

---

## Geração de imagens artificiais e quantização de imagens aplicadas à classificação de imagens

---

***Gabriela Salvador Thumé***

---



SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: \_\_\_\_\_

**Gabriela Salvador Thumé**

**Geração de imagens artificiais e quantização de imagens  
aplicadas à classificação de imagens**

Dissertação apresentada ao Instituto de Ciências  
Matemáticas e de Computação – ICMC-USP,  
como parte dos requisitos para obtenção do título  
de Mestra em Ciências – Ciências de Computação  
e Matemática Computacional. *EXEMPLAR DE  
DEFESA*

Área de Concentração: Ciências de Computação e  
Matemática Computacional

Orientador: Prof. Dr. Moacir Antonelli Ponti  
Coorientador: Prof. Dr. João do Espírito Santo  
Batista Neto

**USP – São Carlos  
Fevereiro de 2016**

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi  
e Seção Técnica de Informática, ICMC/USP,  
com os dados fornecidos pelo(a) autor(a)

T634g Thumé, Gabriela Salvador  
Geração de imagens artificiais e quantização  
de imagens aplicadas à classificação de imagens  
/ Gabriela Salvador Thumé; orientador Moacir  
Antonelli Ponti; coorientador João do Espírito Santo  
Batista Neto. - São Carlos - SP, 2016.  
90 p.

Dissertação (Mestrado - Programa de Pós-Graduação  
em Ciências de Computação e Matemática Computacional)  
- Instituto de Ciências Matemáticas e de Computação,  
Universidade de São Paulo, 2016.

1. Processamento de imagens. 2. bases de dados  
desbalanceados. 3. classificação de imagens.. I.  
Ponti, Moacir Antonelli, orient. II. Neto, João do  
Espírito Santo Batista, coorient. III. Título.

**Gabriela Salvador Thumé**

**Artificial images generation and image quantization applied  
to image classification task**

Master dissertation submitted to the Instituto de Ciências Matemáticas e de Computação – ICMC-USP, in partial fulfillment of the requirements for the degree of the Master Program in Computer Science and Computational Mathematics. *EXAMINATION BOARD PRESENTATION COPY*

Concentration Area: Computer Science and Computational Mathematics

Advisor: Prof. Dr. Moacir Antonelli Ponti  
Co-advisor: Prof. Dr. João do Espírito Santo Batista Neto

**USP – São Carlos**  
**February 2016**







---

## **AGRADECIMENTOS**

---



“”

( )



# RESUMO

THUMÉ, G. S.. **Geração de imagens artificiais e quantização de imagens aplicadas à classificação de imagens.** 2016. 90 f. Dissertação (Mestrado em Ciências – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação (ICMC/USP), São Carlos – SP.

Cada imagem pode ser representada como uma combinação de diversas características, como por exemplo o histograma de intensidades de cor ou propriedades de textura da imagem. Essas características compõe um vetor multidimensional que representa a imagem. É comum esse vetor ser dado como entrada para um método de classificação de padrões que, após aprender através de diversos exemplos, pode indicar o quanto bem separadas as classes de imagens estão. Estudos sugerem evidências de que a preparação das imagens – por meio da especificação cuidadosa da aquisição, pré-processamento e segmentação – pode impactar significativamente a classificação. Além da falta de tratamento das imagens antes da extração de características, o desbalanceamento de classes também se apresenta como um obstáculo para que a classificação seja satisfatória. Imagens possuem características que podem ser exploradas para melhorar a descrição dos objetos de interesse. Entre as possibilidades de melhorias estão: a redução do número de cores das imagens antes da extração de características ao invés de métodos de quantização no vetor já extraído; e a geração de imagens a partir das imagens originais, de forma a promover o balanceamento de bases de dados cujo número de exemplos de cada classe é desbalanceado. Portanto, a proposta desta dissertação é melhorar da classificação utilizando métodos de processamento de imagens antes da extração de características. Busca-se assim estudar a influência de promover o balanceamento de bases de dados e da quantização na classificação. Esse estudo analisa ainda a visualização do espaço de características após os métodos de geração artificial de imagens e de interpolação das características extraídas das imagens originais (SMOTE), contracenando com o espaço original. A ênfase dessa visualização se dá na facilidade de observação da importância do rebalanceamento das classes quando comparado com valores de métricas estatísticas, como a acurácia da classificação. Os resultados indicam que a quantização simplifica as imagens antes da extração de características e posterior redução de dimensionalidade, produzindo vetores mais compactos. Esta dissertação possui resultados que demonstram o potencial da pesquisa e pretende contribuir com a área, investigando métodos que permitam obter melhores espaços de características.

**Palavras-chave:** Processamento de imagens, bases de dados desbalanceados, classificação de imagens..



# ABSTRACT

THUMÉ, G. S.. **Geração de imagens artificiais e quantização de imagens aplicadas à classificação de imagens.** 2016. 90 f. Dissertação (Mestrado em Ciências – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação (ICMC/USP), São Carlos – SP.

An image can be represented as a combination of several features like histograms or texture properties. Those features are composed in a multidimensional vector, which represents the original image. Commonly this vector is given as input to a classification method that can indicate how much separated are the images. The literature suggests that image processing steps like *accute acquisition, preprocessing* and *segmentation* can positively impact such classification. Besides that, *class unbalancing* is also a barrier to achieve good classification accuracy. Some features and methods can be explored to improve objects' description, thus their classification. Possible suggestions include: reducing color's number before feature extraction instead of applying quantization methods to vectors already extracted; and generating synthetic images by means of original ones to balance the number of samples in an uneven dataset. We propose to improve image classification using image processing methods before feature extraction. Specifically we want to analyse the influence of both balancing and quantization methods while applied to datasets in a classification routine. This research also analyses the visualization of feature space after the artificial image generation and feature interpolation (SMOTE), against to original space. Such visualization is used because it allows us to know how important is the rebalacing method when compared with statistical metrics. The main contribution of this research is in methods to improve image classification by obtaining a better feature space.

**Key-words:** Image processing, unbalanced datasets, image classification..



# LISTA DE ILUSTRAÇÕES

---

---

Figura 1 – Etapas canônicas do reconhecimento de padrões desde a aquisição da imagem até sua posterior classificação. . . . .	26
Figura 2 – Visualização pixelizada de uma imagem da base COREL-1000. . . . .	32
Figura 3 – Sobre a imagem RGB de entrada foram realizadas operações de borramento, realce e de equalização de histograma. A imagem à direita é resultante dessas operações. . . . .	33
Figura 4 – Exemplo de filtragem gaussiana como operação de pré-processamento. . . . .	35
Figura 5 – A imagem original, já em escala de cinza, foi realçada utilizando o método <i>unsharp masking</i> . . . . .	36
Figura 6 – Conversão para a escala de cinza com os métodos de Intensidade e MSB. . .	39
Figura 7 – Características latentes de algas verdes. . . . .	43
Figura 8 – Plano computado pelo método de conversão para escala de cinza Intensidade, quando um dos canais de cor possui valor 255. . . . .	52
Figura 9 – Resultado da aplicação de métodos de quantização. A imagem original a resultou em versões de um canal de cor com 232 cores unicas para MSB e 184 cores para os restantes métodos. . . . .	53
Figura 10 – Duas imagens da base de dados Caltech101 com variações no parâmetro de cor utilizando o método MSB. Da esquerda para a direita: imagem original 24-bits e na sequência suas versões quantizadas com: 256, 64, 32, 16 e 8 cores.	54
Figura 11 – Geração artificial a classe minoritária para rebalancear as classes. . . . .	56
Figura 12 – Geração artificial utilizando borramento com filtro bilateral. . . . .	57
Figura 13 – Geração artificial utilizando unsharp masking. . . . .	58
Figura 14 – Geração artificial utilizando adição de ruído de Poisson. . . . .	59
Figura 15 – Geração artificial utilizando o método SMOTE no espaço visual. . . . .	61
Figura 16 – Geração artificial utilizando uma mistura ponderada de duas imagens. . .	62
Figura 17 – Geração artificial utilizando uma mistura limiarizada de duas imagens. .	63
Figura 18 – Geração artificial utilizando uma mistura de duas imagens a partir da saliência da primeira imagem. . . . .	64
Figura 19 – Geração artificial utilizando uma composição de imagens. . . . .	65
Figura 21 – Resultados para Corel(a), Produce(b) e Caltech(c), com todos os métodos de quantização. Para cada método de extração de características a acurácia é resultante da sua aplicação utilizando 256, 128, 64, 32, 16 e 8 cores, da esquerda para a direita. . . . .	71

Figura 22 – Resultados de acurácia para o método de quantização MSB considerando 256, 128, 64 e 32 cores com o descritor BIC. Os boxplots em cinza correspondem às significâncias estatísticas com $p \leq 0.01$ quando comparado a acurácia de 256 cores. . . . .	72
Figura 23 – Resultados de acurácia para o método de quantização Luminância considerando 256, 128, 64 e 32 cores com o descritor Haralick. Os boxplots em cinza correspondem às significâncias estatísticas com $p \leq 0.01$ quando comparado a acurácia de 256 cores. . . . .	72
Figura 24 – Resultados de acurácia para os método MSB (quantização) e LPP para redução de dimensionalidade e BIC para a extração de características. A comparação foi realizada de forma par (LPP versus MSB) com a mesma dimensionalidade. Os boxplots em cinza correspondem às significâncias estatísticas com $p \leq 0.01$ quando comparado a acurácia de 256 cores. . . . .	73
Figura 25 – Comparação da acurácia alcançada com diferentes métodos de quantização: Gleam, Intensidade, Luminância e MSB. Inicialmente com $D = 2310$ e então reduzindo com LPP para $d = 1160, 582, 294$ e $150$ . . . . .	73
Figura 26 – Comparação da acurácia com o uso da projeção LPP e o método MSB para quantização das imagens com o objetivo de redução de dimensionalidade. . . . .	74
Figura 27 – Resultados para a projeção do LPP sobre o espaço de características produzido pelo método de quantização MSB utilizando 256 ( $d = 2310$ ) e 64 cores ( $d = 582$ ). . . . .	74
Figura 28 – Fluxo de operações do experimento realizado. . . . .	76
Figura 29 – Base de imagens COREL-1000. . . . .	76
Figura 30 – Remoção de 50% das imagens de treino da classe Cavalo. . . . .	77
Figura 31 – Comparação dos exemplos de treinamento da geração com SMOTE e no campo visual. Em laranja estão representados os novos exemplos, projetados no plano da base original balanceada. . . . .	77
Figura 32 – Resultado do teste da classificação com 1-NN após o treinamento realizado com as bases rebalanceadas. A cor no interior dos marcadores quadrados representa a classe real dos exemplos e a borda representa a classe predita pelo classificador. . . . .	78
Figura 33 – Região de decisão com K-NN ( $K = 1$ ) . . . . .	79
Figura 34 – Melhores subespaços encontrados após a geração de novos exemplos para o SMOTE e para a geração artificial de imagens, e após a remoção de imagens para a projeção dos dados desbalanceados. . . . .	80
Figura 35 – Visualização do impacto do descritor de características. . . . .	81

---

## **LISTA DE QUADROS**

---



---

# **LISTA DE ALGORITMOS**

---

Algoritmo 1 – Algoritmo do SMOTE . . . . .	45
Algoritmo 2 – Algoritmo de borramento com filtro bilateral . . . . .	57
Algoritmo 3 – Algoritmo de aguçamento . . . . .	58
Algoritmo 4 – Algoritmo da geração com ruído de Poisson . . . . .	60
Algoritmo 5 – Algoritmo da geração com SMOTE visual . . . . .	61
Algoritmo 6 – Algoritmo de mistura ponderada . . . . .	62
Algoritmo 7 – Algoritmo de mistura limiarizada . . . . .	63
Algoritmo 8 – Algoritmo de mistura saliente . . . . .	64
Algoritmo 9 – Algoritmo de composição . . . . .	66



---

## **LISTA DE TABELAS**

---



# SUMÁRIO

---

---

1	<b>INTRODUÇÃO</b>	25
1.1	Contextualização	27
1.2	Contribuições	28
1.3	Estrutura do documento	29
2	<b>FUNDAMENTOS</b>	31
2.1	Considerações iniciais	31
2.2	Pré-processamento de imagens	31
2.2.1	<i>Filtragem espacial e convolução</i>	34
2.2.2	<i>Deconvolução</i>	34
2.2.3	<i>Realce de imagens</i>	35
2.2.4	<i>Restauração</i>	37
2.2.5	<i>Quantização</i>	38
2.3	Extração de características	39
2.4	Características latentes	42
2.5	Desbalanceamento de classes	44
2.5.1	<i>Sobreamostragem</i>	44
2.5.2	<i>Subamostragem</i>	45
2.6	Classificadores de padrões	46
2.6.1	<i>Algoritmo k-Vizinhos Mais Próximos</i>	46
2.7	Redução de dimensionalidade	46
2.7.1	<i>Análise de componentes principais</i>	47
2.7.2	<i>Locality Preserving Projections</i>	48
2.8	Considerações finais	48
3	<b>QUANTIZAÇÃO DE IMAGENS</b>	51
3.1	Considerações iniciais	51
3.2	Quantização de imagens	51
3.3	Considerações finais	53
4	<b>GERAÇÃO ARTIFICIAL DE IMAGENS</b>	55
4.1	Considerações iniciais	55
4.2	Borramento	56
4.3	Aguçamento	58

4.4	Adição de ruído . . . . .	59
4.5	SMOTE visual . . . . .	60
4.6	Mistura ponderada . . . . .	62
4.7	Mistura limiarizada . . . . .	63
4.8	Mistura saliente . . . . .	64
4.9	Composição . . . . .	65
4.10	Considerações finais . . . . .	66
5	<b>RESULTADOS</b> . . . . .	67
5.1	Considerações Iniciais . . . . .	67
5.2	Quantização de Imagens . . . . .	67
5.2.1	<i>Base de Imagens</i> . . . . .	67
5.2.2	<i>Protocolo</i> . . . . .	68
5.2.3	<i>Resultados e Discussão</i> . . . . .	70
5.3	Geração de Imagens Artificiais . . . . .	75
5.3.1	<i>Base de Imagens</i> . . . . .	75
5.3.2	<i>Protocolo</i> . . . . .	75
5.3.3	<i>Resultados e Discussão</i> . . . . .	75
5.4	Considerações Finais . . . . .	79
6	<b>CONCLUSÕES</b> . . . . .	83
6.1	Publicações . . . . .	84
6.2	Trabalhos Futuros . . . . .	84
	<b>Referências</b> . . . . .	87



# INTRODUÇÃO

---

A tarefa de classificação de imagens consiste em predizer corretamente uma imagem como pertencente a uma classe previamente determinada. Um exemplo prático é a classificação da imagem de um *oceano* como parte de uma classe denominada *praia*. Uma forma de definir que certa imagem pertence à uma classe é especificar todas as regras que a caracterizam. Porém, para a maioria dos casos isso é impossível. Considere imagens coloridas, com três canais de cores e de tamanho  $256 \times 256$  pixels onde cada um desses 65536 pixels pode ser representado por  $256^3$  combinações discretas de cores. Essa complexidade pode ser reduzida ao utilizar métodos de extração de características, os quais visam representar uma imagem com um número significativamente menor de valores vetoriais. Utilizando-se tal representação, pode-se desenvolver métodos computacionais que consigam definir e identificar a qual classe pertence a imagem – sem a necessidade de se codificar todas as regras possíveis – por meio de algoritmos de Aprendizado de Máquina. Esses algoritmos possuem capacidade de generalização, crucial para classificar novos exemplos não contidos na base de imagens originalmente utilizada para o seu treinamento. Assim, “aprendem” a determinar a classe correta para as imagens de entrada. Em uma etapa posterior pode-se validar esse aprendizado, aplicando o algoritmo a novos exemplos não contidos no treinamento.

O reconhecimento de padrões em imagens possui aspectos particulares para cada aplicação. Apesar da grande variedade de extratores de características disponíveis, nem sempre é possível representar as imagens de maneira satisfatória. Isso porque existem conjuntos de características que dificultam a diferenciação entre as classes. Um dos objetivos da área de Aprendizado de Máquina é encontrar quais são essas características que melhor discriminam as classes e, dessa forma, obtêm melhores resultados na etapa de reconhecimento. É comum concentrar o maior esforço dessa tarefa ao operar no espaço de características já extraídas. Para lidar com a deficiência da extração dessas características, podem ser utilizadas transformações do espaço ou sistemas de classificação complexos. No entanto, imagens obtidas de diferentes fontes, como imagens naturais, de microscopia, telescopia e tomografia, possuem características que

podem ser exploradas além dos métodos clássicos. Por isso é importante investigar métodos de processamento e preparação de imagens antes da extração dessas características, ao invés de lidar com a má representação das imagens. O uso desses métodos pode revelar *características latentes*, não visíveis nas imagens originais. Tais características podem melhor descrever certas classes, pois melhoram o conjunto de representações de imagens fornecidas à etapa de classificação.

Considerando que é comum realizar a extração de características a partir da imagem original, sem preocupação com a preparação da imagem, o enfoque desse estudo é na etapa de pré-processamento, destacada na Figura 28. Esta ilustra as etapas canônicas do reconhecimento de padrões desde a aquisição da imagem até sua posterior classificação. As etapas de pré-processamento e segmentação — apresentadas em destaque — são normalmente pouco exploradas, quando comparadas com as etapas posteriores.

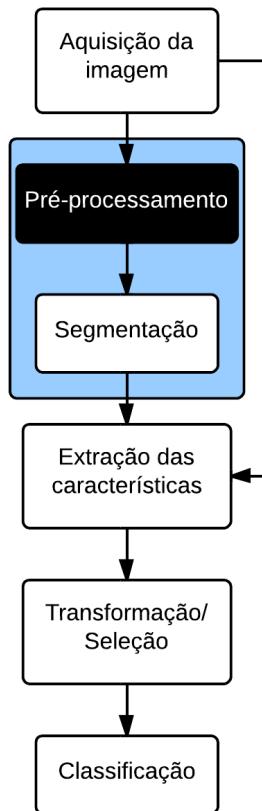


Figura 1 – Etapas canônicas do reconhecimento de padrões desde a aquisição da imagem até sua posterior classificação. As etapas de pré-processamento e segmentação — apresentadas em destaque — são normalmente pouco exploradas, quando comparadas com as etapas posteriores. O enfoque desse estudo é dar maior atenção à etapa de pré-processamento. *Fonte:* Elaborado pela autora.

Ao invés de focar em métodos complexos de transformação do espaço de características, propõe-se a redução da complexidade do problema no início do processo do reconhecimento, ao quantizar as imagens antes da extração de características. Embora a quantização normalmente

faz parte do pipeline, faltam estudos que descrevam o método de quantização e seus parâmetros. Ao negligenciar essa etapa, perde-se a oportunidade de redução da dimensionalidade e/ou o tempo de execução. Ou seja, quer-se estudar se é possível utilizar menos de 8 bits para armazenar as informações de cor para extrair as características com dimensionalidade reduzida.

O desbalanceamento de classes também se apresenta como um obstáculo para que a classificação de imagens seja satisfatória. Esse problema é caracterizado pela diferença entre o número de exemplos disponíveis para cada classe da base de imagens. Em bases médicas, por exemplo, a quantidade de imagens relacionadas com uma doença rara é menor do que as imagens de pacientes sem a doença. Nessas situações, em que as imagens representam eventos importantes porém menos frequentes, o sistema de classificação pode ter problemas para lidar com a classe minoritária. Muitos métodos de transformação do espaço de características e de classificação assumem que as classes da base estão balanceadas, o que nem sempre é verdade. Portanto, é proposto a geração de imagens artificiais a partir do processamento das imagens originais, com o objetivo de rebalancear a base de imagens e consequentemente o modelo criado para a classificação. Esse método foi comparado com o SMOTE, técnica de sobreamostragem dos vetores de características ao interpolar os exemplos mais próximos.

De maneira sumária, esta pesquisa busca melhorar a classificação de imagens, utilizando métodos de processamento com foco na quantização de imagens e no rebalanceamento de classes antes da extração de características. Os resultados obtidos, posteriormente apresentados na Seção 5, demonstram o potencial do processamento de imagens antes da extração de características. Além disso, é fornecido uma visualização do espaço de características após o rebalanceamento das classes, crucial para analisar se as novas características extraídas são relevantes, ou seja, se adicionaram informações que estavam latentes ao aprendizado. Os resultados também demonstram que a quantização das imagens permite ao mesmo tempo obter vetores de características mais compactos e com maior capacidade de discriminação entre classes.

## 1.1 Contextualização

O grupo de pesquisa em Visualização, Imagens e Computação Gráfica (VICG), do Instituto de Ciências Matemáticas e de Computação (ICMC), tem atuado nas áreas de apoio para a classificação de coleções de imagens. Os trabalhos do grupo estão relacionados à visualização de informação com projeções multidimensionais e árvores (JOIA *et al.*, 2011), assim como à extração de características e classificação de coleções de imagens (PAIVA *et al.*, 2011). No que tange o processamento de imagens digitais, PICON, ROSSI e JR (2011) e PONTI (2013) focam no pré-processamento para obter melhores resultados de classificação.

Ainda, PAIVA *et al.* (2011) mostraram que os espaços de características formados por cor e textura podem ser melhorados, porém há um limite até o qual as características podem ser transformadas, ou selecionadas, de forma a garantir a discriminação entre as classes. Tal

projeto atua na investigação de métodos que permitam gerar espaços de características com maior discriminação entre as classes, facilitando a classificação.

