

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ – UTFPR  
DIRETORIA DE GRADUAÇÃO E EDUCAÇÃO PROFISSIONAL  
CURSO DE BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

JHONATTAN SALVADOR GHELLERE

**CLASSIFICAÇÃO DE IMAGENS A PARTIR DO RECONHECIMENTO  
DE OBJETOS GENÉRICOS**

PROJETO

MEDIANEIRA

2015

JHONATTAN SALVADOR GHELLERE

## **CLASSIFICAÇÃO DE IMAGENS A PARTIR DO RECONHECIMENTO DE OBJETOS GENÉRICOS**

Projeto apresentado como requisito parcial de nota junto a disciplina de Trabalho de Conclusão de Curso I, do curso de Ciência da Computação, da Universidade Tecnológica Federal do Paraná – UTFPR – Câmpus Medianeira.

Orientador: Prof. Jean Metz, Dr.  
Co-Orientadora: Prof<sup>a</sup>. Alessandra Bortolotto Garbelotti Hoffmann, Msc.

MEDIANEIRA

2015

## LISTA DE FIGURAS

Figura 1 - Imagem formada por três canais de cores.....	13
Figura 2 - Modelos de Vizinhaça.....	14
Figura 3 - Exemplo de <i>Simple-Matching</i> .....	18
Figura 4 - Etapas para obtenção do descritor SIFT.....	19
Figura 5 - Exemplo da função <i>DoG</i> para cada oitava.....	20
Figura 6 - Exemplo da aplicação DoG em cada oitava. A busca por extremos é realizada apenas nas imagens dentro do intervalo em vermelho.....	21
Figura 7 - Detecção de extremos no espaço escala dos intervalos. ....	22
Figura 8 - Histograma dos gradientes. ....	23
Figura 9 - Exemplo de Histograma para cada uma das 16 regiões.....	23
Figura 10 - Região de aplicação da Gaussiana para enfatizar os pontos vizinhos. ..	24
Figura 11 - Aplicação das derivadas de segunda ordem da Gaussiana e aplicação do Filtro Caixa. ....	25
Figura 12 - Descritor SURF. ....	25
Figura 13 - Padrão de amostragem FREAK.....	27
Figura 14 – Os 512 pares seleccionados divididos em quatro grupos (128 por grupo). ....	28
Figura 15 - Pares seleccionados para computar a orientação. ....	29
Figura 16 - Processo de Descoberta de Conhecimento.....	31

## LISTA DE QUADROS

Quadro 1 - Cronograma 2014 .....	39
Quadro 2 - Cronograma 2015 .....	40

## SUMÁRIO

<b>1 INTRODUÇÃO .....</b>	<b>6</b>
1.1 OBJETIVO GERAL .....	7
1.2 OBJETIVOS ESPECÍFICOS .....	7
1.3 JUSTIFICATIVA .....	8
<b>2 PROCESSAMENTO DIGITAL DE IMAGENS E EXTRAÇÃO DE CARACTERÍSTICAS .....</b>	<b>10</b>
2.1 CARACTERÍSTICAS DA IMAGEM .....	15
2.2 DESCRITORES DA IMAGEM .....	17
2.2.1 Scale Invariant Feature Transforms – SIFT .....	18
2.2.2 Speeded-Up Robust Features – SURF .....	24
2.2.3 Fast Retina Keypoint – FREAK .....	26
<b>3 RECONHECIMENTO DE PADRÕES EM IMAGENS .....</b>	<b>30</b>
3.1 RECONHECIMENTO DE OBJETOS .....	32
3.2 APRENDIZAGEM DE MÁQUINA .....	33
3.2.1 Aprendizagem Supervisionada .....	33
3.2.2 Aprendizagem Não Supervisionada .....	33
3.2.3 Classificação .....	33
3.2.4 Agrupamento .....	33
3.3 BAG OF WORDS .....	33
3.4 AVALIAÇÃO DA CLASSIFICAÇÃO .....	33
3.4.1 Medidas .....	33
3.4.2 Validação Cruzada .....	33
3.4.3 Teste de Hipótese .....	33
<b>4 MATERIAL E MÉTODOS .....</b>	<b>34</b>
4.1 FERRAMENTAS .....	34
4.1.1 Pré-Processamento .....	34
4.1.2 Aprendizagem e Classificação .....	35
4.1.3 Base de Dados .....	35
4.2 ANÁLISE E PROJETO .....	35
4.2.1 Requisitos do Sistema .....	36
4.2.2 Diagrama de Classes .....	36

