



**UNIVERSIDADE FEDERAL DE UBERLÂNDIA**  
**FACULDADE DE ENGENHARIA ELÉTRICA**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM**  
**ENGENHARIA ELÉTRICA**



**ANÁLISE ESTATÍSTICA DAS TAXAS DE RECONHECIMENTO DE**  
**IMAGENS DE EDIFÍCIOS EM AMBIENTES URBANOS UTILIZANDO**  
**QUANTIZAÇÃO VECTORIAL E APRENDIZAGEM PROFUNDA**

**EDUARDO SILVA VASCONCELOS**

**Uberlândia - MG**

**2022**

**EDUARDO SILVA VASCONCELOS**

**ANÁLISE ESTATÍSTICA DAS TAXAS DE RECONHECIMENTO DE  
IMAGENS DE EDIFÍCIOS EM AMBIENTES URBANOS UTILIZANDO  
QUANTIZAÇÃO VECTORIAL E APRENDIZAGEM PROFUNDA**

Tese apresentada ao Programa de Pós-Graduação *Strictu Sensu* da Faculdade de Engenharia Elétrica da Universidade Federal de Uberlândia, como recurso parcial para obtenção do título de Doutor em Ciências.

**BANCA EXAMINADORA**

Orientador - Dr. Gilberto Arantes Carrijo  
Universidade Federal de Uberlândia – UFU

Dr. Antônio Cláudio Paschoarelli Veiga  
Universidade Federal de Uberlândia – UFU

Dr. Alexandre Coutinho Mateus  
Universidade Federal de Uberlândia – UFU

Dr. Leandro Aureliano da Silva  
Universidade de Uberaba - UNIUBE

Dr. Lacordaire Kemel Pimenta Cury  
Instituto Federal Goiano – IF Goiano

**Uberlândia**

**2022**

Ficha Catalográfica Online do Sistema de Bibliotecas da UFU  
com dados informados pelo(a) próprio(a) autor(a).

V331 2022	<p>Vasconcelos, Eduardo Silva, 1967- Análise estatística das taxas de reconhecimento de imagens de edifícios em ambientes urbanos utilizando Quantização Vectorial e Aprendizagem Profunda [recurso eletrônico] / Eduardo Silva Vasconcelos. - 2022.</p> <p>Orientador: Gilberto Arantes Carrijo. Tese (Doutorado) - Universidade Federal de Uberlândia, Pós-graduação em Engenharia Elétrica. Modo de acesso: Internet. Disponível em: <a href="http://doi.org/10.14393/ufu.te.2023.4">http://doi.org/10.14393/ufu.te.2023.4</a> Inclui bibliografia. Inclui ilustrações.</p> <p>1. Engenharia elétrica. I. Carrijo, Gilberto Arantes, 1948-, (Orient.). II. Universidade Federal de Uberlândia. Pós-graduação em Engenharia Elétrica. III. Título.</p> <p style="text-align: right;">CDU: 621.3</p>
--------------	---

Bibliotecários responsáveis pela estrutura de acordo com o AACR2:  
Gizele Cristine Nunes do Couto - CRB6/2091  
Nelson Marcos Ferreira - CRB6/3074



### ATA DE DEFESA - PÓS-GRADUAÇÃO

Programa de Pós-Graduação em:	Engenharia Elétrica				
Defesa de:	Tese de Doutorado, 312, PPGEELT				
Data:	Vinte de dezembro de dois mil e vinte e dois	Hora de início:	8:30	Hora de encerramento:	12:00
Matrícula do Discente:	12023EEL004				
Nome do Discente:	Eduardo Silva Vasconcelos				
Título do Trabalho:	Análise estatística das taxas de reconhecimento de imagens de edifícios em ambientes urbanos utilizando quantização vectorial e aprendizagem profunda				
Área de concentração:	Processamento da Informação				
Linha de pesquisa:	Processamento Digital de Sinais e Redes de Computadores				
Projeto de Pesquisa de vinculação:	Coordenador do projeto: Gilberto Arantes Carrijo Título do projeto: Segmentação de Imagens no reconhecimento Agência financiadora: N/A Número do processo na agência financiadora: N/A Vigência do projeto: 2017 até atual				

Reuniu-se por meio de videoconferência, a Banca Examinadora, designada pelo Colegiado do Programa de Pós-graduação em Engenharia Elétrica, assim composta: Professores Doutores: Alexandre Coutinho Mateus - FEELT/UFU; Antônio Cláudio Paschoarelli Veiga - FEELT/UFU; Lacordaire Kemel Pimenta Cury - IF Goiano; Leandro Aureliano da Silva - UNIUBE; Gilberto Arantes Carrijo - FEELT/UFU, orientador(a) do(a) candidato(a).

Iniciando os trabalhos o(a) presidente da mesa, Dr(a). Gilberto Arantes Carrijo, apresentou a Comissão Examinadora e o candidato(a), agradeceu a presença do público, e concedeu ao Discente a palavra para a exposição do seu trabalho. A duração da apresentação do Discente e o tempo de arguição e resposta foram conforme as normas do Programa.

A seguir o senhor(a) presidente concedeu a palavra, pela ordem sucessivamente, aos(às) examinadores(as), que passaram a arguir o(a) candidato(a). Ultimeada a arguição, que se desenvolveu dentro dos termos regimentais, a Banca, em sessão secreta, atribuiu o resultado final, considerando o(a) candidato(a):

Aprovado(a).

Esta defesa faz parte dos requisitos necessários à obtenção do título de Doutor.

O competente diploma será expedido após cumprimento dos demais requisitos, conforme as normas do Programa, a legislação pertinente e a regulamentação interna da UFU.

Nada mais havendo a tratar foram encerrados os trabalhos. Foi lavrada a presente ata que após lida e achada conforme foi assinada pela Banca Examinadora.



Documento assinado eletronicamente por **Leandro Aureliano da Silva, Usuário Externo**, em 12/12/2022, às 23:11, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Lacordaire Kemel Pimenta Cury, Usuário Externo**, em 20/12/2022, às 10:08, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Alexandre Coutinho Mateus, Professor(a) do Magistério Superior**, em 20/12/2022, às 10:12, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Gilberto Arantes Carrijo, Professor(a) do Magistério Superior**, em 20/12/2022, às 10:18, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Antonio Claudio Paschoarelli Veiga, Professor(a) do Magistério Superior**, em 20/12/2022, às 11:01, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site [https://www.sei.ufu.br/sei/controlador\\_externo.php?acao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=0](https://www.sei.ufu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **4131690** e o código CRC **618DC38B**.

*Dedico esta obra a minha esposa Léia pelo companheirismo e apoiar minhas decisões, a minha filha Débora por caminhar junto nessa jornada e a minha filha Eduarda por me transformar em uma pessoa melhor.*

---

*Agradeço a Deus pelos desafios proporcionados, mas ao mesmo tempo, pela força e perseverança destinadas para transpor todos eles.*

*A minha esposa Léia pelo amor, dedicação e paciência durante a conclusão desta etapa.*

*Às minhas filhas Débora e Eduarda pelo amor incondicional.*

*Aos meus pais José Pedro e Walda por acreditarem que sempre chegaria além das circunstâncias.*

*Ao meu orientador Prof. Dr. Gilberto Carrijo, demonstro aqui todo o meu respeito.*

*À professora Edna, pelo seu apoio inicial para entrar no programa.*

*Aos membros da Banca Examinadora pela contribuição para esta tese.*

*A todos os professores, colaboradores e colegas do Programa de Pós-Graduação da FEELT da Universidade Federal de Uberlândia, por todos os ensinamentos, conhecimentos e amizade neste período.*

*E também a instituição IF Goiano e seus colaboradores pelo apoio.*

---

*“A tarefa não é tanto ver aquilo que ninguém viu, mas pensar o que ninguém ainda pensou sobre aquilo que todo mundo vê” (Arthur Schopenhauer).*

## SUMÁRIO

	Página
CAPÍTULO 1 INTRODUÇÃO	13
1.1 Considerações Iniciais	13
1.2 Objetivos	16
1.2.1 Objetivo Geral	16
1.2.2 Objetivos específicos	16
1.3 Organização do Trabalho	17
CAPÍTULO 2 FUNDAMENTAÇÃO TEÓRICA	18
2.1 Considerações Iniciais	18
2.2 Imagem Digital	18
2.3 Processamento de Imagens	20
2.3.1 Aquisição da imagem	20
2.3.2 Pré processamento	20
2.3.3 Segmentação	21
2.3.4 Pós processamento	21
2.3.5 Extração de atributos	21
2.3.6 Reconhecimento e classificação	22
2.4 Estado da Arte: Reconhecimento de Imagens	22
2.5 Reconhecimento de Padrões	27
2.5.1 Métodos Utilizados no Reconhecimento de Padrões	28
2.6 Algoritmos Genéticos	29
2.7 Rede Neural Artificial	30
2.8 Análise de Componentes Principais	32
2.9 K-Média	33
2.10 Quantização Vetorial	34
2.11 Inteligência artificial e aprendizado de máquina	37
2.12 <i>Deep Learning</i>	38
CAPÍTULO 3 O MODELO PROPOSTO NESTE TRABALHO	40
3.1 Considerações Iniciais	40
3.2 O banco de dados e sua complexidade: Sheffield Building Image Dataset	43
3.3 Descrição geral do método proposto	44
3.3.1 Quantização Vetorial - QV	46
3.3.2 Deep Learning - DL	50
3.4 Considerações finais do capítulo	52
CAPÍTULO 4 ANÁLISE DOS RESULTADOS	53
4.1 Considerações Iniciais	53
4.2 Análise dos resultados por escala de cor	54
4.2.1 Escala de Cinza	54

4.2.2	Escala de Vermelha	55
4.2.3	Escala de Verde	56
4.2.4	Escala de Azul	57
4.3	Análise Comparativa entre os resultados nas escalas de cinza, vermelha, verde e azul	57
4.3.1	Análise com relação a quantidade de centroides	58
4.3.2	Análise com relação as escalas de cores cinza, vermelha, verde e azul	60
4.4	Análise dos resultados em RGB	61
4.5	Análise Comparativa dos resultados nas escalas de cinza, vermelho, verde, azul e RGB	62
4.5.1	Análise comparativa em relação aos centroides	62
4.5.2	Análise comparativa em relação às escalas de cores cinza, vermelha, verde, azul e RGB	65
4.6	Análise dos resultados de pela técnica de DL	66
4.7	Análise Comparativa dos resultados de QV RGB com DL	67
4.8	Comparação com outros trabalhos	68
4.9	Considerações finais do capítulo	69
CAPÍTULO 5	CONSIDERAÇÕES FINAIS	70
5.1	Considerações Finais	70
	REFERÊNCIAS	72

## LISTA DE FIGURAS

### CAPÍTULO 2

- Figura 2.1 – Imagem detalhada mostrando os contadores digitais associados a cada nível de cinza dos pixels
- Figura 2.2 – Processo de transformação da imagem contínua em imagem digital
- Figura 2.3 – Etapas de um sistema de reconhecimento de padrões
- Figura 2.4 – Comparação entre neurônio biológico e neurônio artificial
- Figura 2.5 – Estrutura de rede neural
- Figura 2.6 – Passos de aplicação do algoritmo K-médias
- Figura 2.7 – Sistema de compressão de sinais baseado em Quantização Vetorial simples

### CAPÍTULO 3

- Figura 3.1 – Exemplos de uma imagem aérea e uma imagem de construção
- Figura 3.2 – Fluxograma da classificação de Li et al. (2014), dos sistemas de reconhecimento de edifícios
- Figura 3.3 – Exemplo de imagens do conjunto de dados SBID. Essas três linhas mostram imagens de amostra das categorias 1, 10 e 38, respectivamente
- Figura 3.4 – Diagrama em blocos do processo metodológico para imagens em escala de cinza, vermelha, verde e azul
- Figura 3.5 – Diagrama em blocos do processo metodológico para imagens em RGB
- Figura 3.6 – Exemplo de separação de cada imagem em blocos de 8x8 pixels
- Figura 3.7 – Sumário da rede criada no projeto, utilizando-se um `validation_split = 0,4`
- Figura 3.8 – Processo ilustrativo da implementação da rede

## LISTA DE TABELAS

### CAPÍTULO 3

Tabela 3.1 – Quantidade de Imagens Utilizadas para o Treino e para a Análise de Reconhecimento

### CAPÍTULO 4

Tabela 4.1 – Resultados na escala em cinza

Tabela 4.2 – Resultados na escala em vermelho

Tabela 4.3 – Resultados na escala em verde

Tabela 4.4 – Resultados na escala em azul

Tabela 4.5 – Análise Comparativa entre os resultados nas escalas de cinza, vermelha, verde e azul com relação a quantidade de centroides

Tabela 4.6 – Análise Comparativa entre os resultados nas escalas de cinza, vermelha, verde e azul

Tabela 4.7 – Análise dos Resultados em RGB

Tabela 4.8 – Análise Comparativa entre os resultados nas escalas de cinza, vermelha, verde, azul e RGB com relação a quantidade de centroides

Tabela 4.9 – Análise Comparativa entre os resultados nas escalas de cinza, vermelha, verde, azul e RGB

Tabela 4.10 – Resultados obtidos por DL RGB

Tabela 4.11 – Análise Comparativa dos resultados de QV 32 centroides com DL

Tabela 4.12 – Resultado Obtido Utilizando O Método Proposto E Os Resultados Encontrados Na Literatura

## LISTA DE GRÁFICOS

### CAPÍTULO 4

- Gráfico 4.1 – Diagrama de caixa e caixa estreita comparando a Variabilidade com relação aos centroides
- Gráfico 4.2 – Diagrama de caixa e caixa estreita comparando a Variabilidade das escalas de cores cinza, vermelha, verde e azul
- Gráfico 4.3 – Diagrama de caixa e caixa estreita comparando a Variabilidade dos centroides para as escalas de cores cinza, vermelha, verde, azul e RGB
- Gráfico 4.4 – Diagrama de caixa e caixa estreita comparando a Variabilidade das escalas de cores cinza, vermelha, verde, azul e RGB
- Gráfico 4.5 – Diagrama de caixa e caixa estreita comparando a Variabilidade entre QV RGB 32 Centroides e DL RGB

## RESUMO

O reconhecimento de edifícios é essencial para uma variedade de aplicações tais como detecção automática de alvos, reconstrução de cidades em 3D, navegação digital, etc. Este trabalho visa analisar comparativamente as taxas de reconhecimento de imagens de edifícios, usando a técnica de Quantização Vetorial para compressão de imagens usando o algoritmo Linde-Buzo-Gray, com os resultados obtidos pelo método *Deep Learning*. Foram analisadas 40 classes com 30 imagens por classe, separadamente, na escala de cores cinza, vermelha, verde, azul e em RGB, variando a quantidade de centroides em 16, 32, 64, 128 e 256 para a técnica de quantização vetorial, e também variando a porcentagem da quantidade de imagens para treinamento em 40%, 50% e 60%, com suas respectivas porcentagens da quantidade de imagens para reconhecimento, em ambos os métodos. Para verificar as diferenças, foi realizada a ANOVA, com o post-hoc de Tukey com 5% de significância. Os resultados descritivos obtiveram altas taxas de reconhecimento, em ambos os métodos. Na análise inferencial dos resultados obtidos na Quantização Vetorial, foram encontradas taxas significativas de reconhecimento a partir de 32 centroides. Ao comparar os resultados das duas técnicas, nenhuma diferença significativa foi encontrada.

**Palavras-Chaves:** Reconhecimento de imagem. Quantização Vetorial. Aprendizagem Profunda, inteligência artificial.

## **ABSTRACT**

*Building recognition is essential for a variety of applications such as automatic target detection, 3D city reconstruction, digital navigation, etc. This paper aims to comparatively analyze the recognition rates of building images, using the Vector Quantization technique for image compression using the Linde-Buzo-Gray algorithm, with the results obtained by the Deep Learning method. Forty classes were analyzed, with 30 images per class, separately, in gray, red, green, blue and RGB, varying the number of centroids in 16, 32, 64, 128 and 256 for the vector quantization technique, and also varying the percentage of the amount of images for training in 40%, 50% and 60%, with their respective percentages of the amount of images for recognition, in both methods. To check the differences, ANOVA was performed, with Tukey's post-hoc at 5% significance. The descriptive results showed high recognition rates, in both methods. In the inferential analysis of the results obtained in Vector Quantization, significant recognition rates were found from 32 centroids on. When comparing the results of the two techniques, no significant difference was found.*

**Keywords:** *Image recognition. Vector Quantization. Deep Learning, artificial intelligence.*

## CAPÍTULO 1

### INTRODUÇÃO

#### 1.1 Considerações Iniciais

Um dos grandes desafios atuais na área de reconhecimento de padrões consiste no reconhecimento em larga escala de imagens de edifícios. O reconhecimento de edifícios pode ser utilizado em várias aplicações como: detecção automática e rastreamento de alvo, vigilância, projetos arquitetônicos, navegação de dispositivos móveis, localização de robô, georreferenciamento, dentre outras (BRUCKNER et al., 2012; LI; ALLINSON, 2013, 2009; LIU, 2008; PARK; LEE, 2011; ZHANG; PAN; ZHANG, 2022; ZHANG; LAPIERRE; XIANG, 2013).

A grande dificuldade para esse reconhecimento é a enorme variabilidade de parâmetros encontrados na obtenção dessas imagens. Estas variabilidades podem ser: obtenção das imagens em diferentes ângulos, diversas condições de iluminação ou obstruções parciais de árvores, de veículos ou até mesmo de outros edifícios (LI; ALLINSON, 2013). Como lidar com esses desafios é um problema de pesquisa que se estende a algum tempo e alguns sistemas de reconhecimento de edifícios têm sido propostos nos últimos anos (BEZAK, 2016; CHUNG; HAN; HE, 2009; GROENEWEG et al., 2006; LI; ALLINSON, 2013, 2009; LI; SHAPIRO, 2002; LIU, 2008; ZHANG; PAN; ZHANG, 2022; ZHANG; KOSECKA, 2005).

Um dos grandes problemas encontrados nesses sistemas de reconhecimento de edifícios é a grande quantidade de características a serem extraídas e processadas para a representação de imagens. De acordo com Li e Allinson (2013), existem vários tipos de características robustas para o reconhecimento de objetos, como forma, cor, textura, intensidade, movimento, etc., de tal forma que quanto mais tipos de características forem considerados, maior será a precisão para a classificação.

Por isso a escolha da escala de cores é um aspecto importante para a avaliação qualitativa dos resultados alcançados pela análise quantitativa das taxas de reconhecimento, obtidas pelo processamento. A escala de cores permite exibir o número de amostras em cada pixel convertendo-o ou em uma escala de cinza ou na cor

correspondente, por conseguinte altera a captação e influencia diretamente nas taxas de reconhecimento das imagens (SANTOS et al., 2017).

Alguns trabalhos relacionados à pesquisa se destacam, dentre eles, o trabalho de Zhang, Pan e Zhang (2022) que desenvolveram um sistema de reconhecimento de edifícios para detecção de defeitos de fachadas, utilizando *Deep Learning*. Esse sistema pode ser utilizado como um guia para o desenvolvimento sustentável de cidades. Ele detecta automaticamente elementos de fachadas de edifícios a partir de imagens utilizando o conhecimento prévio de engenharia, por exemplo, janelas, portas, paredes, etc.

Também motivados pelas pesquisas, Li e Allinson (2009) desenvolveram um sistema de reconhecimento de edifícios que após a extração de diversas características, produz um mapa de características para cada imagem. Depois disso, um modelo de saliência é construído pela divisão de cada característica do mapa em um número de sub-regiões e cada sub-região é descrita por uma característica da saliência. Então, para redução de dimensionalidade, essas características de saliências, são submetidas a um desses algoritmos de redução. São eles: análise de componentes principais (PCA), análise de discriminante linear (LDA), análise de discriminante semi-supervisionado (SDA), local preservando as projeções (LPP). Finalmente, aplica-se o resultado a um classificador que no trabalho desses autores, foi o k vizinho mais próximo.

Em suas pesquisas, Li e Shapiro (2002) utilizaram para cada segmento de linha da imagem a cor, a orientação e a informação espacial. Esses segmentos foram integrados e agrupadas em um tipo de característica de nível médio originando os clusters de linhas consistentes. Assim, as relações intracluster e intercluster foram usadas para reconhecer diferentes edifícios.

Zhang e Kosecká (2005) propuseram o reconhecimento de edifícios hierárquico (HBR). O HBR é baseado na detecção do ponto de fuga e nos histogramas de cores localizadas. Nesse sistema os segmentos de linha detectados são agrupados em direções dominantes de fuga e os pontos de fuga são estimados pelo algoritmo de máxima expectativa (EM). Assim, é utilizado um limiar da magnitude do gradiente do pixel da imagem e o resultado é um dos grupos denominados de: esquerdo, direito e vertical. Se a diferença entre sua direção gradiente e as direções dos pontos de fugas principais for menor que um limiar, é então calculado um histograma de cores localizadas apenas nestes pixels e o resultado obtido é indexado sobre um dos vetores definidos como esquerdo, direito e vertical. Finalmente, os histogramas são ajustados pela distância qui-quadrado,

e os resultados do reconhecimento são extraídos do banco de dados ZuBud, que contém 201 edifícios aleatórios de Zurique, Suíça, com 5 imagens por edifício, utilizando o descritor Scale-Invariant Feature (SIFT). Após, aplica-se nestes resultados um modelo probabilístico simples para integrar a evidência do ajuste individual.

