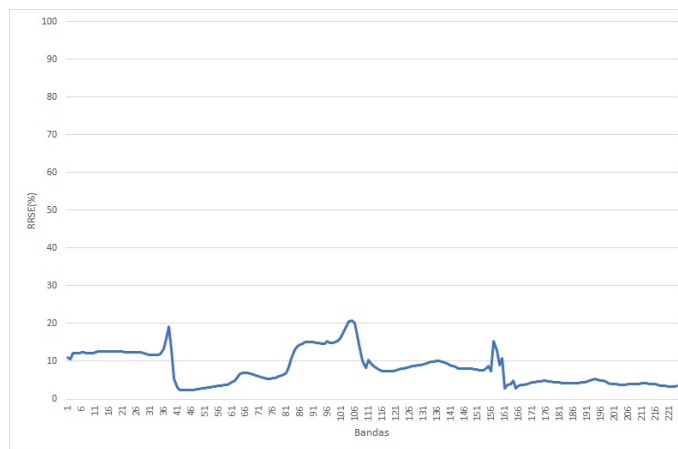


Fonte: Elaborada pelo autor.

A Tabela 4.17 resume as informações anteriormente citadas sobre os RRSEs médios para os espectros da imagem original e para os espectros estimados, confirmando que os espectros estimados pela equação 3.17 são muito mais próximos dos espectros de referência.

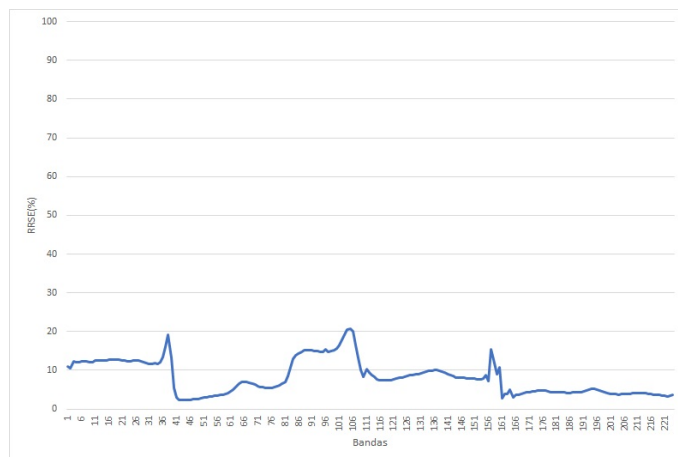
Para efeitos de ilustração dos resultados da estimativa do NIR a partir de uma imagem RGB, imagens NIR foram criadas utilizando os espectros novos gerados a partir dos resultados de classificação com KNN para cada um dos modelos de cor. A imagem Salinas foi criada por completo e uma parte da imagem da plantação no Chile foi escolhida para ilustração dos resultados do método. Para a imagem Salinas, foram recriados apenas os *pixels* que possuem informação contida no *ground truth*. Os *pixels* que não possuem informação foram pintados

Figura 4.30: RRSE entre os espectros gerados para o conjunto de validação Salinas HSV e seus respectivos espectros médios.



Fonte: Elaborada pelo autor.

Figura 4.31: RRSE entre os espectros gerados para o conjunto de validação Salinas CIE Lab e seus respectivos espectros médios.

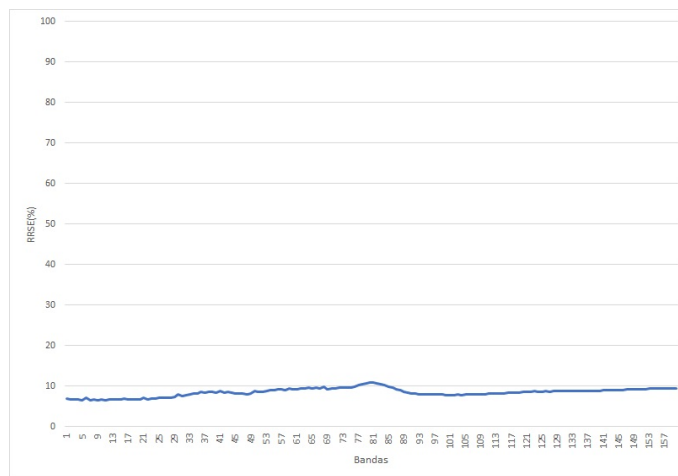


Fonte: Elaborada pelo autor.

com a cor preta. A Figura 4.35 mostra como ficaram as imagens NIR criadas a partir dos espectros novos em comparação as imagens NIR originais da imagem Salinas (banda 80) e Chile (banda 137).

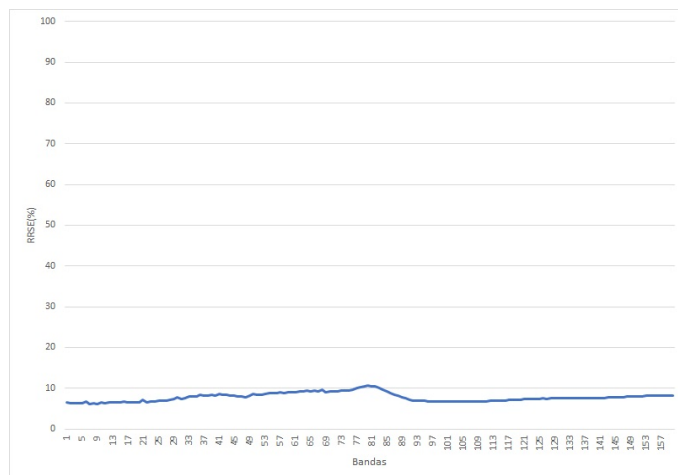
A Tabela 4.18 mostra o resultado do cálculo de similaridade entre as imagens estimadas e as imagens NIR reais presentes na Figura 4.35 para cada um dos três modelos de cor.