Em outros dois trabalhos relacionados é possível ver a diferença da performance para problemas de classificação de imagens. No primeiro, os autores atingem acurácia acima de 98% na classificação de frutas após investigar alterações nos parâmetros de aquisição, realizar pré-processamento e obter a segmentação ([ROCHA et al., 2010](#)). No segundo, os autores indicam que o método utilizado para obter a imagem em escala de cinza (comumente utilizada por algoritmos de extração), pode impactar significativamente a classificação final de diversas bases de imagens ([KANAN; COTTRELL, 2012](#)).

## 1.2 Contribuições

Conforme anteriormente mencionado, muitos aspectos influenciam a performance da classificação de coleções de imagens. É comum encontrar bases cuja extração de características é considerada difícil, ou seja, nas quais algoritmos canônicos de extração não conseguem extrair características que diferenciem bem as classes, prejudicando sua posterior classificação. Nessa situação, normalmente tenta-se lidar com as particularidades das características extraídas através de transformações no espaço de atributos ou mesmo projetando classificadores mais elaborados. Acredita-se que, ao invés disso, é importante investigar métodos de processamento e preparação de imagens antes da extração das características. O objetivo desta pesquisa é explorar as etapas de processamento de imagens com o intuito de melhorar a discriminação entre classes de uma coleção de imagens.

Além disso, o desbalanceamento de classes é um obstáculo para uma classificação satisfatória, e por isso também será estudado. A hipótese, nesse caso, é que a geração de imagens artificiais como preparação para a extração de características pode melhorar a acurácia da classificação, quando comparada à geração de exemplos artificiais no espaço de atributos. Ou seja, gerar novas imagens artificiais — que serão posteriormente reduzidas a atributos — pode apresentar melhores resultados para a classificação do que o *bootstrap* de atributos artificiais. Esse método foi comparado com o SMOTE, técnica de sobreamostragem dos vetores de características ao interpolar os exemplos mais próximos.

Dados tais aspectos, pode-se então diferenciar as contribuições desta pesquisa:

### Geral

Investigar os métodos de pré-processamento de imagens de forma a preparar uma coleção de imagens para a extração de características. Com isso, espera-se ao mesmo tempo obter características latentes e balancear o número de instâncias de diferentes classes.

### Específicas

- A geração de imagens artificiais utilizando métodos de processamento, como filtragem, adição de ruído e mistura pode contribuir com o balanceamento entre classes (em se tratando de problemas de classes desbalanceadas), melhorando a acurácia de algoritmos de classificação, quando comparada à geração de exemplos artificiais no espaço de atributos;
- Demonstrar que é possível obter vetores de características compactos e efetivos ao extrair características de imagens com reduzido níveis de intensidade a um custo computacional baixo. Como a extração de características e posterior redução de dimensionalidade são afetados por diferentes métodos de quantização. Os resultados encontrados apontam para uma alternativa – ou adição – a seleção de características, ao usar os métodos de quantização de imagens. Dado o número de cores limitado na imagem original, a quantidade de possíveis características a serem extraídas é reduzido, especialmente as de cor. A extração de características de textura também é facilitada, considerando que normalmente computa utilizando uma memória proporcional o número de intensidades. Consequentemente, utilizar métodos de descrição de cor após a quantização reduz significativamente o tamanho do vetor de características e para os métodos de descrição de textura pode haver redução no tempo de processamento.

Considerando os objetivos aqui descritos, os resultados esperados desta pesquisa estão destacados na Seção 5.

## 1.3 Estrutura do documento

O conteúdo desta dissertação está estruturado como segue.

**Capítulo 2:** são conceituados os principais fundamentos necessários para o desenvolvimento desta pesquisa: pré-processamento de imagens, extração de características e desbalanceamento de classes.

**Capítulo 3:** a redução do número de intensidades de cor antes da extração de características é descrita.

**Capítulo 4:** descreve-se os métodos para a geração artificial de imagens com o objetivo de rebalancear classes.

**Capítulo 5:** os resultados alcançados são apresentados e discutidos.

**Capítulo 6:** reflete sobre as contribuições deste trabalho e apresenta os trabalhos futuros.





## FUNDAMENTOS

---



---

### 2.1 Considerações iniciais

Neste capítulo são destacados os tópicos mais relevantes para a compreensão do estado da arte de processamento de imagens, com foco na preparação das imagens para a extração de características. O problema do desbalanceamento de classes e seu efeito na classificação satisfatória também são abordados, assim como a extração de características para compreender as propriedades extraídas das imagens. Alguns trabalhos relacionados são utilizados como exemplo, com o objetivo de elucidar tais tópicos.

Inicialmente, a Seção 2.2 apresenta alguns dos principais métodos utilizados para o pré-processamento de imagens (i.e. remoção de ruído, realce de imagens e convolução), relevantes para o desenvolvimento desta pesquisa. Tendo como proposta melhor compreender o que são as características latentes, a Seção 2.4 exemplifica as operações de pré-processamento utilizadas para visualizar tais características em uma aplicação específica. A extração de características é brevemente definida na Seção 2.3. Por fim, o problema do desbalanceamento de classes é apresentado na Seção 2.5. Nessa seção, além de caracterizar o problema, são apresentadas duas vertentes para solucioná-lo: sobreamostragem e subamostragem. Esses conceitos são necessários para a compreensão do capítulo de proposta desta dissertação.

### 2.2 Pré-processamento de imagens

Uma imagem digital  $I$  pode ser definida como uma função  $f(x,y)$ , onde  $x$  e  $y$  são as coordenadas cartesianas de um determinado ponto e  $f$  a intensidade (brilho) da imagem naquele ponto. Essa imagem é composta por finitos elementos chamados de pixels que podem ser diretamente acessados através de sua posição  $x$  e  $y$ . Digitalmente, uma imagem é representada por uma matriz de valores com  $M$  linhas e  $N$  colunas onde cada elemento representa a sua intensidade. Uma imagem colorida  $I$  do sistema RGB possui três canais de cores representantes

das cores vermelha, verde e azul. Cada elemento  $I(x,y)$  corresponde a uma tripla  $(r,g,b)$  de números, com  $0 \leq r \leq 255$ ,  $0 \leq g \leq 255$  e  $0 \leq b \leq 255$ , onde 0 é a intensidade de cor mais escura e 255 a mais clara. A combinação dessas intensidades resulta na cor do pixel (GONZALEZ; WOODS, 2007). A Figura 2 mostra uma visualização ampliada dos pixels de uma imagem digital.



Figura 2 – Visualização pixelizada de uma imagem da base COREL-1000<sup>1</sup>.

*Fonte: Elaborado pela autora.*

O processo de aquisição por um sistema de imageamento pode causar diversas imperfeições nas imagens, como pixels ruidosos, brilho inadequado e outras degradações. O pré-processamento de imagens é caracterizado por receber uma imagem de entrada e fornecer uma imagem de saída. Nessa etapa, efeitos indesejáveis podem ser eliminados e certas características realçadas (Figura 3). Considera-se que um determinado critério utilizado para uma imagem pode não ser o mais eficiente para outra, dependendo assim da área de aplicação.

Assim, técnicas de pré-processamento tornam os dados mais adequados para posterior análise, ao eliminar ou reduzir problemas como ruídos e imperfeições. Em Ponti (2012), o autor relata que a utilização de métodos de restauração na etapa de pré-processamento da imagem, antes da segmentação, resultou em uma qualidade superior para todos os testes, com valores de erro e desvio padrão menores. No referido estudo, métodos de realce causaram perda de informação e por isso não são indicados para uso em imagens obtidas por microscópio. O

<sup>1</sup> Disponível em <http://wang.ist.psu.edu/docs/related/>

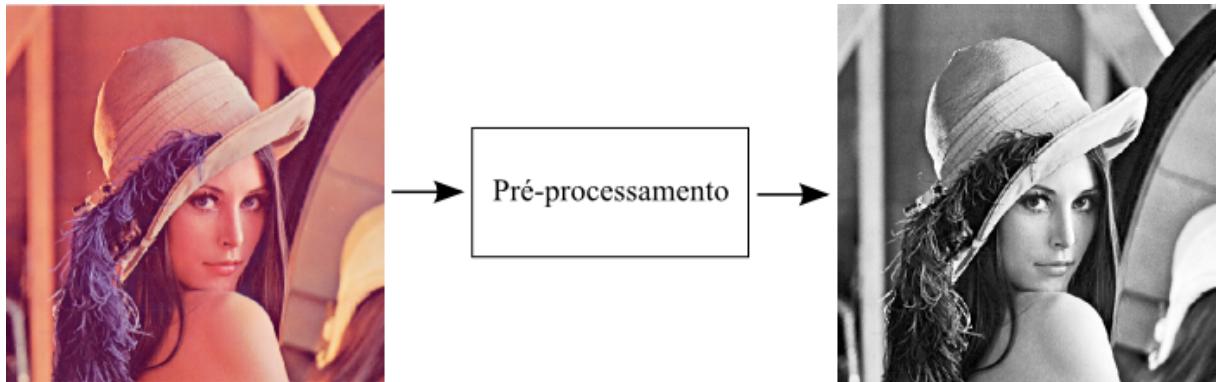


Figura 3 – O pré-processamento de imagens é caracterizado por receber uma imagem como entrada e fornecer uma imagem de saída. Sobre a imagem RGB de entrada (à esquerda) foram realizadas operações de borramento, realce e de equalização de histograma. A imagem à direita é resultante dessas operações.  
Fonte: Elaborado pela autora.

método indicado para evitar a amplificação de ruído nessas imagens é o algoritmo iterativo Richardson-Lucy, que será apresentado na Seção 2.2.2.

Em contrapartida, Ahonen, Hadid e Pietikäinen (2006) propuseram uma representação para imagens faciais baseada em características de textura, sem utilizar pré-processamento. Este aparece somente como sugestão de trabalho futuro, como possível correção de problemas do sistema de captura (i.e. suavização causada pela captura fora de foco). O que implica que, apesar dos bons resultados, a melhoria com a utilização de pré-processamento não foi investigada. Pode-se imaginar, portanto, que o uso de pré-processamento pode melhorar os resultados já obtidos, através do realce de textura e eliminação de imperfeições nas imagens.

Como exemplo do uso de métodos de pré-processamento, considere imagens de algas verdes obtidas por um microscópio de alta resolução. Essas algas são mergulhadas em um líquido que normalmente causa problemas de ruído e pouco contraste. Para a preparação dessas imagens, antes da extração de características, Borges *et al.* (2013) cita algumas etapas comuns em processamento de imagens digitais:

- As imagens – originalmente em RGB – são convertidas para uma escala de cinza;
- A dimensão da imagem é reduzida para diminuir o tempo de execução dos passos subsequentes de processamento;
- O contraste é “ajustado”, para aumentar a diferença das intensidades dos pixels da imagem e corrigir o brilho;
- A imagem é filtrada, removendo ruídos causados pelo processo de captura;
- O contorno é realçado, pois a forma é uma das características mais importantes para discriminar algas (e outros objetos);

- Por fim, o histograma é equalizado.

### 2.2.1 *Filtragem espacial e convolução*

Um filtro espacial, também conhecido como *kernel*, máscara ou janela, consiste em uma matriz de vizinhanças e uma operação a ser realizada nos pixels de uma imagem. A filtragem cria um novo pixel com as mesmas coordenadas do centro da vizinhança, contendo o valor resultante da filtragem. Dessa forma, a imagem filtrada contém os pixels resultantes da passagem do centro do filtro espacial por todos os pixels da imagem original (GONZALEZ; WOODS, 2007).

O processo de percorrer a imagem com um filtro espacial é chamado de correlação. A convolução, que pode ser definida como o operador  $*$  na Equação (2.1), é o mesmo processo, mas com o filtro previamente rotacionado em  $180^\circ$  (GONZALEZ; WOODS, 2007).

$$\text{Mapa de características} = \text{imagem de entrada} * \text{filtro} \quad (2.1)$$

Os métodos de filtragem possuem como objetivo aperfeiçoar certos aspectos da imagem de entrada. Essa filtragem pode ser realizada no domínio da frequência ou no domínio espacial. Um filtro de suavização típico no domínio espacial é o de Gaussiana, que resulta no borramento e redução de ruído, a fim de remover detalhes da imagem (Figura 4). Esse filtro utiliza uma função Gaussiana para calcular a transformação a ser realizada a cada pixel. A equação que representa a função Gaussiana em duas dimensões é definida por

$$G_\sigma(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}},$$

onde  $x, y$  são coordenadas de um determinado pixel da imagem e  $\sigma$  o desvio padrão que determina o raio da distribuição Gaussiana aplicada. Valores altos de variância fazem com que o resultado da função se aproxime da média.

### 2.2.2 *Deconvolução*

O processo de convolução, descrito na seção anterior, também pode ser definido como a passagem de uma imagem por um processo de aquisição que atue como um filtro de passa-baixas, resultando em uma imagem borrada. A deconvolução realiza a inversão desse borramento, o que pode facilitar a segmentação e detecção de características (PONTI, 2012). Para tal, é necessário conhecer previamente ou estimar a função que causou a degradação na imagem, geralmente denotada por  $h(\mathbf{u})$ , onde  $\mathbf{u}$  representa as coordenadas  $(x, y, z)$  para um sinal tridimensional.

Em casos mais simples é possível utilizar o filtro pseudo-inverso, obtendo o filtro a partir da transformada de Fourier da função  $h$ , denotada por  $H$  em

$$W(\mathbf{u}) = \begin{cases} H(\mathbf{u}), & H(\mathbf{u}) > \gamma \\ \gamma, & \text{caso contrário,} \end{cases}$$



Figura 4 – Exemplo de filtragem gaussiana como operação de pré-processamento. *Fonte: Elaborado pela autora.*

onde o limiar  $\gamma$  utilizado é em geral um valor entre 0,0001 e 0,1. O filtro  $W$  é utilizado para realizar a inversão, no domínio da frequências, da imagem  $g$  borrada, obtendo a imagem restaurada a partir de

$$\hat{F}(\mathbf{u}) = \frac{G(\mathbf{u})}{W(\mathbf{u})},$$

onde  $\hat{F}$  é a transformada de Fourier da imagem restaurada,  $G$  é a transformada de Fourier da imagem borrada, e  $W$  é o filtro pseudo-inverso, que realiza a deconvolução.

Um outro exemplo de algoritmo de deconvolução é o Richardson-Lucy ([PONTI et al., 2011](#)), que utiliza um processo iterativo para recuperar uma imagem degradada que foi borrada por algum processo conhecido. Utiliza uma metodologia probabilística, baseada em EM-ML (*Expectation-Maximization Maximum Likelihood*), para encontrar uma imagem que maximize a probabilidade de se visualizar a imagem original sem degradação, considerando um modelo de ruído de Poisson. O algoritmo é descrito na Equação 2.2, onde  $n$  é o número da iteração.

$$\hat{f}_{n+1}(\mathbf{u}) = \left[ \left( \frac{g(\mathbf{u})}{\hat{f}_n(\mathbf{u}) * h(\mathbf{u})} \right) * h(\mathbf{u}) \right] \times \hat{f}_n(\mathbf{u}). \quad (2.2)$$

Algoritmos iterativos como o Richardson-Lucy tem a vantagem de permitir soluções parciais, evitando amplificação de ruído.

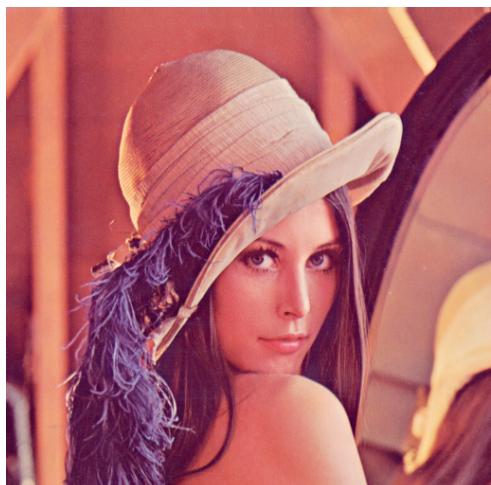
### 2.2.3 Realce de imagens

O realce de imagens é o processo de modificar uma imagem para que se torne mais adequada para uma aplicação específica do que na sua forma original. É subjetivo, porque

depende do sujeito que está analisando a imagem dissennir a qualidade desse realce (GONZALEZ; WOODS, 2007).

Na Figura 5 está demonstrado o efeito do algoritmo de *unsharp masking*, utilizando como borramento um filtro de média. Com o objetivo de realçar imagens, os passos deste método são:

1. Borramento da imagem original;
2. Cálculo da diferença entre a imagem suavizada e a original;
3. Soma dessa diferença à imagem original.



(a) Original



(b) *Unsharp masking*

Figura 5 – A imagem original, já em escala de cinza, foi realçada utilizando o método *unsharp masking*. Fonte: Elaborado pela autora.

Um exemplo clássico de utilização de realce, é para compensar a variação de iluminação em diversas imagens. Em Gross e Brajovic (2003), os autores propuseram um algoritmo para o reconhecimento de faces invariante à iluminação. Eles ressaltam que, desconsiderando a variação da posição, a iluminação é o fator de maior impacto na aparência das faces. A luz varia durante o dia, entre um dia e outro e entre diferentes ambientes. Isso afeta o conjunto de imagens a ser analisado, que passa a conter imagens com diferentes contrastes, o que pode acentuar ou diminuir certas características faciais.

O contraste é a diferença de intensidade entre os níveis de maior e menor intensidade na imagem. Imagens com baixa resolução podem ser geradas a partir de uma iluminação pobre ou outros problemas com a captura. Dessa forma, o processo de “esticar” o contraste expande os níveis de intensidade da imagem (GONZALEZ; WOODS, 2007).

**Explicar melhor histograma e k**

É possível aumentar o contraste de uma imagem ao manipular o seu histograma  $h$ , que pode ser definido como

$$h(i_k) = n_k,$$

onde  $n_k$  é o número de pixels de intensidade  $i_k$ . Ao observar os histogramas de diferentes imagens é possível notar que imagens com alto contraste possuem um histograma com níveis próximos a uma distribuição uniforme. Isso permite que certas operações, como a equalização de histograma, obtenham o melhor contraste de uma imagem. Essa operação é caracterizada pela máxima variância do histograma e pode ser definida como

$$s_k = T(i_k) = \frac{L-1}{MN} \sum_{j=0}^k n_j, \quad (2.3)$$

onde  $L$  é o número de intensidades e  $MN$  as dimensões da imagem. A imagem de saída é obtida ao mapear cada pixel de intensidade  $i_k$  em um nível  $s_k$ , com  $i$  entre  $[0, L-1]$ , sendo  $i = 0$  um pixel preto e  $i = L - 1$ , branco (GONZALEZ; WOODS, 2007).

## 2.2.4 Restauração

**Essa seção é necessária?**

Ao contrário do realce, a etapa da restauração é objetiva. Também procura melhorar os aspectos visuais de uma imagem, mas com base em modelos probabilísticos de degradação de imagens. Dessa forma, tendo-se um modelo de degradação, procura-se recuperar a imagem original (GONZALEZ; WOODS, 2007). Mas um problema comum em algoritmos de remoção de ruído é que alguns detalhes, como textura, irão sofrer alta suavização.

O objetivo dos algoritmos de remoção de ruído é recuperar a imagem original. Assim o modelo estatístico da imagem natural é crucial para a remoção de ruído. O tipo mais simples de ruído é o aditivo, que pode ser caracterizado como

$$p = p_0 + n,$$

onde  $p_0$  é o valor real do pixel e  $n$  é a perturbação de ruído naquele pixel.

Um algoritmo alternativo para remoção de ruído é o de médias não locais. Isso porque pixels similares não necessariamente estão próximos em suas coordenadas cartesianas, assim, informações não locais podem ser utilizadas na redução de ruído. De acordo com Buades, Coll e Morel (2005), dada uma imagem ruidosa  $v = v(i)|i \in I$ , o valor estimado é calculado como uma média ponderada de todos os pixels em uma mesma imagem com

$$v(i) = \sum_{j \in I} w(i, j)v(j),$$

onde os pesos  $w(i, j)$  dependem da similaridade entre pixels.

## 2.2.5 Quantização

Muitos métodos de extração de características são preparados para receber imagens de entrada em escala de cinza. Isso ocorre porque a complexidade de lidar com um pixel representado em três dimensões é muito maior do que em apenas uma dimensão. Assim, os métodos de quantização visam, de alguma forma, reduzir os canais de cores (24 bits) em apenas um (8 bits). Kanan e Cottrell (2012) demonstraram que os métodos para a quantização (conversão de uma imagem colorida para escala de cinza) influenciam a performance no reconhecimento de imagens. Eles também salientam que o método utilizado deveria estar claramente descrito nas publicações da área. Os métodos de conversão para escala de cinza utilizados nessa dissertação foram escolhidos com base em Kanan e Cottrell (2012):

**Intensidade:** método mais simples, consiste em computar a média entre os canais RGB da imagem a partir de

$$Q_{\text{Intensidade}} = \frac{1}{3}(R + G + B)$$

e então realizar a correção por *gamma*.

**Gleam:** ao corrigir por *gamma* cada canal antes de realizar a combinação linear, tem-se o método de quantização *Gleam*:

$$Q_{\text{Gleam}} = \frac{1}{3}(R' + G' + B'),$$

onde  $R'$ ,  $G'$  e  $B'$  são os canais R, G e B corrigidos por *gamma*.

**Luminância:** computa uma soma ponderada dos canais de cor. Esse método foi desenvolido para levar em conta a percepção visual humana. O olho humano percebe verde melhor que vermelho, e vermelho melhor que azul:

$$Q_{\text{Gleam}} = 0.299R + 0.587G + 0.114B.$$

e então realizar a correção por *gamma*.

**Luma:** similar ao anterior, utilizado nas televisões de alta definição.

[achar referencia](#)

$$Q_{\text{Luma}} = 0.2126R' + 0.7152G' + 0.0722B'.$$

**Bits Mais Significativos (MSB):** ao invés de realizar uma combinação linear dos canais de cores, ordena os bits dos canais coloridos em um único canal. Computa quantos bits de cada canal de cor contribuem para a imagem final e extrai os bits do código binário dos canais originais (PONTI; ESCOBAR, 2013).