Li et al. (2014) afirmaram que o sistema desenvolvido por Zhang e Kosecká possui limitações. Essas limitações são: (1) eles assumiram que em cada imagem a ser reconhecida só existe um edifício. No entanto, em cenas reais nem sempre isso acontece; (2) embora eles tenham utilizado um método de indexação rápido, o tempo de processamento ainda é longo para a extração de muitas características das imagens a cores; e (3) o desempenho do algoritmo no reconhecimento só é satisfatório quando o fundo da imagem é simples. Entretanto, o algoritmo é inadequado para sistemas de navegação que necessitam do processamento em tempo real.

Dessa forma, diferentemente do encontrado na literatura para reconhecimento de imagens de edifícios do banco de dados SBID (CHEN et al., 2019), este trabalho justifica-se na inovação de comparar estatisticamente as taxas obtidas pela técnica de Quantização Vetorial (QV), por meio do algoritmo Linde-Buzo-Gray (LBG), com uma técnica de Deep Learning, sendo que essa é uma Rede Neural Convolutiva (CNN) que tem por estratégia obter uma imagem de entrada, atribuir importância a diversos aspectos da imagem e ser capaz de caracterizar uma da outra. E, diante do exposto, apresenta-se como hipótese nula da pesquisa, há não existência de diferença significativa entre os dois métodos de reconhecimento de imagens analisados nesse trabalho.

Para discussão estatística dos dados, faz-se uma apresentação descritiva dos resultados estacando os melhores valores percentuais no reconhecimento para as duas técnicas com as variáveis cores, centroides e treino para Quantização Vetorial (QV) e treino para Deep Learning (DL). Para a estatística inferencial foi feita análise de variância (ANOVA), com teste post-hoc de Tukey, que é um teste paramétrico que testa a hipótese e auxilia a avaliar a importância de um ou mais fatores, comparando as médias das variáveis nos diferentes grupos e se os fatores exercem influência em alguma variável dependente (cor, centroide ou treino).

## **1.2 Objetivos**

### **1.2.1 Objetivo Geral**

Analisar estatisticamente as taxas de reconhecimento de imagens, de edifícios do banco de dados SBID, nas escalas de cores cinza, vermelha, verde, azul e RGB, obtidas pela técnica de Quantização Vetorial (QV), por meio do algoritmo Linde-Buzo-Gray (LBG), e comparar estatisticamente os resultados com maior significância encontrados com as taxas obtidas pela técnica de Deep Learning (DL).

### **1.2.2 Objetivos específicos**

Para alcançar o objetivo geral proposto, os seguintes objetivos específicos foram estabelecidos:

- Levantar o estado da arte referente ao reconhecimento de imagens, buscando conhecer as principais técnicas e padrões disponíveis dada a perspectiva histórica;
- Estudar a fundamentação teórica da Quantização Vetorial e Deep Learning;
- Comparar descritivamente e inferencialmente os percentuais de reconhecimento para as escalas de cores cinza, vermelho, verde, azul e RGB utilizando a técnica de QV por meio do algoritmo LBG para o reconhecimento de imagens classificando por quantidade de centroides e treinos.
- Discriminar os resultados significativos encontrados pela técnica de QV e analisar com os resultados de reconhecimento encontrado pela técnica de DL.

### 1.3 Organização do trabalho

Este trabalho foi escrito em 5 capítulos. Esse, o capítulo 1, contém a Introdução que apresenta o conteúdo de maneira geral, a abordagem do tema, as justificativas e os objetivos a serem alcançados.

O capítulo 2 apresenta o Referencial Teórico, fruto de uma extensa pesquisa nas bases de dados acadêmicas, contém trabalhos relacionados ao tema, e considerações sobre os aspectos de cada trabalho assim como a relação entre eles. No terceiro capítulo é abordada a Metodologia da pesquisa, com a descrição de cada etapa seguida para o desenvolvimento do projeto.

O terceiro capítulo destaca a importância de uma técnica eficiente no reconhecimento de edifícios em ambientes urbanos, apresenta o banco de dados utilizado no trabalho e a sua complexidade e por fim descreve sobre os dois métodos proposta nesse trabalho, Quantização Vetorial e *Deep Learning*.

O quarto capítulo expõe os resultados encontrados e sua discussão, apresentando as análises estatísticas descritiva e inferencial, sendo esse dividido em 7 seções: Análise dos resultados por escala de cor; Análise Comparativa entre os resultados nas escalas de cinza, vermelha, verde e azul; Análise dos resultados em RGB; Análise Comparativa dos resultados nas escalas de cinza, vermelho, verde, azul e RGB; Análise dos resultados de pela técnica de DL; Análise Comparativa dos resultados de QV RGB com DL e por último faz-se a comparação dos resultados finais do trabalho com os resultados encontrados na literatura.

Por fim, o quinto capítulo, apresenta a conclusão com as considerações sobre os resultados obtidos e as análises apresentadas nesse trabalho. Além disso, propõe sugestões que poderão ser realizadas em trabalhos futuros.

## CAPÍTULO 2

### FUNDAMENTAÇÃO TEÓRICA

#### 2.1 Considerações Iniciais

O reconhecimento digital de imagens provê inúmeras possibilidades para aplicações de interação humano-computador e tem atraído a atenção da comunidade científica nas últimas décadas. A estrutura deste capítulo é composta por 12 seções, nas quais apresentam-se definições computacionais sobre processamento de imagens, além de descrever o estado da arte sobre reconhecimento de imagens.

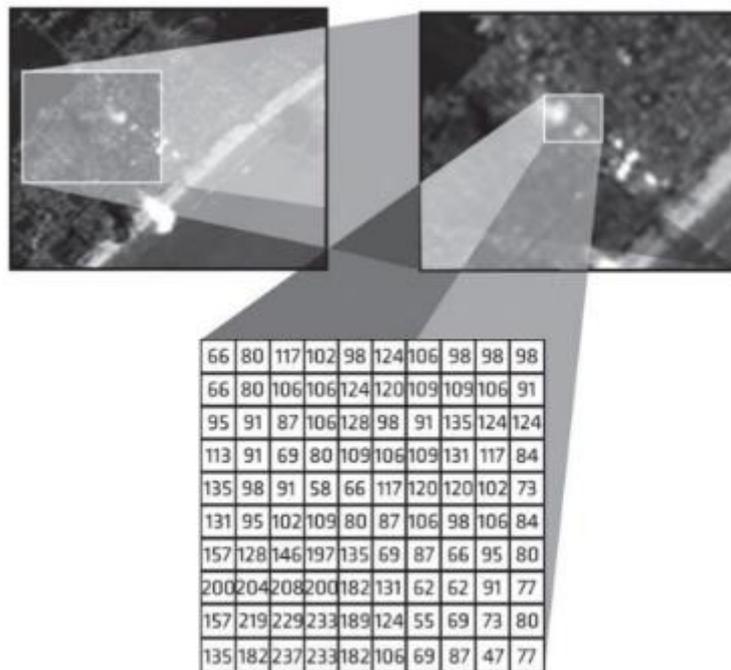
#### 2.2 Imagem Digital

Com a necessidade de se trabalhar com reconhecimento de imagens, é interessante saber como uma imagem é conhecida computacionalmente. Do ponto de vista matemático, conforme Zanotta, Ferreira e Zortea (2019) uma imagem é uma função bidimensional  $f(x,y)$ , armazenada em forma de matriz, onde  $x$  e  $y$  são coordenadas planas, e a amplitude  $f$  é a intensidade ou nível de cinza da imagem no referido ponto  $(x,y)$ .

Quando  $(x,y)$  e a amplitude de  $f$  fazem parte de um conjunto de valores finitos, ou discretos, a imagem é chamada de imagem digital. Contudo, o computador trabalha com números inteiros, o que, por consequência, não consegue representar uma função contínua, apenas simular.

O processo para trazer uma função contínua para o computador é discretizar ou digitalizar. Cada um dos pontos que formam uma imagem digitalizada é chamado de pixel, conforme Figura 2.1.

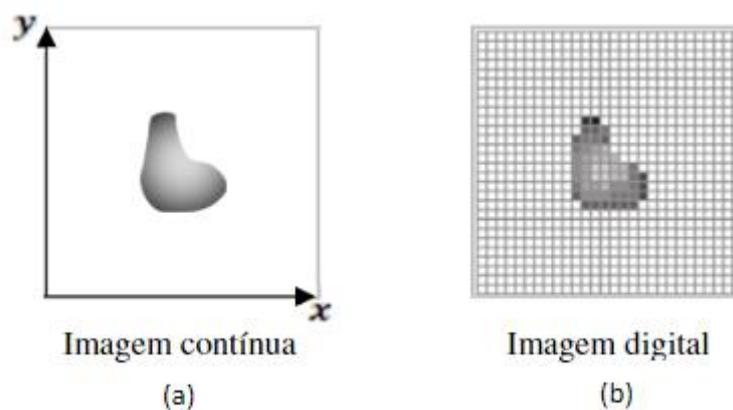
Figura 2.1 – Imagem detalhada mostrando os contadores digitais associados a cada nível de cinza dos pixels



Fonte: ZANOTTA; FERREIRA; ZORTEA (2019).

Um pixel é o elemento básico em uma imagem. O processo de discretização do eixo  $x$ , o domínio, é chamado de Amostragem, e o do eixo  $f(x)$ , o contradomínio, é chamado de Quantização (Figura 2.2).

Figura 2.2 – Processo de transformação da imagem contínua em imagem digital.



Fonte: JÚNIOR; SOARES (2019).

Dois aspectos importantes que definem a imagem são a resolução e a cor. Ao ser digitalizada, a imagem assume um tamanho adimensional, em pixels. A imagem colorida é representada por três matrizes de duas dimensões (2D) após a digitalização em RGB do espaço de cor. Dessa forma, é criada uma matriz para a cor vermelha, outra para a cor verde e por último para a cor azul, respectivamente.

Portanto, para uma imagem colorida tem-se três matrizes 2D do mesmo tamanho da imagem em pixels. E após o processamento tem-se, para cada imagem, três valores de cor, isto é, cada pixel é composto por um conjunto de 3 pontos: verde, vermelho e azul.

## **2.3 Processamento de Imagens**

Nas últimas décadas o campo de processamento de imagens digitais apresentou um extenso crescimento em diversas áreas de atuação, pois a identificação de imagens é muitas vezes necessária. No entanto, não existe, até o momento, um sistema de análise de imagens tão completo que funcione em todos os problemas propostos.

Dessa maneira, o campo do processamento digital de imagens refere-se ao tratamento de imagens digitais por meio de computadores e técnicas para a manipulação de imagens digitais e facilitar a extração das informações desejadas, seguindo uma serie de etapas. Essas etapas envolvem sequências de procedimentos, a saber: aquisição da imagem, pré-processamento, segmentação, pós-processamento, extração de atributos, reconhecimento e classificação.

### **2.3.1 Aquisição da imagem**

Nessa etapa as informações da imagem são convertidas em sinais elétricos por meio de dispositivos óticos, que possuem resolução espacial e de profundidade. A quantização realiza o mapeamento desse sinal, que é contínuo, em um número discreto e organiza espacialmente, de forma a representar também a profundidade do pixel (MELO JÚNIOR et al., 2018).

### **2.3.2 Pré – processamento**

Essa etapa trata os dados brutos, utiliza técnicas de processamento digital de acordo com a finalidade a ser atingida. Nesse etapa que atenua os ruídos, manipula o brilho e contraste da imagem, realça detalhes com algum interesse específico, etc., de forma a atribuir características que permitam obter informações com mais clareza e facilitar a interpretação e o reconhecimento (MELO JÚNIOR et al., 2018).

### 2.3.3 Segmentação

Esta etapa consiste em subdividir uma imagem em regiões como partes independentes, satisfazendo o critério de uniformidade. Como visto por Melo Júnior et al. (2018), a segmentação é uma das etapas mais complexas do processamento de imagens.

Ela busca se adequar às características de cada imagem e às informações que se deseja extrair. As técnicas de segmentação utilizam duas abordagens principais: a similaridade entre os pixels e a sua descontinuidade ou, métodos baseados em similaridade de contorno.

A técnica de similaridade é utilizada quando as amplitudes de níveis de cinza são suficientes para caracterizar um objeto e a saída do sistema é uma imagem com dois níveis de luminância: branco e preto (MELO JÚNIOR et al., 2018). Assim, pixels cujos valores excedem o limiar são alocados em uma determinada região e os pixels que não excedem o limiar são reservados em outra região, criando uma imagem binária (Equação 2.1).

$$b(x, y) = \begin{cases} 1, & \text{se } l(x, y) > T \\ 0, & \text{se } l(x, y) \leq T \end{cases} \quad 2.1$$

Onde:

T = limiar.

### 2.3.4 Pós processamento

O objetivo dessa etapa é melhorar o resultado obtido na segmentação, por meio de equações matemáticas e realizam uma análise quantitativa dos pixels da imagem.

### 2.3.5 Extração de atributos

Aqui se extraem as informações numéricas das imagens, como: tamanho, forma, posição e outras. Retira-se informações quantitativas de interesse, fundamentais para diferenciar ou comparar objetos, tais como edifícios.

### **2.3.6 Reconhecimento e classificação**

Essa etapa é a qual realiza-se de forma automática a identificação de objetos, de forma a determinar propriedades e padrões, e os acomodar em classes específicas.

## **2.4 Estado da Arte: Reconhecimento de Imagens**

Na década de 1970, Casasent e Psaltis (1976) destacam a necessidade de ampliar a flexibilidade das operações possíveis em um processador óptico por meio de transformações geométricas, pré-processamento de imagens por telas de meio-tom, operações não-lineares usando feedback óptico e transformações de Mellin.

Perelmuter et al., (1995) destaca que no começo da década de 1980, as técnicas digitais foram mais empregadas, devido aos avanços na microeletrônica e da arquitetura paralela de processamento de imagens. O reconhecimento de objetos representados em imagens é um problema central em visão computacional, as imagens originais são tratadas como sinais 2D consistindo na distribuição de uma escala de cinzento ou de cor.

Alguns métodos utilizam-se o gradiente da intensidade de brilho ou de cor da imagem, esta técnica é frequentemente utilizada, por exemplo, em algoritmos de reconhecimento de caracteres. No final da década de 1990, com a disseminação crescente de mídias eletrônicas, Carvalho; Sampaio; Mongiovi (1999), exhibe a necessidade das informações manuscritas, como memorandos, cartas, formulários, etc., serem convertidas eletronicamente.

Diante disso, expõe os resultados de uma aplicação de tecnologia de procura de informações, conhecida como regras associativas, para o reconhecimento automático de caracteres produzidos à mão. Na década de 2000, nesse estágio de desenvolvimento da indústria, surge as aplicações multimídia que oferecem sistemas de consultas de imagens-exemplo em grandes bases de imagens, como bibliotecas virtuais e catálogos eletrônicos.

Essas novas tecnologias de visão computacional e processamento de imagens foram utilizadas em muitas aplicações relevantes, principalmente nas áreas da astronomia, medicina, análise de impressões digitais, sensoriamento remoto, multimídia, entretenimento, reconhecimento de assinaturas, manufatura, robótica de manipuladores, robótica móvel, sistemas produtivos, entre outras (RUDEK; COELHO; CANGIOLIERI JUNIOR, 2001).

Contudo a maioria dessas aplicações requerem um tempo de resposta mínimo. Moschetta; Osório; Cavalheiro, (2002), apresentou como foco principal minimizar esse tempo de resposta, através da implementação concorrente e métodos de escalonamento de tarefas aplicados, bem como conceitos básicos sobre busca de imagens, as quais foram validadas por meio de uma análise de desempenho feita para medir tanto a qualidade do *matching* quanto o tempo de resposta da aplicação no reconhecimento de imagens.

Dois anos depois, Quiles e Romero (2004), objetivaram a construir um sistema de controle baseado em redes neurais aplicado em um robô, que o permite a seguir objetos pela cor. Muitas obras que desenvolveram um sistema de visão para controlar robôs, não lidam com o processamento de cores, o que, em alguns casos, pode trazer mais informações do que imagens de nível de cinza.

Koerich (2004), também apresenta um sistema de reconhecimento de manuscritos, o qual é baseado no paradigma segmentação–reconhecimento, onde palavras são livremente segmentadas em caracteres. Devido a crescente automatização dos processos produtivos, ao decorrer dos anos buscou-se tornar automático os sistemas computacionais e de robótica.

Em 2005, foi desenvolvido um sistema computacional de análise de imagens baseado na técnica HartmannShack de aprendizado de máquina para auxiliar o diagnóstico na área oftalmológica (VALERIO NETTO, 2005). A análise de imagens geralmente se refere ao seu processamento com o objetivo de encontrar objetos específicos.

Logo, o desafio foi a caracterização de texturas, que requer uma área de imagem grande para sua análise. Oliveira e Teixeira de Assis (2006), apresenta um processo automático de identificação de impressões digitais usando mapa de coeficientes no espectro de potência baseado no filtro de Gabor.

O núcleo da impressão digital é encontrado pela técnica de vizinhança de pixels, onde são extraídas informações sobre orientação de textura das imagens em 3 diferentes direções do filtro de Gabor.

Nunes e Conci (2007), apresenta um método de segmentação que pode ser usado em texturas naturais ou sintéticas, suaves ou ásperas, possibilitando a distinção com pequenas variações nos mesmos tipos de padrões, a partir do desenvolvimento de um coeficiente novo, denominado CVE. Este coeficiente considera a informação das bandas espectrais, levando em conta qual cor a combinação de bandas gera, bem como as posições espaciais dos pixels.

O CVE permite estimar os limites de regiões muito pequenas ou muito grandes, localizando corretamente as bordas da região de interesse em tempo real. Além disso, pode ser usado para todos os tipos de textura porque as regras do que será identificado são fornecidas pelo usuário e adaptadas à cada situação.

Em 2008, pela primeira vez é proposto um método de limiar de imagem usando uma entropia não extensiva, denominada como entropia *Tsallis*, a qual é uma das técnicas mais eficientes para a segmentação de imagens. Sabendo que, a segmentação da imagem, é uma das tarefas mais críticas na análise automática dessas (ALBUQUERQUE; ESQUEF, 2008).

Ainda nesse ano, teve-se aplicações científicas na descrição da utilização de filtragem Kalman para tratar projeções tomográficas da ciência do solo, para melhorias na razão Sinal-Ruído (LAIA; CRUVINEL, 2008). A validação do resultado foi realizada usando ISNR (melhoria na razão sinal-ruído), bem como a perda dos detalhes das imagens produzidas, que foram geradas utilizando o algoritmo de retroprojeção filtrada. É também apresentada uma combinação Filtragem de Kalman com Redes Neurais Artificiais para tratar as projeções degradadas com ruído Poisson.

Kalva, Enembreck e Koerich (2008), compilaram essas técnicas para a classificação de imagens que combina com informações contextuais. Sua hipótese principal é que a informação contextual relacionada a uma imagem pode contribuir no seu processo de classificação.

Foram coletadas páginas da Web contendo imagens e texto e armazenadas de forma a construir um banco de dados. Das imagens foram extraídas várias características como cor, forma e textura.

Esses recursos combinados formam vetores de características que são usados em conjunto com um classificador de rede neural. A informação contextual é processada e usada em conjunto com um classificador Naïve Bayes.

Os resultados experimentais em um banco de dados de mais de 5.000 imagens mostraram que a combinação de classificadores fornece uma melhoria significativa, de cerca de 16%, na taxa de classificação de imagem em relação aos resultados fornecidos baseados na rede neural que não usa informações contextuais.

Oliveira e Tavares (2009), apresenta uma metodologia para combinar contornos de objetos representados em imagens, usando conjuntos de pontos ordenados extraídos dos contornos externos dos objetos para calcular a transformação geométrica da similaridade que alinha melhor os contornos correspondentes.

Conci, Carvalho e Rauber (2009), mostraram um sistema automático de reconhecimento óptico de caracteres que resolve o problema prático da identificação do carro para cenas reais, isto é, devido a diversos efeitos como nevoeiro, chuva, sombras, condições de iluminação irregular, oclusão parcial, distâncias variáveis, velocidade dos carros, ângulo da cena no quadro, rotação e conservação da placa, número de veículos na cena e outro, o reconhecimento da placa de licença do carro tem características complexas.

Outro trabalho, em 2009, utilizando o processamento automático de reconhecimento de imagens, foi o dos autores Neves et al. (2009), no qual fez-se a caracterização em cheques bancários, que é uma tarefa onde exige eficiência e precisão, pois os bancos adicionaram diferentes padrões de fundo nas verificações para aumentar sua segurança. Para o reconhecimento automático do conteúdo do cheque, os sistemas precisam remover esse fundo, pois não contém nenhuma informação válida.

Isso pode ser feito limitando algoritmos que deixam apenas as cores de tinta permanecerem. Neves et al. (2009), apresentam um algoritmo que usa o conceito de entropia *Tsallis* para encontrar o melhor valor de corte e conseqüentemente um limiar das imagens de cheques bancários.