Figura 4.32: RRSE entre os espectros gerados para o conjunto de validação de cítricos do Chile RGB e seus respectivos espectros médios.



Fonte: Elaborada pelo autor.

Figura 4.33: RRSE entre os espectros gerados para o conjunto de validação de cítricos do Chile HSV e seus respectivos espectros médios.

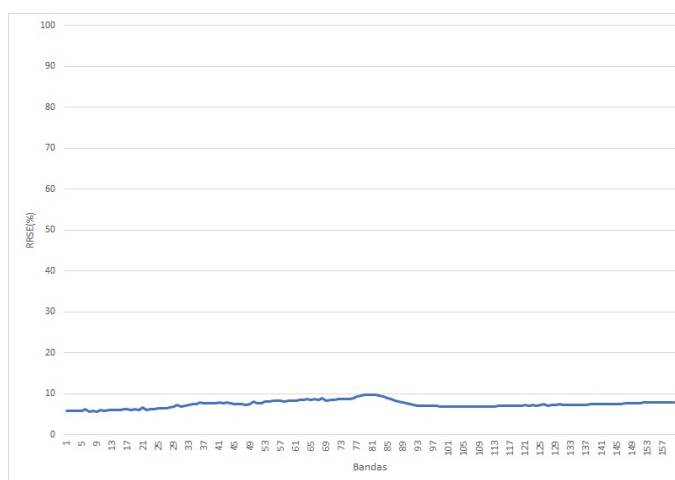


Fonte: Elaborada pelo autor.

4.3 Considerações finais sobre estimativa de imagens NIR utilizando NIRKNN

Os experimentos realizados mostram que o método proposto pode ser utilizado para fazer a estimativa de imagens NIR para aplicação na agricultura, reproduzindo as imagens com erro inferior a 9% em relação aos espectros puros, sendo a utilização dos espectros médios como *endmembers* responsável pelos melhores resultados, obtendo alto grau de similaridade estrutural (MS-SSIM = 0,97) para a imagem devidamente rotulada por especialistas (Salinas). Para a imagem com *ground truth* não definido por especialistas (Chile), as imagens estimadas apresentaram similaridade relativamente alta (MS-SSIM).

Figura 4.34: RRSE entre os espectros gerados para o conjunto de validação de cítricos do Chile CIE Lab e seus respectivos espectros médios.



Fonte: Elaborada pelo autor.

Tabela 4.17: RRSE entre espectros originais e médios X RRSE entre espectros estimados e médios

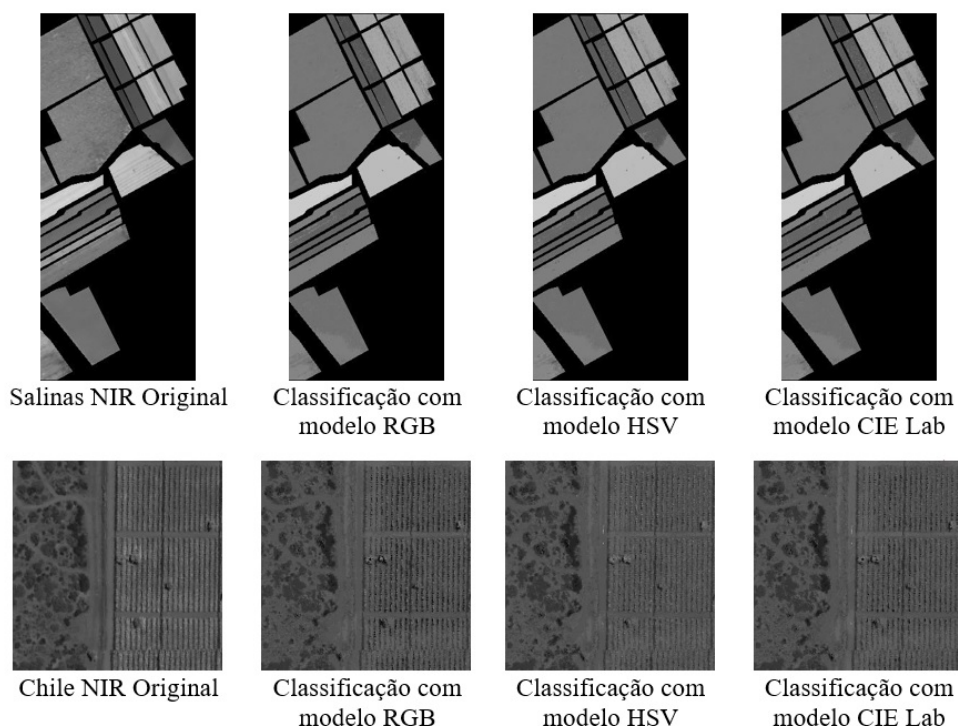
Assinaturas espectrais	RGB	HSV	CIE Lab
Salinas original	58,88%	60,56%	59,87%
Salinas estimado	8,12%	8,12%	8,12%
Chile original	52,73%	52,86%	54,22%
Chile estimado	8,48%	7,77%	7,41%

Tabela 4.18: Resultados de similaridade entre imagens NIR reais e imagens estimadas com o método NIRKNN.