A Figura 6 apresenta a conversão na escala de cinza obtida com o uso destes métodos.

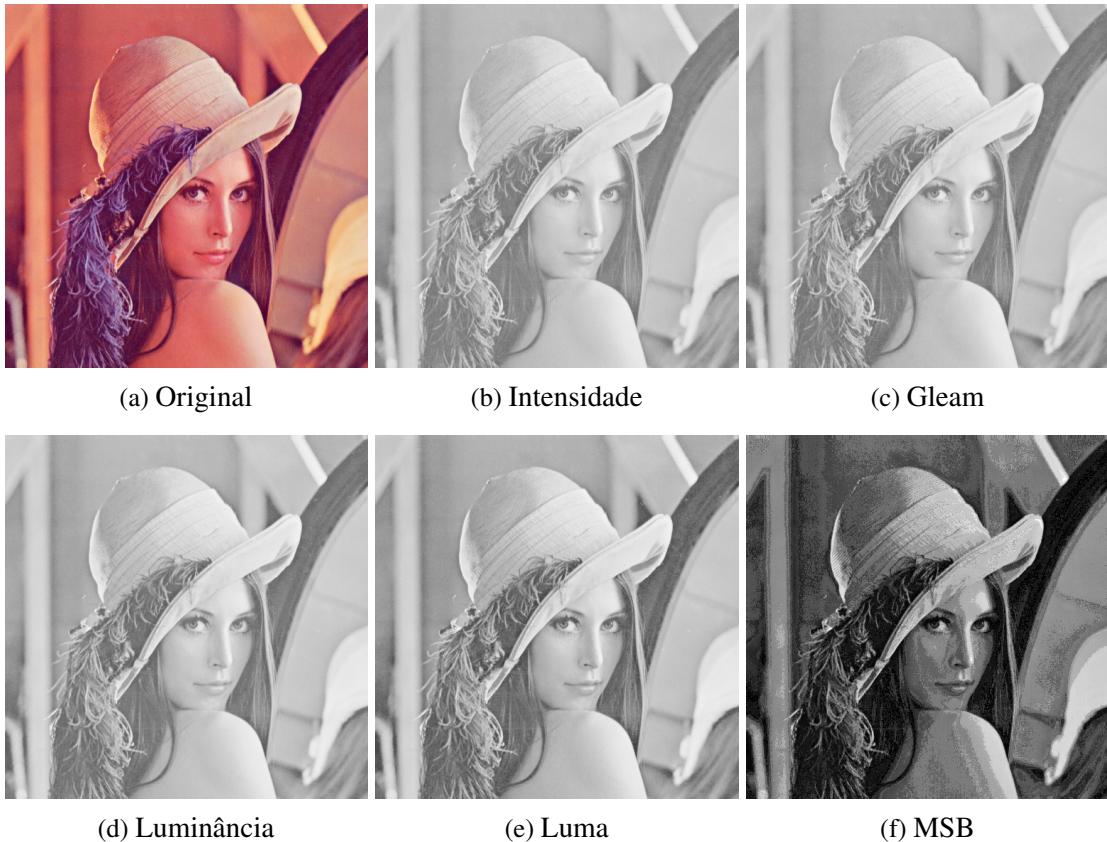


Figura 6 – Conversão para a escala de cinza com os métodos de Intensidade e MSB. *Fonte: Elaborado pela autora.*

## 2.3 Extração de características

O objetivo da extração de características é descrever as informações visuais relevantes em um vetor de características. Esse vetor pode ser utilizado como entrada para um algoritmo de classificação de padrões. Em aplicações que envolvem a classificação de algas, uma informação muito importante para a discriminação entre classes é a forma ([BORGES et al., 2013](#)). Essas características devem salientar as diferenças entre imagens de classes distintas e suavizar possíveis diferenças de imagens da mesma classe. Algumas características, segundo [Gonzalez e Woods \(2007\)](#), são:

**Textura:** na sua descrição estatística, possui propriedades como: suavidade, aspereza e uniformidade. Um exemplo de medida para descrever a textura é a entropia.

**Forma:** representa os objetos em termos de suas características externas, como, por exemplo, a medida da curvatura.

**Cor:** considera a distribuição espacial de cores na imagem. O histograma de uma imagem pode descrever essa configuração de forma global.

Exemplos de métodos conhecidos capazes de descrever forma e outras características são: histogramas de orientação de gradiente (WANG; HAN; YAN, 2009), curvatura, descritores de Fourier, métodos baseados na detecção de SUSAN (SMITH; BRADY, 1997), Harris-Affine (HARRIS; STEPHENS, 1998) e diferença de Gaussianas (LOWE, 2004). Os descritores utilizados no desenvolvimento desta pesquisa para a obtenção dos resultados da Seção 5 estão abaixo descritos.

**Histograma global de cor (GCH):** calcula o histograma global dos níveis de intensidade da imagem. É a alternativa mais simples para representar as informações de uma imagem (GONZALEZ; WOODS, 2007). Produz um vetor de  $N$  dimensões, sendo  $N$  o número de intensidades.

**Vetor de coerência de cor (CCV):** captura informações sobre como as cores são organizadas em regiões conectadas, de acordo com um *threshold*. Classifica os pixels da imagem em coerentes e incoerentes, computa os respectivos histogramas e os concatena (HUANG *et al.*, 1997). Dessa forma, o vetor de características produzido possui  $2N$  dimensões.

O vetor de coerência de cor armazena o número de pixels coerentes e de incoerentes para cada cor. Pixels coerentes fazem parte de uma região contígua enquanto incoerentes não. Seu funcionamento pode ser resumido nos seguintes passos:

1. Borra levemente a imagem ao substituir os pixels pela média do pixel e seus oito vizinhança;
2. Discretiza o espaço de cor de forma que a imagem contenha apenas  $n$  cores distintas;
3. Classifica os pixels entre coerentes e incoerentes: se o tamanho do seu componente conectado for maior ou igual um dado *threshold* é *coerente*; caso contrário, *incoerente*;
4. Computa dois histogramas:
  - Histograma de pixels coerentes;
  - Histograma de pixels incoerentes.
5. Concatena os histogramas.

**Classificação de pixels de borda e interior (BIC):** computa dois histogramas, um para pixels definidos como borda e outro como interior. Se um pixel possuir a mesma cor que seus vizinhos, é pixel de interior; caso contrário, será pixel de borda. Os histogramas são concatenados, gerando um vetor de  $2N$  dimensões (STEHLING; NASCIMENTO; FALCÃO, 2002). Para computar tal vetor, as operações realizadas são:

1. Os pixels são classificados entre borda e interior:

- *Borda*: se está na borda da imagem ou se ao menos um dos seus quatro vizinhos tem uma cor diferente do que o próprio pixel;
  - *Interior*: se seus quatro vizinhos possuem a mesma cor.
2. Computa dois histogramas:
    - Histograma dos pixels classificados como borda;
    - Histograma dos pixels classificados como interior.
  3. Concatena os histogramas.

**Auto-correlograma de cor (ACC):** captura a correlação espacial entre cores idênticas. Para tal, computa a probabilidade de encontrar dois pixels com a mesma cor, a uma distância  $d$  um do outro (PASS; ZABIH; MILLER, 1996). O vetor resultante consiste na concatenação dos auto-correlogramas, um para cada distância. Neste estudo, são consideradas quatro distâncias: 1, 3, 5 e 7; resultando em um vetor com  $4N$  características.

**Haralick-6:** descreve a textura das imagens, ou seja, diferenças locais em níveis de intensidade (HARALICK; SHANMUGAM; DINSTEIN, 1973). O vetor resultante possui seis dimensões que representam as seguintes características:

1. *Probabilidade máxima*: maior resposta na matriz de co-ocorrência. Intervalo:  $[0, 1]$ ;
2. *Correlação*: descreve as correlações entre as linhas e colunas da matrix. Intervalo:  $[-1, 1]$ ;
3. *Contraste*: mede as variações locais dos níveis de cinza da matriz. Intervalo:  $[0, (colors - 1)^2]$ ;
4. *Uniformidade*: soma dos elementos quadrados. Também conhecido como energia ou segundo momento angular. Intervalo:  $[0, 1]$ ;
5. *Homogeneidade*: mede a proximidade da distribuição dos elementos em relação à diagonal. Intervalo:  $[0, 1]$ ;
6. *Entropia*: descreve a aleatoriedade. Intervalo:  $[0, 2 * \log_2 colors]$ .

**Histograma de Gradientes Orientados (HOG):** (DALAL; TRIGGS, 2005).

1. Divide a janela da imagem em células;
2. Computa os gradientes;
3. Cada pixel calcula uma ponderação para um canal do histograma de orientação de bordas baseado na orientação do gradiente do elemento em que esta centrado. Esses valores são acumulados em bins sobre as regiões espaciais de células e formam o histograma;
4. Os votos são interpolados bilinearmente entre os centros de vizinhança do bin em orientação e posição;

5. Normaliza o contraste dos blocos da janela que se sobrepõem. Dessa forma cada célula é normalizada em relação a diferentes blocos;
6. Concatena os histogramas de todas as células.

**LBP utilizando padrões uniformes (LBP):** baseia-se em reconhecer que padrões de textura uniformes são propriedades fundamentais da textura local da imagem e o histograma da sua ocorrência provou-se um bom extrator de características. Computa um histograma de ocorrência dos padrões locais binários em uma vizinhança da imagem, detectando micro-estruturas cuja distribuição é estimada pelo histograma ([OJALA; PIETIKAINEN; MAENPAA, 2002](#)).

1. Divide a janela da imagem em células;
2. Para cada pixel em uma célula, compara com seus vizinhos. Esse passo resulta em um código binário de oito dígitos;
3. Computa o histograma da célula;
4. Normaliza o histograma;
5. Concatena os histogramas de todas as células.

## 2.4 Características latentes

[Deixar essa seção?](#)

Assim como o método de quantização se mostrou relevante na classificação, nesta dissertação são estudados diversos métodos de pré-processamento de imagens, como filtros de borramento e deconvolução, com o objetivo de obter imagens processadas que sejam mais bem caracterizadas para a etapa de classificação (ou seja, aumentando a variância entre as classes, sem aumentar a variância intra-classe). O enfoque está em como realçar determinadas características, como por exemplo cor, textura e forma. A esses atributos pode-se dar o nome de características latentes, não visíveis na imagem original. Identificar e realçar tais características é objetivo deste estudo. Essa abordagem se diferencia das técnicas multi-resolução, pois não pretende encontrar apenas características em versões de diferentes resoluções (convoluídas com filtros passa-baixa), e sim também em versões deconvoluídas ou transformadas por outros operadores.

Se uma das principais características que diferenciam classes de uma certa base utiliza a sua forma, é possível utilizar o método de diferença de Gaussianas para ressaltá-la. O método DoG (*Difference of Gaussians*) se baseia na diferença entre duas imagens filtradas. A operação comumente é feita usando  $\sigma = \sqrt{2}$  e sua definição para dois níveis de filtragem é definida por

$$\begin{aligned} f_1(x, y) &= G_\sigma(f(x, y)) \\ f_2(x, y) &= G_{\sigma\sqrt{2}}(f_1(x, y)) \\ DoG_1(x, y, \sigma) &= f_2 - f_1 \end{aligned}$$

A Figura 7 demonstra uma possível execução de filtragem e restauração seguida pela detecção de bordas por DoG, aplicadas em uma base de imagens de algas verdes.

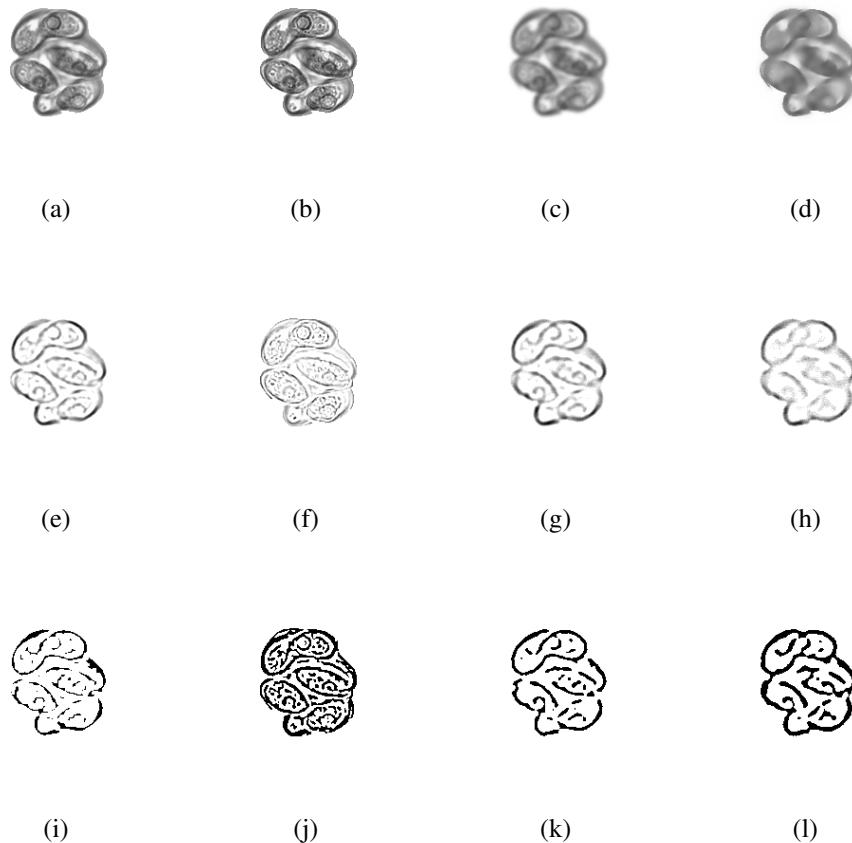


Figura 7 – Características latentes de algas verdes. A primeira imagem (a) é uma imagem original segmentada de alga. As próximas colunas são imagens resultantes da deconvolução da imagem (coluna 2), filtragem Gaussiana (coluna 3) e filtragem Gaussiana seletiva (coluna 4). A primeira linha mostra versões diferentes de imagens de algas, a segunda linha exibe imagens resultantes da diferença de Gaussianas, e a terceira linha demonstra imagens binárias obtidas por limiarização das imagens contidas na segunda linha).  
Fonte: Elaborado pela autora.

A imagem 7a é uma imagem original segmentada de alga. As próximas colunas são imagens resultantes da deconvolução da imagem (coluna 2), filtragem Gaussiana (coluna 3) e filtragem Gaussiana seletiva (coluna 4). As modificações apresentadas na primeira linha geram informações em diferentes planos axiais da imagem de microscopia. Imagens na segunda coluna representam uma versão deconvoluída das imagens originais, ressaltando a textura na superfície das algas, resultantes da diferença de Gaussianas. Já a terceira linha demonstra imagens binárias obtidas por limiarização das imagens da segunda linha, realçando as características da base da célula.

Acredita-se que ao manipular as imagens através de técnicas de pré-processamento, pode-se gerar imagens artificiais com características latentes realçadas, melhorando a extração de características.

## 2.5 Desbalanceamento de classes

Nesta seção é definido o problema do desbalanceamento de classes e apresentados os trabalhos relacionados que possuem duas diferentes abordagens: sobreamostragem (*over-sampling*) e subamostragem (*under-sampling*).

Em conjuntos de dados desbalanceados, determinadas classes possuem um número muito maior de instâncias do que outras. As classes com mais elementos são chamadas de classes majoritárias, e as com menos elementos, de minoritárias. O desempenho de algoritmos de Aprendizado de Máquina é prejudicado quando tratam de bases de dados desbalanceados. Esses algoritmos tendem a favorecer a classificação de um novo objeto à classe majoritária, pois esta fica muito melhor representada após o treinamento do que a minoritária. Considera-se, então, que esse problema é um obstáculo para a classificação satisfatória. Porém, como apontado por (BATISTA; PRATI; MONARD, 2004), o desbalanceamento não é o único responsável por reduzir o desempenho de algoritmos de aprendizagem. Eles sugerem que é possível haver uma ótima classificação mesmo contendo alto desbalanceamento na base de dados. Assim, a motivação do estudo de vários algoritmos para rebalanceamento não é apenas balancear os dados de treinamento, mas obter uma melhor diferenciação entre as classes. Isso porque o desbalanceamento por si só não é um problema, mas em conjunto com a sobreposição de classes pode diminuir significativamente a acurácia da classificação da classe minoritária. Os resultados reportados também mostram que a poda de árvores de decisão raramente levou a alguma melhora na classificação.

(CASTRO; BRAGA, 2011) destacam que duas abordagens têm sido utilizadas para solucionar esse problema: pré-processar os dados de forma a rebalancear as distribuições das classes ao reamostrar os dados; ou então modificar métodos de aprendizado – como através da adição de melhores funções de custo na classificação. Em geral, são reportados melhores resultados obtidos por algoritmos de *over-sampling*, os quais consistem em reamostrar os dados aumentando o número de elementos da classe minoritária (BATISTA; PRATI; MONARD, 2004). Esta pesquisa tem como enfoque o **pré-processamento dos dados**, ao rebalancear as classes através da **geração de imagens artificiais**.

### 2.5.1 Sobreamostragem

Realizar uma sobreamostragem (*over-sampling*) em um determinado conjunto de dados significa aumentar – utilizando alguma estratégia – o número de elementos desse conjunto. Em (CHAWLA; HALL; BOWYER, 2002), a simples replicação de exemplos pertencentes à classe minoritária não melhorou a classificação. Isso se deve ao reconhecimento de regiões muito específicas, causando *overfitting*.

O SMOTE (*Synthetic Minority Over-sampling Technique*) é um método desenvolvido por (CHAWLA; HALL; BOWYER, 2002) para rebalancear classes ao gerar artificialmente novos

elementos, ao invés de apenas replicá-los. É aplicado sobre os vetores de características previamente extraídos, com operações de perturbação dos dados de treino no espaço de características, e não no espaço dos dados. A diferença entre o vetor de características de um elemento e do seu vizinho mais próximo é multiplicada por um número  $0 \leq x \leq 1$ . Esse valor é adicionado ao vetor original, criando um novo elemento. Como pode ser visualizado na Figura ??, essa abordagem provoca a geração de um elemento resultante da interpolação dos dois vetores originais. Os exemplos sintéticos forçam uma região de decisão maior e mais geral para serem aprendidas como exemplos da classe minoritária.

**fazer a imagem do smote**

---

#### Algoritmo 1: Algoritmo do SMOTE

---

**Entrada:** Imagem colorida  $I$  em formato BGR

**Saída:** Exemplos sintéticos

```

1  $N \leftarrow$  vizinhos(minoritária);
2 para cada exemplo  $sample$  faca
3    $nn \leftarrow$  vizinho aleatório de  $N$  para cada característica  $(x,y)$  faca
4     diferença  $\leftarrow nn(x,y) - sample(x,y);$ 
5     gap  $\leftarrow$  número aleatório entre 0 e 1;
6      $G(x,y) \leftarrow sample(x,y) + gap * diferença;$ 
7   fim
8    $S \leftarrow SUG;$ 
9 fim

```

---

Dessa forma, o SMOTE provê mais elementos para o classificador aprender, ao contrário da replicação de dados. Como trabalhos futuros, os autores Chawla, Hall e Bowyer (2002) apontam que diferentes estratégias para criar esses exemplos sintéticos podem melhorar a performance da classificação. Inclusive salientando exemplos que foram errôneamente classificados.

Uma variação desse algoritmo, denominada Borderline-SMOTE1 (HAN; WANG; MAO, 2005), considera que elementos fora da linha de borda de cada classe pouco contribuem para a classificação. Por isso, propõe a geração de elementos sintéticos utilizando apenas elementos de borda. Considera que se os vizinhos mais próximos são da classe majoritária, o exemplo é ruído, e se há mais vizinhos da classe majoritária do que da minoritária, considera esse elemento como sendo de borda. Como trabalho futuro, destacam a necessidade de considerar diferentes estratégias para definir em quais elementos realizar o over-sampling.

### 2.5.2 Subamostragem

Ao contrário da sobreamostragem, a subamostragem visa diminuir o número de elementos de um determinado conjunto. A ideia é eliminar elementos da classe majoritária que estão distantes da fronteira de decisão, isso porque eles são considerados menos relevantes para a

aprendizagem.

Métodos para remoção de exemplos da classe majoritária normalmente apresentam resultados piores do que métodos de sobreamostragem, conforme relatado por [Batista, Prati e Monard \(2004\)](#) e ([JAPKOWICZ; STEPHEN, 2002](#)). Um dos motivos pela preferência natural à sobreamostragem é o fato de que ao realizar subamostragem pode-se remover informações essenciais dos dados originais. Mas não há uma estratégia única que funcione melhor para todos os cenários.

## 2.6 Classificadores de padrões

### 2.6.1 Algoritmo *k*-Vizinhos Mais Próximos

K-NN é um classificador supervisionado que considera a proximidade entre os dados para realizar previsões. Baseia-se na premissa de que os objetos do mesmo conceito são semelhantes. Na fase de treinamento, apenas armazena os exemplos rotulados do conjunto de dados de treinamento. Quando um novo exemplo deve ser classificado, calcula a distância entre os vetores de características do novo exemplo e aqueles já rotulados. O novo exemplo é então classificado como sendo da classe do exemplo de treinamento com menor distância ([BOIMAN; SHECHTMAN; IRANI, 2008](#)).

1. Escolhe o valor de  $k$  e uma métrica de distância;
2. Encontra os  $k$  vizinhos mais próximos do exemplo que queremos classificar;
3. Rotula o sample como pertencente à classe que recebeu mais votos.