4.2.3 Diagrama de Componentes.....	36
4.2.4 Diagrama de Atividades .....	36
4.3 IMPLEMENTAÇÃO E VALIDAÇÃO .....	37
<b>5 RESULTADOS ESPERADOS .....</b>	<b>38</b>
<b>6 CRONOGRAMA .....</b>	<b>39</b>
<b>REFERÊNCIAS.....</b>	<b>41</b>

## 1 INTRODUÇÃO

A visão computacional pode ser definida como um conjunto de técnicas e teorias para obtenção de informação por meio de imagens, sendo de modo geral, o estudo da capacidade das máquinas de visualizarem o mundo real (tridimensional), por meio de sensores, câmeras e outros dispositivos que extraem informação dos ambientes (ALVES et al., 2005).

Essa área é repleta de problemas com soluções parciais. Dentre estes problemas pode-se citar o reconhecimento de objetos, o qual possui inúmeras soluções, mas todas implementadas para resolução de situações específicas, não existindo uma solução que contemple uma resolução genérica para o reconhecimento de objetos.

Este problema no contexto da visão computacional, por exemplo, consiste em identificar a existência de objetos específicos em um cenário e, a partir desses objetos, reconhecê-los ou categorizá-los de acordo com uma necessidade qualquer, como por exemplo, identificá-lo em outras imagens.

Este problema pode ser resolvido de maneira específica, por meio do reconhecimento da forma do objeto desejado, como a identificação desta mesma forma no cenário de destino. Frequentemente essa técnica é chamada de *simple-matching* ou correspondência perfeita entre as imagens. No entanto, essa solução não permite o reconhecimento de outro objeto similar, por exemplo identificar a presença de uma face humana genérica em uma imagem a não ser que seja exatamente a face desejada.

Neste contexto, apresenta-se a intrínseca relação do processo de reconhecimento e a área de inteligência artificial. A aplicação das técnicas presentes nesta área, como aprendizado supervisionado, permite utilizar as informações obtidas pela visão computacional para fornecer ao sistema a capacidade de interpretação e aprendizado, e com isso, a possibilidade de se gerar conhecimento com base naquilo que foi 'visto'; conhecimento este que pode ou não ser usado em processos de tomada de decisão automática.

Com intuito de resolver o problema de classificar imagens a partir do reconhecimento de objetos genéricos, pretende-se com este projeto estudar e avaliar métodos de extração de características a partir do processamento digital de imagens, bem como técnicas de inteligência computacional para a tarefa de classificação, e com isso

encontrar uma combinação satisfatória dessas técnicas para solucionar este problema com base em imagens provenientes de redes sociais.

## 1.1 OBJETIVO GERAL

Identificar e aplicar técnicas de Inteligência Computacional e Processamento Digital de Imagens, na análise e localização de objetos genéricos em imagens provenientes de redes sociais.

## 1.2 OBJETIVOS ESPECÍFICOS

Para organizar o processo de desenvolvimento do trabalho, definiu-se os seguintes objetivos específicos:

- a) Identificar os métodos de processamento digital de imagens para a detecção de pontos de interesse em imagens e, com base na literatura, selecionando o de maior relevância para a solução do problema proposto.
- b) Avaliar os métodos de extração de características de imagens com ênfase nos que possuem invariância a transformações.
- c) Identificar e selecionar bibliotecas de processamento de imagens e reconhecimento de padrões.
- d) Avaliar os métodos de reconhecimento de padrões.
- e) Analisar, determinar e avaliar uma abordagem para a tarefa de classificação das imagens extraídas considerando os paradigmas de Inteligência Artificial.
- f) Analisar, projetar e desenvolver um componente responsável pelo reconhecimento dos objetos e classificação das imagens.