Modelo de Cor	Salinas		Chile	
	MS-SSIM	RMSE	MS-SSIM	RMSE
RGB	0,97	9,74	0,78	17,26
CIE Lab	0,96	9,86	0,81	14,65
HSV	0,96	9,84	0,80	14,25

Apesar do bom resultado, para que o método apresente bom desempenho, é preciso que exista uma grande base de dados espectral, com várias classes bem definidas da região em que se deseja fazer o mapeamento aéreo. A partir destas classes serão escolhidos os espectros de referência para a realização da estimativa do novo espectro. No caso deste trabalho, o cálculo do espectro médio se mostrou a melhor alternativa para definição de espectros de referência. Os materiais presentes na região correspondente à imagem (tipos diferentes de vegetação, tipos de solo, etc) também devem ser bem conhecidos para definir a base de dados da região de uma fazenda, coletando uma grande quantidade de espectros "puros" para cada uma das classes.

Figura 4.35: Comparação entre imagens NIR originais e os resultados da estimativa de espectro para os modelos RGB, HSV e CIE Lab.



Fonte: Elaborada pelo autor.

4.4 Comparação entre os dois métodos - NIRDLearn e NIRKNN

Foi realizado um experimento de modo a verificar o desempenho dos dois métodos utilizando o mesmo conjunto de imagens. Como um dos métodos (NIRKNN) necessita de *endmembers* para a realização da estimativa, foi escolhida uma imagem hiperespectral (cítricos do Chile, escolhida por ter maiores dimensões, possibilitando extração de mais amostras para treinamento), uma vez que as imagens multiespectrais utilizadas para o treinamento da rede neural (NIRDLearn) possuem apenas cinco bandas espectrais. Foram selecionadas 10 amostras de 256×256 pixels para a realização da validação para os dois métodos. A base de dados utilizada para a classificação do KNN no método NIRKNN é a mesma mostrada na Tabela 4.15. Os *endmembers* escolhidos foram os espectros médios e o modelo de cor escolhido foi o RGB. Pelo motivo da imagem hiperespectral ter sido capturada por um sensor diferente e com resolução espacial maior (cada *pixel* da imagem de cítricos do Chile equivale a 50 cm de largura do objeto), a rede NIRDLearn precisou ser treinada com amostras desta imagem. A rede foi treinada com 100 amostras de 256×256 pixels, convertidas para o modelo CIE Lab, extraídas da imagem hiperespectral de cítricos do Chile, sendo a função \mathcal{L}_{Mix} escolhida para o cálculo do custo. Como a maior parte das amostras foi utilizada para o treinamento da rede, foram utilizadas ape-

nas 10 amostras para realização do estudo comparativo, uma vez que a rede gera resultados de estimativas com grande similaridade para amostras já conhecidas.

Para avaliar a similaridade entre as imagens estimadas e as imagens NIR reais, foram utilizadas as métricas MS-SSIM e RMSE. A Figura 4.36 mostra a comparação de alguns resultados de estimativa para os dois métodos propostos. As médias dos resultados de similaridade são mostrados na Tabela 4.19;

Tabela 4.19: Média da similaridade entre imagens NIR reais e estimadas para os dois métodos.

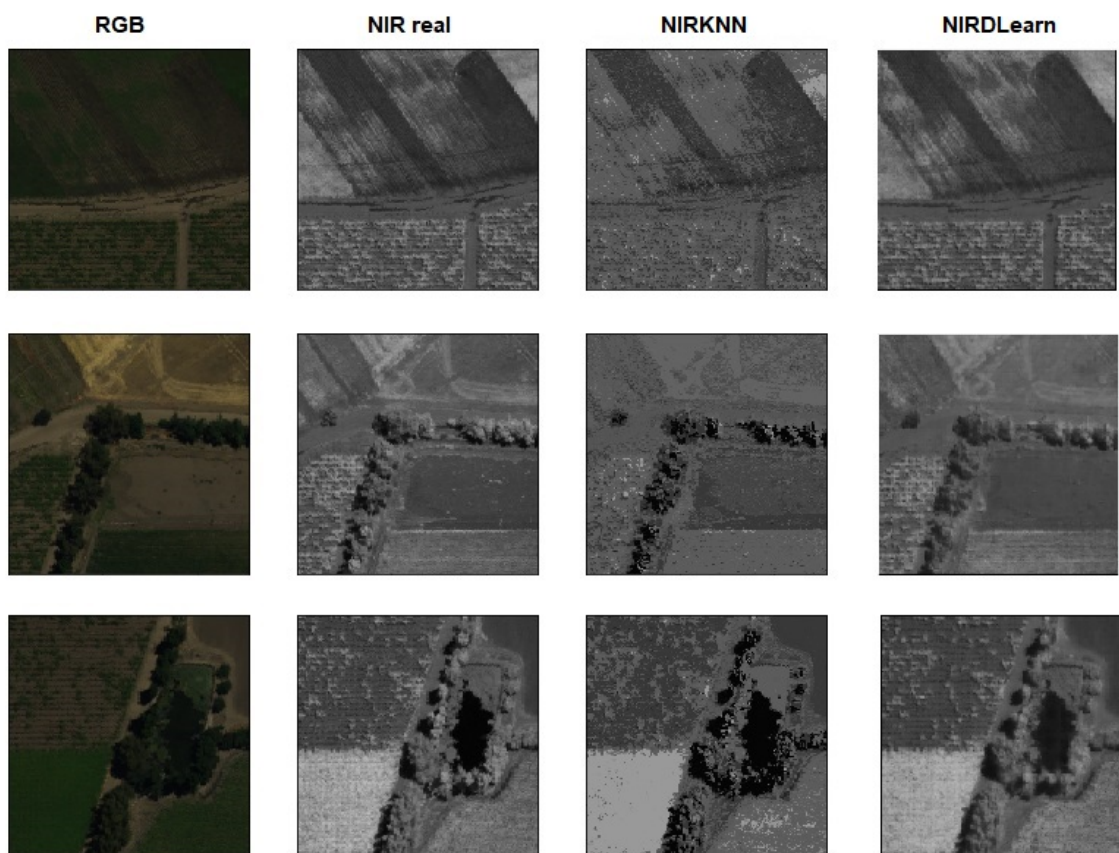
Método	MS-SSIM	RMSE
<i>NIRKNN</i>	0,775	20,64
<i>NIRDLearn</i>	0,912	10,358