Nos últimos anos, os esforços voltaram para aumentar a acessibilidade e incluir digitalmente pessoas com deficiências físicas. Digiampietri et al. (2012), desenvolveu um conjunto de módulos especificamente para o processamento de imagens e vídeos de forma a permitir o reconhecimento automático de Libras, baseado no uso de *workflows* científicos e composto por métodos para a segmentação e classificação de imagens.

Sousa (2012) apresenta a construção de um aplicativo destinado a auxiliar os deficientes visuais no cruzamento de ruas usando dispositivos móveis. Este trabalho também examinou os desafios e algumas soluções para desenvolver interfaces móveis disponíveis que executam o processamento de imagem e reconhecimento de voz para a detecção de passagem de pedestres.

Teve-se ainda aplicações nas atividades de comércio e preservação, com a identificação de espécies. Devido à escassez de dados e bases de imagens florestais e técnicas computacionais para essa tarefa, Schenatto, De Paula Filho e Bazzi (2012), propôs uma abordagem de identificação onde foram construídas duas bases de imagens macroscópicas a partir de amostras de madeira de espécies florestais, considerando dois métodos diferentes: abordagem tradicional em laboratório e abordagem em campo.

As imagens são divididas em subimagens e são extraídas informações de cor e textura, que são utilizadas para a construção de classificadores. Para extração desses atributos são avaliadas diversas técnicas: análises de cor, GLCM, histograma de borda, Fractais, LBP, LPQ e Gabor.

Pereira, Santos e Cozman (2012), propuseram uma abordagem em que o contexto de objetos é modelado por meio de uma lógica de descrição probabilística, onde os dados extraídos por visão computacional são fontes de informação para deduções sobre objetos. Através do algoritmo *Speed Up Robust Features* obtiveram descritores de objetos em imagens capturadas de dois ambientes distintos.

Amaral, Giraldi e Thomaz (2012) propuseram e implementaram uma comparação entre as técnicas *Local Binary Pattern* (LBP) e *Principal Components Analysis* (PCA), para entender o comportamento de cada método de extração de recursos ao usar imagens de rosto previamente processadas e previamente especializadas, o qual mostrou uma boa performance pela LBP apenas em expressão facial neutra e pose frontal. Com isso, em 2013, os autores propõem e implementam uma nova abordagem para aprimorar o processo de extração e classificação do LBP em análise de imagens frontais de faces.

Com base na definição de grupos, compostos por faces com expressão neutra ou sorrindo, realizou-se um mapeamento das regiões estatisticamente significantes e utilizou-se esta informação para identificar as características faciais mais relevantes.

Em 2014, desenvolveu-se um aplicativo para o sistema ANDROID utilizando um sistema de reconhecimento óptico de caracteres para detectar e reconhecer a sinalização vertical de regulamentação de trânsito, especificamente para o sinal velocidade máxima permitida nos padrões brasileiros (GOMES; REBOUÇAS; FILHO, 2014). Para o desenvolvimento deste aplicativo foram utilizadas técnicas de Rede Neural Artificial.

A evolução tecnológica da atualidade impulsiona a economia em nível mundial, especialmente no que diz respeito ao progresso das indústrias eletrônicas. Neste contexto, a inovação computacional e dos dispositivos de imagem têm disposto regularmente uma crescente aplicação na visão computacional aliado a uma maior segurança e um sistema de reconhecimento.

Atualmente, a maioria das investigações feitas nesse campo tem sido na detecção automática de sinais de trânsito e sistemas de reconhecimento, com uma vasta gama de aplicações tais como: condução automática, sistemas de assistência a condutores e inventário automático de estradas. Vungo (2017) desenvolveu um método de reconhecimento e classificação de sinais de trânsito, nomeadamente da sinalização de

perigo e alguns da regulamentação, previamente identificados em imagens digitais obtidas nas bases de dados *German Traffic Sign Recognition Benchmark* e da Faculdade de Ciências da Universidade do Porto.

A metodologia consiste em extrair as características dos sinais. A metodologia consiste em três fases; a aquisição da imagem, que é a extração da imagem na base de dados, feita por uma função criada no MATLAB, a qual utiliza um processo de segmentação a partir de histogramas das componentes binárias das cores vermelha, verde e azul da imagem original e conseqüentemente a sua binarização. A fase da classificação, que tem como base o reconhecimento da forma geométrica de cada região e sua cor. E a fase de reconhecimento, onde identifica a sinalização que teve entrada e classifica conforme a correspondência dos pixels presentes no sinal.

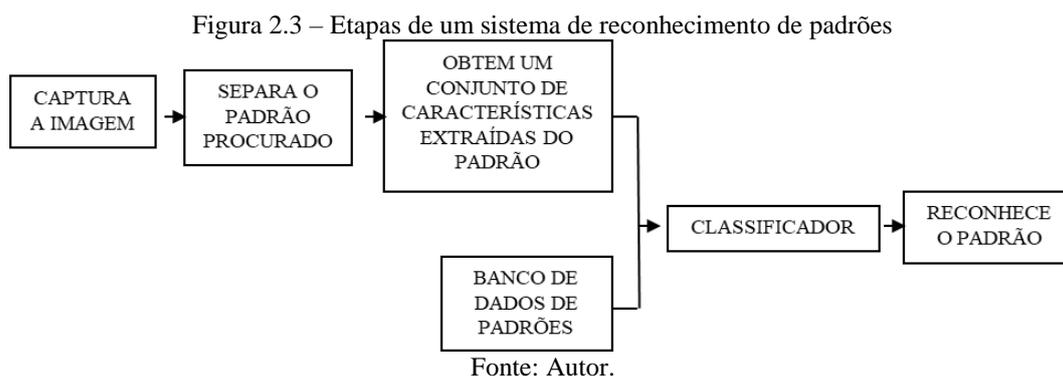
## 2.5 Reconhecimento de Padrões

Um padrão pode ser considerado como um conjunto de características semelhantes. Suas técnicas de reconhecimento podem ser divididas em duas classes: não supervisionadas e supervisionadas (BONAFINI; PERLIN, 2018).

As técnicas não supervisionadas, têm a finalidade de verificar a formação natural de agrupamentos; e os métodos supervisionados, utilizam informação prévia sobre um conjunto de amostras conhecidas (BONAFINI; PERLIN, 2018).

Contudo, em sistemas de reconhecimento de padrões, diferentes formatos de representação dos dados podem trazer mudanças significativas para o desempenho da obtenção de resultados (ARAÚJO; MOYA; JUNIOR, 2017).

Um projeto de reconhecimento de padrões envolve normalmente extração de características dos objetos a descrever; seleção das características mais discriminativas; e construção de um classificador (Figura 2.3).



Araújo; Moya; Junior (2017) apresenta que esse processo se inicia pelo sensor que obtém as observações que serão descritas, chamado de filtragem de entrada, cujo objetivo principal é eliminar dados desnecessários ou distorcidos. Em seguida, o mecanismo de extração de características efetua a análise dos dados de entrada de forma a extrair e derivar essas informações úteis para o processo de reconhecimento.

Desta maneira, as informações serão registradas. Por último, os resultados serão categorizados conforme seus aspectos de forma, dimensão, cor, textura, etc., através do classificador. Para funcionarem corretamente, estes classificadores não podem ser nem generalizados, pois sua classificação será muito vaga e nem específicos, pois não analisarão elementos relevantes por considerarem apenas um tipo de constante (ARAÚJO; MOYA; JUNIOR, 2017).

### **2.5.1 Métodos Utilizados no Reconhecimento de Padrões**

Fu (1982) destaca que as técnicas matemáticas usadas para resolver os problemas de Reconhecimento de Padrões podem ser agrupadas em dois procedimentos gerais: o método sintático, linguístico ou estrutural e o método de decisão teórica ou discriminante.

No método sintático, linguístico ou estrutural, a informação que descreve cada padrão é importante, e o processo de reconhecimento inclui, além da capacidade de designar uma classe, a capacidade de descrever os aspectos do padrão é de uma tal forma, que não possa ser designado para outra classe.

O método sintático ou estrutural trabalha com problemas onde os padrões são considerados complexos e o número de características é alto. Cada padrão é descrito em termos de sub padrões e cada um destes em suas partes, chegando até mais simples chamados de primitivos. Seu funcionamento é análogo a gramática de uma linguagem, onde forma frases e sentenças pela concatenação de palavras e estas por letras.

No método estatístico, o conjunto de medidas características são extraídas dos padrões. A designação do reconhecimento de padrões para determinadas classes é, normalmente, feita pelo particionamento do espaço característico. A ferramenta básica é a teoria da decisão estatística.

Mais recentemente o desenvolvimento do estudo em Redes Neurais considera esta uma terceira abordagem de Reconhecimento de Padrões, conceituado mais adiante.

## 2.6 Algoritmos Genéticos

Os algoritmos genéticos são uma técnica de busca e otimização introduzida por John Holland em 1975 (SILVA, 2019). Esses algoritmos buscam através de métodos heurísticos uma solução baseada na teoria da evolução de Darwin.

Fazem parte dos algoritmos evolucionários, que são algoritmos que usam modelos computacionais, a partir de operadores genéticos, dos processos naturais de evolução como uma ferramenta para resolver problemas (SILVA, 2019).

Operadores genéticos podem ser determinados como aproximações computacionais de fenômenos vistos na natureza, tais como reprodução sexuada, mutação genética, seleção de indivíduos, etc. É necessário que um indivíduo mostre características que lhe atribuam destaque dentro do grupo para que ele possa, conseqüentemente, sobreviver para uma próxima geração.

Após a seleção dos indivíduos em uma população, são aplicados os operadores genéticos, obtendo-se conseqüentemente uma população com melhores indivíduos. Esse processo é repetido até que se encontre uma solução ou até que se perceba que não serão alcançadas melhores soluções (JÚNIOR; ABREU, 2018).

Os algoritmos genéticos básicos são geralmente constituídos de dois operadores: cruzamento e mutação (JÚNIOR; ABREU, 2018). Eles buscam gerar uma população inicial aleatória para a melhor solução usando métodos probabilísticos que usam a evolução, de acordo com a capacidade de adaptação.

Como um processo inteligente de otimização na busca de soluções de boa qualidade para problemas que não contém algoritmos específicos. Isto é, seleciona os melhores indivíduos da população para servir como solução do problema, onde não existe um algoritmo conhecido, gerando-se uma população inicial e, de acordo com critérios de avaliação, selecionando os melhores indivíduos dessa população, que servirão como solução para o problema ou, caso contrário, serão combinados para obter uma nova geração.

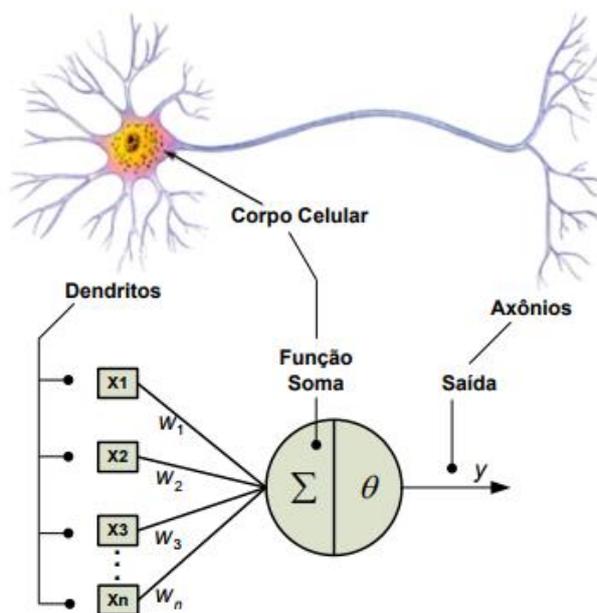
Os algoritmos genéticos não consistem em um processo baseado à obtenção da solução ótima. Mas em fazer competir um conjunto de indivíduos, fazendo com que sobrevivam aqueles que são mais aptos (JÚNIOR; ABREU, 2018; SILVA, 2019).

## 2.7 Rede Neural Artificial

As redes neurais artificiais são sistemas computacionais paralelos constituídos por unidades de processamento simples, também denominadas neurônios artificiais ou nodos, conectadas entre si de maneira específica para desempenhar determinada tarefa (BELTRÃO, 2022). Além de imitar as habilidades do sistema nervoso biológico, apresentam estrutura maciça e paralelamente distribuída em camadas; habilidade de aprender e generalizar, que as tornam capazes de resolver problemas complexos; são tolerantes a falhas e ruídos e; podem modelar diversas variáveis e suas relações não lineares.

A construção de redes neurais artificiais tem inspiração nos neurônios biológicos e nos sistemas nervosos (Figura 2.4).

Figura 2.4 – Comparação entre neurônio biológico e neurônio artificial.



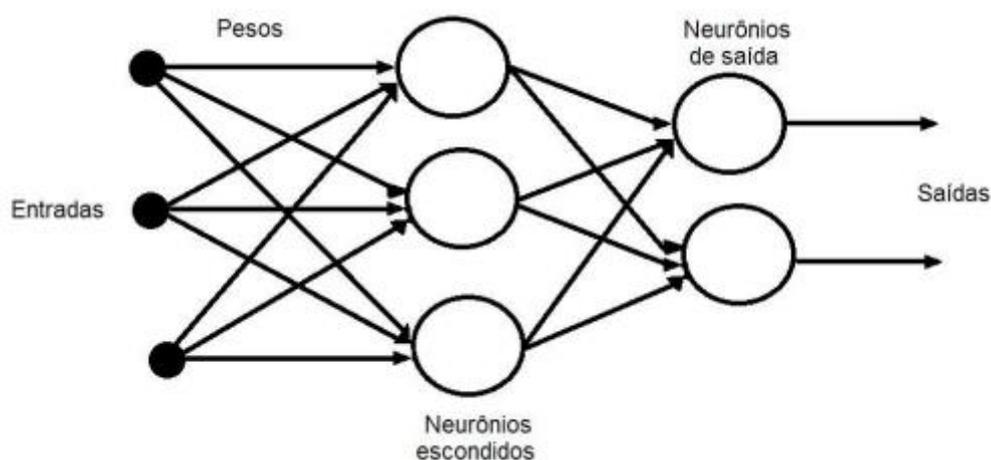
Fonte: (CLAUDIO, 2014).

Os três elementos básicos são: 1) um conjunto de conexões de entrada, cada uma delas possuindo um peso associado; dessa forma, cada entrada  $x$  que está ligada a um dado neurônio deve ser multiplicada pelo seu respectivo peso  $w$ ; 2) um somador, para realizar a soma dos sinais de entrada pelos seus respectivos pesos  $e$ ; 3) uma função de transferência para gerar uma saída. O objetivo da função de transferência é limitar a amplitude da saída do neurônio.

Seu processo de funcionamento consiste na distribuição de três tipos de neurônios, os que recebem excitações do ambiente externo e são chamados neurônios de entrada, relativos aos dendritos. Os que alteram as respostas recebidas pelo anterior que são os neurônios de saída e correspondem aos axônios. E os conhecidos como neurônios internos. Os neurônios internos têm grande importância à rede, são chamados, na literatura, como *hidden*, traduzido como escondidos (SOUZA; ARAÚJO, 2021).

Desse modo, uma rede neural passa a ser formada por três tipos básicos de neurônios: neurônios de entrada, intermediários e de saída, conforme demonstrado na Figura 2.5.

Figura 2.5 – Modelo de uma estrutura de rede neural alimentada por múltiplas camadas.



Fonte: (STEIM, 2017).

Na camada de entrada são recebidos os valores iniciais de dados, processados pelos seus respectivos pesos e transmitidos para as camadas intermediárias. Por fim, a camada de saída, onde recebe os valores da camada intermediária e calcula o valor de saída.

Com isso, as características de uma rede neural, são, a capacidade de estabelecer relações não lineares e a de aprendizado. O aprendizado pode ser definido como a mudança no comportamento, ou seja, é um processo de antes e depois, em que a performance é medida antes e depois do processo. A diferença entre eles indica o quanto a rede aprendeu (BELTRÃO, 2022).

Em uma rede neural, logo que a entrada é processada pelo sistema, ela pode ser comparada com um valor de saída que é desejado, dessa forma, os pesos são modificados proporcionalmente, aumentando ou diminuindo, dependendo da direção da diferença. O

objetivo passa a ser, então, processar um grande número de casos para minimizar a diferença existente entre o valor de saída, calculado pela rede, e o valor desejado.

## 2.8 Análise de Componentes Principais

Nos modelos de regressão múltipla, é muito comum a ocorrência de variáveis independentes altamente correlacionadas, resultando em coeficientes de regressão estimados com baixa precisão. Por isso, para que se tenha um resultado mais confiável, é necessário que esses dados passem por alguns tratamentos estatísticos.

Para investigar as relações entre um conjunto de  $p$  variáveis correlacionadas, pode ser útil transformar o conjunto de variáveis originais em um novo conjunto de variáveis não-correlacionadas, chamadas componentes principais. Entre as várias alternativas que existem para reduzir a dimensionalidade dos dados, uma delas consiste na utilização dessas componentes principais.

Como nos modelos de regressão, cujo propósito é a explicação da variável dependente, devem-se reter aquelas componentes principais que têm altas correlações com a variável dependente. Os principais objetivos das componentes principais são: a) reduzir o número de variáveis; b) analisar quais as variáveis ou quais conjuntos de variáveis explicam a maior parte da variabilidade total, revelando que tipo de relacionamento existe entre eles (TRAMONTIN, 2016).

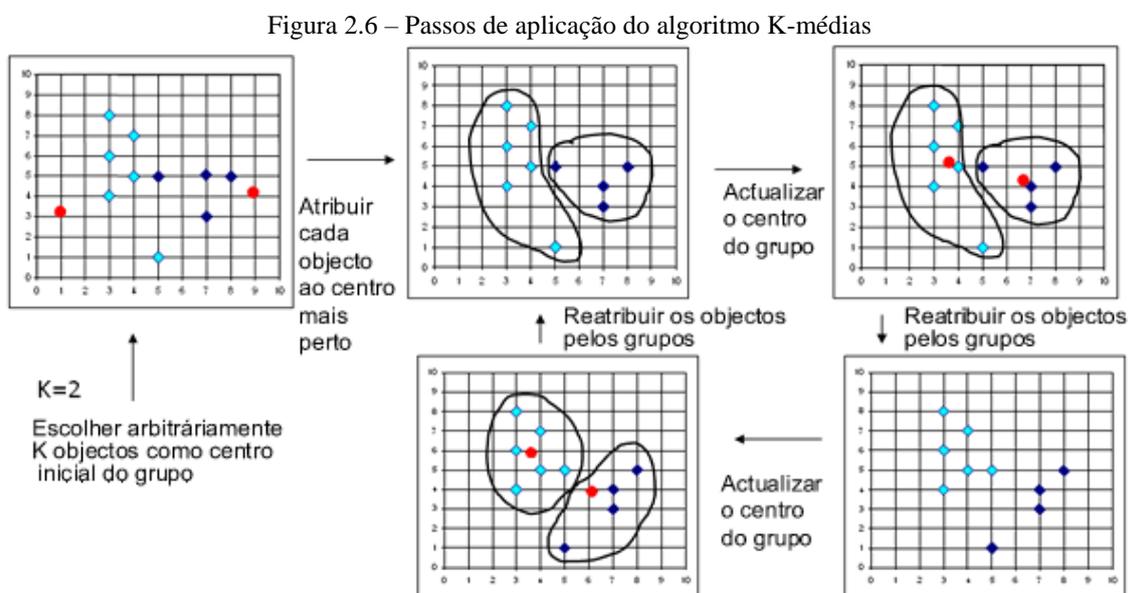
O método das componentes principais permite uma redução significativa no número de variáveis. Portanto, a análise de componentes principais é a base para diversos métodos de reconhecimento de padrões, classificação e calibração multivariada.

Menezes Junior et al. (2017) definiu-a como uma técnica estatística de análise múltipla utilizada para reduzir a dimensionalidade de um conjunto de dados com a preservação do máximo de variabilidade nas covariáveis. Normalmente, essa análise é empregada com o objetivo de visualizar a estrutura dos dados, encontrar similaridades entre amostras, detectar amostras anômalas, os outliers, e reduzir a dimensionalidade dos dados (TRAMONTIN, 2016).

## 2.9 K-Média

Os métodos automatizados, quando aplicados em um conjunto de dados, conseguem revelar padrões, semelhanças e diferenças que inicialmente estavam ocultas na massa de dados (TEIXEIRA, 2016). Estes são utilizados para destacar as características da massa de dados e obter agrupamentos, chamados de *clusters*.

A partir desses *clusters* pode-se visualizar como é o comportamento da maioria dos dados amostrais. É possível ainda que a partir deles sejam empregados métodos estatísticos para determinar a tipologia da amostra e reduzir uma grande quantidade de curvas características em uma curva típica a cada grupo, que representa todos os dados com características semelhantes (Figura 2.6).



Fonte: GFBioinfo.