**figura aqui**

Esse classificador foi utilizado por ser suscetível à diferenciação entre as classes. Com  $K = 1$ , a predição da classe corresponde ao exemplo mais próximo.

## 2.7 Redução de dimensionalidade

A visualização do espaço de características obtido após a geração artificial de imagens pode ajudar a verificar se as novas imagens melhoram a definição da classe minoritária em relação ao espaço original (inclusive antes de imagens serem removidas para provocar o desbalanceamento). Ou seja, se o método utilizado revelou características latentes. Dessa forma, ao projetar os novos vetores no espaço das imagens originais, é possível analisar qual método (entre SMOTE ou geração de imagens no campo visual) mais se assemelha à distribuição original dos dados.

Considerando que um vetor de características extraído com extratores comuns pode possuir entre 6 (e.g. Haralick) e 512 (e.g. BIC) características, a visualização de um exemplo requer que seja realizado o mapeamento desses valores em apenas duas dimensões. Para isso, uma redução de dimensionalidade mapeia os dados de  $N$  dimensões para um espaço 1D, 2D ou 3D. A partir desses novos dados, pode então ser criada alguma representação visual que tente manter a relação de distância entre os novos e os originais (PAULOVICH; OLIVEIRA; MINGHIM, 2007).

### 2.7.1 Análise de componentes principais

O PCA (Análise de Componentes Principais) é uma técnica não supervisionada que pode ser utilizada para reduzir a dimensionalidade dos dados com a máxima variância possível. Cada imagem, originalmente representada por um vetor com  $N$  características, pode então ser representada por apenas um ou mais valores. Essa redução permite a projeção dessas imagens no espaço de características. O objetivo é extrair as informações mais importantes dos dados e representá-las como um conjunto de variáveis ortogonais chamadas de componentes principais. Para isso encontra-se uma outra base: uma combinação linear da base original, que melhor representa os dados ao assumir que as direções das maiores variâncias são as mais importantes. Ou seja, a variância associada com cada direção quantifica o quanto principal é aquela direção (ABDI; WILLIAMS, 2010). Pode-se, portanto, enumerar os passos necessários para o PCA sendo:

1. Centraliza todos os atributos em zero ao subtrair a média de cada dimensão;
2. Calcula a matriz de covariância  $C_x$  dada por

$$C_x = XX^T, \quad (2.4)$$

onde  $X$  é a matriz de dados original e  $X^T$  sua transposta;

3. Encontra os autovalores e autovetores de  $C_x$ . Um autovetor  $u$  de uma matriz  $A$  pode ser definido por  $Au = \lambda u$ , onde  $\lambda$  é um autovalor escalar associado ao autovetor. Um vetor  $u$  é um autovetor da matriz  $A$  se o tamanho do vetor – e não sua direção – é modificado quando multiplicado por  $A$ . Os autovalores podem ser representados na diagonal de uma matriz  $\lambda$  (com outros valores como zero) e o conjunto dos autovetores de  $A$  em uma matriz  $U$ . Assim,

$$A = U\lambda U^T; \quad (2.5)$$

4. Então, os autovetores são ordenados de forma decrescente de acordo com seus autovalores correspondentes e escolhe-se os  $k$  principais autovetores (i.e. maiores autovalores) para

formar uma matriz  $P$  de dimensão  $n \times k$ , onde cada coluna representa um autovetor. O valor  $k$  será a quantidade de dimensões do novo espaço de atributos;

5. O novo subespaço pode ser encontrado multiplicando essa matriz  $P$  pela matriz original, de acordo com a equação  $Y = PX$ , onde  $X$  representa o conjunto de dados original,  $Y$  é uma nova representação desses dados e  $P$  a matriz ortonormal que transforma  $X$  em  $Y$ . As linhas de  $P$  são os componentes principais de  $X$ .

### 2.7.2 Locality Preserving Projections

LPP é um algoritmo linear de redução de dimensionalidade com propriedades de preservação da estrutura local dos dados. Não te a dificuldade dos algoritmos tradicionais (como PCA) de manter o manifold não linear dos dados originais ([ZHUO; CHENG; ZHANG, 2014](#)). Embora o método mais utilizado para redução da dimensionalidade de forma não-supervisionada seja o PCA, métodos como esse produzem melhores projeções em termos de separação das classes. Em ([ZHUO; CHENG; ZHANG, 2014](#)) o LPP alcançou a melhor relação entre complexidade computacional e a redução da dimensionalidade, enquanto manteve a acurácia. Seu algoritmo segue três passos principais ([HE; NIYOGI, 2004](#)):

1. Constrói um grafo de adjacências. Os nós  $i$  e  $j$  possuem uma aresta entre eles se fazem parte do conjunto de  $k$ -vizinhos mais próximos de cada nó (sendo  $k$  um parâmetro do algoritmo);
2. Encontra os pesos:  $W_{ij} = 1$  se os vértices  $i$  e  $j$  estão pertos, ou seja, conectados por uma aresta e  $W_{ij} = 0$  caso contrário;
3. Computa os autovalores e autovetores

$$XLX^T a = \lambda XDX^T a, \quad (2.6)$$

onde  $D$  é a matriz diagonal na qual a entrada são as somas das colunas de  $W$

A matriz de projeção corresponde aos autovetores, sendo os d menores autovalores.

## 2.8 Considerações finais

Deu-se destaque à discussão das etapas de pré-processamento e realce de características latentes, ambas foco deste estudo, assim como a geração de imagens artificiais para o balançamento de classes. Este capítulo apresentou diversos métodos para exemplificação, além de trabalhos similares.

A extração de características foi abordada, apresentando os principais descritores utilizados nesta pesquisa. A lacuna destacada é que existem características não passíveis de extração

por descriptores convencionais. A geração dessas imagens visa rebalancear classes que diferem em número de imagens, e detalhes sobre esse problema também foram fundamentados nesse capítulo.

Esses fundamentos permitem compreender em que contexto esta dissertação de mestrado está inserida. O próximo capítulo abrangerá a proposta deste trabalho.





# QUANTIZAÇÃO DE IMAGENS

---



---

## 3.1 Considerações iniciais

Sistemas de reconhecimento de imagens comumente utilizam uma imagem em níveis de cinza 8-bits com 256 intensidades para a extração de características. Ao aplicar a quantização (redução de cores) na etapa de pré-processamento, é esperada a redução da complexidade do vetor de características logo no início, beneficiando todos os passos subsequentes.

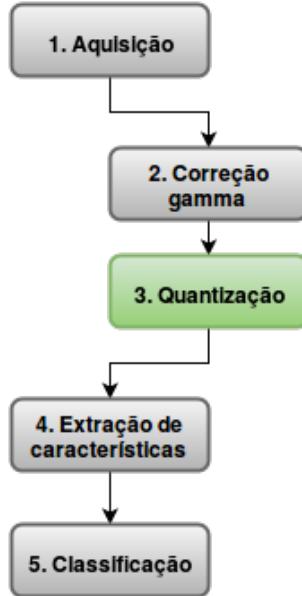
Para analisar o impacto da utilização da quantização, são utilizados diferentes parâmetros de quantização combinados com quatro métodos de extração de cor e um de textura. Esses extractores foram escolhidos de acordo com os resultados apresentados por [Penatti, Valle e Torres \(2012\)](#).

Este capítulo trata dessa ideia de quantizar as imagens antes da extração de características, ligando com os métodos apresentados nos fundamentos.

## 3.2 Quantização de imagens

O pipeline de reconhecimento de imagens envolve um passo de converter imagens coloridas em imagens com apenas um canal, obtendo uma imagem quantizada que pode ser então processada por métodos de extração de características. Dessa forma, cada imagem – originalmente no espaço de cor RGB – é convertida a um único canal com  $C$  níveis de intensidade. Após, são utilizados os métodos apresentados na Seção 2.3 para extrair as características. A Figura ?? ilustra esses passos, da aquisição até a classificação das imagens.

Cada método de quantização se comporta diferentemente para uma dada imagem RGB. Por exemplo, o método Intensidade mapeia todas as permutações dos mesmos valores em RGB para a mesma cor. Assim, produz um plano no cubo RGB conforme mostrado na Figura 8. O Gleam também é similar, mas spanning uma superfície curva, dada a natureza da função



gamma. O mesmo efeito é alcançado utilizando Intensidade'. Em todos os casos, o resultado é o mapeamento de características cromáticas bem diferentes em valores de intensidades similares. O método Luminância tenta melhorar isso ao adicionar um peso na combinação linear dos canais.

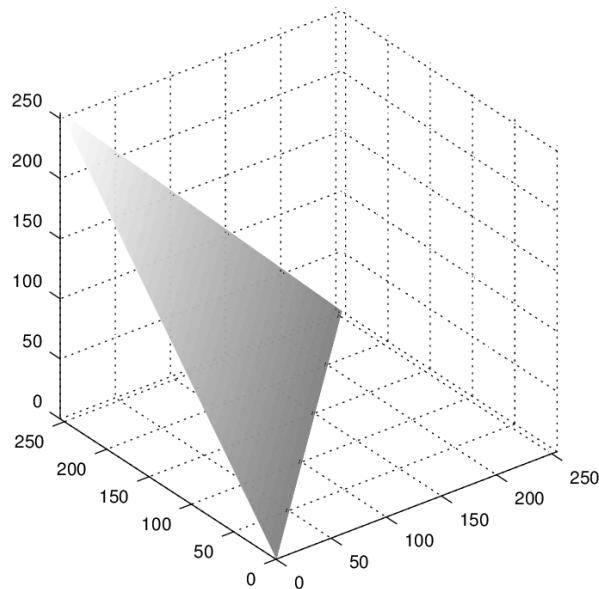


Figura 8 – Plano computado pelo método de conversão para escala de cinza Intensidade, quando um dos canais de cor possui valor 255. Fonte: ([PONTI; NAZARÉ; THUMÉ, 2016](#)).

Um exemplo das imagens obtidas após os métodos de quantização apresentados anteriormente pode ser visto na Figura 9. Nesse caso é possível notar que tanto Luminância quanto MSB conseguem melhor discriminar cores. Além disso, o mapa de cores MSB obteve um maior número de cores únicas. A barra de gradientes abaixo da imagem dos pinceis demonstra como os métodos de quantização se comportam dada a variação de cor.

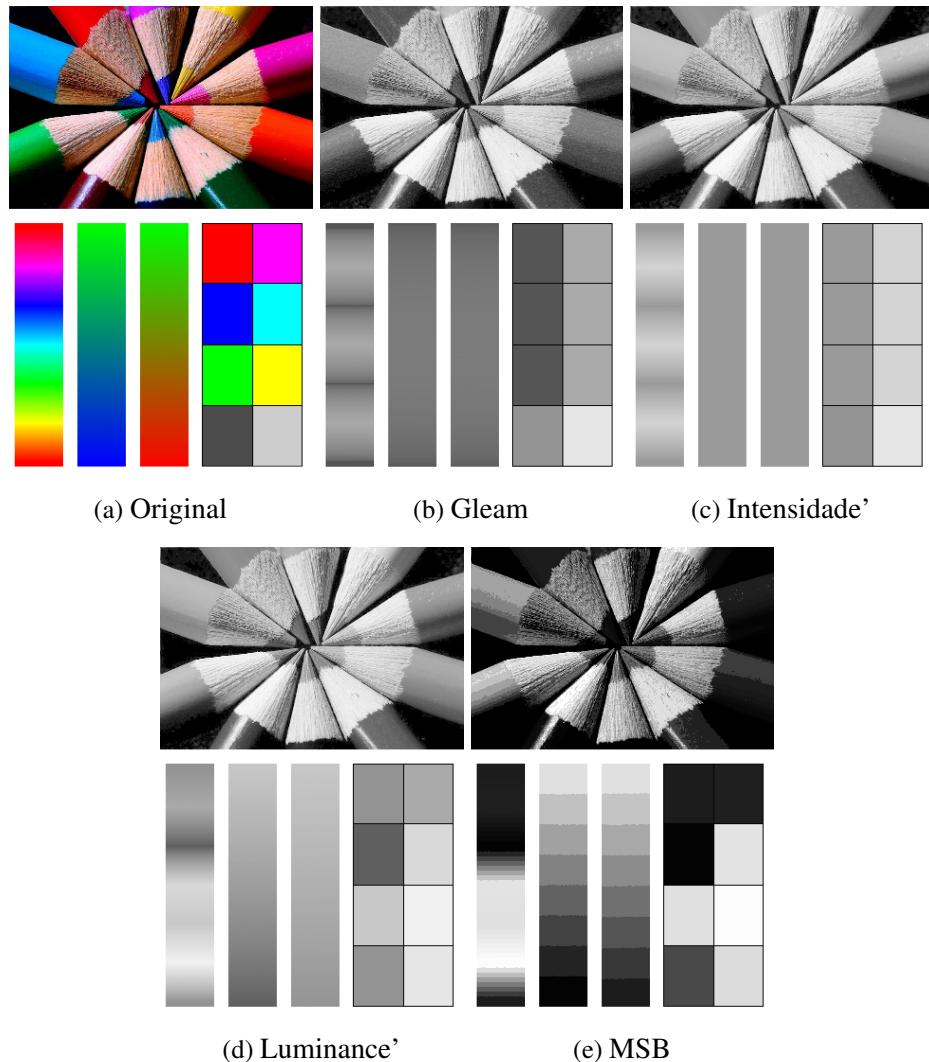


Figura 9 – Resultado da aplicação de métodos de quantização. A imagem original a resultou em versões de um canal de cor com 232 cores unicas para MSB e 184 cores para os restantes métodos. *Fonte: (PONTI; NAZARÉ; THUMÉ, 2016).*

Na Figura 10 é demonstrado um exemplo de redução de cores utilizando o método MSB para um par de imagens da base de dados Caltech-101. Importante destacar que há preservação das cores, especialmente entre 64 e 256. Ao utilizar apenas 32 cores, as imagens ainda são satisfatórias mas há uma perda de informação considerável.

### 3.3 Considerações finais

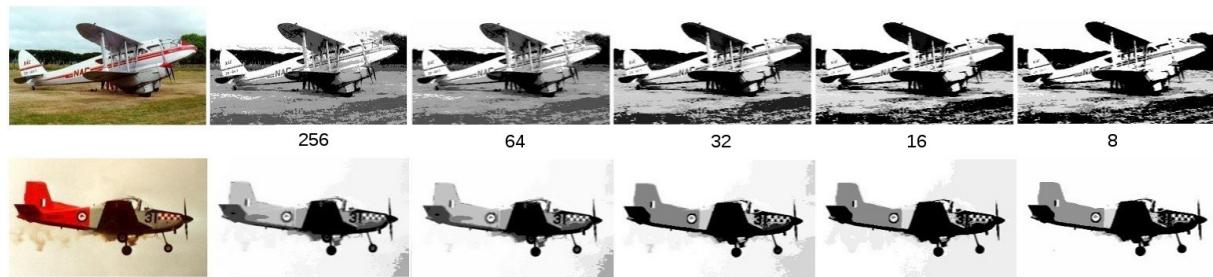


Figura 10 – Duas imagens da base de dados Caltech101 com variações no parâmetro de cor utilizando o método MSB. Da esquerda para a direita: imagem original 24-bits e na sequência suas versões quantizadas com: 256, 64, 32, 16 e 8 cores. *Fonte:* ([PONTI; NAZARÉ; THUMÉ, 2016](#)).



# GERAÇÃO ARTIFICIAL DE IMAGENS

---

## 4.1 Considerações iniciais

O processo de manipular imagens para que elas se tornem mais satisfatórias para um determinado objetivo depende do domínio de aplicação. Ou seja, não existe uma teoria geral para melhorar qualquer tipo de imagem (GONZALEZ; WOODS, 2007): um método que processa melhor uma imagem bem definida pelas suas cores difere do processamento de imagens texturizadas, às quais um processamento sobre a intensidade dos pixels da imagem – como uma operação de borramento – pode ocasionar perda da textura. Assim, justifica-se a exploração de um vasto número de métodos de processamento de imagens e bases.

Nesta pesquisa oito métodos de processamento de imagens são aplicados nas imagens minoritárias originais, gerando imagens artificiais. Isso é realizado a fim de permitir a extração de informações latentes com o objetivo de melhorar a classificação com alguma técnica de Aprendizado de Máquina, o que reflete a melhora da diferenciação entre as classes. Dada a quantidade de imagens necessárias para rebalancear a base original, são geradas imagens utilizando cada um dos métodos, além de uma versão combinando todos eles (ou seja, compondo um conjunto com algumas imagens processadas por cada método) e outra apenas replicando as imagens como *baseline*. Como demonstrado na Figura 11, dado o conjunto de treinamento da classe (ou classes) com menor número total de imagens, é realizado o rebalanceamento ao aplicar os métodos descritos neste capítulo e posteriormente essas imagens resultantes são utilizadas como treinamento.

Neste capítulo os métodos de geração artificial para o rebalanceamento de classes de imagens são descritos. Os experimentos posteriormente destacados no Capítulo 5 foram realizados utilizando: borramento; mistura ponderada; *unsharp masking*; composição; combinação de thresholds; combinação com saliência; visual SMOTE; e adição de ruído.

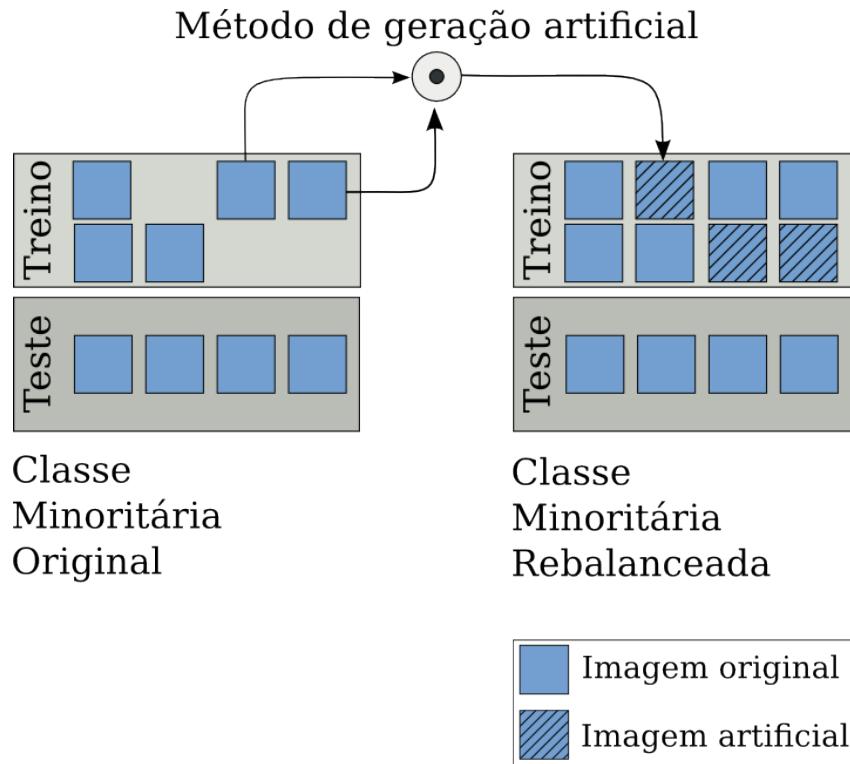


Figura 11 – Geração artificial a classe minoritária para rebalancear as classes. *Fonte: Elaborado pela autora.*

## 4.2 Borramento

Também conhecido como filtro de suavização, o borramento é uma operação de processamento comumente utilizada com o objetivo de filtrar as baixas frequências de uma imagem, removendo ruídos e detalhes não relevantes. Normalmente esse tipo de filtro provoca também um certo borramento das bordas, como pode ser observado na Figura 4.

Esse comportamento não é esperado quando devemos gerar novas imagens, pois informações relevantes podem ser removidas. Dessa forma, a operação de borramento utilizada é a de filtro bilateral. Ela substitui o valor do pixel  $(x,y)$  por uma média dos pixels de intensidade similar na imagem e dos pixels vizinhos (TOMASI; MANDUCHI, 1998). Ou seja, é uma média ponderada das intensidades. O Algoritmo 2 descreve os passos desse filtro na sua versão mais simples, de força bruta. A Figura 12 exemplifica o seu funcionamento: à esquerda está demonstrada a imagem original e à direita a imagem borrada.

**Algoritmo 2:** Algoritmo de borramento com filtro bilateral

---

**Entrada:** Imagem colorida  $I$  em formato BGR

**Entrada:** Diâmetro  $d$  de pixinhança de pixels

**Entrada:** Sigma do espaço de cor

**Entrada:** Sigma do espaço de coordenadas

**Saída:** Imagem gerada  $BF$

```

1 para cada pixel  $(x,y)$  faça
2    $BF(x,y) \leftarrow 0;$ 
3    $W(x,y) \leftarrow 0;$ 
4   para cada pixel  $(i,j)$  faça
5      $w \leftarrow G_{\sigma_s}(\|(x,y) - (i,j)\|)G_{\sigma_x}(|I(x,y) - I(i,j)|);$ 
6      $BF(x,y) \leftarrow BF(x,y) + wI(i,j);$ 
7      $W(x,y) \leftarrow W(x,y) + w;$ 
8   fim
9    $BF(x,y) \leftarrow I(x,y)/W(x,y);$ 
10 fim
11 
```

---

**Conferir algoritmo**



(a) Original



(b) Imagem artificial

Figura 12 – Geração artificial utilizando borramento com filtro bilateral. *Fonte: Elaborado pela autora.*

**Parâmetros e suas variações** Conforme descrito no Algoritmo 2, os parâmetros para essa geração são: o diâmetro  $d$  de pixinhança de pixels, o  $\sigma$  do espaço de cor e o  $\sigma$  do espaço de coordenadas. Esses parâmetros dependem das propriedades das imagens e dos resultados pretendidos. Dessa forma, o tamanho do filtro é um valor escolhido arbitrariamente para cada aplicação em específico (TOMASI; MANDUCHI, 1998). Como o nosso objetivo com a geração das imagens não foi especializar no comportamento de uma classe de imagens específica, um valor foi escolhido aleatoriamente e a partir dele os parâmetros de entrada foram definidos.