Os resultados, tanto visuais, quanto numéricos, mostraram que o desempenho da rede NIRDLearn é superior na tarefa de estimativa de imagens NIR, gerando imagens mais similares às imagens reais. O resultado pode ter sido favorável à rede NIRDLearn pelo motivo da imagem hiperespectral não ter um *ground truth* preciso. A imagem Salinas, que possui *ground truth* bem definido, não foi utilizada para a realização do estudo comparativo por impossibilitar a extração de muitas amostras para o treinamento da rede (a imagem possui 217 x 512 *pixels*, enquanto a imagem de cítricos do Chile possui 8119 x 1763 *pixels*).

4.5 Considerações Finais

Neste capítulo foram apresentados os resultados de estimativa de imagens NIR para os dois métodos propostos, NIRDLearn e NIRKNN. O próximo capítulo apresenta conclusões sobre a tese, assim como proposta de trabalhos futuros e trabalhos publicados, que em função desta tese.

Figura 4.36: Resultados de estimativa de imagens NIR para os dois métodos propostos.



Fonte: Elaborada pelo autor.

Capítulo 5

CONCLUSÕES

Os estudos e experimentos conduzidos nesta tese tornaram possíveis avaliar dois métodos para a realização da estimativa de imagens NIR a partir de imagens RGB, NIRKNN, sendo um método original (estimativa de espectros a partir de assinaturas espectrais de referência previamente obtidas) e a substituição da função de *Loss* da rede *Pix2Pix* e alteração do modelo de cor das imagens de entrada, NIRDLearn.

O método NIRKNN apresenta uma média de erros entre bandas de espectros puros e estimados abaixo de 9%, com grande similaridade em relação às imagens NIR reais, porém, as imagens NIR geradas possuem baixa resolução e necessitam de uma base de dados muito grande para extração de *endmembers*, além de ser extremamente necessário ter o conhecimento da área em que se deseja realizar o mapeamento aéreo, criando um *ground truth* preciso para que se possa garantir alta qualidade do resultado de estimativa. A disponibilidade de assinaturas espectrais para a realização da estimativa por este método torna-se um fator limitante, uma vez que é necessário um sensor hiperespectral para aquisição de dados, ou então utilizar imagens de satélites, que possuem resolução espacial muito baixa e, por consequência, possuem mais mistura de materiais terrestres por *pixel*, dificultando a obtenção de espectros de referência que representam cada material e levando a estudos de métodos adequados de *spectral unmixing*.

O resultado do NIRDLearn, com a arquitetura de *Deep Learning Pix2Pix*, mostrou-se uma solução mais simples para estimativa de imagens NIR a partir de uma imagem RGB comum, necessitando apenas dos pares RGB e imagem NIR correspondente para a realização do treinamento; e apenas de imagem RGB para a estimativa. Além de utilizar menos informação e não necessitar da realização de uma classificação prévia da área que será sobrevoada, a rede *Pix2Pix*, utilizando funções de *loss* que consideram informações estruturais e perceptivas (\mathcal{L}_{Mix} e \mathcal{L}_{PMix} , respectivamente), apresentou imagens muito mais similares às reais, com média de 0,91 para o índice MS-SSIM, frente a média de 0,77 obtida com o método NIRKNN. O maior ganho

obtido para a estimativa de imagens NIR com a *Pix2Pix* foi a conversão das imagens de entrada do padrão RGB para os modelos de cor CIE Lab e HSV, resultando em uma diminuição da média do RMSE entre *pixels* reais e estimados de 17,77 para 5,94. Portanto, pode-se concluir que a utilização da arquitetura de *Deep Learning* utilizada nos experimentos é a alternativa melhor para estimar imagens NIR, possibilitando a oferta da tecnologia da utilização de *drones* para monitoramento de propriedades rurais a valores mais atraentes e acessíveis para pequenos produtores, sem inviabilizar porém o outro método desenvolvido.