O método de *K*-médias é um classificador não – supervisionado que utiliza uma abordagem de agrupamento. Foi desenvolvido para agrupar elementos em *k* grupos, onde *k* é a quantidade de grupos. Isto é, a imagem é partida em *K* grupos, cada pixel da imagem é alocado ao centro mais próximo segundo a distância euclidiana (TEIXEIRA, 2016).

Segundo Lucena et al. (2016), o *k*-médias é um método de partição que fornece indicações mais precisas sobre o número de grupos a ser formado. Esse método é bastante utilizado para agrupar objetos similares.

Esse algoritmo foi desenvolvido por J. B. MacQueen em 1967 e um algoritmo ótimo para utilização do *k*-médias foi desenvolvido por Hartigan e Wong em 1979, no

qual consiste em encontrar a melhor partição iniciando com um conjunto de observações indicando o número de grupos (TEIXEIRA, 2016). No passo seguinte encontram-se os centroides de cada grupo, a partir do método dos mínimos quadrados (Equação 2.2), e coloca cada elemento ao grupo do centroide mais próximo.

$$J = \sum_{j=1}^k \sum_{n \in S_j} \|x_n - c_j\|^2 \quad 2.2$$

Feito isto, recalcula-os de acordo com os elementos alocados (Equação 2.3).

$$C_{i,j} = \frac{\sum_{n \in S_j} x_{n,i}}{Q(c_j)} \quad 2.3$$

Onde:

$Q(c_j)$  é o número de objetos que pertence ao grupo  $c_j$

$c_{i,j}$  é a dimensão  $i$  do centroide  $j$ , para  $i = \{1, \dots, n\}$  e  $j = \{1, \dots, k\}$

$x_{n,i}$  é a dimensão  $i$  do exemplo de treino  $x_n$  pertencente ao centroide  $j$ , para  $i = \{1, \dots, n\}$  e  $j = \{1, \dots, k\}$

Isto é feito até encontrar um grupo ótimo. A grande vantagem desta heurística é a simplicidade e conseqüentemente a alta velocidade em que apresenta os resultados. Entretanto, o método é sensível aos *outliers*, dados que apresentam um afastamento dos demais, que são pontos que não pertencem a nenhum agrupamento. Assim, podendo distorcer os resultados e diminuir a qualidade das informações obtidas (LUCENA et al., 2016).

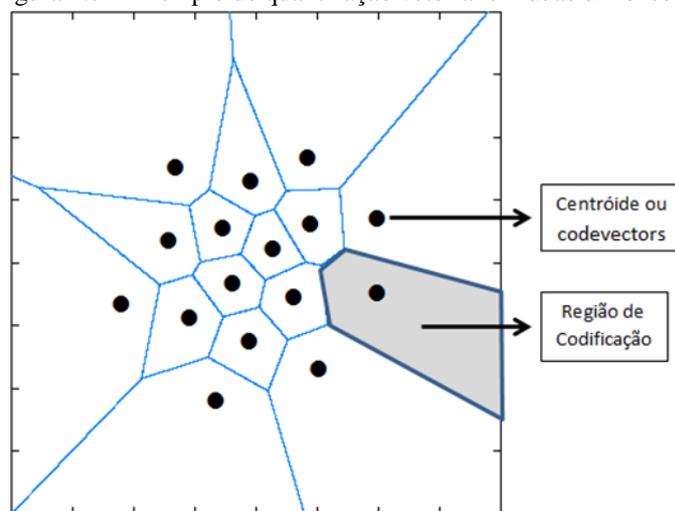
## 2.10 Quantização Vetorial

Silva et al. (2022) destacam que o alvo do processamento de imagens é a digitalização de valores de amostras de uma imagem por meio da compressão de seus dados e agrupamento, isto é mapeamento de valores de um conjunto maior para um conjunto menor, conhecido como clusterização, que é o processo de agrupar um conjunto de dados em classes ou grupos, também chamados de clusters, de forma que amostras do mesmo grupo apresentem alta similaridade entre si. Para tanto, existe a possibilidade de

melhoria do desempenho dessa compressão utilizando técnicas de Quantização Vetorial (QV), uma técnica relevante em sistemas de mapeamento e imagens (SILVA et al., 2022).

A QV traz como objetivo representar distribuições de dados utilizando um número de referência de padrões significativamente menor que o número de dados (VASCONCELOS; MELO; CARRIJO, 2022). Dessa forma, a compressão de um conjunto de dados busca diminuir a quantidade de informação, tanto para transmissão quanto para armazenamento (LINDE; BUZO; GRAY, 1980), pois apenas os vetores-referências ou centroides precisam ser armazenados, ao invés da base de dados inteira (Figura 2.7).

Figura 2.7 – Exemplo de quantização vetorial em duas dimensões.



Fonte: VASCONCELOS et al. (2022).

As regiões definidas pelas linhas de fronteiras são intituladas regiões de codificação. A região destacada exemplifica uma região de codificação, sendo que essa possui um centróide correspondente a ela, de forma que para mapear qualquer dado dessa região o centróide é suficiente. O conjunto de todos os centroides é conhecido por *codebook* ou livro de códigos, que é uma lista previamente computada, e o conjunto de todas as regiões de codificação é chamado de partição do espaço.

Os centroides servem como elemento representativo dos agrupamentos obtidos, sua quantidade é escolhida com base na Equação 2.4 (LINDE; BUZO; GRAY, 1980).

$$C = 2^m \quad 2.4$$

Onde:

C = quantidade de centroides;

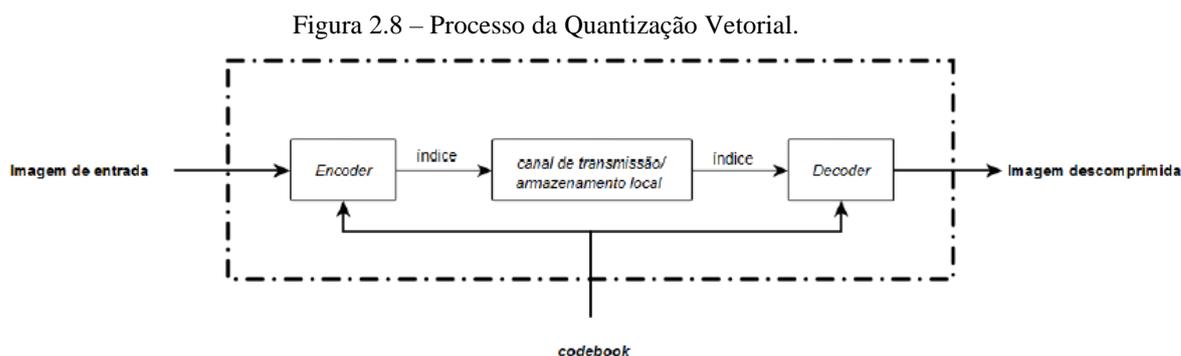
$m = \epsilon$  é um valor de entrada pré-estabelecido.

O conjunto de todos os centroides é conhecido por *codebook* ou livro de códigos, que é uma lista previamente computada, e o conjunto de todas as regiões de codificação é chamado de partição do espaço. A similaridade da imagem processada com a imagem original depende diretamente do *codebook*. A distância entre os vetores e os centroides pode ser medida de diversas formas, sendo a mais usada, a Distância Euclidiana ao Quadrado (Equação 2.5).

$$d(x, \hat{x}) = \sum_{i=0}^{k-1} |x_i - \hat{x}_i|^2 \quad 2.5$$

Ressalta-se que, além da compressão de imagens, há um amplo espectro de aplicações para a QV, como, por exemplo, esteganografia (VASCONCELOS; MELO; CARRIJO, 2022), marca d'água digital, identificação vocal e classificação de sinais de voz com patologias (ROCHA; SILVA; BARROS, 2019).

Na QV a imagem de entrada para o treino é analisada a fim de determinar o *codebook*, contudo a complexidade computacional, na fase de codificação dos vetores a serem quantizados, constitui alguns desafios, tais como o projeto de criação de *codebooks* e a sensibilidade da técnica aos erros de transmissão (Figura 2.8).



Fonte: (GROENEWEG et al., 2006).

Sendo assim, alguns algoritmos podem ser empregados. Esses algoritmos quando incorporados leva a uma redução do tempo de execução no processamento de imagens, além de encontrar um *codebook* ótimo para a representação vetorial do conjunto de dados processados, de forma a minimizar o erro inerente quando utilizado para comprimir imagens. Entre esses algoritmos de compressão de imagens destaca-se o Linde-Buzo-

Gray (LBG). O LBG é uma técnica que infere a maneira como os dados estão organizados e relacionados em cada grupo, ele divide a imagem em centroides, criando um novo *codebook* que é iterativamente atualizado, até alcançar a menor distância, ou seja, o *codebook* ótimo (LINDE; BUZO; GRAY, 1980).

## 2.11 Inteligência artificial e aprendizado de máquina

A Inteligência Artificial (IA) busca desenvolver sistemas artificiais inteligentes de tomada de decisões e resolução de problemas. É utilizada para mineração de banco de dados, diagnósticos médicos e odontológicos, robótica, reconhecimento de voz e de imagem, processamento de linguagem natural, na área agrícola, entre outros (NIU et al., 2016).

Já o Aprendizado de Máquina é uma área da IA que compreende o desenvolvimento de algoritmos, técnicas e programas que possibilitam que os computadores sejam capazes de adquirir conhecimento e aperfeiçoar seu desempenho (OLIVEIRA, 2018). Ele aborda conceitos de ciência da computação, matemática, estatística, física e ciência dos dados, além de ser um dos campos que mais cresce em visão computacional.

Segunda Oliveira (2018) o AM pode ser realizado por duas abordagens distintas, o Aprendizado Dedutivo, que usa uma regra geral para casos particulares, baseando-se no raciocínio dedutivo; e o Aprendizado Indutivo que trata de inferências lógicas de forma iterativa para a obtenção de conclusões genéricas a partir de um conjunto particular de exemplos.

Como visto por Oliveira (2018) a IA possui cinco paradigmas clássicos de AM, os quais podem ser utilizados na construção de sistemas individuais ou híbridos, a saber:

- Simbólico: o aprendizado acontece por meio de representações simbólicas, utilizando exemplos e contraexemplos, como expressões lógicas, regras de produção, árvore de decisão ou rede semântica;
- Estatístico: usa modelos estatísticos para a aproximação do resultado induzido, podendo ser métodos paramétricos que assumem alguma forma de modelo a fim de encontrar os valores apropriados para estes parâmetros;

- Baseado em treinos: sistemas que classificam novos exemplos com base em exemplos conhecidos. Treina o banco de dados armazenado previamente para que a classificação do novo exemplo aconteça;
- Evolucionário: baseia-se na evolução natural e na reprodução genética, simulando modelos biológicos para evoluir soluções;
- Conexionista: trata-se da dinâmica das conexões entre nós, chamados também de neurônios. Simula funções dos neurônios humanos, a partir do conceito de Redes Neurais Artificiais (RNA), que são capazes de simular características do cérebro humano como aprendizado, associação, generalização e abstração. Tais redes são caracterizadas como estruturas de Aprendizado Indutivo Supervisionado Conexionista.

## 2.12 *Deep Learning*

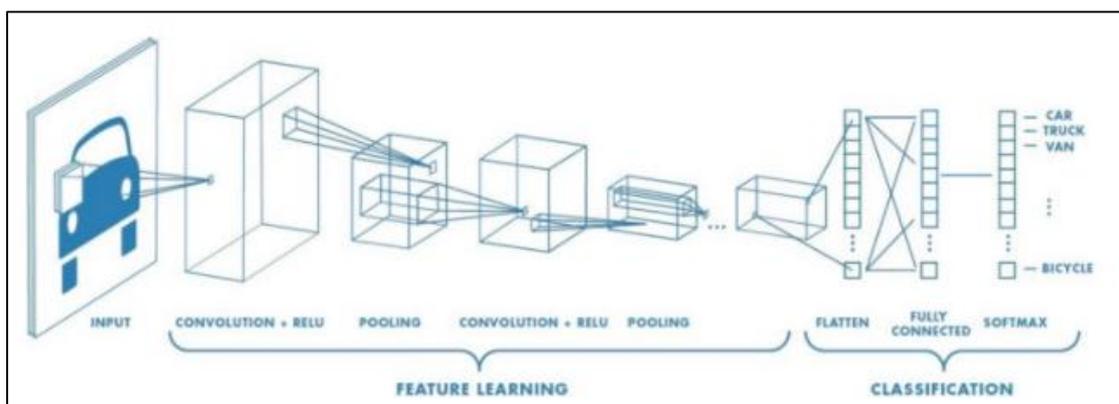
A transformação de grandes quantidades de dados em reconhecimento está tornando cada vez mais importante e utilizada em várias áreas (OTTONI; NOVO, 2021). Para extrair informações dessas imagens o *Deep Learning*, que é um ramo do aprendizado de máquina, tem sido uma metodologia amplamente utilizada.

O *Deep Learning*, é uma técnica de treinamento de arquiteturas profundas, que processa os dados de forma rápida. Dentre essas arquiteturas tem-se, as redes neurais convolucionais (CNNs), as LSTMs, as redes bayesianas, as convolucionais totalmente conectadas, entre outras (CHEN et al., 2021).

Por sua vez, a CNN, procura padrões em grandes quantidades de dados, em outras palavras, ela treina o computador a aprender sozinho por meio do reconhecimento de padrões. A CNN consiste em uma rede neural de camada profunda com características de desempenho em problemas de visão computacional que se destacam, sendo muito aplicada em problemas de classificação de imagens, detecção de objetos, etc. (CHEN et al., 2019).

A CNN é constituída de algumas camadas, sendo as principais: camadas de convolução, camadas de agrupamento e camadas densas (OTTONI; NOVO, 2021), como ilustrado na Figura 2.9.

Figura 2.9 – Camadas de uma CNN



Fonte: (VASCONCELOS et al., 2022).

A camada de convolução é responsável por transformar a imagem de entrada com o objetivo de extrair características que possibilitem uma distinção correta. Para a extração dessas informações, utiliza-se filtros de convolução, através de matrizes denominadas máscaras ou kernel, sendo que em algumas aplicações existe a possibilidade de utilização de mais de um kernel de convolução em uma mesma imagem, a fim de extrair o máximo de características possíveis.

A camada de agrupamento ou *pooling* é responsável por reduzir o tamanho da imagem, mantendo-se apenas os recursos mais importantes e removendo a área restante da imagem, o que na prática, reduz o custo computacional. Na camada de *pooling*, o tamanho da matriz de agrupamento é quem determinará a taxa de redução da imagem, por exemplo, uma matriz 2x2 reduzirá o tamanho da imagem em 50% (LEE et al., 2020).

A camada densa é executada após a rede neural estar completamente conectada, é nela que se informa o tamanho da imagem de saída e a função de ativação. Para o reconhecimento de imagens usualmente constrói-se a camada densa com a função de ativação unidade linear retificada (ReLU), sendo essa aplicada aos valores convolutos a fim de aumentar a não linearidade dos mesmos. Após é construída outra camada densa, com dimensão igual a quantidade de classes, a serem caracterizadas com a função de ativação “softmax”.

## CAPÍTULO 3

### O MODELO PROPOSTO NESTE TRABALHO

Este capítulo está estruturado da seguinte forma: a seção 3.1 traz as considerações iniciais sobre o reconhecimento de prédios; a seção 3.2 apresenta o banco de dados utilizado e sua complexidade. E por fim a seção 3.3 descreve o método proposto usado neste trabalho.

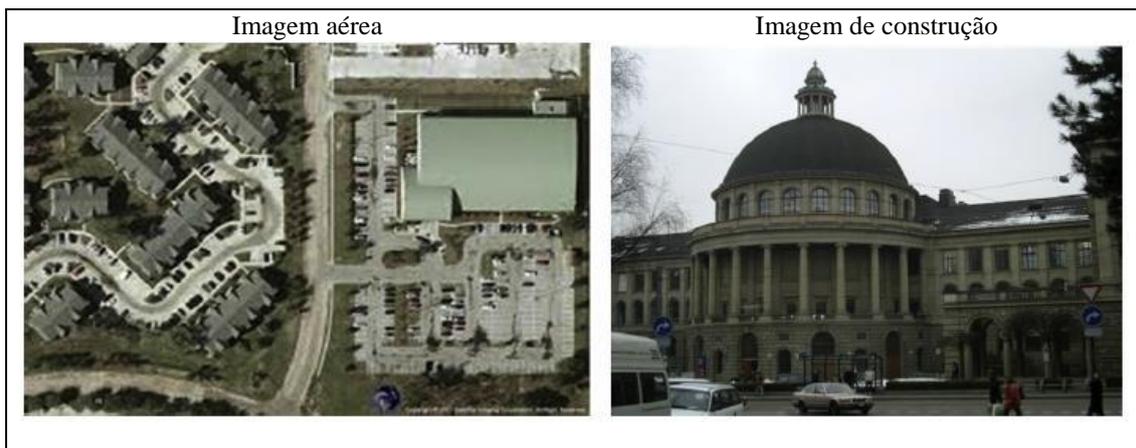
#### 3.1 Considerações Iniciais

O reconhecimento de edifícios em ambientes urbanos objetiva diferenciar edifícios em um conjunto de imagens de larga escala (LI et al., 2014). Ele permite que robôs se localizem com facilidade e precisão em cenas ao ar livre, além de poder ser aplicado na monitorização de tráfego, onde os veículos em movimento são normalmente detectados; permite que os projetistas de construção ou desenvolvedores pesquisem, eficientemente, em bancos de dados os modelos de construções mais almejavéis ao projetar um novo prédio; também desempenha um papel fundamental na reconstrução de cidades em 3D para o planejamento urbano ou realidade aumentada em videogames.

Em comparação com as tarefas gerais de reconhecimento de objetos, tanto em imagens como em vídeos, o reconhecimento de edifícios urbanos é mais complexo porque a maioria das imagens de construção contém objetos - paredes, portas e janelas - cenas naturais, e por muitas vezes exibem repetições e superfícies planares. Além disso, as imagens tiradas do mesmo edifício podem ser tomadas de diferentes pontos de vista e condições de iluminação, sofrer de interrupções parciais de árvores, veículos em movimento ou outros edifícios.

As imagens dos edifícios podem ser separadas em dois tipos: (i) imagens aéreas e (ii) imagens de construção. Estes dois tipos de são essencialmente diferentes, como mostrado na Figura 3.1.

Figura 3.1 – Exemplos de uma imagem aérea e uma imagem de construção



Fonte: LI et al., 2014.

Portanto, uma técnica de reconhecimento de edifícios ideal deve ser suficientemente eficiente para aplicações em tempo real em bancos de dados em larga escala. O que leva à investigação de seguintes aspectos: (1) modelos visuais que podem descrever com precisão edifícios e ser sensíveis à pequenas alterações na imagem e (2) rápida compressão para melhorar a eficiência de um algoritmo, enquanto diminui o espaço de armazenamento de dados e complexidade computacional.

O reconhecimento de edifícios consiste em três partes: (i) representação das características; (ii) correspondência das características; e (iii) classificação (LI et al., 2014). A representação das características descreve diferentes objetos em uma imagem, extraíndo propriedades globais, que são aquelas removidas de todos os pixels em uma imagem inteira, tais como, cor, forma, textura etc.; e locais, que se refere aos padrões de imagem, por exemplo, uma região de imagem de interesse ou um objeto específico, que são diferentes das suas vizinhanças. Pode, também, descrever informação local de uma imagem e tendem a ser invariantes a alterações, por exemplo, paredes, portas, janelas, etc.

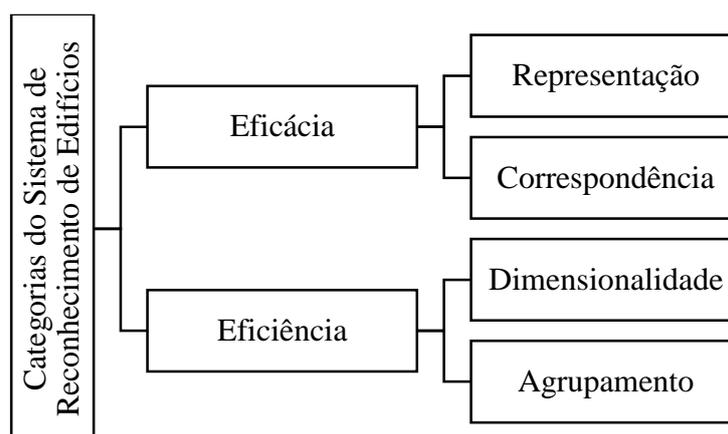
Após essa representação a correspondência de características é conduzida para encontrar semelhanças entre as imagens, isto é, a imagem de consulta e uma imagem de referência no banco de dados por meio de métricas de distância, por exemplo, distância euclidiana. Finalmente, a classificação é conduzida para determinar a melhor correspondência, onde os classificadores em relação aos modelos estatísticos combinam as saídas dos vetores de características globais ou locais de aparência para maximizar a qualidade da saída em um conjunto de treinamento (LI et al., 2014).

O desempenho do reconhecimento é geralmente avaliado por precisão e recordação, onde a precisão é definida como a porcentagem das imagens relevantes nas

imagens recuperadas e a recuperação é definida como a porcentagem das imagens relevantes em todas as imagens positivas no banco de dados.