### 1.3 JUSTIFICATIVA

A comunicação atualmente feita por intermédio das redes sociais, não ocorre somente em texto, visto que o número de pessoas que acabam utilizando imagens para resumir pensamentos e até mesmo atos é crescentemente notório. Este fato está atrelado a necessidade natural do ser humano em relacionar-se, e que pode acabar levando-o a exposição pública de suas informações, gostos e opiniões seja por meio de textos como também imagens ou vídeos. Um exemplo dessa exposição, se tem verificado com a popularização dos *selfies* (LAGAREIRO, 2014), que é o ato de fotografar a si mesmo estando ou não em companhia de outras pessoas.

Diversos estudos têm sido desenvolvidos (LECUN et al., 2014; LOWE, 2004; ALAHI et al., 2012, BAY et al., 2006) de modo a descobrir técnicas de reconhecimento e classificação de imagens, mesmo após elas terem sofrido algum tipo de transformação (escala e perspectiva por exemplo).

O processo de reconhecimento de objetos, entretanto, não é trivial. Segundo LECUN et al. (2014), o reconhecimento genérico com invariância a perspectiva, iluminação e a presença de ruídos é um dos maiores desafios da área de Visão Computacional.

Normalmente a correspondência entre duas imagens é feita usando uma cópia transformada, como por exemplo na detecção de semi-réplicas, que consiste na consulta das imagens sofreram alguma transformação a partir da imagem original (BUENO, 2011). Com o presente trabalho procura-se realizar a classificação de imagens a partir da correspondência da mesma com imagens que representam objetos genéricos, que poderão estar presentes dentro das imagens que sofreram transformações de escala, perspectiva, iluminação e rotação.

Sendo assim, com a associação (por meio da geração e compartilhamento) de imagens nos mais diversos assuntos todos os dias cria-se a necessidade de classificá-las (BOSCH et al., 2007). A partir das características extraídas, surge a importância prática de detectar objetos, que uma vez reconhecidos podem revelar posteriormente a descoberta de grupos de pessoas que compartilham mesmos interesses criando-se perfis ou até mesmo informações relevantes a segurança pública, alguns exemplos são a detecção de conteúdo impróprio como nudez (CARVALHO, 2012), objetos 3D

dentro de bagagens por meio de imagens de raios-x (FLITTON et al., 2010) e/ou de armas de fogo, o que possibilitará a identificação do usuário associado a esse perfil.

## 2 PROCESSAMENTO DIGITAL DE IMAGENS E EXTRAÇÃO DE CARACTERÍSTICAS

Processamento Digital de Imagens é uma área que se dedica ao estudo de teorias, modelos e algoritmos para a manipulação de imagens, feito por meio de computadores (BEZDEK et al., 2005), tendo origem, portanto, junto ao desenvolvimento dos mesmos. O primeiro registro de sua aplicação foi na utilização de técnicas para corrigir deformações no que foi a primeira imagem capturada da lua, ocorrida na década de 1960 com o desenvolvimento de programas espaciais norte-americanos (GONZALEZ e WOODS, 2002).

A partir deste ponto juntamente com a maior acesso aos computadores devido a diminuição do custo e popularização, assim como o aumento de desempenho do processamento, cresceu o número de áreas com aplicações, dentre elas:

- *Medicina e Biologia*: O PDI tornou possível a criação de diversos aparelhos que hoje são utilizados na medicina para aumentar a eficiência e rapidez do diagnostico, como por exemplo a medicina nuclear, que utiliza a emissão coletada da detecção de raios gama para gerar imagens e localizar tumores em paciente (GONZALEZ e WOODS, 2002; ACHARYA e RAY, 2005; POLAKOWSKI et al., 1997; DOUGHERTY, 2009). Processar imagens de microscópio para contagem de células ou aumentar a precisão das tarefas de laboratório (DOUGHERTY, 2009; PEDRINI e SCHWARTZ, 2008) identificação de doenças no coração, realizar exames de mamografias digitais (ACHARYA e RAY, 2005). Classificação e análise do genoma (DOUGHERTY, 2009).
- *Astronomia, Sensoriamento Remoto e Meteorologia*: muito estudos tem estão sendo realizados na área de reconhecimento de padrões com aplicações na astronomia, como por exemplo, reconhecimento de supernovas por meio de aprendizagem de máquina (ROMANO et al., 2006), reconhecimento de objetos (TAGLIAFERRI et al., 1999), classificação de galáxias (DE LA CALLEJA e FUENTES, 2004), além disso, imagens astronômicas podem ser obtidas da emissão de raios gama de estrelas que explodiram a milhares de anos (STARCK e MURTAGH, 2006), e imagens recebidas de satélites a partir do espaço podem ser processadas para análise da situação

ambiental e climática do planeta (ACHARYA e RAY, 2005). Detecção automática de tempestades solares (DOUGHERTY, 2009).