**Limitações** Esse filtro tende a remover texturas e a criar novos contornos. Dependendo dos valores, pode gerar uma imagem “cartoonizada”.

**Métodos relacionados** São diversos os métodos de borramento descritos na literatura, como a filtragem Gaussiana, a de mediana e a de médias.

**Visualização** É interessante notar na Figura ?? o comportamento da adição de imagens borradas para o rebalanceamento de classes bem descritas pela propriedade da cor.

colocar aqui alguma figura do espaço...

## 4.3 Aguçamento

Diferentemente da suavização, o processamento de aguçamento procura enfatizar as transições de intensidade. Um processo bem conhecido para atingir tal objetivo é o *unsharp mask*. Ele borra a imagem, subtrai a imagem borrada da original e adiciona essa diferença na imagem original (Ver Algoritmo 3). A imagem resultante, ilustrada na Figura 13, é uma versão realçada da imagem original, dado que soma à imagem justamente o que é removido com um filtro de suavização.

---

### Algoritmo 3: Algoritmo de aguçamento

---

**Entrada:** Imagem colorida  $I$  em formato BGR

**Saída:** Imagem gerada  $G$

```

1  $borrada \leftarrow$  filtro de suavização( $I$ ) para cada pixel  $(x,y)$  faça
2   |   diferença  $\leftarrow I(x,y) - borrada(x,y);$ 
3   |    $G(x,y) \leftarrow I(x,y) + k * diferença;$ 
4 fim

```

---



(a) Original



(b) Imagem artificial

Figura 13 – Geração artificial utilizando unsharp masking. Fonte: Elaborado pela autora.

**Parâmetros e suas variações** Pode-se variar o parâmetro  $k$  de forma a ponderar a soma dessa diferença. Para a geração das imagens da classe minoritária, foi utilizado  $k = 1$ .

**Limitações** É possível que existam pixels com valor negativo no resultado final. Isso pode causar o aparecimento de uma áurea em volta das bordas, efeito não desejado (GONZALEZ; WOODS, 2007).

**Métodos relacionados** Outros algoritmos de aguçamento conhecidos são: utilizar primeira derivada (gradiente) ou a segunda derivada da imagem (Laplaciano).

### Visualização

## 4.4 Adição de ruído

O ruído de Poisson ocorre na contagem de fótons de dispositivos ópticos. Ele segue a distribuição de Poisson, que expressa a probabilidade de um certo número de eventos ocorrerem em um intervalo fixo de tempo e/ou espaço se esses eventos ocorrem com uma taxa média conhecida. O efeito da adição de ruído pode ser visto na Figura 14.



(a) Original



(b) Imagem artificial

Figura 14 – Geração artificial utilizando adição de ruído de Poisson. *Fonte: Elaborado pela autora.*

### descrever o algoritmo

A distribuição de Poisson segue a equação:

Uma possível implementação para encontrar os valores de Poisson foi desenvolvida por Knuth e pode ser vista no Algoritmo 4.

---

**Algoritmo 4:** Algoritmo da geração com ruído de Poisson
 

---

**Entrada:** Imagem colorida  $I$  em formato BGR

**Saída:** Imagem gerada  $G$

```

1 para cada pixel  $(x,y)$  faça
2    $L \leftarrow \exp(-I(x,y));$ 
3    $p \leftarrow 1;$ 
4    $k \leftarrow 0;$ 
5   faça
6      $k \leftarrow k + 1;$ 
7      $p \leftarrow p * \text{número aleatório entre } 0 \text{ e } 1;$ 
8   enquanto  $p > L;$ 
9    $G(x,y) \leftarrow k - 1;$ 
10 fim

```

---

## Parâmetros e suas variações

**Limitações** A adição de ruído é normalmente indesejável e a utilizamos para englobar mais um processamento de imagens. (??)

## Métodos relacionados

### Visualização

## 4.5 SMOTE visual

Conforme visto na Seção 2.5.1, o SMOTE é um método de rebalanceamento aplicado após a extração de características. A ideia dessa geração, chamada de SMOTE visual, é imitar esse funcionamento no nível de pixels. A diferença é que não é feito entre as imagens mais próximas, mas sim entre duas imagens escolhidas de forma aleatória do conjunto de treinamento da classe minoritária.

Para cada pixel é calculado a diferença entre as duas imagens. Essa diferença é então multiplicada por um número aleatório no intervalo  $[0 - 1]$  e adicionado na imagem original (Ver Algoritmo 5). O efeito que esse processamento causa na imagem pode ser visualizado na Figura 15.

---

**Algoritmo 5:** Algoritmo da geração com SMOTE visual

---

**Entrada:** Imagem colorida  $I$  em formato BGR

**Entrada:** Imagem colorida  $I_2$  em formato BGR

**Saída:** Imagem gerada  $G$

```

1 para cada pixel  $(x,y)$  faça
2   diferença  $\leftarrow I(x,y) - I_2(x,y);$ 
3   gap  $\leftarrow$  número aleatório entre 0 e 1;
4    $G(x,y) \leftarrow I(x,y) + gap * \text{diferença};$ 
5 fim
6 mínimo  $\leftarrow$  menor valor de G;
7 máximo  $\leftarrow$  maior valor de G;
8 para cada pixel  $(x,y)$  faça
9    $| G(x,y) \leftarrow G(x,y) - \text{mínimo};$ 
10 fim
11 para cada pixel  $(x,y)$  faça
12    $| G(x,y) \leftarrow G(x,y) * (255 / (\text{máximo} - \text{mínimo}));$ 
13 fim

```

---



(a) Original



(b) Original



(c) Imagem artificial

Figura 15 – Geração artificial utilizando o método SMOTE no espaço visual. *Fonte: Elaborado pela autora.*

**Limitações** Esse método adiciona texturas e bordas que não estavam originalmente nas imagens.

**Métodos relacionados** Esse método é visualmente parecido com o de mistura ponderada, apresentado na próxima seção.

## Visualização

## 4.6 Mistura ponderada

Essa geração calcula a soma ponderada de duas imagens, de acordo com o Algoritmo 6. O efeito dessa mistura pode ser visto na Figura 16, onde dadas duas imagens como entrada, a imagem da direita corresponde a soma delas.

---

### Algoritmo 6: Algoritmo de mistura ponderada

---

**Entrada:** Primeira imagem colorida  $I$  em formato BGR

**Entrada:** Segunda imagem colorida  $I_2$  em formato BGR

**Saída:** Imagem gerada  $G$

- 1  $\alpha \leftarrow$  número aleatório entre 10 e 80;
  - 2  $\beta \leftarrow 100 - \alpha;$
  - 3 **para cada** pixel  $(x, y)$  **faça**
  - 4   |  $G(x, y) \leftarrow \beta \cdot I(x, y) + \alpha \cdot I_2(x, y);$
  - 5 **fim**
- 



(a) Original



(b) Original



(c) Imagem artificial

Figura 16 – Geração artificial utilizando uma mistura ponderada de duas imagens. *Fonte: Elaborado pela autora.*

**Parâmetros e suas variações** Os parâmetros  $\alpha$  e  $\beta$  são escolhidos de forma aleatória. Um valor entre 10% e 80% é escolhido para  $\alpha$ ; e o  $\beta$  é o restante para completar 100%.

**Limitações** Assim como todas as gerações artificiais que envolvem a mistura de imagens, efeitos são adicionados às imagens originais. Dependendo da combinação dos métodos de descrição, quantização e classificação, isso pode piorar a acurácia da classificação.

**Métodos relacionados** É um método de combinação de imagens primitivo. Algoritmos similares são muito mais complexos, como os de threshold e saliência descritos a seguir.

## Visualização

## 4.7 Mistura limiarizada

A combinação de *thresholds* é uma composição do fundo (*background*) de uma imagem e do objeto da cena (*foreground*) de outra imagem. A Figura 17 mostra a mistura dos *thresholds* de duas imagens originais para compor uma nova imagem. O Algoritmo 7 descreve as operações necessárias para realizar tal processamento.



Figura 17 – Geração artificial utilizando uma mistura limiarizada de duas imagens. *Fonte:* Elaborado pela autora.

---

### Algoritmo 7: Algoritmo de mistura limiarizada

---

**Entrada:** Imagem colorida  $I$  em formato BGR

**Entrada:** Imagem colorida  $I_2$  em formato BGR

**Saída:** Imagem gerada  $G$

- 1  $I_{cinza} \leftarrow$  escala de cinza( $I$ );
  - 2  $I_{threshold} \leftarrow OTSU(I_{cinza})$ ;
  - 3  $I_{morfologica} \leftarrow$  abertura e dilatação( $I_{threshold}$ );
  - 4  $I_{foreground} \leftarrow$  aplica máscara( $I_{morfologica}, I$ );
  - 5  $I_{morfologica} \leftarrow$  oposto( $I_{morfologica}$ );
  - 6  $I_{background} \leftarrow$  aplica máscara( $I_{morfologica}, I_2$ );
  - 7  $G \leftarrow I_{background} + I_{foreground}$ ;
- 

**Parâmetros e suas variações** No âmbito desta pesquisa, os parâmetros estão fixos, mas é possível modificar o tamanho dos elementos estruturantes que fazem as operações de abertura e dilatação para remover pequenas regiões.

**Limitações** Dependendo da quantidade de informações da imagem, o *threshold de OTSU* pode não conseguir extrair nenhuma informação relevante ou mesmo a imagem toda.

**Métodos relacionados** Essa geração está fortemente correlacionada com a mistura a partir da saliência da imagem, apresentada a seguir.

### Visualização

## 4.8 Mistura saliente

A combinação de regiões salientes é muito similar com o método anterior de combinação de *thresholds*, porém, utiliza um algoritmo mais robusto que detecta a saliência da imagem a partir do método SLIC. A Figura 18 mostra a combinação da região saliente da imagem original à esquerda com a imagem central, resultando na imagem combinada à direita.

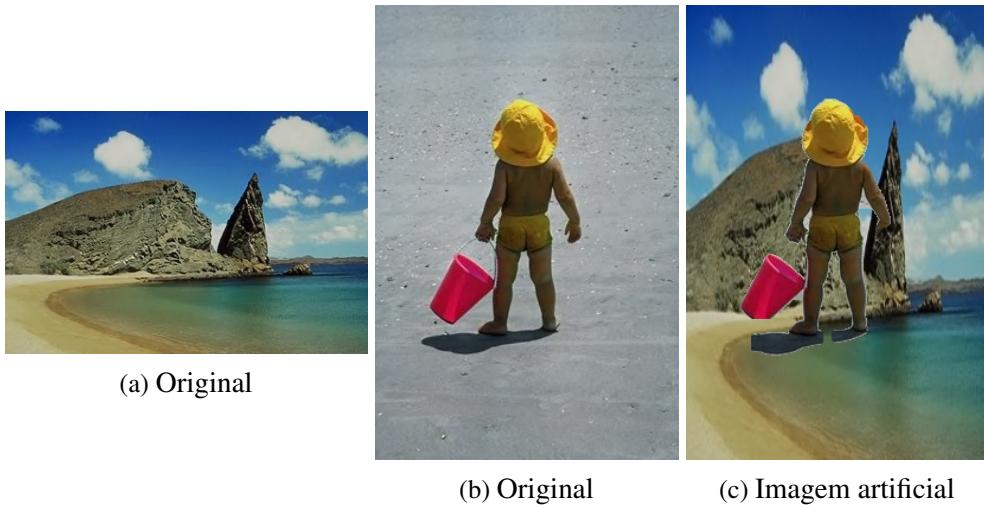


Figura 18 – Geração artificial utilizando uma mistura de duas imagens a partir da saliência da primeira imagem.  
Fonte: Elaborado pela autora.

As operações aplicadas na imagem para extrair a região mais saliente são: SLIC; rotulação por conectividade; *threshold de OTSU*; e operações morfológicas. O Algoritmo 8 apresenta os passos para o cálculo do *background* e *foreground*.

---

### Algoritmo 8: Algoritmo de mistura saliente

---

**Entrada:** Imagem colorida  $I$  em formato BGR

**Entrada:** Imagem colorida  $I_2$  em formato BGR

**Saída:** Imagem gerada  $G$

- 1  $I_{\text{rotulada por segmento}} \leftarrow \text{SLIC}(I);$
  - 2  $I_{\text{mapa de saliência}} \leftarrow \text{rotulação por conectividade}(I_{\text{rotulada por segmento}});$
  - 3  $I_{\text{threshold}} \leftarrow \text{OTSU}(I_{\text{mapa de saliência}});$
  - 4  $I_{\text{morfologica}} \leftarrow \text{abertura e dilatação}(I_{\text{threshold}});$
  - 5  $I_{\text{foreground}} \leftarrow \text{aplica máscara}(I_{\text{morfologica}}, I);$
  - 6  $I_{\text{morfologica}} \leftarrow \text{oísto}(I_{\text{morfologica}});$
  - 7  $I_{\text{background}} \leftarrow \text{aplica máscara}(I_{\text{morfologica}}, I_2);$
  - 8  $G \leftarrow I_{\text{background}} + I_{\text{foreground}};$
- 

**Parâmetros e suas variações** Assim como no método anterior, os parâmetros são relacionados

ao tamanho do elemento estruturante para a abertura e dilatação e estão fixos.

**Limitações** Não é garantido que o algoritmo de saliência consiga extrair a melhor região, ou mesmo que sempre haja uma região.

**Métodos relacionados** Similar à mistura por *thresholds*.

### Visualização

## 4.9 Composição

Essa geração pretende compor informações de diversas imagens em uma única imagem. Assim é feito um mosaico com várias imagens, conforme pode ser visto na Figura 19. Se as imagens possuem um elemento centralizado, essa geração pode resultar em uma imagem de um objeto que parece uma mutação dos objetos centrais, conforme pode ser visualizado na Figura ??.

**figura da mistura central**



Figura 19 – Geração artificial utilizando uma composição de imagens. *Fonte: Elaborado pela autora.*

Para cada quadrado a ser preenchido sorteia uma imagem do conjunto de treinamento, realiza uma operação de borramento, aguçamento, mistura ponderada ou visual SMOTE e adiciona essa imagem no quadrado respectivo. Os passos para tal composição estão descritos no Algoritmo 9.

---

**Algoritmo 9:** Algoritmo de composição

---

**Saída:** Imagem gerada  $G$

```

1 enquanto  $total < \text{número de quadrados } q$  faça
2    $I \leftarrow$  imagem aleatória do conjunto de treinamento;
3    $operação \leftarrow 1 + (\text{rand}() \% 3);$ 
4   selecione  $operação$  faça
5     caso 1
6        $I \leftarrow \text{borramento}(I);$ 
7     caso 2
8        $I \leftarrow \text{mistura ponderada}(I);$ 
9     caso 3
10     $I \leftarrow \text{aguçamento}(I);$ 
11    caso 4
12     $I \leftarrow \text{visual SMOTE}(I);$ 
13  fim
14   $x \leftarrow \text{posição aleatória em } x \text{ de } I;$ 
15   $y \leftarrow \text{posição aleatória em } y \text{ de } I;$ 
16   $qx \leftarrow \text{posição atual para o quadrado em } x \text{ de } G;$ 
17   $qy \leftarrow \text{posição atual para o quadrado em } y \text{ de } G;$ 
18   $G(qx, qy) \leftarrow I(x, y);$ 
19   $total ++;$ 
20 fim

```

---

**Parâmetros e suas variações** O parâmetro  $q$  controla quantos quadrados serão criados na nova imagem. Nesta pesquisa foram realizados testes com 4 e 16.

**Limitações** O término brusco de uma imagem para início da outra ao formar a grade de imagens tenha efeitos colaterais de inserção de textura que não excedam a vantagem de compor uma mesma imagem com várias cores, texturas e formas das imagens originais.

### Métodos relacionados

#### Visualização

## 4.10 Considerações finais



# RESULTADOS

---



---

## 5.1 Considerações Iniciais

## 5.2 Quantização de Imagens

Diferentes métodos de quantização produzem diferentes resultados em termos de acurácia. Quando comparado com métodos mais complexos, esse procedimento permite uma redução de dimensionalidade significante, enquanto preserva ou melhora a acurácia do sistema.

Na Figura ?? estão demonstrados o diagrama do fluxo das operações e os métodos utilizados nos experimentos. O objetivo dessa seção de resultados é mostrar os efeitos da etapa de quantização e como ela pode ser utilizada para reduzir a dimensionalidade do espaço de características.

Inicialmente, as imagens são quantizadas em 256, 128, 64, 32 e 16 cores. Após, suas características são extraídas e dois tipos de experimentos são realizados:

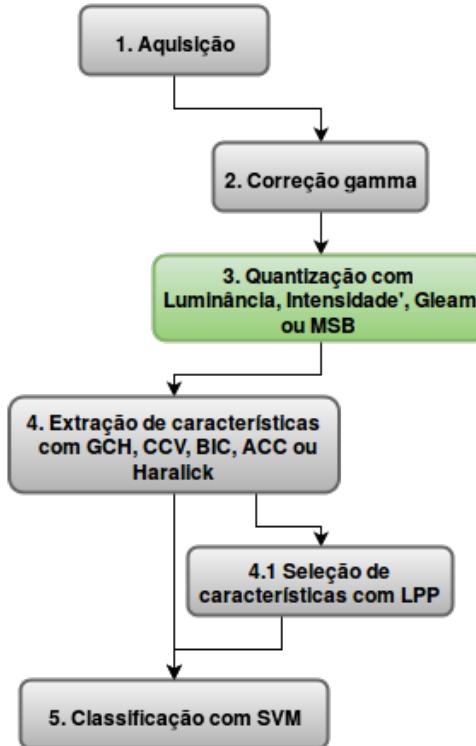
1. Experimentos com um método de extração de características, seguido pela classificação com seleção de características;
2. Experimentos com o vetor resultante da concatenação de todos os métodos de extração, seguido pela classificação com e sem seleção de características.

### 5.2.1 Base de Imagens

Três bases de imagens foram utilizadas nos experimentos, ilustradas na Figura ??:

**Corel-1000**<sup>1</sup>: consiste em dez classes balanceadas de imagens naturais, com algumas bem definidas e algumas não;

<sup>1</sup> Disponível em <http://www.wang.ist.psu.edu/docs/related/>



**Caltech101-600**<sup>2</sup>: contém fotos e desenhos. Foram utilizados seis classes balanceadas: aviões, bonsais, candelabros, tartarugas, motocicletas e relógios;

**Produce**<sup>3</sup> (também conhecido como base de vegetais e frutas tropicais): composta por imagens com fundo similar, mas mudanças na iluminação, no número de objetos e na escala. Apesar da oclusão parcial de objetos ser observada, essa classe possui dados bem comportados.

Considerando que esses experimentos possuem foco na redução na dimensionalidade e com a motivação de evitar o problema do desbalanceamento, as bases *Produce* e *Caltech101* foram modificadas. Para tal, as classes disponíveis foram balanceadas ao remover imagens das classes majoritárias.

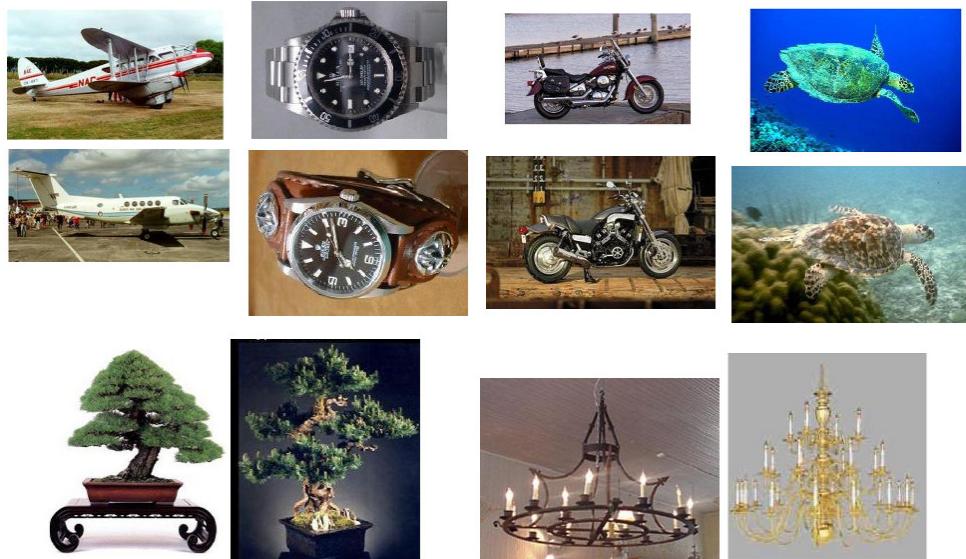
### 5.2.2 Protocolo

O seguinte protocolo foi seguido para a obtenção dos resultados:

1. **Quantização:**
2. **Extração de características:** os parâmetros para os métodos foram escolhidos com base nas recomendações dos artigos originais:

<sup>2</sup> Disponível em <http://www.vision.caltech.edu/ImageDatasets/Caltech101/>

<sup>3</sup> Disponível em ??



(a) Base de imagens Caltech101



(b) Base de imagens Caltech101



(c) Base de imagens Caltech101

Figura 20 – Fonte: ([PONTI; NAZARÉ; THUMÉ, 2016](#)).

- ACC: utilizando um conjunto de quatro distâncias  $D$  e a distância xadrez;
- BIC: com uma vizinhança de quatro pixels;
- CCV: adotando um valor de *threshold* de 25 para a classificação dos pixels entre coerentes e incoerentes;
- Haralick-6: o pixel vizinho para o qual computar a matriz de correlação foi definido como sendo o pixel a direita.