Li et al. (2014), em sua pesquisa classifica os sistemas de reconhecimento de edifícios em duas categorias: (i) abordagens de **eficácia** que se concentram principalmente na melhoria do desempenho do sistema de reconhecimento e (ii) métodos de **eficiência** que aceleram o sistema de reconhecimento. As abordagens de eficácia podem ser ainda categorizadas em dois grupos diferentes: (i) recurso de algoritmos baseados em **representação** e (ii) ampla linha de base de métodos de **correspondência**. Já os métodos de eficiência são divididos em: (i) métodos baseados na redução da **dimensionalidade** e (ii) algoritmos baseados em **agrupamento**. Esquema ilustrado no fluxograma da Figura 3.2.

Figura 3.2 – Fluxograma da classificação dos sistemas de reconhecimento de edifícios



Fonte: Adaptado de Li et al. (2014).

Com isso, conforme a classificação de Li et al. (2014), nesta pesquisa, o método proposto se enquadra na categoria de métodos de eficiência, tanto baseado na redução da dimensionalidade, quanto em agrupamento (clusters), que buscam descobrir as relações entre diferentes estruturas de imagem, visto que, a proposta dessa técnica é o processamento eficiente.

### 3.2 O banco de dados e sua complexidade: Sheffield Building Image Dataset

Conforme Li et al. (2014), o banco de dados *Sheffield Building Image Dataset* (SBID) torna a tarefa de reconhecimento de edifícios mais desafiadora, pois combina diferentes condições de iluminação e pontos de vista, conforme ilustrado na Figura 3.3.

Figura 3.3 – Exemplo de imagens do conjunto de dados SBID. Essas três linhas mostram imagens de amostra das categorias 1, 10 e 38, respectivamente.



Fonte: LI et al., 2014.

As imagens do SBID possuem, rotação, escalonamento, diferentes condições de iluminação, alterações de pontos de vista, oclusões e vibração. É composta por 3192 imagens tiradas de 40 edifícios e distribuídas por classes. Sendo que em cada classe a quantidade de imagens é diferente.

As imagens incluem igrejas e uma variedade de edifícios modernos, como salas de exposições e prédios de escritórios. Essas imagens foram tiradas de prédios da Universidade de Sheffield e do centro da cidade de Sheffield em épocas diferentes e em dias separados do ano de 2008, onde as épocas diferentes cobrem o começo da manhã, o meio do dia, à parte da tarde e o começo da noite.

As imagens fixas para cada edifício foram tiradas de diferentes pontos de vista, variando de três a nove vistas. Esse banco de dados torna a tarefa de reconhecimento de edifícios desafiadora, pois combina diferentes condições de iluminação e pontos de vista (LI et al., 2014). O tamanho das imagens é fixado em 160x120 pixels, a fim de garantir eficiência computacional e requisitos de memória baixa.

### 3.3 Descrição geral do método proposto

O método proposto compreende o processamento das imagens dos edifícios nas escalas de cinza, vermelha, verde, azul e RGB utilizando para reconhecimento a QV, sendo que o projeto dos codebooks do quantizador vetorial é realizado pelo algoritmo tradicional LBG. O reconhecimento por meio da técnica de Deep Learning trabalha com as imagens em RGB.

Os resultados foram analisados a partir das variáveis: centroides e escala de cores, para a QV, e porcentual de treino, para as duas técnicas. Foi extraído informações nas primeiras 30 imagens por classe, para cada uma das 40 classes do banco de dados.

Estipulou-se a análise com as primeiras imagens em cada classe pois a distribuição dessas imagens encontra-se distribuídas de forma aleatória em cada classe. Estabeleceu-se a análise para 30 imagens por classe, porque essas classes têm diferentes quantidades de imagens, sendo uma delas com somente 34 imagens. Apesar de que no site que disponibiliza as imagens argumenta que o conjunto de dados consiste em mais de 3.000 imagens de baixa resolução de quarenta prédios diferentes - geralmente entre 70 e 120 imagens por prédio.

As imagens em escala cinza têm extensão .png e as em escala vermelha, verde, azuis e RGB extensão .jpeg. As imagens dentro das classes estão dispostas aleatoriamente, não seguem uma sequência da direita para a esquerda e nem do claro para o escuro, motivo pelo qual não houve necessidade de escolha aleatória das imagens para o processamento.

As imagens utilizadas para construir os centroides não são as mesmas usadas para reconhecimentos, no entanto fazem parte da mesma classe. Cada classe possui imagens de um mesmo prédio, variando a angulação e a iluminação.

Em cada etapa o treino foi executado, separadamente, por padronização, com 40%, 50% e 60% das primeiras imagens das classes, e o restante dessas imagens foram utilizadas para o reconhecimento, sendo que essas não são as mesmas trabalhadas no treino. A quantidade de imagens utilizada para o treino e teste (verificação de reconhecimento) está apresentado na Tabela 3.1.

Tabela 3.1 – Quantidade de Imagens Utilizadas para o Treino e para a Análise de Reconhecimento

Quantidade de imagens por classe	Treino	Quantidade total de imagens utilizadas para treino	Quantidade de Treino por Classe	Quantidade total de imagens utilizadas para teste	Quantidade de Teste por Classe	Total de imagens processadas
30	40%	480	12	720	18	1.200
	50%	600	15	600	15	
	60%	720	18	480	12	

Para a estatística descritiva, os resultados foram analisados por meio de frequência porcentual. Para a estatística inferencial foi feita análise de variância (ANOVA), com teste post-hoc de Tukey, que é um teste paramétrico que testa a hipótese e auxilia a avaliar a importância de um ou mais fatores, comparando as médias das variáveis nos diferentes grupos.

A ANOVA ou análise de variância é um método estatístico que informa se existe, ou não, diferença discrepante, significativa, entre os grupos tratados, isto é, uma análise de variância (ANOVA) rejeita ou não a hipótese de igualdade das médias de todos os grupos tratados, mas não determina quais grupos têm médias estatisticamente diferentes.

Após a análise de variância aplica-se o teste de Tukey para comparar as médias duas a duas (*pairwise comparison*). O teste de Tukey é um teste conhecido como *a posteriori* ou *post-hoc*, pois faz comparações não planejadas com antecedência (*unplanned comparisons*), pois são selecionadas após conhecer as médias amostrais.

Aplica-se o teste Tukey para manter o nível de significância para o experimento (*experimentwise Type I error rate* ou *familywise Type I error rate*), que é a probabilidade de cometer um erro Tipo I, quando são feitas as checagens entre os pares de médias do grupo de médias. A probabilidade de cometer um erro do tipo I num teste de hipóteses é denominada significância do teste, e é representada pela letra grega  $\alpha$ .

O erro do tipo I consiste em rejeitar a hipótese nula quando ela for verdadeira. Um  $\alpha$  de 0,05 indica que se aceita uma chance de 5% de se estar errado ao rejeitar a hipótese nula. Para reduzir este risco, usa-se um valor inferior para  $\alpha$ . A hipótese nula é a hipótese que irá ser testada.

Realiza-se o teste de hipótese fixando-se o nível de significância. Adotou-se para esse trabalho a letra p em substituição à letra  $\alpha$ . O nível de significância utilizado nesse trabalho é de 5%, isto é, considerou-se como significantes os resultados que apresentaram  $p \leq 0,05$ .

Adotou-se como hipótese nula do trabalho que não existe diferença significativa entre os dois métodos de reconhecimento de imagens analisados. Sendo que para a QV adotou-se o processo mais significativo após as análises comparativas entre todas as variáveis e esse resultado melhor encontrado foi comparado inferencialmente com o método de DL.

Os dados foram inseridos e processados por meio de banco de dados do programa SPSS® 20.0 e do programa Excel®, analisados por meio de estatística descritiva e inferencial.

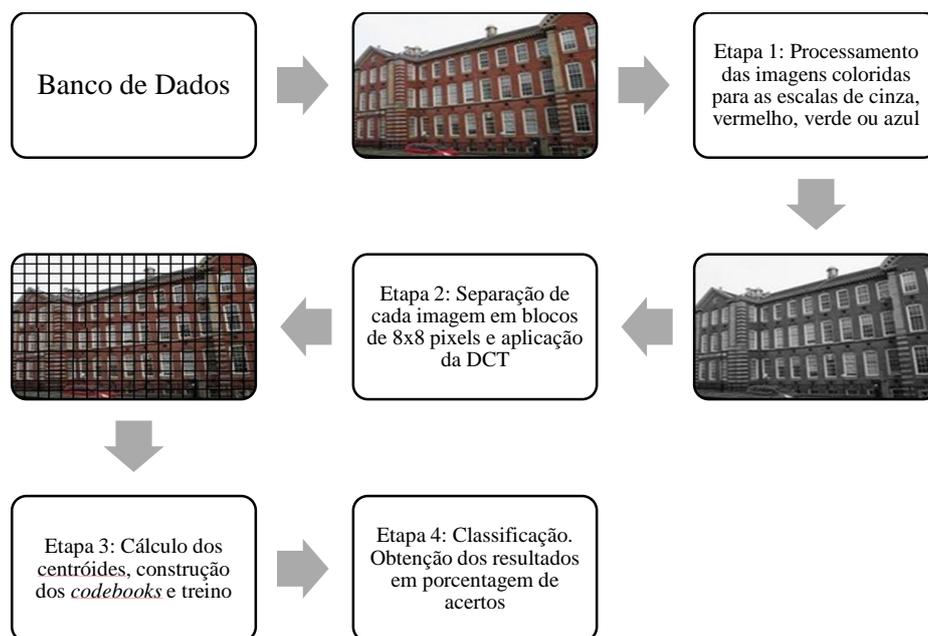
### **3.3.1 Quantização Vetorial - QV**

Todo o procedimento foi processado e analisado para 16, 32, 64, 128 e 256 centroides. Os resultados encontrados são apresentados em tabelas.

No processo de trabalho com imagens em escala cinza, vermelha, verde e azul tem-se um procedimento a mais do que na escala RGB, pois as imagens entram em RGB e são processadas as mudanças de escala conforme observa-se na Figura 3.4 que mostra as etapas do procedimento de reconhecimento.

O método consiste, basicamente, de quatro etapas (Figura 3.4). A primeira versa no processamento de todas as imagens coloridas editando-as para as escalas de cinza, vermelho, verde ou azul. O algoritmo já se inicia usando as imagens em suas respectivas escalas.

Figura 3.4 – Diagrama em blocos do processo metodológico para imagens em escala de cinza, vermelha, verde e azul



Fonte: Autor.

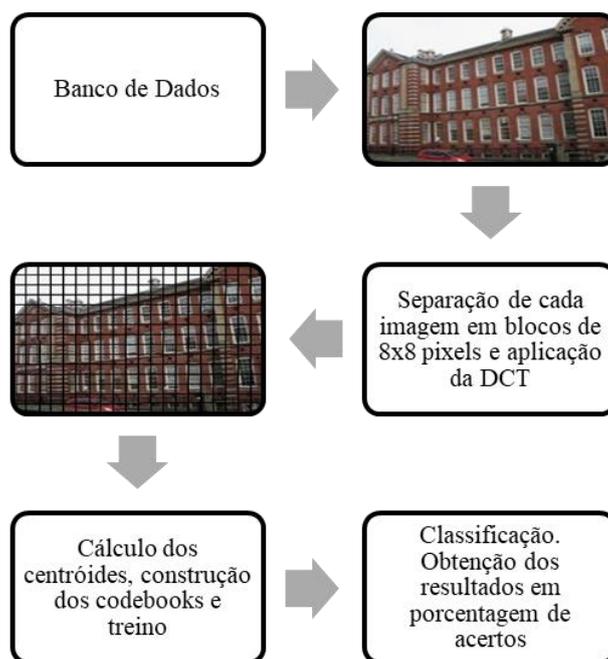
O algoritmo do projeto é iterativo e requer um codebook inicial, o qual é obtido pelo método de divisão, proposto na segunda das etapas que realiza a separação de cada imagem em blocos de 8x8 pixels (Figura 3.6). Cada bloco será um vetor de 64 ordenadas, pois é considerada somente uma escala de cor (Equação 3.1).

$$X_K = \{X_m; m = 1,2, \dots, 64\}; K = 1,2, \dots, 300 \quad 3.1$$

As imagens a princípio são RGB, por isso usa-se funções específicas do programa Matlab para convertê-las para as escalas de cor cinza, vermelha, verde ou azul.

Para imagens em RGB ocorre o processamento a partir da imagem original, as etapas do processo são apresentadas na Figura 3.5

Figura 3.5 – Diagrama em blocos do processo metodológico para imagens em RGB



Fonte: Autor.

Para as imagens em RGB cada bloco é um vetor com dimensão de 64 (Equação 3.2), em que  $X_m$  corresponde às ordenadas da escala da cor vermelha,  $Y_m$  às ordenadas da escala da cor verde e  $Z_m$  às ordenadas da escala da cor azul.

$$X_K = \{X_m, Y_m, Z_m; m = 1, 2, \dots, 64\}; K = 1, 2, \dots, 300 \quad 3.2$$

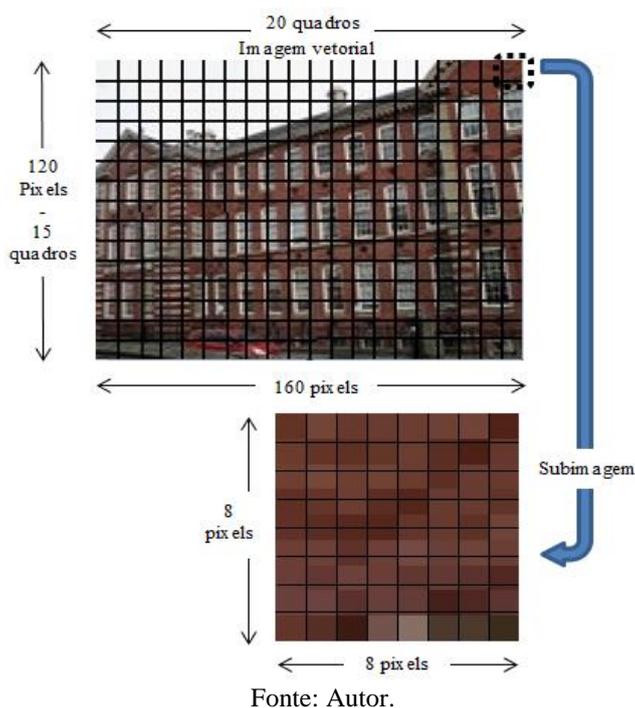
No processamento da QV, o método de reconhecimento divide a imagem em blocos quadrados de 8 x 8 pixels, totalizando 300 blocos por imagem, 20 blocos na horizontal e 15 blocos na vertical (Figura 3.6). O que torna, de acordo com Duda et al. (2001), a quantização um método de compressão assimétrica.

Em seguida aplica-se a Transformada Discreta do Cosseno (DCT) para a conversão dos dados de amplitude espacial em coeficientes de frequências espaciais para cada bloco 8x8 pixels.

Visto que a transformação da imagem como um todo, com uma única multiplicação por matriz, exigiria uma matriz imensa, o que diminuiria a eficácia. Pois, matematicamente, a DCT corresponde a uma multiplicação por uma matriz, que tem a propriedade de converter dados de amplitude espacial, que são os valores dos pixels, em coeficientes representando frequências espaciais.

A DCT tem por função nos sistemas de compressão de imagens, identificar a presença de redundância espacial, isto é, semelhança entre um pixel e os de sua vizinhança, em cada imagem, sendo um processo sem perda (*lossless*) e reversível. A possibilidade de desprezar os coeficientes menos significativos sem muita perda da qualidade da imagem é justamente a razão principal de se usar a DCT na codificação de imagens, pois permite economia de bits.

Figura 3.6 – Exemplo de separação de cada imagem em blocos de 8x8 pixels



Na sequência executa-se o algoritmo LBG dando início aos testes no processo de reconhecimento e a classificação da imagem em sua respectiva classe. O resultado final consiste na obtenção do percentual de acertos do reconhecimento (Equação 3.3).

$$\text{Porcentagem} = \frac{Qa}{Qc \times (Qic - Qit)} \times 100 \quad 3.3$$

Em que  $Qa$  equivale a quantidade de acerto,  $Qc$  à quantidade de classes,  $Qic$  à quantidade de imagens em cada classe e  $Qit$  à quantidade de imagens para treino.

### 3.3.2 Deep Learning - DL

Utilizou-se a estratégia “max-pooling” que consiste em verificar se um determinado recurso é encontrado em qualquer lugar de uma região da imagem e em seguida, elimina-se a informação posicional exata. Para que a rede CNN tenha uma precisão suficiente, leva-se em consideração a dimensão do conjunto de dados amostrados.

Na construção da rede, utilizou-se a plataforma Google Colab, devido a esta suportar hardware habilitado para GPU, o que auxilia no desempenho durante o treinamento da rede. Na implementação, seguiu-se a seguinte ordem:

1 - Download e carregamento de dados: os dados foram hospedados no Google Drive, e por meio do comando “gdown”, foi possível realizar o download e a importação dos mesmos.

2 - Configuração da rede: introduz-se um conjunto de configuração compreendendo largura (W) e altura (H) das imagens (160 x 120), dimensão da matriz dos filtros (3 x 3), dimensão do “downscaling”, o “max pooling” (2 x 2) e o caminho relativo para os ficheiros.

3 - Leitura das imagens: procede-se à leitura das imagens, utilizando uma técnica de gerar novas imagens a partir das existentes (aplicando transformações aleatórias: distorção, rotação, zoom, etc.), para compensar a quantidade relativamente reduzida de imagens utilizadas do banco de dados.

Ao final obtém-se a lista de classes a caracterizar. A fim de verificar a precisão da rede de acordo com a variação na quantidade de dados de treino, utilizou-se o comando “validation\_split”, que assume os seguintes valores:

- 0,6: para utilização de 40% dos dados para treino;
- 0,5: para utilização de 50% dos dados para treino; e
- 0,4: para utilização de 60% dos dados para treino.

4 - Criação da CNN: criou-se a CNN composta por duas camadas de convolução, onde aplicou-se 32 filtros na primeira e 64 filtros na segunda camada. Ao

término da varredura dos filtros, aplicou-se a função de ativação “ReLU” e em seguida o “max pooling”.

Adotou-se a função de ativação ReLU, devido ao fato de a mesma não ativar todos os neurônios ao mesmo tempo, o que significa que quando a entrada for negativa, ela será convertida em zero e o neurônio não será ativado, conseqüentemente, apenas alguns neurônios são ativados, o que torna a rede mais eficiente e compreensível [20]. Após, efetuou-se o “flatten”, que é a camada utilizada na divisão das duas partes da CNN para a extração de características da rede neural.

Gerou-se o vetor de entrada para a rede neural, dispondo de uma camada oculta de 128 neurônios ativados novamente por um “ReLU”, e que utiliza o “Adam Optimizer” para aplicação do “stochastic gradient descente”, e o “categorical cross entropy”, como função de erro. A camada de saída é constituída por 40 neurônios, cada um correspondente a uma das 40 classificações possíveis, ativada por um “softmax” (Figura 3.7).

Figura 3.7 – Sumário da rede criada no projeto, utilizando-se um validation\_split = 0,4

```

Model: "sequential_9"
-----
Layer (type)                Output Shape              Param #
-----
conv2d_18 (Conv2D)          (None, 158, 118, 32)     896
-----
max_pooling2d_18 (MaxPooling (None, 79, 59, 32)     0
-----
conv2d_19 (Conv2D)          (None, 77, 57, 64)      18496
-----
max_pooling2d_19 (MaxPooling (None, 38, 28, 64)     0
-----
flatten_9 (Flatten)         (None, 68096)            0
-----
dense_18 (Dense)            (None, 128)              8716416
-----
dense_19 (Dense)            (None, 40)                5160
-----
Total params: 8,740,968
Trainable params: 8,740,968
Non-trainable params: 0

```

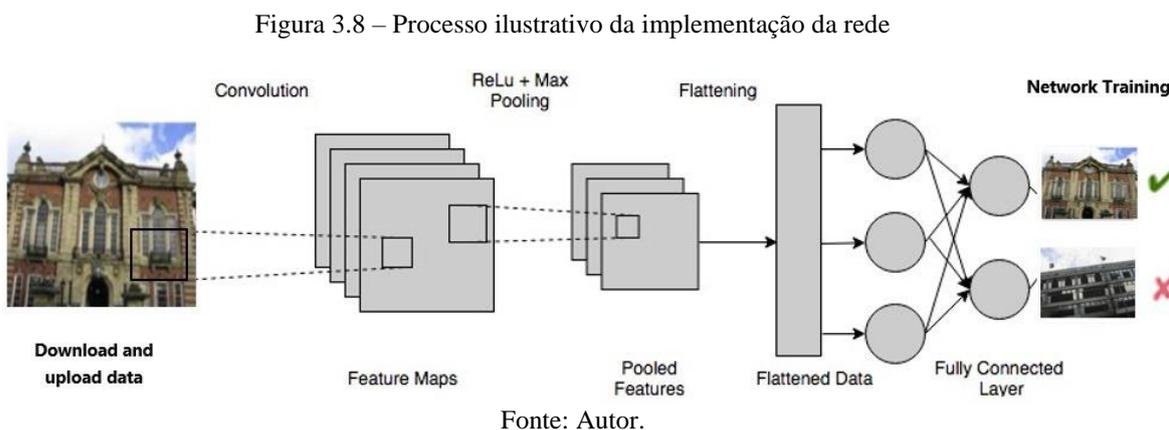
Fonte: Autor.