- *Segurança*: o uso de PDI na área de segurança é bastante útil para barrar o acesso não autorizado a locais ou recursos, e sua aplicação é evidente por meio do uso de reconhecedores de impressão digital, íris e até mesmo face, que servem como uma alternativa automatizada e até mesmo mais segura que os métodos tradicionais, para controlar acessos não autorizados, já que uma impressão digital é mais difícil de ser reproduzida que um cartão eletrônico (BOURIDANE, 2009).

Cada vez mais tem crescido o número de pesquisas na área para aumentar o desempenho na extração de características das imagens para posterior reconhecimento de padrões, porém, segundo GONZALEZ e WOODS (2002) não há consenso na literatura para estabelecer um limiar entre o Processamento Digital de Imagens, Visão Computacional e a Inteligência Artificial. A relação entre, quando uma começa e outra termina ou se, uma está contida na outra, ainda não é muito clara; Para tentar compreender melhor a relação entre as áreas, eles imaginaram uma linha contínua com o processamento digital de imagens em um extremo e a visão computacional no outro, dividida em três tipos de processos: processo nível baixo, médio e alto.

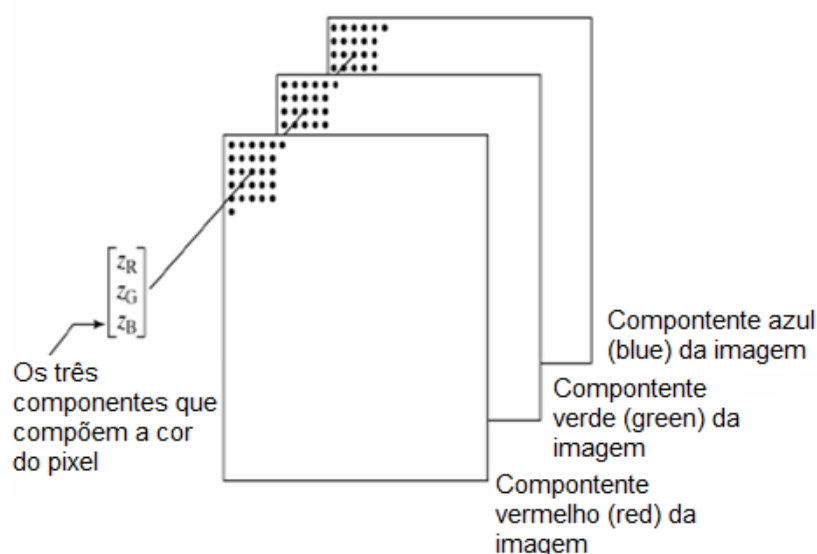
- *Processos de Nível Baixo*: Dentro desse nível se encontram as tarefas que vão desde a aquisição da imagem até filtragens, realces e outras tarefas de pré-processamento (GONZALEZ e WOODS, 2002). A aquisição da imagem depende muito do domínio do problema, elas podem ser geradas ou obtidas a partir de diversas fontes (i.e. raios gama, raios-x, sonar, ondas de rádio, micro-ondas, espectro visível, infravermelho, dentre outras) e meios (sensores, câmeras, satélites etc), e dentre as particularidades que devem ser levadas em conta nessa tarefa estão na escolha do meio de capturar a imagem, as cores, níveis de cinza e condições do ambiente (PEDRINI e SCHWARTZ, 2008; GONZALEZ e WOODS, 2002). Geralmente a aquisição de imagens a partir dessas fontes envolve alguma forma de pré-processamento, como por exemplo, redimensionar a imagem, aumentar ou diminuir o contraste para adequar a imagem à percepção da visão humana, visando corrigir imperfeições que podem ter ou não sido geradas pela etapa anterior (PEDRINI e SCHWARTZ, 2008). O pré-processamento das imagens é res-

ponsável por operações primitivas para redução ou inclusão de ruídos, realce e restauração da imagem para que a mesma possa se enquadrar melhor no domínio do problema, tendo ao final apenas resultados com relação a qualidade da imagem (CONCI et al., 2008). Sendo assim, um processo de nível baixo é caracterizado por tanto a entrada quanto a saída do processo serem uma imagem (GONZALEZ e WOODS, 2002).