3. **Redução da dimensionalidade:** a projeção LPP foi realizada com o parâmetro  $k = 128$ , 64, 32 e 16 dimensões e 10 vizinhos. Esse parâmetro foi determinado empiricamente e não influencia consideravelmente a acurácia.
4. **Classificação:** para essa tarefa foi utilizado o classificador SVM (*Support Vector Machines*). Os parâmetros para tal foram encontrados utilizando uma *grid search* no conjunto de treino.

Os experimentos foram realizados com uma validação cruzada de *10-fold*. Considerando que as bases estão balanceadas e que a seleção de exemplos para a validação cruzada é estratificada, a medida estatística de *acurácia* foi utilizada para avaliar a performance da classificação.

### 5.2.3 Resultados e Discussão

A Figura ilustra a acurácia média para o primeiro conjunto de experimentos. Para cada base de dados e método de extração, estão demonstrados seis resultados de acurácia, correspondendo a quantização para 256, 128, 64, 32, 16 e 8 cores. Com base nessa figura é possível identificar que o método para obter a imagem quantizada tem impacto significativo na acurácia da classificação. Uma importante observação é que a redução de 256 para um menor número de cores normalmente mantém as acurácia e pode em alguns casos até melhorar um pouco ela, especialmente para os níveis de 128 e 64.

A partir dessa análise geral, uma análise mais específica foi realizada com a combinação dos métodos BIC e MSB; e Haralick e Luminância. Considerando que a utilização de apenas 16 e 8 cores resultou em uma acurácia muito inferior, para o restante dos resultados o número de cores ficou em 256, 128, 64 e 32.

#### descrever o teste estatístico e o resultado

O teste estatístico ANOVA foi realizado para comparar as acurácia dos experimentos da Figura ?? e 23. O boxplot para 256, 128, 64 e 32 cores com os métodos BIC e MSB está demonstrado na Figura ?? . De acordo com o teste estatístico representado, utilizar características de cor e níveis de quantização providos pelo método MSB demonstrou resultados melhores do que com 256 cores para as bases Corel (128, 64 e 32 cores) e Caltech (64 cores). O único

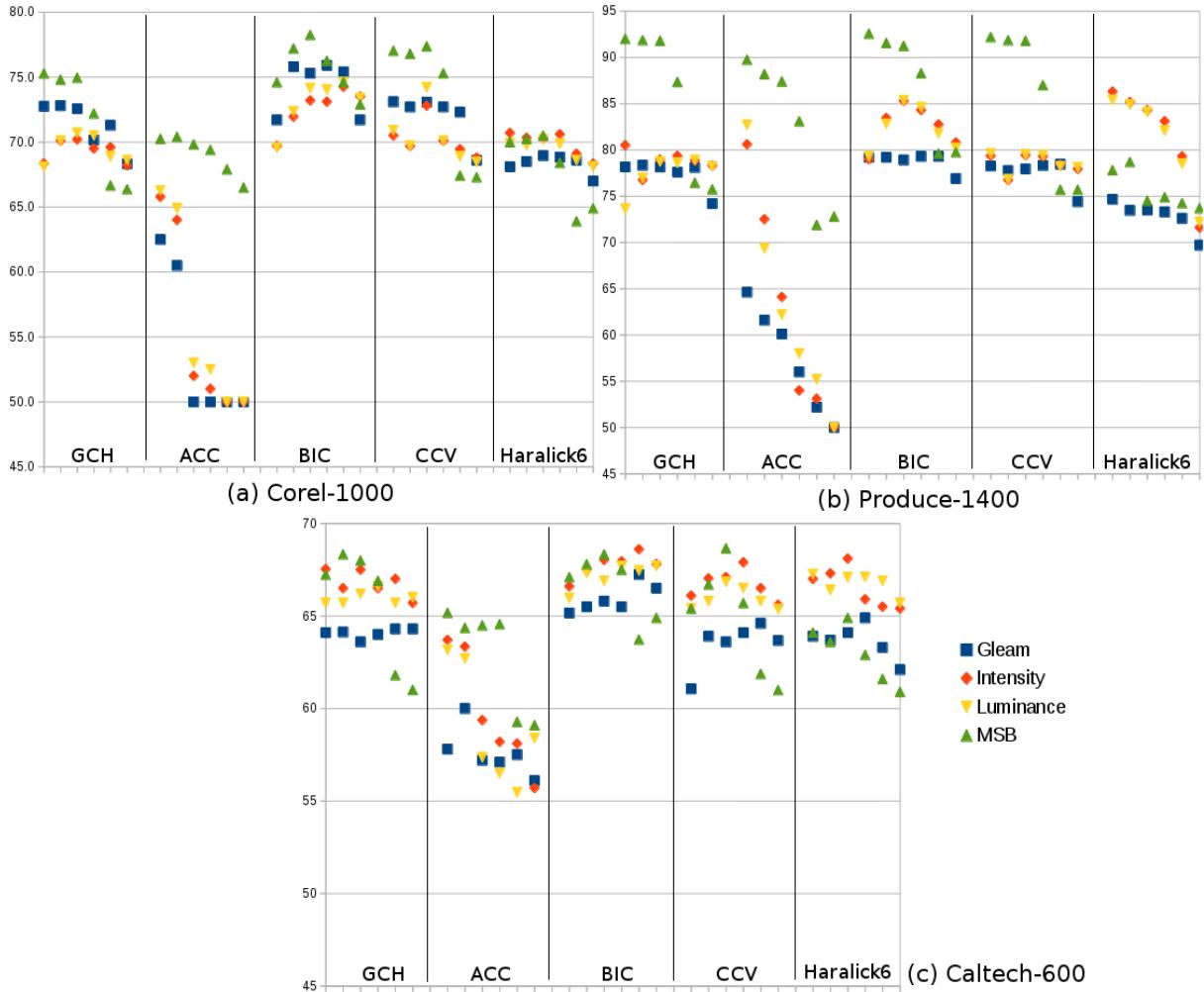


Figura 21 – Resultados para Corel(a), Produce(b) e Caltech(c), com todos os métodos de quantização. Para cada método de extração de características a acurácia é resultante da sua aplicação utilizando 256, 128, 64, 32, 16 e 8 cores, da esquerda para a direita. *Fonte: (PONTI; NAZARÉ; THUMÉ, 2016).*

resultado que piorou significativamente foi para 32 cores da base de imagens *Produce*. Portanto, reduzir converter as imagens de 3 canais de cores para um e reduzir os 256 possíveis valores para apenas 64 provou uma boa escolha de processamento anterior a extração de características. Menores valores podem degradar os resultados em características de textura, como mostrado na Figura 23.

Outra comparação interessante é a redução de dimensionalidade obtida utilizando os métodos de quantização e LPP ilustrada na Figura 24. A imagem de entrada foi quantizada com MSB para 256 cores. Essa imagem foi dada como entrada para o método de extração de características BIC que resultou em um vetor dado como entrada para o LPP. Esse último passo teve o objetivo de produzir versões reduzidas desse vetor para 256, 128 e 64 dimensões. As acurárias obtidas foram comparadas com a classificação dos vetores reduzidos apenas pela quantização. O método de quantização obteve valores de acurácia menores à utilização do LPP em três experimentos: de 256 dimensões com a base *Corel* e com 256 e 64 na base *Produce*. Para

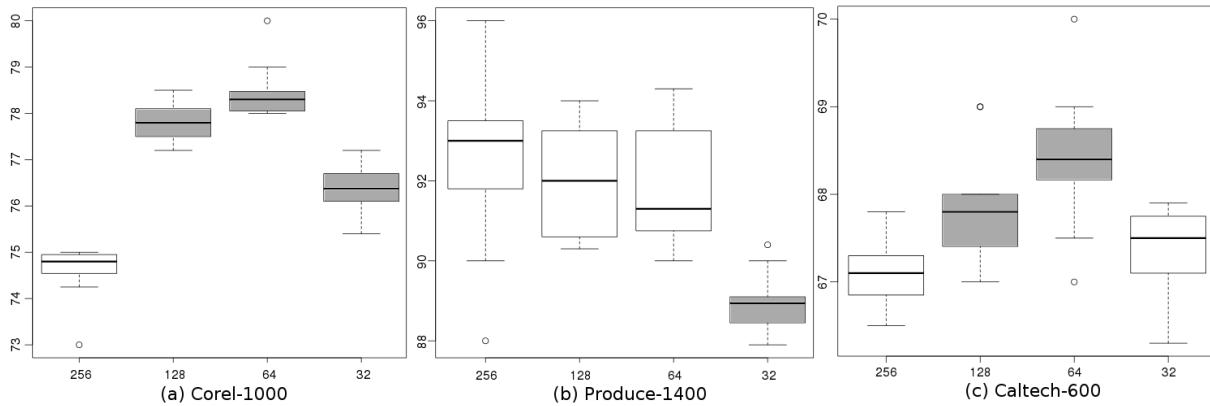


Figura 22 – Resultados de acurácia para o método de quantização MSB considerando 256, 128, 64 e 32 cores com o descritor BIC. Os boxplots em cinza correspondem às significâncias estatísticas com  $p \leq 0.01$  quando comparado a acurácia de 256 cores. Fonte: ([PONTI; NAZARÉ; THUMÉ, 2016](#)).

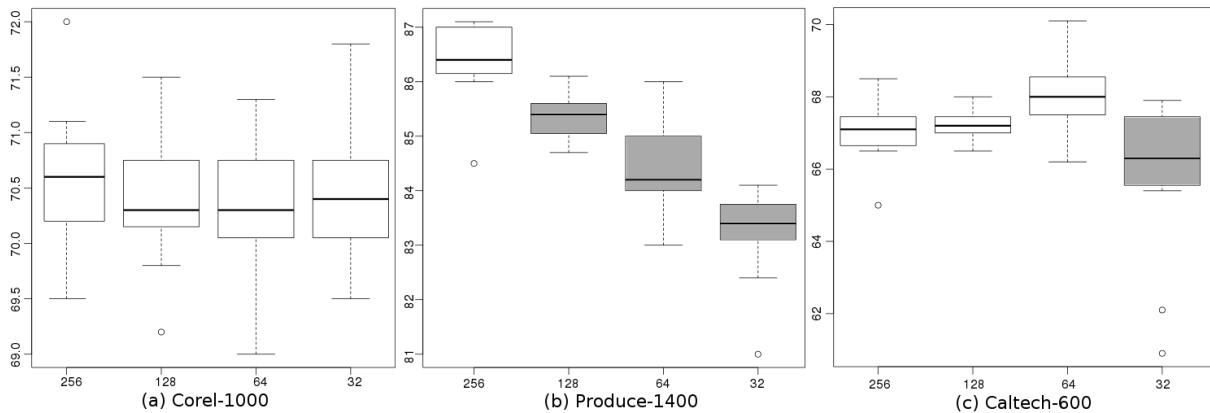


Figura 23 – Resultados de acurácia para o método de quantização Luminância considerando 256, 128, 64 e 32 cores com o descritor Haralick. Os boxplots em cinza correspondem às significâncias estatísticas com  $p \leq 0.01$  quando comparado a acurácia de 256 cores. Fonte: ([PONTI; NAZARÉ; THUMÉ, 2016](#)).

a base *Caltech* a quantização foi melhor com 256 e 128 dimensões. O restante dos experimentos não apresentaram diferença estatística relevante. Apesar da perda de acurácia em alguns casos, é importante notar que – se utilizado um número de cores correto – é possível manter ou até mesmo melhorar as acurárias após a redução da dimensionalidade. Isso pode ser observado na Figura 24 referente à base de dados *Caltech*.

O problema da dimensionalidade de um vetor resultante de apenas um método de extração de características pode ser considerado baixo. Ainda mais que é comum extrair diversos descritores para uma situação, já que normalmente não está claro qual método deveria ser utilizado em cada caso. Por conta disso, os próximos experimentos utilizaram da concatenação de características.

O objetivo desses experimento é verificar se a concatenação de todos os descritores pode melhorar os resultados. Além disso, comparar os resultados encontrados com os experimentos anteriores, afim de verificar se a quantização pode ser uma alternativa a redução da dimensional-

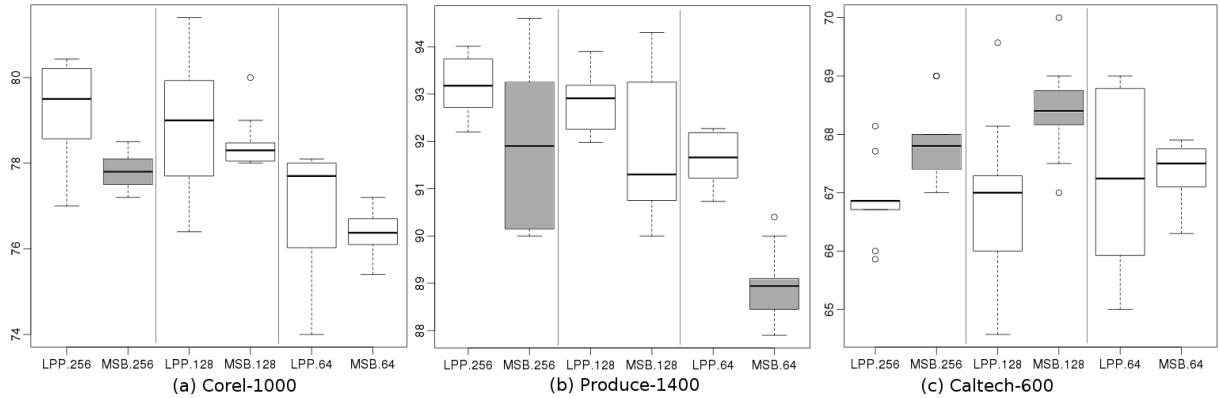


Figura 24 – Resultados de acurácia para os método MSB (quantização) e LPP para redução de dimensionalidade e BIC para a extração de características. A comparação foi realizada de forma par (LPP versus MSB) com a mesma dimensionalidade. Os boxplots em cinza correspondem às significâncias estatísticas com  $p \leq 0.01$  quando comparado a acurácia de 256 cores. Fonte: (PONTI; NAZARÉ; THUMÉ, 2016).

dade com métodos convencionais (LPP, neste caso). A melhor configuração encontrada entre tamanho do vetor e acurácia foi utilizando 128 e 64 cores.

O número de características em relação ao número de cores, concatenando todos os vetores resultantes dos métodos de extração de características, é: 256 cores – 2310 características; 128 cores – 1160 características; 64 cores – 582 características; 32 cores — 294 características; e 16 cores – 150 características.

Inicialmente, foi testada a configuração de um vetor  $D = 2310$  com LPP para redução de dimensionalidade com  $d = 1160, 582, 294$  e  $150$ . Ou seja, produzindo vetores com o mesmo tamanho dos obtidos apenas com a quantização como redução da dimensão. A Figura 25 mostra os resultados utilizando LPP. Note que o método de quantização MSB resultou em acurárias melhores que os outros métodos.

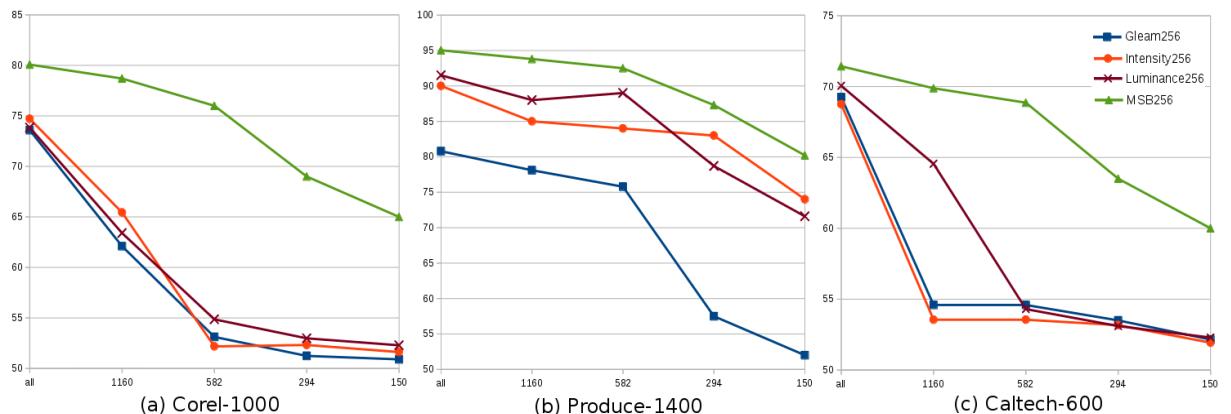


Figura 25 – Comparação da acurácia alcançada com diferentes métodos de quantização: Gleam, Intensidade, Luminância e MSB. Inicialmente com  $D = 2310$  e então reduzindo com LPP para  $d = 1160, 582, 294$  e  $150$ . Fonte: (PONTI; NAZARÉ; THUMÉ, 2016).

A utilização da concatenação de todos os vetores melhorou a acurácia em relação ao

melhor descritor individual. A Figura 26 apresenta a comparação do espaço original com LPP e MSB para redução da dimensionalidade. O teste estatístico ANOVA foi realizado, seguido do teste da significância diferença honesta de Tukey. Os dois testes utilizaram  $\alpha = 0.01$  como nível de significância. Os resultados que não mudaram significativamente as acurácia foram MSB com 582 características para a base de dados Corel e MSB com 1160 nas três bases. O único resultado de piora foi para 32 níveis da base Produce. Assim, 64 cores parece ser uma boa escolha de parâmetro de quantização.

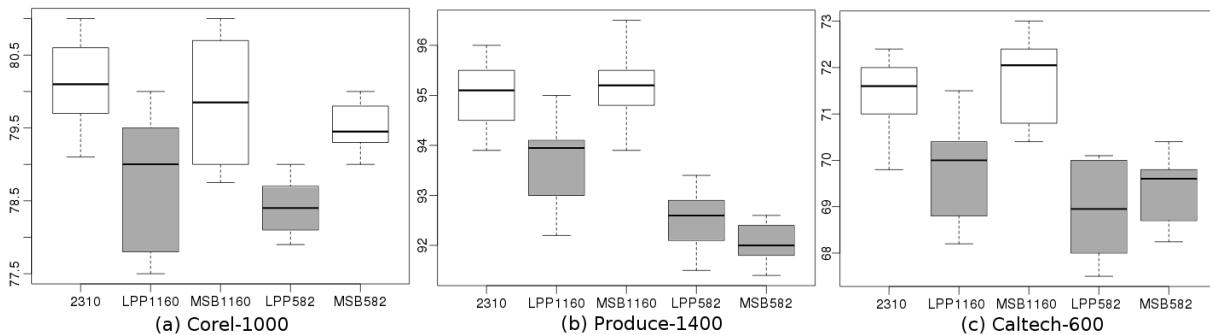


Figura 26 – Comparação da acurácia com o uso da projeção LPP e o método MSB para quantização das imagens com o objetivo de redução de dimensionalidade. Fonte: (PONTI; NAZARÉ; THUMÉ, 2016).

Os resultados indicam que a quantização pode ser utilizada como redução da dimensão para dados visuais, especialmente com 128 e 64 cores. Como experimento, a Figura 27 mostra os resultados da utilização do LPP sob o vetor obtido após a quantização com MSB utilizando 256 e 64 cores ( $d = 2310$  e  $d = 582$ , respectivamente). É interessante notar que as projeções LPP em geral foram melhores com as imagens quantizadas em 64 cores com MSB ao invés da original em 256. A razão para isso deve estar no fato da quantização remover informações confusas: ela simplifica as imagens de forma que as cores restantes possam melhor descrever uma certa classe.

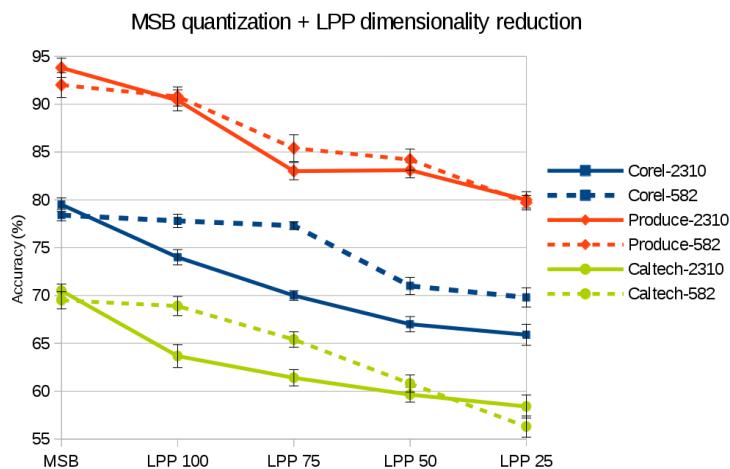


Figura 27 – Resultados para a projeção do LPP sobre o espaço de características produzido pelo método de quantização MSB utilizando 256 ( $d = 2310$ ) e 64 cores ( $d = 582$ ). Fonte: (PONTI; NAZARÉ; THUMÉ, 2016).

O vetor concatenado com todos os descritores possui  $9C + 6$  dimensões, onde  $C$  é o número de cores da imagem de entrada. E o tempo de execução para a extração de todas as características é  $f(N) = 42N + 6C^2$ , onde  $N$  é o número de pixels. Para cada imagem são necessárias  $D^2 + kD + d^2$  operações para computar o vetor reduzido com LPP, onde  $D$  é o tamanho do vetor original,  $d$  o tamanho do vetor de saída e  $k$  é o número de vizinhos utilizados no algoritmo.