Os parâmetros de rede adotados são provenientes, de forma empírica, a partir de testes realizados na rede (tentativa e erro), pois devido à utilização de um banco de dados desbalanceado, a previsão desses valores se torna mais complexa. Após a determinação

dos parâmetros satisfatórios, deu-se início aos processos de treino e teste, podendo futuramente alterar-se tais parâmetros a fim de melhorar ainda mais os resultados obtidos.

5 - Treino da rede: por fim executa-se o processo de treino da rede, sobre o “training set” e “test set”, variando o número de épocas e passos por época, a fim de obter a melhor acurácia do modelo. Para facilitar, grava-se a rede treinada para utilizá-la novamente, já treinada.

A Figura 3.8, a seguir, traz uma representação ilustrativa dos 5 passos apresentados para o processo de implementação da rede neural utilizada: (1) Download e carregamento de dados; (2) Configuração da rede; (3) Leitura das imagens; (4) Criação da CNN e (5) Treino da rede.



### 3.4 – Considerações finais do capítulo

Esse capítulo apresentou o modelo proposto nesse trabalho e para o capítulo 4, a seguir, far-se-á as análises dos resultados e discussões para imagens nas escalas cinza, vermelha, verde, azul e RGB utilizando Quantização Vetorial e análise com os resultados obtidos pela técnica Deep Learning. Por fim comparar-se-á os resultados e apresentará as discussões a respeito.

## CAPÍTULO 4

### ANÁLISE DOS RESULTADOS

Esse capítulo está dividido em 7 seções que apresentam as tabelas dos resultados dos percentuais com a análise inferencial dos percentuais acertos no reconhecimento obtidos utilizando a técnica de Quantização Vetorial nas escalas cinza, vermelho, verde, azul e RGB, e as taxas de reconhecimento utilizando a técnica de DL com as imagens em RGB e as discussões alusivas a essas.

Na Seção 4.2 será feito a análise dos resultados do reconhecimento das imagens em escalas de cores cinza, vermelha, verde e azul. Na Seção 4.3 será feito a análise comparativa entre os resultados nas escalas de cinza, vermelha, verde e azul. Na Seção 4.4 a análise será com os resultados do reconhecimento na escala RGB. Na seção 4.5 será feito a análise comparativa dos resultados nas escalas de cinza, vermelho, verde, azul e RGB. Na seção 4.6 será feito a análise dos resultados de pela técnica de DL. Na seção 4.7 será feito a análise comparativa dos resultados de QV 32 centroides com DL-RGB. E por último, na seção 4.8, faz-se a comparação dos resultados finais do trabalho com os resultados encontrados na literatura

#### 4.1 Considerações Iniciais

As apresentações dos resultados terão foco na estatística inferencial das frequências percentuais de acerto para reconhecimento utilizando análise de variância (ANOVA), com teste paramétrico post- hoc de Tukey, considerando significantes os resultados que apresentaram  $p \leq 0,05$ .

Os gráficos utilizados (Gráfico 4.1, Gráfico 4.2, Gráfico 4.3, Gráfico 4.4 e Gráfico 4.5), que antecedem as tabelas 4.5, 4.6, 4.8, 4.9 e 4.11, respectivamente, são diagramas de caixas (*boxplot*). Essa ferramenta estatística gráfica permitem a visualização das distribuições de quartis, máximos, mínimos e valores discrepantes (*outliers*) dos dados estudados.

O diagrama de caixa fornece uma perspectiva da distribuição dos dados e um arranjo gráfico comparativo entre os grupos analisados, pois proporciona uma análise visual da posição e da dispersão do conjunto de dados estudado.

A análise do intervalo interquartil (diferença entre o terceiro e o primeiro quartil – tamanho da caixa), como ferramenta estatística observada nos diagramas de caixas, permite ressaltar a dispersão do grupo. O intervalo interquartil, utilizado para análise, é uma estatística para medir a variabilidade, sendo que não sofre influência de *outliers*.

Em algumas análises foi destacado a amplitude dos dados, que é obtida pela diferença do valor máximo pelo valor mínimo dos percentuais de reconhecimento encontrado. Observou-se, também, a existência de possíveis valores discrepantes (*outliers*).

## **4.2 Análise dos resultados por escala de cor**

Esta seção foi dividida em quatro subseções. Foram construídas cinco tabelas (Tabela 4.1, 4.2, 4.3, 4.4, e 4.5), uma em cada subseção, apresentando os resultados do reconhecimento de prédios, processados com QV na escala em cinza, vermelha, verde, azul e RGB, obtidos com 16, 32, 64, 128 e 256 centroides, e análise inferencial, relativa a esses resultados. Foram analisadas 40 classes, cada classe com 30 imagens, com os treinos de 40%, 50% e 60%.

Na Tabela 4.5 e feito uma análise comparativa entre todos os resultados com o objetivo que verificar em qual quantidade mínima de centroides pode-se estabelecer a análise com resultados significativos. E, por fim será apresentado o maior resultado percentual de reconhecimento com a quantidade de centroides e o percentual utilizado para treino.

### **4.2.1 Escala de Cinza**

Nessa subseção foi construída uma tabela (Tabela 4.1) apresentando os resultados do reconhecimento de prédios, processados com QV na escala de cinza e feito análise da estatística inferencial.

Tabela 4.1 – Resultados na escala em cinza

<b>Treino</b>			
<b>Centroides</b>	<b>40%</b>	<b>50%</b>	<b>60%</b>
<b>16</b>	54,72b	62,83b	65,21b
<b>32</b>	61,53ab	72,67ab	72,71ab
<b>64</b>	69,17ab	77,67ab	78,96ab
<b>128</b>	70,69a	79,67a	82,08a
<b>256</b>	74,44a	82,0a	84,17a

Fonte: Autor.

Na análise inferencial foi observado que não ocorreu diferença significativa entre os treinos, isto é, taxas de reconhecimento de imagens com relação aos treinos são estatisticamente iguais. No entanto, entre os centroides houve essa diferença conforme observado na Tabela 4.1.

Encontrou-se diferença significativa entre as taxas de reconhecimento processadas com 16 centroides e as taxas de 128 e 256 centroides. As imagens processadas a partir de 32 centroides apresentam taxa de reconhecimento significativa, sendo assim, pode-se utilizar para processamento de reconhecimento de imagens, na escala em cinza, 32 centroides, isto é, taxas de reconhecimento de imagens com 32, 64, 128 e 256 são estatisticamente iguais.

#### 4.2.2 Escala de vermelha

Nessa subseção foi construída uma tabela (Tabela 4.2) apresentando os resultados do reconhecimento de prédios, processados com QV na escala de vermelho e feito análise da estatística inferencial.

Tabela 4.2 – Resultados na escala em vermelho

<b>Treino</b>			
<b>Centroides</b>	<b>40%</b>	<b>50%</b>	<b>60%</b>
<b>16</b>	55,83b	58,17b	62,5b
<b>32</b>	65,14ab	71,5ab	71,04ab
<b>64</b>	68,47a	77a	77,29a
<b>128</b>	72,08a	77,67a	80,63a
<b>256</b>	73,19a	81,33a	82,71a

Fonte: Autor.

Na análise inferencial foi observado que não ocorreu diferença significativa entre os treinos, isto é, taxas de reconhecimento de imagens com relação aos treinos são estatisticamente iguais. No entanto, entre os centroides houve essa diferença conforme observado na Tabela 4.2.

Encontrou-se diferença significativa entre as taxas de reconhecimento processadas com 16 centroides e as taxas de 64, 128 e 256 centroides. As imagens processadas a partir de 32 centroides apresentam taxa de reconhecimento significativa, sendo assim, pode-se utilizar para processamento de reconhecimento de imagens, na escala em vermelho, 32 centroides, isto é, taxas de reconhecimento de imagens com 32, 64, 128 e 256 são estatisticamente iguais.

### 4.2.3 Escala de verde

Nessa subseção foi construída uma tabela (Tabela 4.3) apresentando os resultados do reconhecimento de prédios, processados com QV na escala de verde e feito análise da estatística inferencial.

Tabela 4.3 – Resultados na escala em verde

<b>Treino</b>	<b>40%</b>	<b>50%</b>	<b>60%</b>
<b>Centroides</b>			
<b>16</b>	68,47c	70,83c	69,79c
<b>32</b>	70,14bc	75,17bc	77,71bc
<b>64</b>	73,89abc	81,5abc	78,13abc
<b>128</b>	77,64ab	82,83ab	84,58ab
<b>256</b>	79,72a	84,67a	85,0a

Fonte: Autor.

Na análise inferencial foi observado que não ocorreu diferença significativa entre os treinos, isto é, taxas de reconhecimento de imagens com relação aos treinos são estatisticamente iguais. No entanto, entre os centroides houve essa diferença conforme observado na Tabela 4.3.

Encontrou-se diferença significativa entre às taxas de reconhecimento processadas com 16 centroides em relação às taxas 128 e 256 centroides. As imagens processadas a partir de 32 centroides apresentaram diferença significativa somente em relação às taxas de 256 centroides e 16 centroides. Para 32, 64 e 128 centroides não

ocorreu diferença significativa. E, para 128 e 256 centroides, também, não ocorreu diferença significativa.

#### 4.2.4 Escala de azul

Nessa subseção foi construída uma tabela (Tabela 4.4) apresentando os resultados do reconhecimento de prédios, processados com QV na escala de azul e feito análise da estatística inferencial.

Tabela 4.4 – Resultados na escala em azul

<b>Treino</b>	<b>40%</b>	<b>50%</b>	<b>60%</b>
<b>Centroides</b>			
<b>16</b>	68,19a	75,5a	79,38a
<b>32</b>	75,28a	80,17a	83,54a
<b>64</b>	78,61a	84,17a	88,13a
<b>128</b>	79,31a	84,83a	90,21a
<b>256</b>	81,11a	86,17a	89,79a

Fonte: Autor.

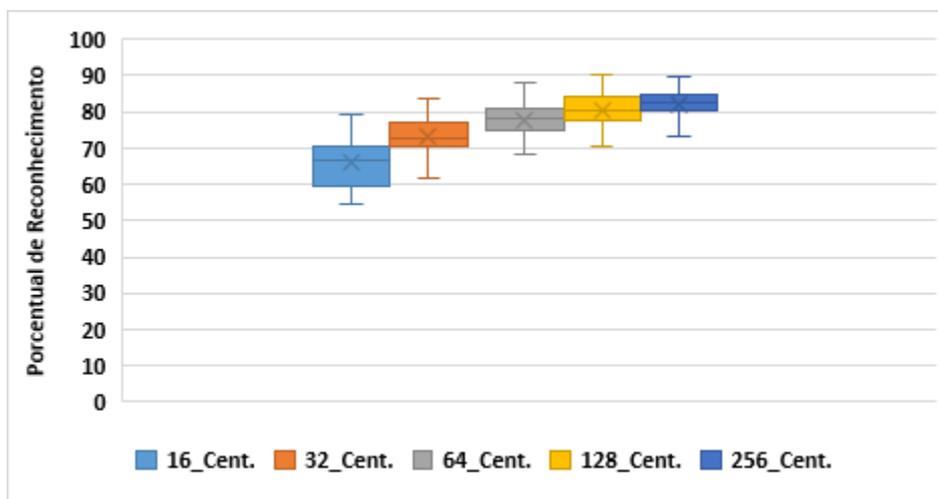
Na análise inferencial foi observado que não ocorreu diferença significativa entre os treinos e entre os centroides. Isto é, taxas de reconhecimento de imagens com 16, 32, 64, 128 e 256 centroides independente do treino são estatisticamente iguais.

#### 4.3 Análise Comparativa entre os resultados nas escalas de cinza, vermelha, verde e azul

Nessa subseção iremos apresentar uma análise descritiva e inferencial entre os resultados encontrados quando utilizado QV para diferentes quantidades de centroides, de escalas de cor e treino.

### 4.3.1 – Análise com relação a quantidade de centroides

Gráfico 4.1 – Diagrama de caixa e caixa estreita comparando a Variabilidade com relação aos centroides



Fonte: Autor.

Observa-se no Gráfico 4.1 (diagrama de caixa e caixa estreita) que existe uma tendência que quanto maior a quantidade de centroides maior a taxa de reconhecimento e menor a variabilidade (intervalo interquartílico) de onde se tem que o grupo de 256 centroides é menos dispersão em relação aos demais grupos de centroides. Observa-se a não existência de *outliers*.

No entanto, para verificar se é estatisticamente significativo foi feita análise de variância (ANOVA), com teste post- hoc de Tukey. Considerou-se como significantes os resultados que apresentaram  $p \leq 0,05$  (Tabela 4.5).

Tabela 4.5 – Análise Comparativa entre os resultados nas escalas de cinza, vermelha, verde e azul com relação a quantidade de centroides

	Treino	Centroides				
		16_Cent.	32_Cent.	64_Cent.	128_Cent.	256_Cent.
Cinza	40 %	54,72c	61,53b	69,17ab	70,69a	74,44a
	50 %	62,83c	72,67b	77,67ab	79,67a	82a
	60 %	65,21c	72,71b	78,96ab	82,08a	84,17a
Vermelha	40 %	55,83c	65,14b	68,47ab	72,08a	73,19a
	50 %	58,17c	71,5b	77ab	77,67a	81,33a
	60 %	62,5c	71,04b	77,29ab	80,63a	82,71a
Verde	40 %	68,47c	70,14b	73,89ab	77,64a	79,72a
	50 %	70,83c	75,17b	81,5ab	82,83a	84,67a
	60 %	69,79c	77,71b	78,13ab	84,58a	85a
Azul	40 %	68,19c	75,28b	78,61ab	79,31a	81,11a
	50 %	75,5c	80,17b	84,17ab	84,83a	86,17a
	60 %	79,38c	83,54b	88,13ab	90,21a	89,79a

Fonte: Autor.

Na análise inferencial foi observado que não ocorreu diferença significativa entre os treinos, isto é, taxas de reconhecimento de imagens com relação aos treinos são estatisticamente iguais.

Observa-se na análise inferencial apresentada na Tabela 4.5 que não foi encontrado taxa de reconhecimento significativa a partir da quantidade de 64 centroides, nas imagens processadas pela técnica de QV com as cores nas escalas de cinza, vermelha, verde e azul. Isto é, taxas com 64, 128 ou 256 centroides e 40%, 50% ou 60% de treino são estatisticamente iguais.

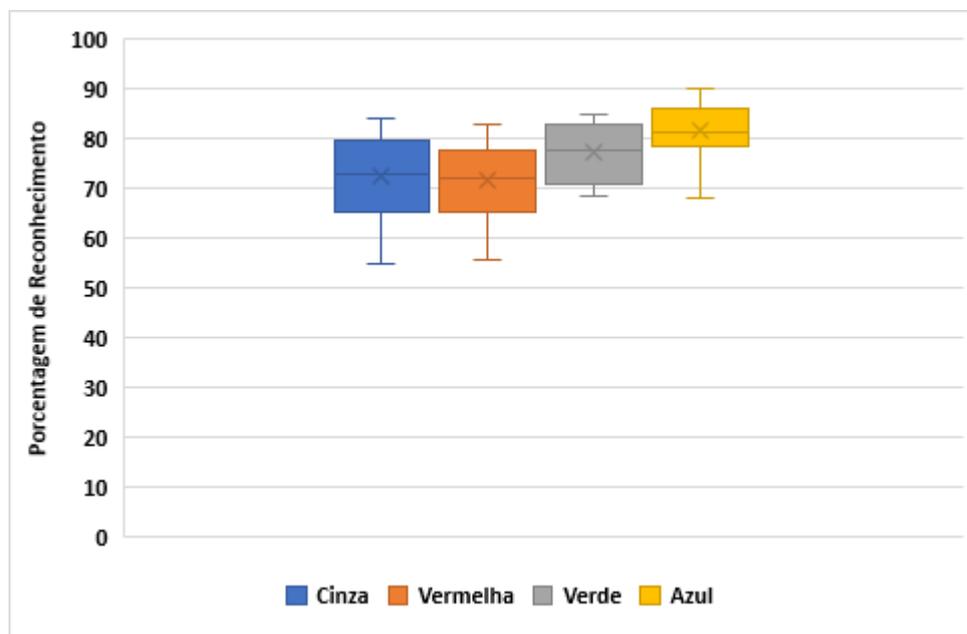
No entanto, o maior porcentual descritivo de reconhecimento entre as escalas de cores e os treinos, foi de 90,21% encontrado na escala de cor azul para 50% de treino e 256 centroides.

Independentemente do treino e da escala de cor, para os resultados com 16 centroides foi encontrado valor significativo em relação aos demais grupos de centroides. Isto é, o grupo de 16 centroides é estatisticamente diferente dos demais grupos de centroides. E, os valores para 32 e 64 centroides não obtivam significância entre si, os consideramos estatisticamente iguais entre si.

Apesar de observado que, para o reconhecimento das imagens trabalhadas, utilizando QV, conforme se aumenta a quantidade de centroides os valores percentuais das taxas de reconhecimento aumentam, obtém-se resultados estatisticamente melhores com 64, 128 e 256, sendo que esses grupos são considerados estatisticamente iguais.

### 4.3.2 – Análise com relação as escalas de cores cinza, vermelha, verde e azul

Gráfico 4.2 – Diagrama de caixa e caixa estreita comparando a Variabilidade das escalas de cores cinza, vermelha, verde e azul



Fonte: Autor.

Observa-se no Gráfico 4.2, que existe uma variabilidade entre as taxas de reconhecimento com relação as cores. Para a cor azul tem-se um valor mínimo, conforme observado no gráfico, bem abaixo do primeiro quartil. No entanto, a dispersão da cor azul é menor que as demais cores e o cinza tem a maior dispersão. A mediana e a média das cores azul e verde ficam próximas e possuem um valor aproximado de 80%. Enquanto que o cinza e o vermelho possuem medianas e médias próximas de 70%. Observa-se a não existência de *outliers*.

No entanto, para verificar se é estatisticamente significativo foi feito a análise de variância (ANOVA), com teste post-hoc de Tukey. Considerou-se como significantes os resultados que apresentaram  $p \leq 0,05$  (Tabela 4.6).

Tabela 4.6 – Análise Comparativa entre os resultados nas escalas de cinza, vermelha, verde e azul

Treino	Centroides	Cinza	Vermelha	Verde	Azul
40%	16	54,72b	55,83b	68,47 <b>ab</b>	68,19 <b>a</b>
	32	61,53b	65,14b	70,14 <b>ab</b>	75,28 <b>a</b>
	64	69,17b	68,47b	73,89 <b>ab</b>	78,61 <b>a</b>
	128	70,69b	72,08b	77,64 <b>ab</b>	79,31 <b>a</b>
	256	74,44b	73,19b	79,72 <b>ab</b>	81,11 <b>a</b>
50%	16	62,83b	58,17b	70,83 <b>ab</b>	75,5 <b>a</b>
	32	72,67b	71,5b	75,17 <b>ab</b>	80,17 <b>a</b>
	64	77,67b	77b	81,5 <b>ab</b>	84,17 <b>a</b>
	128	79,67b	77,67b	82,83 <b>ab</b>	84,83 <b>a</b>
	256	82b	81,33b	84,67 <b>ab</b>	86,17 <b>a</b>
60%	16	65,21b	62,5b	69,79 <b>ab</b>	79,38 <b>a</b>
	32	72,71b	71,04b	77,71 <b>ab</b>	83,54 <b>a</b>
	64	78,96b	77,29b	78,13 <b>ab</b>	88,13 <b>a</b>
	128	82,08b	80,63b	84,58 <b>ab</b>	90,21 <b>a</b>
	256	84,17b	82,71b	85 <b>ab</b>	89,79 <b>a</b>

Fonte: Autor.

Observa-se na análise inferencial apresentada na Tabela 4.6 que não foi encontrado taxa de reconhecimento significativa entre as cores cinza, vermelha e verde, isto é, estatisticamente são iguais. E, entre as cores verde e azul, também, não há significância, sendo que essas duas cores possuem taxas maiores e se enquadram no mesmo grupo estatístico.

Logo, para o reconhecimento das imagens trabalhadas, utilizando QV, obtém-se resultados estatisticamente melhores nas escalas de cores verde e azul em relação às escalas cinza e vermelha.

#### 4.4 Análise dos resultados em RGB

Nessa seção foi construída uma tabela (Tabela 4.7) apresentando os resultados do reconhecimento de prédios, processados com QV na escala em RGB, obtidos com 16, 32, 64, 128 e 256 centroides, e análise inferencial, relativa a esses resultados. Foram analisadas 40 classes, cada classe com 30 imagens, com os treinos de 40%, 50% e 60%.

Tabela 4.7 – Análise dos Resultados em RGB

<b>Treino</b>			
	<b>40%</b>	<b>50%</b>	<b>60%</b>
<b>Centroides</b>			
<b>16</b>	93,33b	89,17b	86,46b
<b>32</b>	98,19a	95,67a	93,96a
<b>64</b>	99,72a	98,83a	98,33a
<b>128</b>	100a	100a	100a
<b>256</b>	100a	100a	100a

Fonte: Autor.