Os trabalhos futuros envolvem o estudo de outros classificadores para realização de experimentos similares aos do NIRKNN, verificando a possibilidade da criação de novos métodos que utilizem o conceito de espectros de referência, sendo estes obtidos por meio de imagens hiperspectrais ou sensores portáteis, na geração de imagens NIR. Com relação ao NIRDLearn, serão investigadas novas arquiteturas de *Deep Learning* usadas para tradução de imagem para imagem, como novos métodos de GAN inspirados no *Pix2Pix*. Outros trabalhos a serem realizados são obter mapas NDVI calculados a partir das imagens estimadas e verificar a diferença em comparação aos mapas NDVI calculados com dados reais; e realizar treinamentos com validação cruzada (*cross validation*).

5.1 Trabalhos publicados ou submetidos

Durante o período referente ao doutorado, os seguintes trabalhos foram produzidos:

Artigos publicados em conferência:

- SAQUI, D.; SAITO, J. H. ; LIMA, D. C. ; JORGE, L. A. C. ; ATAKY, S. . Approach Based on SPEA-2-Band Selection and Random Forest Classifier to Generate Thematic Maps from Hyperspectral Images.. In: AMCIS 2019 - Americas Conference on Information Systems, 2019, Cancun. Twenty-fifth Americas Conference on Information Systems, 2019. p. 1-10.
- LIMA, D. C.; SAQUI, D. ; ATAKY, S. ; JORGE, L. A. C. ; FERREIRA, E. J. ; SAITO, J. H. . Estimating Agriculture NIR Images from Aerial RGB Data.. In: ICCS 2019 - International Conference on Computational Science, 2019, Faro, Algarve, 2019. v. 11536. p. 562-574.
- SILVA, D. A. P.; LIMA, D. C. ; SAITO, JOSE H. . Tolerance Near Sets and tNM Application in City Images.. In: ICCS 2019 - International Conference on Computational Science, 2019, Faro, Algarve, 2019. v. 11537. p. 564-579.

- SAQUI, D. ; SAITO, J. H. ; JORGE, L. A. DE C. ; FERREIRA, E. J. ; **LIMA, D. C.** ; HERRERA, J. P. . Methodology for Band Selection of Hyperspectral Images Using Genetic Algorithms and Gaussian Maximum Likelihood Classifier. In: 2016 International Conference on Computational Science and Computational Intelligence (CSCI), 2016, Las Vegas. 2016 International Conference on Computational Science and Computational Intelligence (CSCI), 2016. p. 733.

Artigos publicados em periódicos:

- SAQUI, D.; SAITO, J. H.; **DE LIMA, D. C.** ; DEL VAL CURA, L. M. ; ATAKY, S. T. M. . Incorporated Decision-maker-based Multiobjective Band Selection for Pixel Classification of Hyperspectral Images. *Advances in Electrical and Computer Engineering*, v. 19, p. 21-28, 2019.
- **LIMA, D. C.** ; SAQUI, D. ; ATAKY , S; JORGE, L. A. DE C. ; FERREIRA, E. J. ; SAITO, J. H. . Estimating Agriculture NIR Images from Aerial RGB Data. *Lecture Notes in Computer Science*, v. 11536, p. 562, 2019.

Artigo submetido para periódico:

Título: Using Pix2Pix Network to Estimate Agriculture NIR images from RGB Data.

Periódico: *Geoscience and Remote Sensing Letters*.

REFERÊNCIAS

APOSTOL, S. et al. Laser-induced fluorescence signatures as a tool for remote monitoring of water and nitrogen stresses in plants. *Canadian Journal of Remote Sensing*, v. 29, p. 57–65, 2003.

ARAI, K. et al. Method for nir reflectance estimation with visible camera data based on regression for ndvi estimation and its application for insect damage detection of rice paddy fields. *International Journal of Advanced Research in Artificial Intelligence*, v. 5, n. 11, p. 17–22, 2016.

AREL, I.; ROSE, D. C.; KARNOWSKI, T. P. Deep machine learning-a new frontier in artificial intelligence research [research frontier]. *IEEE computational intelligence magazine*, IEEE, v. 5, n. 4, p. 13–18, 2010.

BAUER, M. E.; CIPRA, J. E. Identification of agricultural crops by computer processing of erts mss data. *LARS Technical Reports*, 1973. Disponível em: <http://docs.lib.purdue.edu/larstech/20/>.

BAUMGARDNER, M. F.; BIEHL, L. L.; LANDGREBE, D. A. *220 Band AVIRIS Hyperspectral Image Data Set: June 12, 1992 Indian Pine Test Site 3*. 2015. Disponível em: <https://purr.purdue.edu/publications/1947/1>.

BEDORD, L. *12 Potential uses for UAS in agriculture*. Successful Farming, 2015. Disponível em: http://www.agriculture.com/technology/robotics/uas/12-potential-uses-f-uas-in-agriculture_587-ar51680.

BEDORD, L. *DroneDeploy partners with Sentera and SlantRange to provide accurate crop health analysis*. Successful Farming, 2017. Disponível em: <http://www.agriculture.com/dronedeploy-partners-with-sentera-and-slantrange-to-provide-accurate-crop-health-analysis>.

BERNARDI, A. C. C. et al. *Agricultura de Precisão: Resultados de um Novo Olhar*. [S.l.: s.n.], 2014.