Considere o seguinte exemplo: 100 imagens com 256 cores demandam 231.6 milhões de instruções do computador para extrair as características e reduzir o vetor utilizando o método LPP (com  $k = 10$  e  $d = 50$ ). Se ao invés disso, fossem utilizadas 64 cores, esse número cairia para 58.7 milhões, o que corresponde a uma redução de 74,6%. Dessa forma, independente de utilizar um método de seleção de características, ao escolher um método de quantização apropriado e seus parâmetros, é possível reduzir a dimensionalidade e acelerar computacionalmente as etapas que precedem o reconhecimento de imagens.

## 5.3 Geração de Imagens Artificiais

Essa seção descreve os resultados encontrados ao rebalancear as classes de imagens. Na figura 28 é possível observar o fluxo de operações realizadas para analisar o impacto da geração de imagens no rebalanceamento de classes. O mesmo protocolo de conversão para escala de cinza, extração de características e classificação foi seguido para três experimentos: base desbalanceada; base rebalanceada com interpolação dos vetores de características; e base rebalanceada com a geração artificial de imagens. Portanto inicia-se com a descrição destas bases de imagens, a descrição do protocolo adotado e por fim os resultados obtidos a partir de seu uso são mostrados e discutidos.

### 5.3.1 Base de Imagens

Os resultados foram obtidos utilizando a base de imagens COREL-1000<sup>4</sup>, composta por fotografias que representam classes variadas: tribos africanas, praia, construções, ônibus, dinossauros, elefantes, flores, cavalos, montanhas e tipos de comidas. São 10 classes balanceadas com 100 imagens cada. Para fins de exemplificação, são apresentadas amostras das imagens que representam essas classes na Figura 29.

### 5.3.2 Protocolo

### 5.3.3 Resultados e Discussão

As classes Elefante e Cavalo possuem 100 imagens cada. O primeiro passo é remover imagens de uma das classes, tornando a base desbalanceada. Como o foco é na visualização

<sup>4</sup> Disponível em <http://wang.ist.psu.edu/docs/related/>

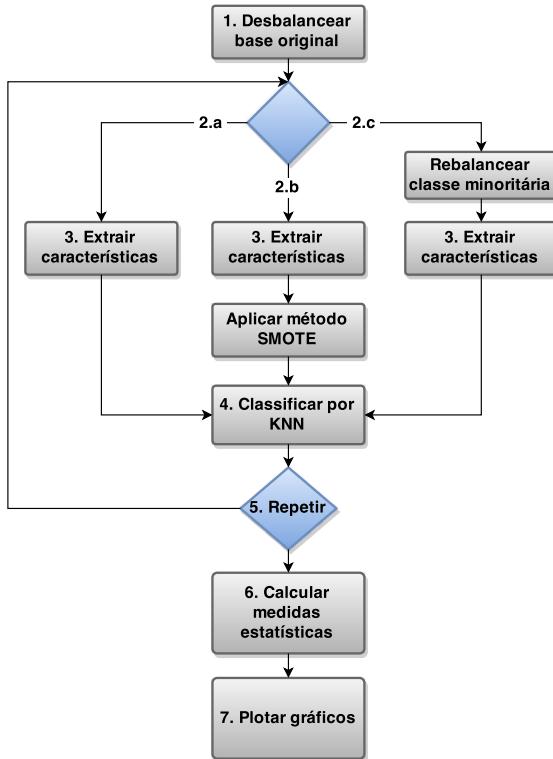


Figura 28 – Fluxo de operações do experimento realizado.



Figura 29 – Base de imagens COREL-1000 utilizada. Estão representadas as 10 classes da base. *Fonte: Elaborado pela autora.*

do espaço de características, é relevante ter o modelo do espaço ideal das classes balanceadas, por isso esse experimento em específico não trata de uma base naturalmente desbalanceada. Na Figura 30 está ilustrada a remoção de 50% das imagens de treino da classe Cavalo, originalmente balanceada. Essa e as próximas projeções desta seção foram obtidas com a técnica para redução de dimensionalidade PCA, descrita na Seção 2.7.1 e são referentes aos dois componentes principais com maiores autovalores.

A classificação dos três experimentos utilizando KNN reportou que o *f-score* da geração artificial de imagens utilizando o método de mistura teve um ganho de mais de 10% em relação ao rebalanceamento no espaço de características com o SMOTE. Para confirmar que a geração aqui proposta inseriu mais informação na classe minoritária do que apenas povoar os espaços entre

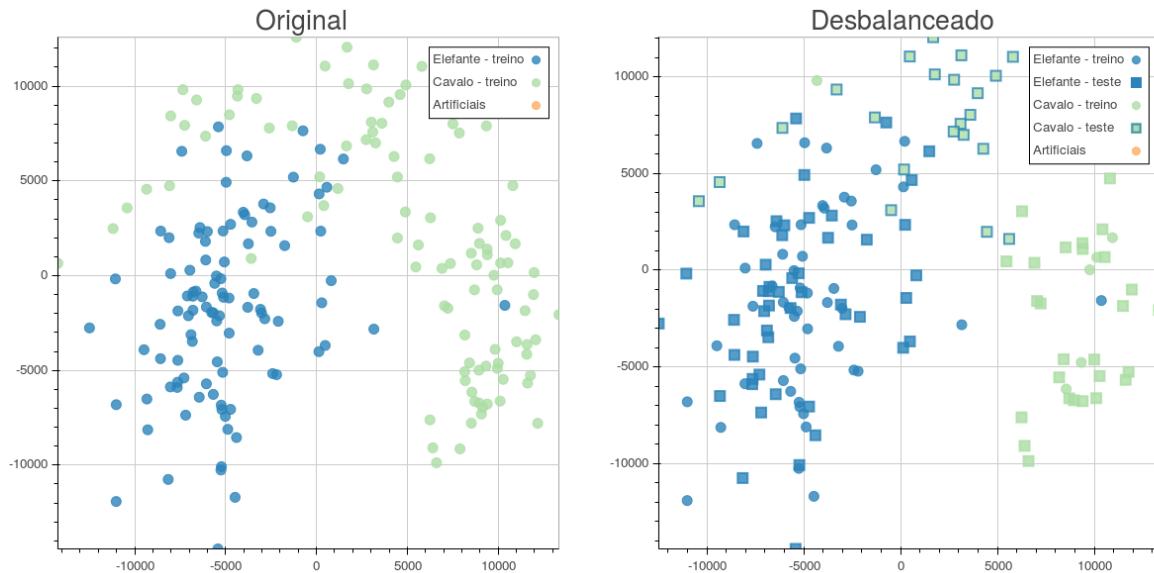


Figura 30 – Remoção de 50% das imagens de treino da classe Cavalo.

os exemplos (i.e. SMOTE), a classe rebalanceada utilizando ambos métodos está demonstrada na Figura 31. Em laranja estão representados os novos exemplos de treinamento, projetados no plano da base original balanceada.

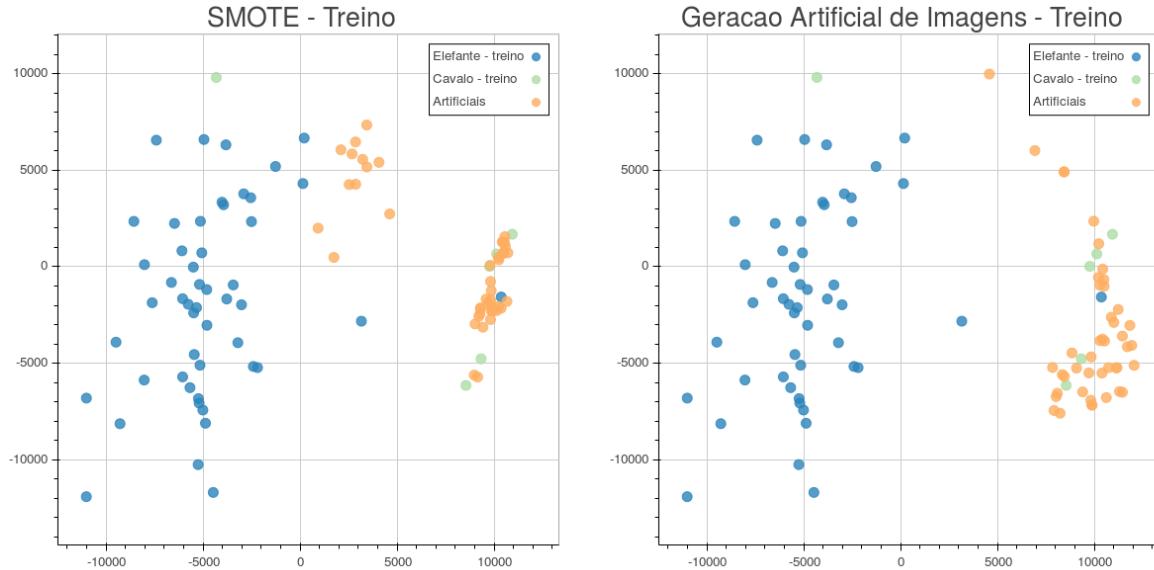


Figura 31 – Comparação dos exemplos de treinamento da geração com SMOTE e no campo visual. Em laranja estão representados os novos exemplos, projetados no plano da base original balanceada.

Após o treinamento realizado com as novas imagens e exemplos, o conjunto de teste foi fornecido ao classificador 1-NN e o resultado das predições está ilustrado na Figura 32. A cor no interior dos marcadores quadrados representa a classe real dos exemplos e a borda representa a classe predita pelo classificador. Nota-se que a melhoria na classificação com a geração de imagens fica visível e corresponde ao aumento do  $f$ -score.

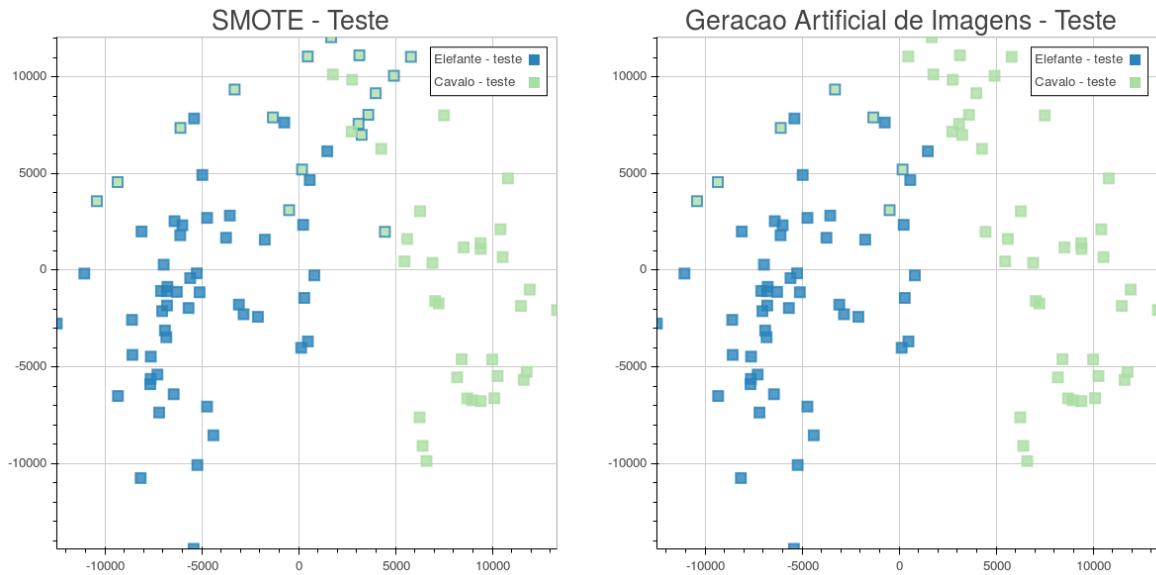


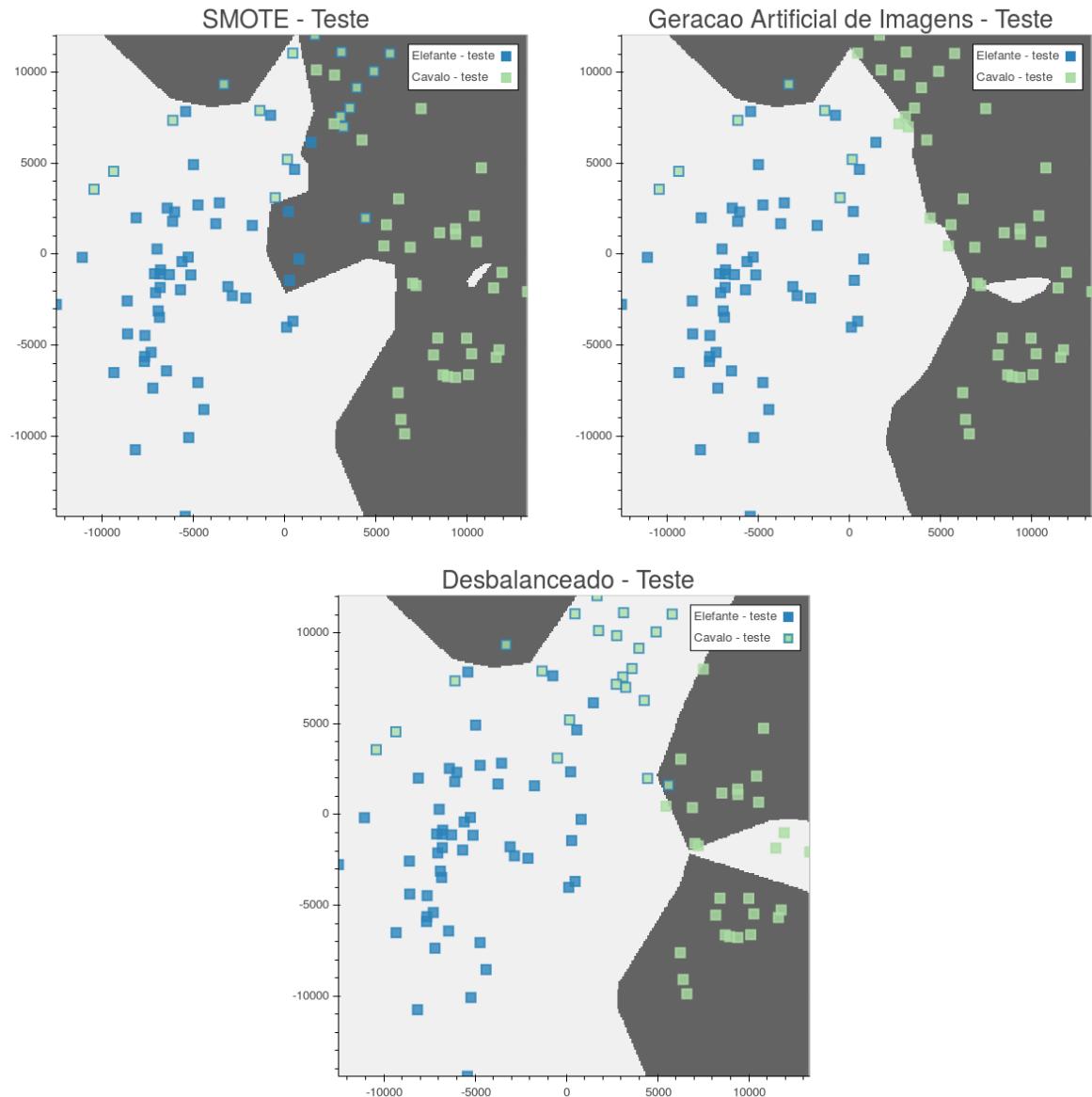
Figura 32 – Resultado do teste da classificação com 1-NN após o treinamento realizado com as bases rebalanceadas. A cor no interior dos marcadores quadrados representa a classe real dos exemplos e a borda representa a classe predita pelo classificador.

De uma forma geral, pode-se dizer que a geração de imagens melhorou a definição da classe minoritária e foi o método que mais se assemelhou à distribuição dos dados originais. Além disso, um dos problemas do SMOTE pode ser verificado nessas projeções: ao realizar a interpolação dos vetores de características originais, pode-se criar exemplos em regiões do espaço que fazem parte da outra classe. Ficou claro também que o método SMOTE não possui capacidade de extrapolar a sua região, como pode ser observado no grupo de exemplos gerados à direita do espaço. O SMOTE gerou novos elementos em linha reta, enquanto a geração de imagens proporcionou uma abrangência maior em volta desse espaço, com maior dispersão.

É válido também visualizar a região de decisão, observando suas modificações frente aos métodos. Pode ser observado que em ambas técnicas a região da classe minoritária apresenta-se melhor representada. Além disso, é possível verificar que o SMOTE ocasionou uma certa “invasão” do espaço de características da classe majoritária.

Nas figuras anteriores os exemplos foram projetados no plano criado pelas suas componentes principais com maior autovalores da base original balanceada. Se após a geração de novos exemplos essas componentes forem recalculadas (Figura 34), pode-se notar que a geração de imagens artificiais proporciona a criação de um subespaço que melhor discretiza as classes, quando comparado com SMOTE ou com a base desbalanceada.

Como relatado no início dessa seção, o extrator de características utilizado foi o BIC. Fundamentalmente ele captura informações de intensidade de cor das imagens. Na Figura 35 as próprias imagens foram utilizadas como marcadores na projeção do melhor subespaço após a geração artificial com o método de mistura. É nítido o impacto da etapa de extração de

Figura 33 – Região de decisão com K-NN ( $K = 1$ )

características na separação das classes e também no método de geração de imagens antes dessa extração.

## 5.4 Considerações Finais

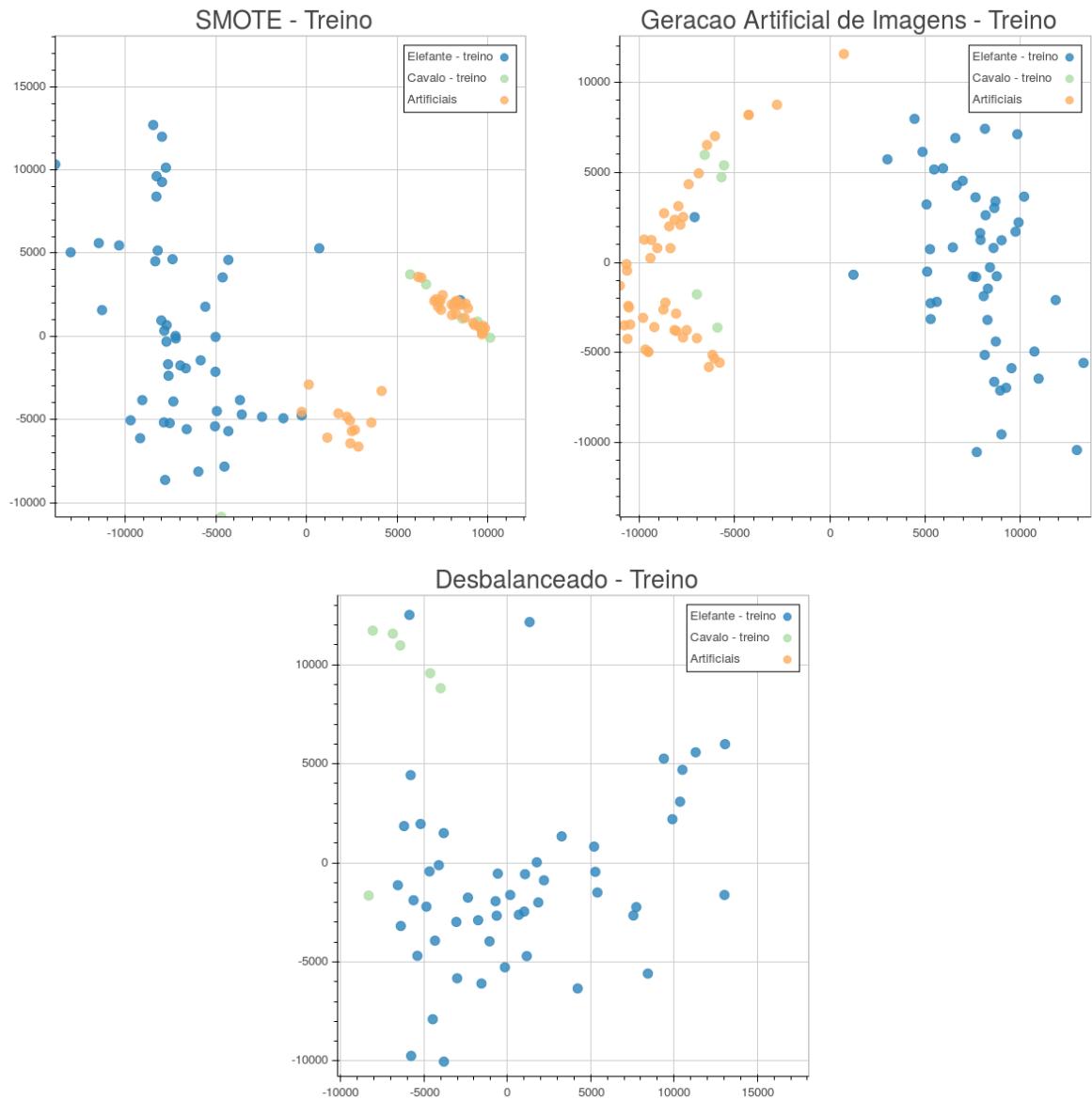


Figura 34 – Melhores subespaços encontrados após a geração de novos exemplos para o SMOTE e para a geração artificial de imagens, e após a remoção de imagens para a projeção dos dados desbalanceados.