Na análise inferencial foi observado que não ocorreu diferença significativa entre os treinos, isto é, taxas de reconhecimento de imagens com relação aos treinos são estatisticamente iguais. No entanto, entre os centroides encontrou-se diferença significativa entre as taxas de reconhecimento processadas com 16 centroides e as taxas dos demais centroides. As imagens processadas a partir de 32 centroides não apresentam taxa de reconhecimento significativa, sendo assim, pode-se utilizar para processamento de reconhecimento de imagens, na escala em RGB, 32 centroides, isto é, taxas de reconhecimento de imagens com 32, 64, 128 e 256 são estatisticamente iguais.

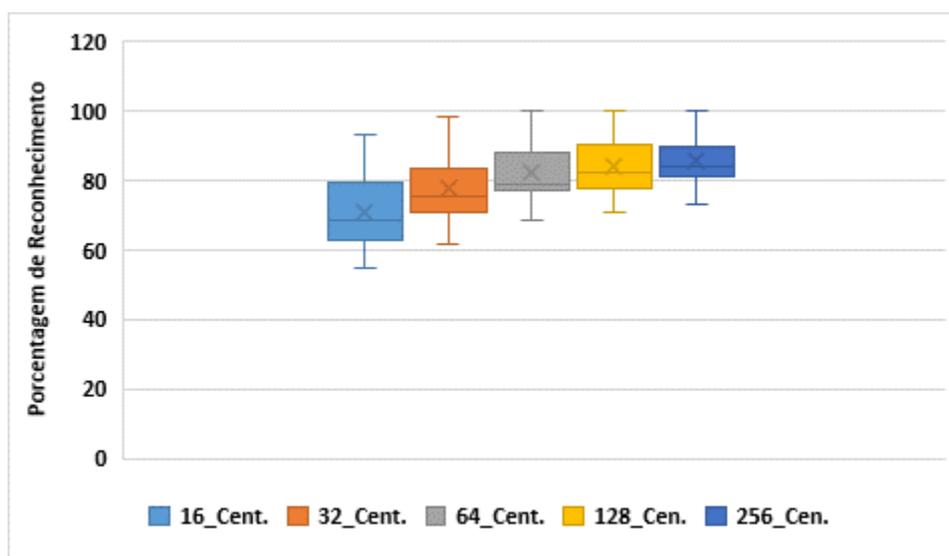
#### **4.5 Análise Comparativa dos resultados nas escalas de cinza, vermelho, verde, azul e RGB**

##### **4.5.1 – Análise comparativa em relação aos centroides**

Nessa seção foram construídas duas tabelas (Tabela 4.7 e 4.8) apresentando os dos resultados descritivos do reconhecimento das imagens propostas na pesquisa, processadas utilizando QV na escala de cores cinza, vermelha, verde, azul e RGB, obtidos com 16, 32, 64, 128 e 256 centroides, e análise inferencial, relativa a esses resultados. Foram analisadas 40 classes, cada classe com 30 imagens, com os treinos de 40%, 50% e 60%.

O Gráfico 4.3 apresenta a variabilidade entre as taxas de reconhecimento com relação aos centroides para as escalas de cores cinza, vermelha, verde, azul e RGB.

Gráfico 4.3 – Diagrama de caixa e caixa estreita comparando a Variabilidade dos centroides para as escalas de cores cinza, vermelha, verde, azul e RGB



Fonte: Autor.

No Gráfico 4.3 inseriu-se as taxas de reconhecimento em RGB em relação aos trabalhados no Gráfico 4.1.

A variabilidade para 256 centroides foi menor em relação às demais. A mediana e a média para 128 e 256 centroides ficaram acima de 80%. E, a mediana e a média para 16 e 32 centroides ficaram abaixo de 80%.

Ocorreu diferenças significativamente grandes na amplitude dos grupos analisados (valor máximo menos valor mínimo). Pode-se inferir que a diferença foi em virtude da inclusão dos valores percentuais de reconhecimento da cor RGB para a análise.

Observa-se, também, que as taxas de reconhecimentos se manteve próximas para 64, 128 e 256 centroides. Destaca-se que o reconhecimento para 16 centroides é o que possui a maior distribuição de amplitude e dispersão dos dados.

No entanto, para verificar a existência de fato dessa variabilidade entre as taxas de reconhecimento com relação aos centroides e feito análise de variância (ANOVA), com teste post- hoc de Tukey. Considerou-se como significantes os resultados que apresentaram  $p \leq 0,05$  (Tabela 4.8).

Tabela 4.8 – Análise Comparativa entre os resultados nas escalas de cinza, vermelha, verde, azul e RGB com relação a quantidade de centroides

	Treino	Centroides				
		16_Cent.	32_Cent.	64_Cent.	128_Cent.	256_Cent.
<b>Cinza</b>	40 %	54,72b	61,53ab	69,17a	70,69a	74,44a
	50 %	62,83b	72,67ab	77,67a	79,67a	82a
	60 %	65,21b	72,71ab	78,96a	82,08a	84,17a
<b>Vermelha</b>	40 %	55,83b	65,14ab	68,47a	72,08a	73,19a
	50 %	58,17b	71,5ab	77a	77,67a	81,33a
	60 %	62,5b	71,04ab	77,29a	80,63a	82,71a
<b>Verde</b>	40 %	68,47b	70,14ab	73,89a	77,64a	79,72a
	50 %	70,83b	75,17ab	81,5a	82,83a	84,67a
	60 %	69,79b	77,71ab	78,13a	84,58a	85a
<b>Azul</b>	40 %	68,19b	75,28ab	78,61a	79,31a	81,11a
	50 %	75,5b	80,17ab	84,17a	84,83a	86,17a
	60 %	79,38b	83,54ab	88,13a	90,21a	89,79a
<b>RGB</b>	40 %	93,33b	98,19ab	99,72a	100,0a	100,0a
	50 %	89,17b	95,67ab	98,83a	100,0a	100,0a
	60 %	86,46b	93,96	98,33a	100,0a	100,0a

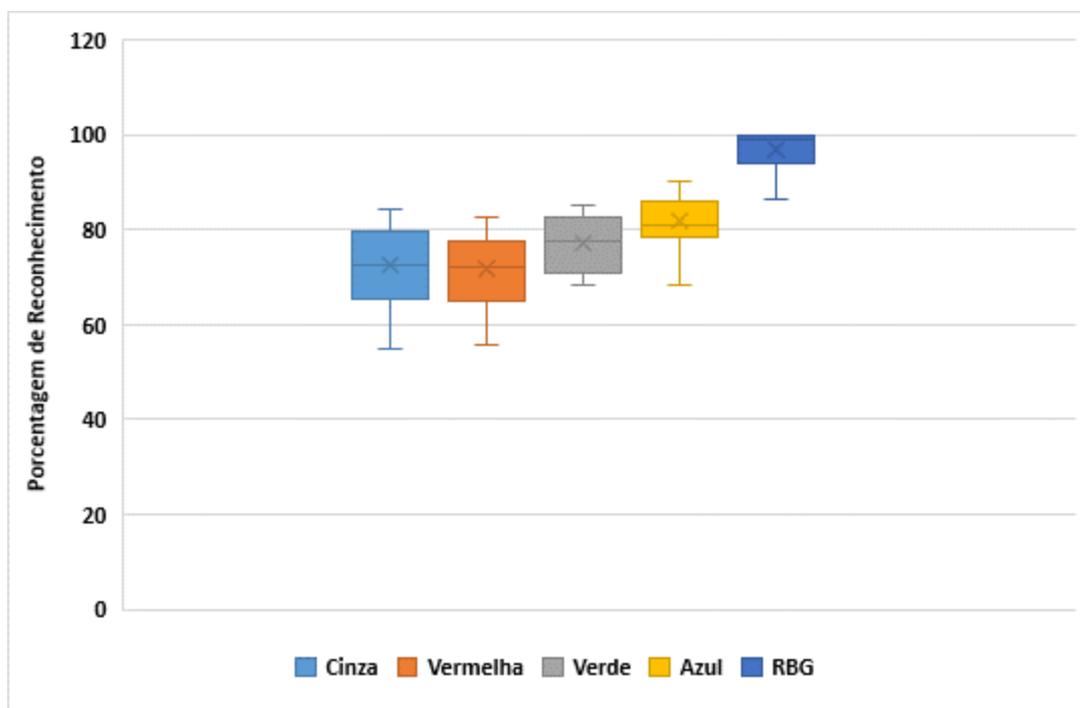
Fonte: Autor.

Apresenta-se que para a variável treino não houve significância quando utilizado o cálculo da ANOVA, pelo post-hoc de Tukey, com o nível de significância de  $p \leq 0,05$ .

Tem -se que quanto maior o número de centroides melhor a taxa de reconhecimento, sabendo disso. No entanto, conforme o teste de ANOVA com post-hoc de tukey tem-se que os resultados encontrados para 32 centroides não são significativos em relação à 64, 128 e 256 centroides, então para as escalas de cinza, vermelho, verde, azul e RGB, isto é, esses grupos são estatisticamente iguais.

#### 4.5.2 – Análise comparativa em relação às escalas de cores cinza, vermelha, verde, azul e RGB

Gráfico 4.4 – Diagrama de caixa e caixa estreita comparando a Variabilidade das escalas de cores cinza, vermelha, verde, azul e RGB



Fonte: Autor.

Observa-se no Gráfico 4.4, que para a análise da variabilidade entre as taxas de reconhecimento com relação as cores, a cor RGB se destaca com uma dispersão muito pequena, sendo que o valor máximo é de 100% de reconhecimento e a média fica logo abaixo de 100%. No entanto as cores cinza, vermelha e verde possuem médias e medianas abaixo de 80%. E a cor azul mantém mediana e média logo acima de 80% de reconhecimento. As cores cinza e vermelho possuem valores mínimos abaixo de 60%.

No entanto, mesmo a cor RGB se destacando nas taxas de reconhecimento, como observado no gráfico, para verificar se o grupo de cores são estatisticamente significativos foi feita análise de variância (ANOVA), com teste post- hoc de Tukey. Considerou-se como significantes os resultados que apresentaram  $p \leq 0,05$  (Tabela 4.9).

Tabela 4.9 – Análise Comparativa entre os resultados nas escalas de cinza, vermelha, verde, azul e RGB

Treino	Centroides	Cinza	Vermelha	Verde	Azul	RGB
40%	16	54,72c	55,83c	68,47bc	68,19b	93,33a
	32	61,53c	65,14c	70,14bc	75,28b	89,17a
	64	69,17c	68,47c	73,89bc	78,61b	86,46a
	128	70,69c	72,08c	77,64bc	79,31b	98,19a
	256	74,44c	73,19c	79,72bc	81,11b	95,67a
50%	16	62,83c	58,17c	70,83bc	75,5b	93,96a
	32	72,67c	71,5c	75,17bc	80,17b	99,72a
	64	77,67c	77c	81,5bc	84,17b	98,83a
	128	79,67c	77,67c	82,83bc	84,83b	98,33a
	256	82c	81,33c	84,67bc	86,17b	100a
60%	16	65,21c	62,5c	69,79bc	79,38b	100a
	32	72,71c	71,04c	77,71bc	83,54b	100a
	64	78,96c	77,29c	78,13bc	88,13b	100a
	128	82,08c	80,63c	84,58bc	90,21b	100a
	256	84,17c	82,71c	85bc	89,79b	100a

Fonte: Autor.

Observa-se na análise inferencial apresentada na Tabela 4.8 que não foi encontrado taxa de reconhecimento significativa entre as cores cinza, vermelha e verde, isto é, estatisticamente são iguais. E, entre as cores verde e azul, também, não há significância, sendo que essas duas cores se enquadram no mesmo grupo estatístico.

No entanto, o grupo de cores RGB se sobressai separadamente com significância e taxas maiores do que os demais grupos, da onde pode-se destacar que para imagens em RGB tem-se taxas de reconhecimento melhores, verificado descritivamente e pelo teste adotado nesse trabalho.

Portanto, para o reconhecimento das imagens trabalhadas, utilizando QV, obtém-se resultados estatisticamente melhores na escala de cor RGB, conforme apresentado na Tabela 4.8 e para otimização do reconhecimento utilizar-se-á imagens com 32 centroides, conforme discutido no item 5.4.1 e apresentado na Tabela 4.6.

#### 4.6 Análise dos resultados de pela técnica de DL

Para utilização da técnica de Deep Learning utilizou-se 30 imagens na escala RGB, em cada uma das 40 classes e adotou-se treinos de 40%, 50% e 60% de imagens por classe, para que posteriormente possa-se obter a comparação com os resultados obtidos pela técnica de Quantização Vetorial com imagens em RGB, destacados na seção anterior.

A técnica de Deep Learning também alcançou altas taxas de reconhecimento conforme apresentado na Tabela 4.10.

Tabela 4.10 – Resultados obtidos por DL RGB

Treino \ Técnica	40%	50%	60%
Deep Learning RGB	93	91,06	95,81

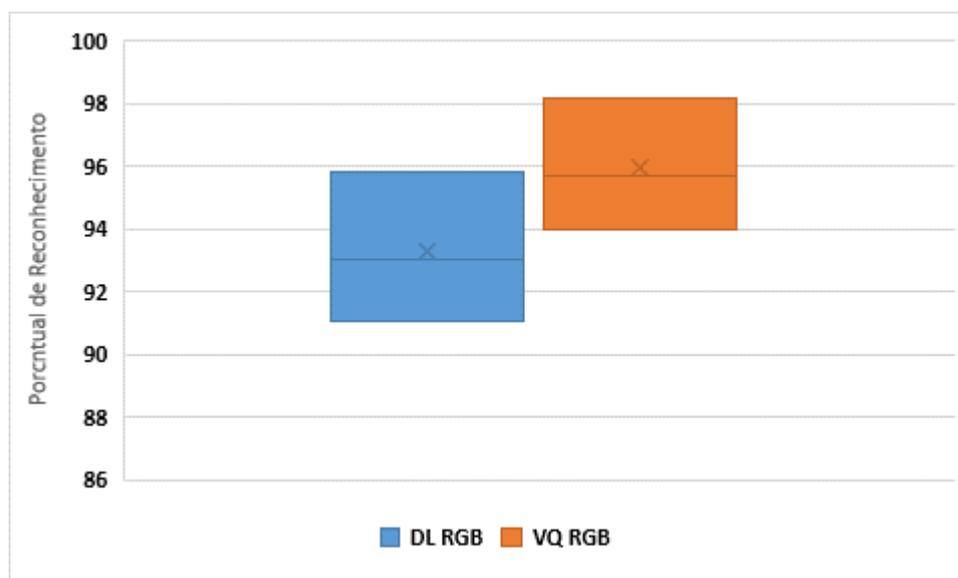
Fonte: Autor.

#### 4.7 Análise Comparativa dos resultados de QV RGB com DL

Para a comparação com os resultados obtidos pelo processamento do Deep Learning, adotou-se os reconhecimentos alcançados com 32 centroides, pois foi encontrado uma taxa de reconhecimento significativa a partir dessa quantidade de centroides, nas imagens processadas pela técnica de QV na escala RGB, conforme apresentado na seção 4.5.2.

Foi feita uma distribuição desses resultados utilizando o diagrama de caixa e caixa estreita, indicando a variabilidade da amostra o Gráfico 4.5.

Gráfico 4.5 – Diagrama de caixa e caixa estreita comparando a Variabilidade entre QV RGB 32 Centroides e DL RGB



Fonte: Autor.

No geral, como notado no Gráfico 4.4 tem-se uma variação das taxas de reconhecimento para as técnicas de DL-RGB e QV com 32 centroides na cor RGB. Os

valores máximo e mínimo para os dois grupos ficaram dentro do intervalo interquartilico. Nota-se percentuais descritivos de reconhecimento maiores para a técnica de QV em relação à de DL.

A mediana e a média para os resultados de QV com 32 centroides na cor RGB ficaram próximos a 96% e os valores da mediana e da média para os resultados de DL na cor RGB ficaram abaixo de 94%.

Diante das diferenças apresentadas pelos dois métodos em estudo, foi feita a inferência dos dados para verificar se existe diferenças significativas entre os resultados dessas técnicas (Tabela 4.11).

Tabela 4.11 – Análise Comparativa dos resultados de QV 32 centroides com DL

<b>Treino</b>	<b>40</b>	<b>50</b>	<b>60</b>
<b>Técnica</b>			
<b>QV RGB</b>	98,19a	95,67a	93,96a
<b>Deep Learning RGB</b>	93a	91,06a	95,81a

Fonte: Autor.

Para tanto, observou-se neste estudo que não houve diferença significativa entre as técnicas, quando variado os treinos (40%, 50% e 60%) e as técnicas de reconhecimento, para cada quantidade de imagens por classes.

A média de reconhecimento obtida entre as imagens por classe e todos os treinos, da técnica de Deep Learning RGB é 93,29% e a média da técnica de QV RGB 32 centroides é de 95,94%, para quando utilizado 40 classes, sendo 30 imagens por classe.

#### **4.8 Comparação com outros trabalhos**

Para fins de comparação dos resultados obtidos nessa pesquisa com outros trabalhos encontrados na literatura foi utilizado a média percentual dos resultados apresentados e discutidos na seção 4.7. A Tabela 4.12 mostra a relação de alguns trabalhos encontrados na literatura em função da taxa de reconhecimento.

Tabela 4.12 – Resultado Obtido Utilizando O Método Proposto E Os Resultados Encontrados Na Literatura

<b>Trabalhos correlatos</b>	<b>Base de dados</b>	<b>Taxa de reconhecimento %</b>
Método Proposto - QV	SBID	95,94%
Método Proposto – DL	SBID	93,29%
Li e Allinson [1]	SBID	94,6%
Li e Allinson [7]	SBID	85,25%
Groeneweg. et al [12]	ZuBud	91%
Li e Shapiro [8]	977 <i>color image</i>	94,2%
Zhang e Kosecká [9]	ZuBud	95%
Chung, Han e He [11]	ZuBud	81%

Fonte: Autor.

Na comparação dos resultados médios apresentado pelos métodos propostos no trabalho (QV e DL) tem-se que o método da Quantização Vetorial obteve um valor percentual descritivamente maior dos que aos apresentados na literatura até o momento, conforme observa-se na Tabela 4.12.

#### 4.9 – Considerações finais do capítulo

Esse capítulo apresentou as análises dos resultados e discussões para imagens nas escalas cinza, vermelha, verde, azul e RGB utilizando Quantização Vetorial. Destacou os resultados com 32 centroides na cor RGB para a técnica de Quantização Vetorial e comparou com os resultados obtidos pela técnica Deep Learning por meio de análise de variância (ANOVA), com teste paramétrico post-hoc de Tukey, considerando significantes os resultados que apresentaram  $p \leq 0,05$ .

Por fim comparou-se os resultados obtidos com os encontrados na literatura até esse momento, de onde se destaca que o método proposto pela Quantização Vetorial obtém, em média, valores percentuais para o reconhecimento, descritivamente, melhores do que os encontrados na literatura.

## CAPÍTULO 5

### CONSIDERAÇÕES FINAIS

Mediante o exposto no trabalho destaca-se nesse capítulo, os melhores resultados encontrados com suas respectivas técnicas. Apresenta os melhores resultados nas escalas de cores e centroides para QV e a comparação com os resultados em DL. Mostra os melhores resultados obtidos com a análise de variância para verificar a existência de diferença significativa entre as médias e se os fatores exercem influência em alguma variável: cor, centroide e treino. E, por fim esclarece a confirmação da hipótese nula apresentada na pesquisa.

#### 5.1 Considerações Finais

Avanços no campo computacional têm promovido técnicas novas e aprimoradas para o reconhecimento de edifícios. No entanto, destaca-se que a técnica de QV, utilizando o algoritmo LBG, teve um excelente desempenho para reconhecimento de edifícios, sendo que foi encontrado, para 128 e 256 centroides, taxas de reconhecimento de 100% para a escala de cor RGB independentemente do porcentual de treino adotados nesse trabalho. E, que para 64 centroides os resultados foram acima de 98%.

Quando analisado as escalas de cores cinza, vermelha, verde e azul obteve-se pelos resultados da estatística inferencial que as escalas verde e azul obtiveram resultados melhores e estatisticamente iguais.

Observou-se também que as imagens processadas pelo método de QV, quando analisadas com as escalas de cores cinza, vermelha, verde, azul e RGB apresentaram resultados estatisticamente significativos a partir de 32 centroides. E, quando analisadas em relação as cores, a escala de cor RGB se sobressai separadamente com significância e taxas maiores do que as demais escalas, da onde pode-se destacar que para imagens em RGB tem-se taxas de reconhecimento melhores, verificado descritivamente e pela estatística inferencial das frequências percentuais de acerto para reconhecimento utilizando análise de variância (ANOVA), com teste paramétrico post- hoc de Tukey, considerando significantes os resultados que apresentaram  $p \leq 0,05$ . Dessa forma, adotou-se essas taxas de reconhecimentos para comparar com os resultados encontrados pelo método de Deep Learning com as imagens em RGB.