BHATTI, A. U.; MULLA, D. J.; FRAZIER, B. E. Estimation of soil properties and wheat yields on complex eroded hills using geostatistics and thematic mapper images. *Remote Sensing of Environment*, v. 37, p. 181–191, 1991.

BILLMEYER, F. W.; SALTZMAN, M. Principles of color technology. In: _____. 2nd. ed. [S.l.]: Wiley, 1981. cap. Describing Colors, p. 25–66.

BLACKBURN, G. A. Hyperspectral remote sensing of plant pigments. *Journal of experimental botany*, Oxford University Press, v. 58, n. 4, p. 855–867, 2006.

- BRASIL. *Principais Conceitos E Expressões Utilizados Na Agricultura De Precisão*. 2012. Disponível em: <http://www.agricultura.gov.br/assuntos/sustentabilidade/tecnologia-agropecuaria/agricultura-de-precisao-1/arquivos-de-agricultura-de-precisao/conceitos-relacionados-a-agricultura-de-precisao-1.pdf>.
- CHANG, C.-I. An information-theoretic approach to spectral variability, similarity, and discrimination for hyperspectral image analysis. *IEEE Transactions on information theory*, IEEE, v. 46, n. 5, p. 1927–1932, 2000.
- CHANG, C.-I.; PLAZA, A. A fast iterative algorithm for implementation of pixel purity index. *IEEE Geoscience and Remote Sensing Letters*, IEEE, v. 3, n. 1, p. 63–67, 2006.
- CHAVES-GONZÁLEZ, J. M. et al. Detecting skin in face recognition systems: A colour spaces study. *Digital Signal Processing*, Elsevier, v. 20, n. 3, p. 806–823, 2009.
- COHEN, Y. et al. Estimation of leaf water potential by thermal imagery and spatial analysis. *Journal of Experimental Botany*, v. 56, p. 1843–1852, 2005.
- COOTES, T. F.; EDWARDS, G. J.; TAYLOR, C. J. Active appearance models. *IEEE Transactions on pattern analysis and machine intelligence*, IEEE, v. 23, n. 6, p. 681–685, 2001.
- CROOKSTON, K. A top 10 list of developments and issues impacting crop management and ecology during the past 50 years. *Crop Science*, v. 46, p. 2253–2262, 2006.
- FRANK, E.; HALL, M. A.; WITTEN, I. H. *The Weka Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques"*. Fourth edition. [S.l.: s.n.], 2016.
- FRÖHLICH, B. et al. Land cover classification of satellite images using contextual information. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, v. 3, 2013.
- FRÖHLICH, B.; RODNER, E.; DENZLER, J. As time goes by—anytime semantic segmentation with iterative context forests. *Pattern Recognition*, Springer, p. 1–10, 2012.
- GONZALEZ, R. C.; WOODS, R. *Processamento Digital de Imagens*. 3. ed. São Paulo: Pearson Prentice Hall, 2010.
- GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep Learning*. [S.l.]: MIT Press, 2016. <http://www.deeplearningbook.org>.
- GOODFELLOW, I. et al. Generative adversarial nets. In: *Advances in neural information processing systems*. [S.l.: s.n.], 2014. p. 2672–2680.
- HAN, T. Y.; KIM, Y. J.; SONG, B. C. Convolutional neural network-based infrared image super resolution under low light environment. In: *IEEE. Signal Processing Conference (EUSIPCO), 2017 25th European*. [S.l.], 2017. p. 803–807.
- HINTON, G. E. et al. Improving neural networks by preventing co-adaptation of feature detectors. *arXiv preprint arXiv:1207.0580*, 2012.
- HUETE, A. et al. Overview of the radiometric and biophysical performance of the modis vegetation indices. *Remote sensing of environment*, Elsevier, v. 83, n. 1, p. 195–213, 2002.

- HUNG, C.; BRYSON, M.; SUKKARIEH, S. Vision-based shadow-aided tree crown detection and classification algorithm using imagery from an unmanned airborne vehicle. *34th International Symposium on Remote Sensing of Environment (ISRSE)*, 2011.
- IOFFE, S.; SZEGEDY, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*, 2015.
- ISOLA, P. et al. Image-to-image translation with conditional adversarial networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2017. p. 1125–1134.
- JACKSON, R. D.; HUETE, A. R. Interpreting vegetation indices. *Preventive veterinary medicine*, Elsevier, v. 11, n. 3-4, p. 185–200, 1991.
- JOHNSON, J.; ALAHI, A.; FEI-FEI, L. Perceptual losses for real-time style transfer and super-resolution. In: SPRINGER. *European conference on computer vision*. [S.l.], 2016. p. 694–711.
- JORGE, L. d. C.; INAMASU, R. Y. Uso de veículos aéreos não tripulados (vant) em agricultura de precisão. *Embrapa Instrumentação-Capítulo em livro técnico-científico (ALICE)*, In: BERNARDI, AC de C.; NAIME, J. de M.; RESENDE, AV de; BASSOI, LH; INAMASU, RY (Ed.). *Agricultura de precisão: resultados de um novo olhar*. Brasília, DF: Embrapa, 2014. p. 109-134., 2014.
- KAJI, S.; KIDA, S. Overview of image-to-image translation by use of deep neural networks: denoising, super-resolution, modality conversion, and reconstruction in medical imaging. *Radiological Physics and Technology*, Springer, p. 1–14, 2019.
- KINGMA, D. P.; BA, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- KRUSE, F. A. et al. The spectral image processing system (sips)-interactive visualization and analysis of imaging spectrometer data. In: AIP. *AIP Conference Proceedings*. [S.l.], 1993. v. 283, n. 1, p. 192–201.
- LANDGREBE, D. A. *Signal theory methods in multispectral remote sensing*. [S.l.]: John Wiley & Sons, 2005. v. 29.
- LAWRANCE, R.; LABUS, M. Early detection of douglas-fir beetle infestation with sub canopy resolution hyperspectral imagery. *Western Journal of Applied Forestry*, v. 18, p. 202–206, 2003.
- LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. *Nature*, Nature Research, v. 521, n. 7553, p. 436–444, 2015.
- LIMMER, M.; LENSCH, H. P. Infrared colorization using deep convolutional neural networks. In: IEEE. *Machine Learning and Applications (ICMLA), 2016 15th IEEE International Conference on*. [S.l.], 2016. p. 61–68.
- LINHARES, M. M. A.; ROCHA, N. C. C.; AMARAL, B. A. S. Análise do índice mpri como indicador vegetativo através da correlação do mesmo com o índice ndvi. *Anais XVI Simpósio Brasileiro de Sensoriamento Remoto - SBSR*, INPE, Foz do Iguaçu, Paraná, Brasil, p. 8254–8260, 2013. Disponível em: <http://www.dsr.inpe.br/sbsr2013/files/p1190.pdf>.