- *Processos de Nível Médio:* Neste nível estão tarefas como: segmentação, descrição e o reconhecimento. O processo de segmentação consiste em identificar regiões ou objetos, que sejam pertinentes ao domínio do problema e separá-los da imagem, com base em características como bordas e texturas, por exemplo (PEDRINI e SCHWARTZ, 2008). Realizar esta tarefa de forma autônoma, é um dos problemas mais difíceis dentro da área, pois ainda não há uma implementação genérica que realize a segmentação identificando objetos e regiões de forma a cobrir todos os que sejam de interesse, e esta tarefa é especialmente importante para se identificar objetos separados (GONZALEZ e WOODS, 2002). Extrair e descrever as características da imagem, é próximo passo no processamento, regiões ou objetos que podem ter sido segmentados tem dados extraídos que possibilitem comparar ou diferenciar um objeto do outro, representados normalmente por vetores com atributos numéricos (PEDRINI e SCHWARTZ, 2008). Após a extração das informações que representam as características da imagem ou objeto é realizado o reconhecimento. Esta etapa consiste em aplicar um tipo ao objeto com base nos dados extraídos de modo a dar um significado ao mesmo, como por exemplo aplicar o rotulo “animal” ou “avião” (PEDRINI e SCHWARTZ, 2008; GONZALEZ e WOODS, 2002). Os processos de nível médio são caracterizados por entradas serem imagens mas as saídas serem dados (GONZALEZ e WOODS, 2002).
- *Processos de Nível Alto:* Grande parte das pesquisas de visão computacional são voltadas para problemas “o que” e “onde”, como por exemplo detecção e classificação de objetos, porém é possível pensar além dessas tarefas, como classificar imagem se baseando em conceitos abstratos, como pacificidade (CIPOLLA et al., 2012), tarefas como esta estão ligadas com a área de cognição e de Inteligência Artificial. Elas buscam dar sentido ao que a máquina vê, ao mesmo tempo que tentam emular a inteligência humana,

de modo a possibilitar que a mesma aprenda, faça inferências e tire conclusões com base no que está enxergando no ambiente (GONZALEZ e WOODS, 2002).

Uma imagem digital, formada por um único canal de cor, pode ser definida como uma função bidimensional  $f(x, y)$  sendo que  $(x, y)$  indica a coordenada espacial pertencente à  $\mathbb{R}^2$  que representa a posição de um ponto, e o valor de  $f$  representa a intensidade ou nível de cinza do ponto na imagem (BALLARD e BROWN, 1982). Uma imagem que possui mais de um canal de cor terá uma função representando cada canal de cor, como por exemplo o padrão BGR (*Blue, Green, Red*) (LIBERMAN, 1997; BALLARD e BROWN, 1982), que é esquematizado na Figura 1. Nota-se que, um ponto  $(x, y)$  no espaço da imagem é representado por vetor no qual cada componente é um valor escalar representando respectivamente a intensidade de cada cor.



**Figura 1 - Imagem formada por três canais de cores.**  
**Fonte: Paula Filho (2014)**

A representação da imagem pode ser feita computacionalmente por meio de uma matriz  $M(i, j)$  formada por um número de elementos finitos denominados *pixels*, que possuem um valor escalar e uma localização  $(i, j)$  a partir da origem da imagem.

Cada *pixel* possui uma relação com seu vizinho. Segundo GONZALEZ e WOODS (2002) cada pixel  $p$  possui quatro vizinhos horizontais e verticais representados pelas coordenadas:

$$(x + 1, y), (x - 1, y), (x, y + 1), (x, y - 1) \quad (1)$$

Esse conjunto de pontos é também chamado de *vizinhança-4* de  $p$ , e pode ser expresso por  $N_4(p)$ .