A técnica de Deep Learning também alcançou altas taxas de reconhecimento. E, foi feita a inferência dos dados para verificar se existe diferenças significativas entre os resultados dessas técnicas, onde concluiu-se para análise de variância que não existe diferença significativa entre os resultados obtidos pelas técnicas de QV trabalhadas com 32 centroides com imagens RGB e DL com imagens em RGB.

Observou-se neste estudo que não houve diferença significativa entre as técnicas, quando variado os treinos (40%, 50% e 60%).

Na análise com 32 centroides obteve-se, pela QV com imagens em RGB, uma média de 95,94% de reconhecimento das imagens, e para a técnica de DL com imagens em RGB a média foi de 93,29%.

Conclui-se que as duas técnicas quando utilizadas para reconhecimento de edifícios em 40 classes do banco de dados SBID, com 30 imagens por classe, não possui diferença significativa entre as taxas de reconhecimento. Portanto, aceita-se a hipótese nula apresentada no trabalho.

Com a utilização dos testes estatísticos gerou-se valor à pesquisa, pois não foi encontrado trabalhos que compara estatisticamente as técnicas de processamento utilizadas, visto que isto é pouco explorado na literatura, mas que merece destaque. Logo, a inovação do trabalho compreende-se na proposta da originalidade da comparação, utilizando o teste paramétrico da ANOVA, com post-hoc de Tukey, entre as técnicas utilizadas para o reconhecimento de edifícios.

Esse trabalho não fecha nenhuma discussão a respeito de reconhecimento de edifícios com o uso de QV e DL e sim abre novas portas para futuras pesquisas.

Como sugestão para futuros trabalhos pode-se comparar as diversas técnicas utilizadas para reconhecimento em um banco de dados com uma quantidade maior de imagens e analisar se essas técnicas obtém resultados de reconhecimento estatisticamente significativas entre si.

## REFERÊNCIAS

- ALBUQUERQUE, M.; ESQUEF, I. A. Image Segmentation using nonextensive relative entropy. **IEEE Latin America Transactions**, v. 6, n. 5, p. 477–483, set. 2008. <https://doi.org/10.1109/TLA.2008.4839118>
- AMARAL, V. DO; GIRALDI, G. A.; THOMAZ, C. E. **LBP Estatístico Aplicado ao Reconhecimento de Expressões ~ Faciais**. Proceedings of the X Encontro Nacional de Inteligencia Artificial e Computacional. Anais... Em: ENIAC 13. 2012.
- ARAÚJO, P.; MOYA, M.; JUNIOR, I. **Classificação do óleo de algodão em processo de semirrefino usando técnicas de processamento de imagem**. Anais de XXXV Simpósio Brasileiro de Telecomunicações e Processamento de Sinais. **Anais...** Em: XXXV SIMPÓSIO BRASILEIRO DE TELECOMUNICAÇÕES E PROCESSAMENTO DE SINAIS. Sociedade Brasileira de Telecomunicações, 2017. Disponível em: <<http://biblioteca.sbrrt.org.br/articles/646>>. Acesso em: 7 out. 2022. <https://doi.org/10.14209/sbrrt.2017.166>
- BELTRÃO, F. **Uso de Rede Neural Convolutacional no Reconhecimento de Artrópodes da Classe Insecta**. [s.l.] UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ, 2022.
- BEZAK, P. **Building recognition system based on deep learning**. 2016 Third International Conference on Artificial Intelligence and Pattern Recognition (AIPR). Anais... Em: 2016 THIRD INTERNATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE AND PATTERN RECOGNITION (AIPR). set. 2016. <https://doi.org/10.1109/ICAIPR.2016.7585230>
- BONAFINI, B. L.; PERLIN, H. A. Um Estudo Sobre Reconhecimento De Padrões Aplicado A Detecção De Câncer Do Tipo Melanoma Maligno. **Revista Científica Interdisciplinar INTERLOGOS**, v. 3, n. 1, p. 137–154, 29 jan. 2018.
- BRUCKNER, D. et al. Hierarchical Semantic Processing Architecture for Smart Sensors in Surveillance Networks. **IEEE Transactions on Industrial Informatics**, v. 8, n. 2, p. 291–301, maio 2012. <https://doi.org/10.1109/TII.2012.2186142>
- CASASANT, D.; PSALTIS, D. Position, rotation, and scale invariant optical correlation. **Applied Optics**, v. 15, n. 7, p. 1795–1799, 1 jul. 1976. <https://doi.org/10.1364/AO.15.001795>
- CHEN, C. et al. Learning a two-stage CNN model for multi-sized building detection in remote sensing images. **Remote Sensing Letters**, v. 10, n. 2, p. 103–110, 1 fev. 2019. <https://doi.org/10.1080/2150704X.2018.1528398>
- CHEN, K. et al. Automated crack segmentation in close-range building façade inspection images using deep learning techniques. **Journal of Building Engineering**, v. 43, p. 102913, 1 nov. 2021. <https://doi.org/10.1016/j.jobee.2021.102913>
- CHUNG, Y.-C.; HAN, T. X.; HE, Z. **Building recognition using sketch-based representations and spectral graph matching**. 2009 IEEE 12th International Conference on Computer Vision. Anais... Em: 2009 IEEE 12TH INTERNATIONAL CONFERENCE ON COMPUTER VISION. set. 2009.
- CLAUDIO, R. A. **Aplicação de redes neurais artificiais à estimação de curtíssimo prazo do preço...** São Carlos: USP, 2014.

CONCI, A.; CARVALHO, J.; RAUBER, T. A complete system for vehicle plate localization, segmentation and recognition in real life scene. | **IEEE Journals & Magazine** | **IEEE Xplore**. v. 7, n. 5, p. 497–506, 2009. <https://doi.org/10.1109/TLA.2009.5361185>

DE CARVALHO, J. V.; SAMPAIO, M. C.; MONGIOVI, G. Utilização de Técnicas de “Data Mining” para o Reconhecimento de Caracteres Manuscritos. p. 16, 1999.

DIGIAMPIETRI, L. A. et al. **Um sistema de informação extensível para o reconhecimento automático de LIBRAS**. Anais do Simpósio Brasileiro de Sistemas de Informação (SBSI). **Anais...** Em: ANAIS DO VIII SIMPÓSIO BRASILEIRO DE SISTEMAS DE INFORMAÇÃO. SBC, 16 maio 2012. Disponível em: <<https://sol.sbc.org.br/index.php/sbsi/article/view/14410>>. <https://doi.org/10.5753/sbsi.2012.14410>

GOMES, S.; REBOUÇAS, E.; FILHO, P. P. Reconhecimento Optico de Caracteres para Reconhecimento das Sinalizações Verticais das vias de trânsito. **Soluções para Desenvolvimento do País**, v. 9, p. 9, 1 maio 2014.

GONZAGA DE OLIVEIRA, S. L.; TEIXEIRA DE ASSIS, J. A methodology for identification of fingerprints based on Gabor filter. **IEEE Latin America Transactions**, v. 4, n. 1, p. 1–6, mar. 2006. <https://doi.org/10.1109/TLA.2006.1642442>

GROENEWEG, N. J. C. et al. **A Fast Offline Building Recognition Application on a Mobile Telephone**. (J. Blanc-Talon et al., Eds.)Advanced Concepts for Intelligent Vision Systems. Anais...: Lecture Notes in Computer Science.Berlin, Heidelberg: Springer, 2006. [https://doi.org/10.1007/11864349\\_102](https://doi.org/10.1007/11864349_102)

JÚNIOR, F.; ABREU, K. **Estimação dos Parâmetros de uma SVM utilizando um Algoritmo Genético para o Reconhecimento de Caracteres Manuscritos**. Anais da Escola Regional de Informática do Piauí (ERI-PI). **Anais...** Em: ANAIS DA IV ESCOLA REGIONAL DE INFORMÁTICA DO PIAUÍ. SBC, 16 out. 2018. Disponível em: <<https://sol.sbc.org.br/index.php/eripi/article/view/5141>>. Acesso em: 14 out. 2022.

JÚNIOR, J. L. O. L.; SOARES, J. B. Desenvolvimento de metodologia para avaliação da adesividade agregado-ligante com o uso de processamento digital de imagem. **TRANSPORTES**, v. 27, n. 1, p. 82–95, 30 abr. 2019. <https://doi.org/10.14295/transportes.v27i1.1552>

KALVA, P. R.; ENEMBRECK, F.; KOERICH, A. L. WEB Image Classification using Classifier Combination. **IEEE Latin America Transactions**, v. 6, n. 7, p. 661–671, dez. 2008. <https://doi.org/10.1109/TLA.2008.4917439>

KOERICH, A. L. Handwritten Word Recognition Using Markov Models. **IEEE Latin America Transactions**, v. 2, n. 2, p. 132–141, jun. 2004. <https://doi.org/10.1109/TLA.2004.1468632>

LAIA, M.; CRUVINEL, P. Soil Science Tomographic Projections Filtering using Discret Kalman and Neural Networks. **IEEE Latin America Transactions**, v. 6, n. 1, p. 114–121, mar. 2008. <https://doi.org/10.1109/TLA.2008.4461640>

LEE, K. et al. MultiDefectNet: Multi-Class Defect Detection of Building Façade Based on Deep Convolutional Neural Network. **Sustainability**, v. 12, n. 22, p. 9785, jan. 2020. <https://doi.org/10.3390/su12229785>

- LI, J. et al. Building recognition in urban environments: A survey of state-of-the-art and future challenges. **Information Sciences**, v. 277, p. 406–420, 1 set. 2014. <https://doi.org/10.1016/j.ins.2014.02.112>
- LI, J.; ALLINSON, N. Building Recognition Using Local Oriented Features. **IEEE Transactions on Industrial Informatics**, v. 9, n. 3, p. 1697–1704, ago. 2013. <https://doi.org/10.1109/TII.2013.2245910>
- LI, J.; ALLINSON, N. M. Subspace learning-based dimensionality reduction in building recognition. **Neurocomputing**, Timely Developments in Applied Neural Computing (EANN 2007) / Some Novel Analysis and Learning Methods for Neural Networks (ISNN 2008) / Pattern Recognition in Graphical Domains. v. 73, n. 1, p. 324–330, 1 dez. 2009. <https://doi.org/10.1016/j.neucom.2009.08.016>
- LI, Y.; SHAPIRO, L. G. **Consistent line clusters for building recognition in CBIR**. 2002 International Conference on Pattern Recognition. **Anais...** Em: 2002 INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION. ago. 2002.
- LINDE, Y.; BUZO, A.; GRAY, R. An Algorithm for Vector Quantizer Design. **IEEE Transactions on Communications**, v. 28, n. 1, p. 84–95, jan. 1980. <https://doi.org/10.1109/TCOM.1980.1094577>
- LIU, H. A Fuzzy Qualitative Framework for Connecting Robot Qualitative and Quantitative Representations. **IEEE Transactions on Fuzzy Systems**, v. 16, n. 6, p. 1522–1530, dez. 2008. <https://doi.org/10.1109/TFUZZ.2008.2005004>
- LUCENA, L. R. R. DE et al. LACUNARITY ANALYSIS OF CLIMATIC PHENOMENA IN THE STATE OF PERNAMBUCO. **Brazilian Journal of Biometrics**, v. 34, n. 2, p. 334–346, 28 jun. 2016.
- MELO JÚNIOR, C. M. et al. Geração de mapas de danos de fachadas de edifícios por processamento digital de imagens capturadas por Vant e uso de fotogrametria digital. **Ambiente Construído**, v. 18, n. 3, p. 211–226, set. 2018. <https://doi.org/10.1590/s1678-86212018000300277>
- MENEZES JUNIOR, J. M. P. D. et al. **ANÁLISE DO RECONHECIMENTO DE FACES UTILIZANDO O ALGORITMO DE KNN**. Em: CONGRESSO TÉCNICO CIENTÍFICO DA ENGENHARIA E DA AGRONOMIA. Belém - PA: 2017.
- MOSCHETTA, E.; OSÓRIO, F. S.; CAVALHEIRO, G. G. H. **Reconhecimento de Imagens em Aplicações Críticas**. Anais do Simpósio em Sistemas Computacionais de Alto Desempenho (WSCAD). Anais... Em: ANAIS DO III WORKSHOP EM SISTEMAS COMPUTACIONAIS DE ALTO DESEMPENHO. SBC, 28 out. 2002. Disponível em: <<https://sol.sbc.org.br/index.php/wscad/article/view/20763>>. Acesso em: 13 out. 2022. <https://doi.org/10.5753/wscad.2002.20763>
- NEVES, R. et al. Thresholding the Courtesy Amount of Brazilian Bank Checks Based on Tsallis Entropy. **IEEE Latin America Transactions**, v. 7, n. 6, p. 726–731, dez. 2009. <https://doi.org/10.1109/TLA.2009.5419371>
- NIU, J. et al. Global Research on Artificial Intelligence from 1990–2014: Spatially-Explicit Bibliometric Analysis. **ISPRS International Journal of Geo-Information**, v. 5, n. 5, p. 66, maio 2016. <https://doi.org/10.3390/ijgi5050066>
- OLIVEIRA NUNES, E.; CONCI, A. Texture Segmentation and Edge Location of Regions in Multiband Images. **IEEE Latin America Transactions**, v. 5, n. 3, p. 184–191, jun. 2007. <https://doi.org/10.1109/TLA.2007.4378503>

- OLIVEIRA, F. P. M.; TAVARES, J. M. R. S. **Alinhamento Computacional de Imagens de Pedobarografia Estática e Dinâmica**. p. 3, 2009.
- OLIVEIRA, W. D. **Software Para Reconhecimento De Espécies Florestais A Partir De Imagens Digitais De Madeiras Utilizando Deep Learning**. p. 103, 2018.
- OTTONI, A. L. C.; NOVO, M. S. A Deep Learning Approach to Vegetation Images Recognition in Buildings: a Hyperparameter Tuning Case Study. **IEEE Latin America Transactions**, v. 19, n. 12, p. 2062–2070, 26 maio 2021. <https://doi.org/10.1109/TLA.2021.9480148>
- PARK, J.; LEE, J. A Beacon Color Code Scheduling for the Localization of Multiple Robots. **IEEE Transactions on Industrial Informatics**, v. 7, n. 3, p. 467–475, ago. 2011. <https://doi.org/10.1109/TII.2011.2158833>
- PEREIRA, V. F.; SANTOS, P. E.; COZMAN, F. G. **Reconhecimento de Objetos Baseado em Contexto Utilizando a Lógica de Descrição Probabilística CRALC**. p. 11, 2012.
- PERELMUTER, G. et al. **Reconhecimento de imagens bidimensionais utilizando redes neurais artificiais**. v. Anais do VIII SIBGRAPI, p. 197–203, 1995.
- PRISCILA ROCHA; WASHINGTON SILVA; ALLAN BARROS. **Hierarchical Expert Neural Network System for Speech Recognition** | SpringerLink. n. 30, p. 347–359, 2019. <https://doi.org/10.1007/s40313-019-00459-w>
- QUILES, M. G.; ROMERO, R. A. F. **Um Sistema de Visão Computacional Baseado em Cores Aplicado ao Controle de um Robô Móvel**. p. 5, 2004.
- RODRIGUES, C. H. M. **Extração de fachadas de edifícios em imagens obtidas com UAV**. masterThesis—[s.l: s.n.].
- RUDEK, M.; COELHO, L.; CANGIOLIERI JUNIOR, O. **Visão Computacional Aplicada a Sistemas Produtivos: Fundamentos e Estudo de Caso**. 2001.
- SANTOS, W. et al. **Influência da escala de cores na avaliação qualitativa em cintigrafia de perfusão cerebral**. v. 18, p. 36–43, 2017.
- SCHENATTO, K.; DE PAULA FILHO, P. L. D. P.; BAZZI, C. L. Reconhecimento de doenças em folhas de laranjeiras utilizando técnicas de processamento de imagens. **Revista Eletrônica Científica Inovação e Tecnologia**, v. 3, n. 5, p. 15–21, 14 ago. 2012.
- SILVA, A. L. R. DA. **Seleção de atributos para apoio ao diagnóstico do câncer de mama usando imagens termográficas, algoritmos genéticos e otimização por enxame de partículas**. Recife: UFPE, 2019.
- SILVA, J. V. S. et al. Combinação de técnicas de compressão de modelos profundos. **IEEE Latin America Transactions**, v. 20, n. 3, p. 458–464, 2022.
- SILVA, L. A. da, CARRIJO, G.A., MOREIRA, F.S., JUNIOR, A.C.L., GOULART, C.S., & SOUSA, L.M. (2022). Reconhecimento de edifícios utilizando o filtro de gabor, a wavelet transformada e a rede neural perceptron de camadas múltiplas. **Brazilian Journal of Development**, 8 (5), 35182–35199.. <https://doi.org/10.34117/bjdv8n5-164>
- SOUSA, K. A. O. **Uso de visão computacional em dispositivos móveis para auxílio à travessia de pedestres**. São Paulo: UNIVERSIDADE PRESBITERIANA MACKENZIE, 2012.

SOUZA, F. R. A. DE; ARAÚJO, J. DE S. **Uma Metodologia para o Reconhecimento de Padrões em Imagens Faciais com Redes Neurais Artificiais**. Anais da Escola Regional de Alto Desempenho Norte 2 (ERAD-NO2) e Escola Regional de Aprendizado de Máquina e Inteligência Artificial Norte 2 (ERAMIA-NO2). Anais... Em: ANAIS DA I ESCOLA REGIONAL DE ALTO DESEMPENHO NORTE 2 E I ESCOLA REGIONAL DE APRENDIZADO DE MÁQUINA E INTELIGÊNCIA ARTIFICIAL NORTE 2. SBC, 18 nov. 2021. Disponível em: <<https://sol.sbc.org.br/index.php/erad-eramia-no2/article/view/18675>>. Acesso em: 14 out. 2022. <https://doi.org/10.5753/erad-no2.2021.18675>

STEIM, I. P. I. **Uma proposta de arquitetura de rede neural convolucional intervalar para o processamento de imagens intervalares**. Pelotas: UNIVERSIDADE FEDERAL DE PELOTAS, 2017.

SULEIMAN, W.; JOLIVEAU, T.; FAVIER, E. A New Algorithm for 3D Isovists. Em: TIMPF, S.; LAUBE, P. (Eds.). *Advances in Spatial Data Handling: Geospatial Dynamics, Geosimulation and Exploratory Visualization*. **Advances in Geographic Information Science**. Berlin, Heidelberg: Springer, 2013. p. 157–173. [https://doi.org/10.1007/978-3-642-32316-4\\_11](https://doi.org/10.1007/978-3-642-32316-4_11)

TEIXEIRA, A. M. D. A. **Método De Atualização Nas Escalas 1:100000 E 1:250000 Do Mapeamento Sistemático A Partir De Imagem De Sensoriamento Remoto: Reservatório De Itaparica**. Recife: UNIVERSIDADE FEDERAL DE PERNAMBUCO, 2016.

TRAMONTIN, E. D. **Análise e Aplicação de Reconhecimento Facial em Sistema Embarcado**. Araranguá: UNIVERSIDADE FEDERAL DE SANTA CATARINA, 2016.

VALERIO NETTO, A. System Based on Computational Intelligence for Ophthalmology Image Understanding. **Latin America Transactions, IEEE (Revista IEEE America Latina)**, v. 3, p. 14–22, 2005. <https://doi.org/10.1109/TLA.2005.1642434>

VASCONCELOS, E. S. et al. Utilização de Quantização Vetorial para Reconhecimento de Edifícios em Ambientes Urbanos. **Conjecturas**, v. 22, 3 out. 2022. <https://doi.org/10.53660/CONJ-1763-2K69>

VASCONCELOS, E. S.; MELO, D. V.; CARRIJO, G. A. Reconhecimento de edifícios em ambientes urbanos com a utilização de quantização vetorial / Recognition of buildings in urban environments using vector quantization. **Brazilian Journal of Development**, v. 8, n. 4, p. 29268–29278, 22 abr. 2022. <https://doi.org/10.34117/bjdv8n4-425>

VUNGO, M. **Reconhecimento Automático de Sinais de Trânsito em Imagens Digitais**. [s.l.] UNIVERSIDADE DO PORTO, 2017.

ZANOTTA, D. C.; FERREIRA, M. P.; ZORTEA, M. **Processamento de imagens de satélite**. [s.l.] Oficina de Textos, 2019.

ZHANG, G.; PAN, Y.; ZHANG, L. Deep learning for detecting building façade elements from images considering prior knowledge. **Automation in Construction**, v. 133, p. 104016, 1 jan. 2022. <https://doi.org/10.1016/j.autcon.2021.104016>

ZHANG, Q.; LAPIERRE, L.; XIANG, X. Distributed Control of Coordinated Path Tracking for Networked Nonholonomic Mobile Vehicles. **IEEE Transactions on Industrial Informatics**, v. 9, n. 1, p. 472–484, fev. 2013. <https://doi.org/10.1109/TII.2012.2219541>

ZHANG, W.; KOSECKA, J. **Localization Based on Building Recognition**. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Workshops. Anais... Em: 2005 IEEE COMPUTER SOCIETY CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR'05) - WORKSHOPS. set. 2005.