- LU, Y. M. et al. Designing color filter arrays for the joint capture of visible and near-infrared images. In: *IEEE. Image Processing (ICIP), 2009 16th IEEE International Conference on*. [S.l.], 2009. p. 3797–3800.
- MAAS, A. L.; HANNUN, A. Y.; NG, A. Y. Rectifier nonlinearities improve neural network acoustic models. In: *Proc. icml*. [S.l.: s.n.], 2013. v. 30, n. 1, p. 3.
- MANOLAKIS, D.; MARDEN, D.; SHAW, G. A. Hyperspectral image processing for automatic target detection applications. *Lincoln laboratory journal*, v. 14, n. 1, p. 79–116, 2003.
- MEDEIROS, F. A. *Desenvolvimento de um veículo aéreo não tripulado para aplicação em agricultura de precisão*. Santa Maria: [s.n.], 2007.
- MENESES, P. R.; ALMEIDA, T. d. Introdução ao processamento de imagens de sensoriamento remoto. *Embrapa Cerrados-Livros técnicos (INFOTECA-E)*, Brasília, DF: UnB, 2012., 2012.
- MIRZA, M.; OSINDERO, S. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014.
- MULLA, D. J. Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. *Biosystems Engineering*, v. 114, p. 358–371, 2013.
- NASA. *NASA's Imagine the Universe: The Electromagnetic Spectrum*. 2013. Disponível em: <https://imagine.gsfc.nasa.gov/science/toolbox/emspectrum1.html>.
- OJALA, T.; PIETIKAINEN, M.; MAENPAA, T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 24, n. 7, p. 971–987, 2002.
- PAPP, A. C. *Embrapa e Qualcomm querem potencializar drones no campo*. O Estado de São Paulo, 2016. Disponível em: <http://economia.estadao.com.br/noticias/agro,embrapa-e-qualcomm-querem-potencializar-drones-no-campo,10000060122>.
- PINTER, P. J. et al. Remote sensing for crop management. *Photogrammetric Engineering and Remote Sensing*, v. 69, p. 647–664, 2003.
- PIXFORCE. *Drones na agricultura: Tudo sobre a tecnologia que está mudando o setor*. 2016. Disponível em: <http://pixforce.com.br/drones-na-agricultura/>.
- PLAZA, J. et al. On endmember identification in hyperspectral images without pure pixels: A comparison of algorithms. *Journal of Mathematical Imaging and Vision*, v. 42, n. 2, p. 163–175, Feb 2012. ISSN 1573-7683. Disponível em: <https://doi.org/10.1007/s10851-011-0276-0>.
- PRATT, W. K. *Digital Image Processing: PIKS Scientific inside*. 4th edition. ed. [S.l.]: Wiley-Interscience, 2006.
- RABATEL, G.; GORRETTA, N.; LABBÉ, S. Getting ndvi spectral bands from a single standard rgb digital camera: a methodological approach. *Advances in Artificial Intelligence*, Springer, p. 333–342, 2011.
- REID, A.; RAMOS, F.; SUKKARIEH, S. Multi-class classification of vegetation in natural environments using an unmanned aerial system. *2011 IEEE International Conference on Robotics and Automation (ICRA)*, p. 2953–2959, 2011.

REITER, M. et al. 3d and infrared face reconstruction from rgb data using canonical correlation analysis. In: IEEE. *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*. [S.l.], 2006. v. 1, p. 425–428.

REN, H.; CHANG, C.-I. Automatic spectral target recognition in hyperspectral imagery. *IEEE Transactions on Aerospace and Electronic Systems*, IEEE, v. 39, n. 4, p. 1232–1249, 2003.

RIBEIRO, C. J. *Agricultura: Drone ou Satélite?* 2017. Disponível em: <http://sensix.com.br/agricultura-drone-ou-satelite/>.

RONNEBERGER, O.; FISCHER, P.; BROX, T. U-net: Convolutional networks for biomedical image segmentation. In: SPRINGER. *International Conference on Medical image computing and computer-assisted intervention*. [S.l.], 2015. p. 234–241.