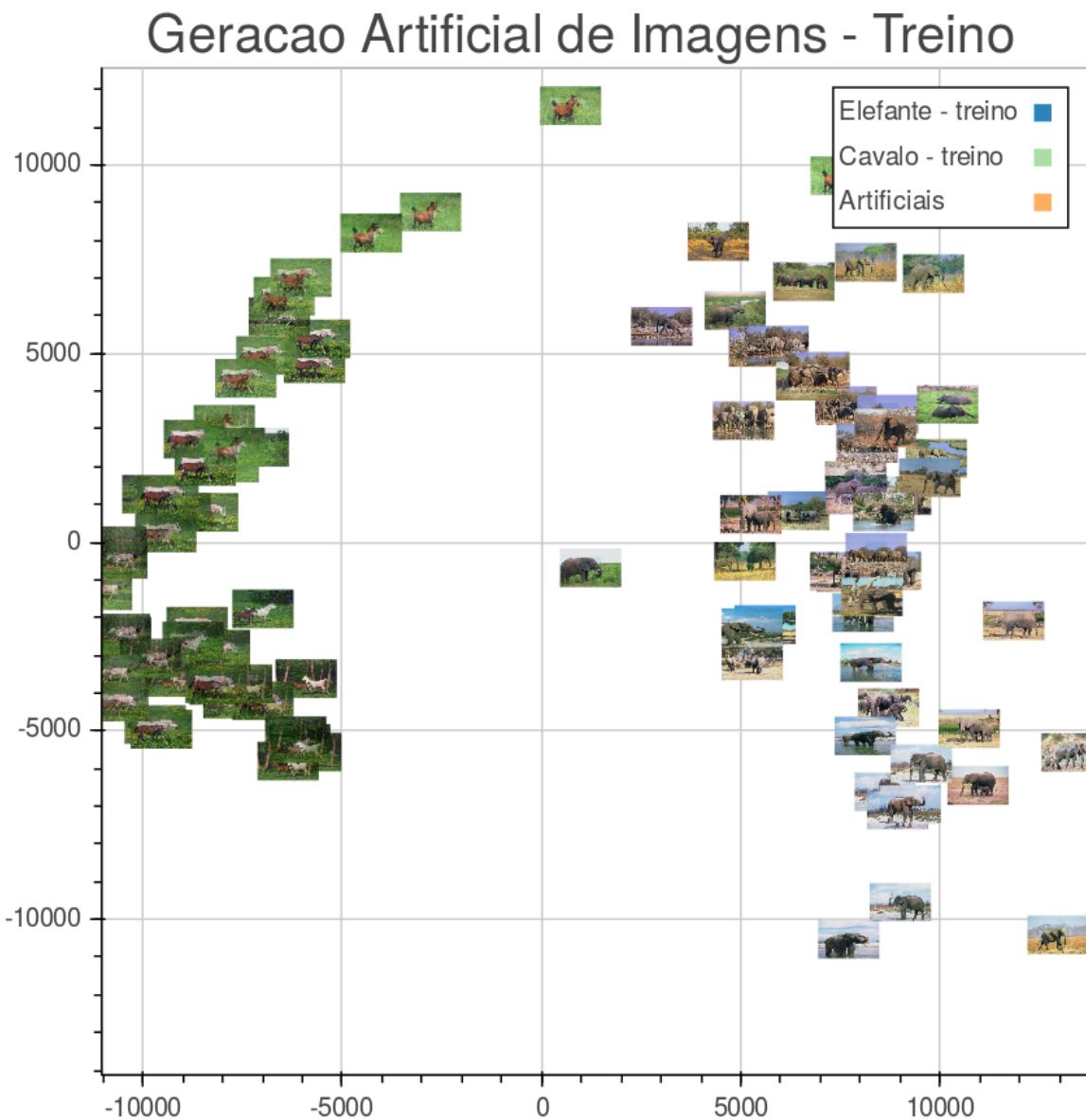


Figura 35 – Visualização do impacto do descritor de características.



CAPÍTULO

---

**6**

## CONCLUSÕES

---

Ficou constatado que um vetor original de  $D$  dimensões pode ser reduzido a  $d \approx D/4$  modificando apenas o parâmetro de quantização e produzindo bons resultados. Outra possibilidade é utilizar esse métodos como um primeiro passo de redução e então utilizar o LPP para obter apenas 100 características que melhor representam os dados, atingindo acurácia maiores ou similares. É importante ressaltar que realizar a quantização de imagens não requer treinamento e já faz parte de uma tarefa do pipeline de reconhecimento. Por esta razão, seu uso não aumenta o custo computacional do sistema, e ainda simplifica os passos subsequentes. Isso reduz a dimensão do vetor de características para os vetores de cor e o tempo de computação para os descritores de textura. Outra observação importante é que a quantização é usada especialmente para dados visuais, então não é um método geral de redução de dimensionalidade.

Com os experimentos realizados foi possível notar que a geração de imagens artificiais pode gerar novas informações para a classificação das imagens. Assim a geração de elementos no espaço visual provou contribuir com o balanceamento entre classes (em se tratando de problemas de classes desbalanceadas), melhorando a acurácia de algoritmos de classificação, quando comparada à geração de exemplos artificiais no espaço de atributos (i.e. SMOTE).

A visualização do espaço de características utilizando a técnica de análise de componentes principais (PCA) se mostrou crucial para confirmar visualmente que a melhora da acurácia da classificação se deve à melhor definição da classe minoritária.

Para validar a ideia da geração artificial de imagens, as características das novas imagens – extraídas com o método BIC – e os exemplos resultantes da interpolação de vetores originais foram projetados no plano das imagens originais antes do desbalanceamento. Dessa forma foi possível visualizar que a geração de imagens artificiais proposta foi capaz de ocupar uma região do espaço mais abrangente do que o SMOTE. Este último, comprovadamente, possui o ponto negativo de não extrapolar os limites da classe minoritária. Ainda, está suscetível à criação de novos exemplos em regiões da classe majoritária, o que também prejudica a classificação.

A projeção forneceu uma análise intuitiva sobre o comportamento das técnicas que não é facilmente analisando quando olhado apenas os valores de acurácia ou *f-score*.

Como extensão dos experimentos reportados pode ser feita a análise dos espaços encontrados para os diferentes métodos de geração artificial de imagens. Além disso, o impacto de tais métodos em diferentes extratores de características pode sugerir quais são as características latentes percebidas com cada extrator.

## 6.1 Publicações

## 6.2 Trabalhos Futuros

Ao usar imagens com reduzido número de cores (quantizadas), os métodos de extração de características baseados em orientação (HoG, SIFT), *bag of visual words* e *Fisher vectors*, seriam provavelmente mais esparços.

Em deep learning, pode ser investigado se o uso de imagens quantizadas ajudaria o aprendizado de características.

**Porque deep leaning? Mais descritores de forma e textura (dependentes da base) Explicar melhor o comportamento do SMOTE x ACC Bases maiores Algoritmos genéticos para geração de imagens artificiais?**

Atualmente, o estado da arte de extração e classificação de imagens corresponde ao uso de redes neurais de convolução, conhecidas por CNN (SCHMIDHUBER, 2014). Essas redes são compostas por camadas de neurônios que têm por objetivo aprender quais são as melhores características que diferenciam as classes de imagens. O aprendizado, nesse caso, corresponde ao ajuste dos parâmetros para reduzir a diferença entre a saída esperada – classe verdadeira – e a produzida. Dessa forma, tais redes aprendem quais são as características latentes nas imagens de entrada. As redes neurais são discutidas na Seção ??, onde as máquinas de Boltzmann restritas (RBM) também são apresentadas. A representação das imagens de entrada, aprendida pela etapa de treinamento da RBM, será utilizada para definir quais imagens são relevantes para o aprendizado ou não. De certa forma essas técnicas produzem versões processadas das imagens de entrada, indicando que os filtros aprendidos são os que melhor diferenciam as classes de imagens.

Investigar, também, o aprendizado de redes neurais de convolução, com o objetivo de observar quais são as características relevantes ao treiná-la com bases de imagens bem discriminadas em termos de cor, textura e forma;

Para definir quais são as imagens que acrescentam informações na base, primeiramente será treinada uma máquina de Boltzmann restrita (RBM). A partir da matriz de características aprendida (memória associativa), é possível verificar se uma imagem acrescenta informações à base ou não.

O próximo passo é utilizar a matriz de características aprendida por máquinas de Boltzmann restritas para verificar se as imagens artificialmente geradas são relevantes para o aprendizado ou não, além de melhor escolher as imagens originais para a geração dessas imagens.

A extração de características foi abordada, apresentando os principais descritores utilizados nesta pesquisa. A lacuna destacada é que existem características não passíveis de extração por descritores convencionais. Para isso, as redes de convolução são apresentadas, pois possuem capacidade de aprender as características relevantes das imagens de entrada. Justificando, assim, seu uso para análise das propriedades das bases de imagens. Podem também indicar possíveis operações para auxiliar na geração de imagens artificiais.

Ainda, a memória associativa aprendida por máquinas de Boltzmann restritas pode ser indicadora de quais imagens geradas adicionam informações ao aprendizado.

Esta pesquisa pretende dar continuidade a esse trabalho, ao analisar técnicas de *deep learning*. Essas técnicas realizam múltiplas operações sobre imagens de entrada de forma a aprender quais operações permitem gerar características capazes de discriminar as classes (BENGIO, 2009).

Com o objetivo de confirmar tais hipóteses, uma das propostas é analisar as características aprendidas por uma rede neural de convolução (CNN). Essa rede permite encontrar as características mais relevantes da base de imagens, que os extractores de características canônicos não capturaram. Isso porque ela possui uma hierarquia de camadas, desde a imagem original até uma etapa de classificação, com o objetivo de aprender qual o melhor processamento para as imagens de modo a melhor discriminar as classes.



## REFERÊNCIAS

---



---

- ABDI, H.; WILLIAMS, L. J. Principal Component Analysis. **Wiley Interdisciplinary Reviews: Computational Statistics**, v. 2, n. 4, p. 433–459, 2010. ISSN 19395108. Disponível em: <<http://dx.doi.org/10.1002/wics.101>>. Citado na página 47.
- AHONEN, T.; HADID, A.; PIETIKÄINEN, M. Face description with local binary patterns: application to face recognition. **IEEE transactions on pattern analysis and machine intelligence**, v. 28, n. 12, p. 2037–2041, dec 2006. ISSN 0162-8828. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/17108377>>. Citado na página 33.
- BATISTA, G. E.; PRATI, R. C.; MONARD, M. C. A study of the behavior of several methods for balancing machine learning training data. **ACM Sigkdd Explorations Newsletter**, v. 6, n. 1, p. 20–29, 2004. Disponível em: <<http://dl.acm.org/citation.cfm?id=1007735>>. Citado 2 vezes nas páginas 44 e 46.
- BENGIO, Y. Learning Deep Architectures for AI. **Foundations and Trends in Machine Learning**, Now Publishers Inc., v. 2, n. 1, p. 1–127, jan 2009. ISSN 1935-8237. Disponível em: <<http://dl.acm.org/citation.cfm?id=1658423.1658424>>. Citado na página 85.
- BOIMAN, O.; SHECHTMAN, E.; IRANI, M. In defense of Nearest-Neighbor based image classification. **IEEE Conference on Computer Vision and Pattern Recognition**, n. i, p. 1–8, 2008. Disponível em: <<http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4587598>>. Citado na página 46.
- BORGES, V. R. P.; OLIVEIRA, M. de; FERREIRA, T. G.; VIEIRA, A. A. H.; OLIVEIRA, M. C. F. de. Feature Extraction and Interactive Visualization to Assist Green Algae Taxonomic Classification. In: **XXVI SIBGRAPI - Conference on graphics, patterns and images**. Arequipa, Peru: [s.n.], 2013. p. 4. Disponível em: <<http://www.ucsp.edu.pe/sibgrapi2013/eproceedings/wip/115151.pdf>>. Citado 2 vezes nas páginas 33 e 39.
- BUADES, A.; COLL, B.; MOREL, J. M. A Review of Image Denoising Algorithms, with a New One. **Multiscale Modeling & Simulation**, Society for Industrial and Applied Mathematics, v. 4, n. 2, p. 490–530, jan 2005. ISSN 1540-3459. Disponível em: <<http://pubs.siam.org/doi/abs/10.1137/040616024>>. Citado na página 37.
- CASTRO, C. L.; BRAGA, A. P. Aprendizado supervisionado com conjuntos de dados desbalanceados. **Sba Controle & Automação**, v. 22, n. 5, p. 441 – 446, 2011. Citado na página 44.
- CHAWLA, N. V.; HALL, L. O.; BOWYER, K. W. SMOTE: Synthetic Minority Over-sampling Technique. **Journal of Artificial Intelligence Research**, v. 16, p. 321–357, 2002. Disponível em: <[http://link.springer.com/chapter/10.1007/3-540-45428-4{\\\_\}5http://arxiv.org/abs/1106.1](http://link.springer.com/chapter/10.1007/3-540-45428-4{\_\}5http://arxiv.org/abs/1106.1)>. Citado 2 vezes nas páginas 44 e 45.
- DALAL, N.; TRIGGS, B. Histograms of Oriented Gradients for Human Detection. In: **2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition**

(CVPR'05). IEEE, 2005. v. 1, p. 886–893. ISBN 0-7695-2372-2. ISSN 1063-6919. Disponível em: <<http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1467360>>. Citado na página 41.

GONZALEZ, R. C.; WOODS, R. E. **Digital Image Processing**. 3. ed. [S.l.]: Prentice-Hall, 2007. ISBN 013168728X. Citado 8 vezes nas páginas 32, 34, 36, 37, 39, 40, 55 e 59.

GROSS, R.; BRAJOVIC, V. An image preprocessing algorithm for illumination invariant face recognition. **Audio and Video-Based Biometric Person Authentication**, p. 10–18, 2003. Citado na página 36.

HAN, H.; WANG, W.-Y.; MAO, B.-H. Borderline-SMOTE: A new over-sampling method in imbalanced data sets learning. **Advances in intelligent computing**, Alvey Vision Club, v. 17, n. 12, p. 878–887, 2005. Citado na página 45.

HARALICK, R. M.; SHANMUGAM, K.; DINSTEIN, I. Textural Features for Image Classification. **IEEE Transactions on Systems, Man, and Cybernetics**, v. 3, n. 6, p. 610–621, nov 1973. ISSN 0018-9472. Disponível em: <<http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4309314>>. Citado na página 41.

HARRIS, C.; STEPHENS, M. A combined corner and edge detector. In: **Proceedings of the Alvey Vision Conference**. Alvey Vision Club, 1998. p. 147—152. Disponível em: <<http://www.bmva.org/bmvc/1988/avc-88-023.html>>. Citado na página 40.

HE, X.; NIYOGI, P. Locality preserving projections. In: **Neural information processing systems**. [S.l.: s.n.], 2004. v. 16, p. 153–160. ISBN 0262201526. Citado na página 48.

HUANG, J.; KUMAR, S. R.; MITRA, M.; ZHU, W.-J.; ZABIH, R. Image indexing using color correlograms. In: **Computer Society Conference on Computer Vision and Pattern Recognition**. IEEE, 1997. p. 762–768. ISBN 0-8186-7822-4. ISSN 1063-6919. Disponível em: <<http://ieeexplore.ieee.org/articleDetails.jsp?arnumber=609412>>. Citado na página 40.

JAPKOWICZ, N.; STEPHEN, S. The class imbalance problem : A systematic study. **Intelligent data analysis**, IOS Press, v. 6, n. 5, p. 429–449, 2002. Citado na página 46.

JOIA, P.; PAULOVICH, F. V.; COIMBRA, D.; CUMINATO, J. A.; NONATO, L. G. Local Affine Multidimensional Projection. **IEEE Transactions on Visualization and Computer Graphics**, v. 17, n. 12, p. 2563–2571, 2011. Citado na página 27.

KANAN, C.; COTTRELL, G. W. Color-to-grayscale: does the method matter in image recognition? **PloS one**, v. 7, n. 1, p. e29740, jan 2012. ISSN 1932-6203. Disponível em: <<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3300000/>>. Citado 2 vezes nas páginas 28 e 38.

LOWE, D. G. Distinctive image features from scale-invariant keypoints. **International Journal of Computer Vision**, v. 60, n. 2, p. 91–110, 2004. Disponível em: <<http://link.springer.com/article/10.1023/B:VISI.0000029664.99615.94>>. Citado na página 40.

OJALA, T.; PIETIKAINEN, M.; MAENPAA, T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 24, n. 7, p. 971–987, jul 2002. ISSN 0162-8828. Disponível em: <<http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1017623>>. Citado na página 42.

- PAIVA, J. G. S.; FLORIAN, L.; PEDRINI, H.; TELLES, G. P.; MINGHIM, R. Improved similarity trees and their application to visual data classification. **IEEE Transactions on Visualization and Computer Graphics**, IEEE, v. 17, n. 12, p. 2459–2468, 2011. Citado na página 27.
- PASS, G.; ZABIH, R.; MILLER, J. Comparing images using color coherence vectors. In: **Proceedings of the fourth ACM international conference on Multimedia**. New York, USA: ACM Press, 1996. p. 65–73. ISBN 0897918711. Disponível em: <<http://dl.acm.org/citation.cfm?id=244130.244148>>. Citado na página 41.
- PAULOVICH, F. V.; OLIVEIRA, M. C. F.; MINGHIM, R. The Projection Explorer: A Flexible Tool for Projection-based Multidimensional Visualization. In: **XX Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI 2007)**. IEEE, 2007. p. 27–36. ISBN 0-7695-2996-8. ISSN 1530-1834. Disponível em: <<http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4368165>>. Citado na página 47.
- PENATTI, O. A.; VALLE, E.; TORRES, R. d. S. Comparative study of global color and texture descriptors for web image retrieval. **Journal of Visual Communication and Image Representation**, v. 23, n. 2, p. 359–380, feb 2012. ISSN 10473203. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1047320311001465>>. Citado na página 51.
- PICON, C. T.; ROSSI, I.; JR, M. P. P. Análise da classificação de imagens por descritores de cor utilizando várias resoluções. In: **Workshop of Undergraduate Works**. SIBGRAPI, 2011. Disponível em: <[http://www.icmc.usp.br/{~}moacir/papers/Picon{\\\_}WUW201](http://www.icmc.usp.br/{~}moacir/papers/Picon{\_}WUW201)>. Citado na página 27.
- PONTI, M.; ESCOBAR, L. Compact color features with bitwise quantization and reduced resolution for mobile processing. In: **Global Conference on Signal and Information Processing**. Austin, TX, Estados Unidos: [s.n.], 2013. p. 751–754. Disponível em: <<http://link.springer.com/article/10.1007/s11760-011-0216-x>>. Citado na página 38.
- PONTI, M.; MASCARENHAS, N.; FERREIRA, P.; SUAZO, C. Three-dimensional noisy image restoration using filtered extrapolation and deconvolution. **Signal, Image and Video Processing**, v. 7, n. 1, p. 1–10, feb 2011. ISSN 1863-1703. Disponível em: <<http://link.springer.com/10.1007/s11760-011-0216-x>> e <<http://link.springer.com/article/10.1007/s11760-011-0216-x>>. Citado na página 35.
- PONTI, M.; NAZARÉ, T. S.; THUMÉ, G. S. Image quantization as a dimensionality reduction procedure in color and texture feature extraction. **Neurocomputing**, v. 173, p. 385–396, jan 2016. ISSN 09252312. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0925231215012771>>. Citado 8 vezes nas páginas 52, 53, 54, 69, 71, 72, 73 e 74.
- PONTI, M. P. Microscope volume segmentation improved through non-linear restoration. **Nature-Inspired Computing Design, Development, and Applications**, IGI Global, 2012. Disponível em: <<http://www.igi-global.com/article/microscope-segmentation-improved-through-non/52614>>. Citado 2 vezes nas páginas 32 e 34.
- \_\_\_\_\_. Segmentation of Low-Cost Remote Sensing Images Combining Vegetation Indices and Mean Shift. **Geoscience and Remote Sensing Letters**, IEEE, v. 10, n. 1, p. 67–70, jan 2013. ISSN 1545-598X. Disponível em: <<http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6202674>>. Citado na página 27.

ROCHA, A.; HAUAGGE, D. C.; WAINER, J.; GOLDENSTEIN, S. Automatic fruit and vegetable classification from images. **Computers and Electronics in Agriculture**, v. 70, n. 1, p. 96–104, jan 2010. ISSN 01681699. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S016816990900180X>>. Citado na página 28.

SCHMIDHUBER, J. Deep learning in neural networks: An overview. **Neural Networks**, p. 88, oct 2014. ISSN 08936080. Disponível em: <<http://arxiv.org/abs/1404.7828>>. Citado na página 84.

SMITH, S. M.; BRADY, J. M. SUSAN—a new approach to low level image processing. **International journal of computer vision**, v. 23, n. 1, p. 45–78, 1997. Disponível em: <<http://link.springer.com/article/10.1023/A:1007963824710>>. Citado na página 40.

STEHLING, R. O.; NASCIMENTO, M. A.; FALCÃO, A. X. A compact and efficient image retrieval approach based on border/interior pixel classification. In: **Proceedings of the eleventh international conference on Information and knowledge management**. New York, USA: ACM Press, 2002. p. 102–109. ISBN 1581134924. Disponível em: <<http://dl.acm.org/citation.cfm?id=584792.584812>>. Citado na página 40.

TOMASI, C.; MANDUCHI, R. Bilateral filtering for gray and color images. In: **Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271)**. Narosa Publishing House, 1998. p. 839–846. ISBN 81-7319-221-9. Disponível em: <<http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=710815>>. Citado 2 vezes nas páginas 56 e 57.

WANG, X.; HAN, T. X.; YAN, S. An HOG-LBP human detector with partial occlusion handling. **IEEE 12th International Conference on Computer Vision**, IEEE, 2009. Citado na página 40.

ZHUO, L.; CHENG, B.; ZHANG, J. A comparative study of dimensionality reduction methods for large-scale image retrieval. **Neurocomputing**, Elsevier, v. 141, p. 202–210, oct 2014. ISSN 09252312. Disponível em: <<http://linkinghub.elsevier.com/retrieve/pii/S0925231214004238>>. Citado na página 48.