RUSSAKOVSKY, O. et al. Imagenet large scale visual recognition challenge. *International journal of computer vision*, Springer, v. 115, n. 3, p. 211–252, 2015.

SANTOS, E. *Tecnologia de imagens hiperespectrais reforça trabalhos de pesquisa da Embrapa Algodão. Pesquisa, Desenvolvimento e Inovação*. EMBRAPA, 2014. Disponível em: <https://www.embrapa.br/buscadenoticias/-/noticia/2016342/tecnologia-de-imagens-hiperespectrais-reforca-trabalhos-de-pesquisa-da-embrapa-algodao>.

SHIPPERT, P. Why use hyperspectral imagery? *Photogrammetric engineering and remote sensing*, ASPRS AMERICAN SOCIETY FOR PHOTOGRAMMETRY AND, v. 70, n. 4, p. 377–396, 2004.

SHIRATSUCHI, L. S. et al. Sensoriamento remoto: conceitos básicos e aplicações na agricultura de precisão. *Embrapa Solos-Capítulo em livro técnico (INFOTECA-E)*, In: BERNARDI, AC de C.; NAIME, J. de M.; RESENDE, AV de; BASSOI, LH; INAMASU, RY (Ed.). *Agricultura de precisão: resultados de um novo olhar*. Brasília, DF: Embrapa, 2014. cap. 4, p. 58-73., 2014.

SHOTTON, J. et al. Textonboost: Joint appearance, shape and context modeling for multi-class object recognition and segmentation. *Computer Vision–ECCV 2006*, p. 1–16, 2006.

SILVA, J. *Pesquisador da Embrapa fala sobre tipos de drones e softwares para tratamento de imagens em São Paulo*. EMBRAPA, 2015. Disponível em: <https://www.embrapa.br/busca-de-noticias/-/noticia/6689079/pesquisador-da-embrapa-fala-sobre-tipos-de-drones-e-softwares-para-tratamento-de-imagens-em-sao-paulo>

SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

SOLTANI-FARANI, A.; RABIEE, H. R.; HOSSEINI, S. A. Spatial-aware dictionary learning for hyperspectral image classification. *IEEE Transactions on geoscience and remote sensing*, IEEE, v. 53, n. 1, p. 527–541, 2015.

SOLUTIONS, H. G. *How does ENVI's Pixel Purity Index Work?* 1999. Disponível em: <http://www.harrisgeospatial.com/Company/PressRoom/TabId/190/ArtMID/786/ArticleID/1631/1631.aspx>.

SOUZA, J. R. et al. Automatic detection of ceratocystis wilt in eucalyptus crops from aerial images. *IEEE International Conference on Robotics and Automation (ICRA)*, 2015.

- SOUZA, V. L. C. A. O uso de imagens hiperespectrais com efeitos de reflectância bidirecional na melhoria da classificação de culturas agrícolas. *Revista Eletrônica: Tempo-Técnica-Território*, v. 4, n. 1, p. 001–020, 2013.
- TKALCIC, M.; TASIC, J. F. Color spaces: Perceptual, historical and applicational background. In: *Proc. IEEE International Conference on Computer as a Tool (IEEE EUROCON 2003)*, Sept. [S.l.: s.n.], 2003. v. 1, p. 304–308.
- WANG, Z. et al. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, v. 13, n. 4, p. 600–612, 2004.
- WANG, Z.; SIMONCELLI, E. P.; BOVIK, A. C. Multiscale structural similarity for image quality assessment. In: IEEE. *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*. [S.l.], 2003. v. 2, p. 1398–1402.
- WEBB, A. R.; COPSEY, K. D. *Statistical pattern recognition*. Third edition. Malvern, UK: John Wiley & Sons, 2011.
- WINTER, M. E. N-findr: An algorithm for fast autonomous spectral end-member determination in hyperspectral data. In: INTERNATIONAL SOCIETY FOR OPTICS AND PHOTONICS. *Imaging Spectrometry V*. [S.l.], 1999. v. 3753, p. 266–276.
- YANG, M. C.; CHANG, Y. C. Assessing disease severity of rice bacterial leaf blight with canopy hyperspectral reflectance. *Precision Agriculture*, v. 11, p. 61–81, 2010.
- YANG, Z.; WILLIS, P.; MUELLER, R. Impact of band-ratio enhanced awifs image to crop classification accuracy. *Pecora 17,—The Future of Land Imaging::: Going Operational Denver*, p. 18–20, 2008.
- ZHANG, X. et al. Application of hyperspectral imaging and chemometric calibrations for variety discrimination of maize seeds. *Sensors 2012*, p. 17234–17246, 2012.
- ZHAO, H. et al. Loss functions for image restoration with neural networks. *IEEE Transactions on Computational Imaging*, IEEE, v. 3, n. 1, p. 47–57, 2016.
- ZHENG, H. et al. Evaluation of rgb, color-infrared and multispectral images acquired from unmanned aerial systems for the estimation of nitrogen accumulation in rice. *Remote Sensing*, Multidisciplinary Digital Publishing Institute, v. 10, n. 6, p. 824, 2018.
- ZHU, J.-Y. et al. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: *Proceedings of the IEEE international conference on computer vision*. [S.l.: s.n.], 2017. p. 2223–2232.