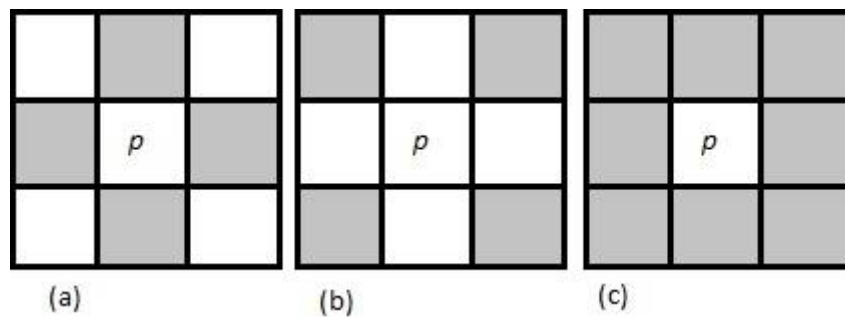
Há também o conjunto de vizinhos diagonais de  $p$ , expresso por  $N_D(p)$  e é representado pelas coordenadas:

$$(x + 1, y + 1), (x + 1, y - 1), (x - 1, y - 1), (x - 1, y + 1) \quad (2)$$

Seguindo essa ideia, há também a chamada vizinhança-8 de  $p$ , descrita por (PEDRINI e SCHWARTZ, 2008) como:

$$N_8(p) = N_4(p) \cup N_D(p)$$

Onde  $N_8(p)$  contém todas as coordenadas de  $N_4(p)$  e  $N_D(p)$ , a não ser que o pixel  $p$  esteja localizado na borda da imagem, neste caso algumas coordenadas não aparecerão no conjunto (GONZALEZ e WOODS, 2002). Por meio da Figura 2, é apresentada os modelos de vizinhança, sendo (a) a vizinhança-4 de  $p$ , (b) vizinhança-D de  $p$  e (c) a vizinhança-8 de  $p$ .



**Figura 2 - Modelos de Vizinhança.**  
**Fonte: Adaptado de GONZALEZ e WOODS (2002)**

## 2.1 CARACTERÍSTICAS DA IMAGEM

O método mais conhecido para procurar por imagens é compará-las diretamente. Isso é feito com a correspondência dos valores de um pixel ou um conjunto de pixels (região) de uma imagem com os valores de uma outra imagem (DESELAERS, 2003).

Esses conjuntos de pixels uma vez analisados podem revelar as características de uma imagem, que são propriedades cujos valores devem ser semelhantes para os objetos em uma classe particular, e diferentes para os objetos em outra classe ou mesmo em relação ao fundo da imagem (DOUGHERTY, 2013).

Portanto, para que o computador consiga processar os dados de forma a reconhecer semanticamente o que está presente na imagem, o primeiro passo é detectar e extrair características de forma eficiente e efetiva (TIAN, 2013). Entre os exemplos dos detalhes e características que os humanos conseguem extrair de uma imagem com percepção visual estão a identificação de picos em montanhas, cantos e portas de construções, caminhos através da neve, (SZELISKI, 2010), uma região de textura.

A escolha das características corretas da imagem depende muito do domínio do problema ou da própria imagem, mas geralmente procura-se por características que sejam invariantes a possíveis transformações que a imagem possa sofrer, como translação, mudança no tamanho e rotação (DOUGHERTY, 2013; NIXON e AGUADO, 2002).

As características de uma imagem podem ser extraídas a partir de informações locais e/ou globais. Características globais tem a propriedade de generalizar um objeto inteiro com apenas um vetor (LISIN et al., 2005) e geralmente algoritmos que extraem informações globais tendem a ser mais rápidos e simples (PENATTI, 2009) e com boa tolerância a ruídos (DE ARAÚJO, 2009). Alguns exemplos de características globais extraídas de uma imagem pode ser: cor (RAOUI et al., 2011; PAULA FILHO, 2013) geralmente extraídas por meio de um histograma (PENATTI, 2009), textura (PAULA FILHO, 2013; LIBERMAN, 1997) e forma (DESELAERS, 2003; LISIN et al., 2005; RAOUI et al., 2011).

Contudo, características globais localizam poucos detalhes e são ineficientes quando a imagem sofre alguma alteração em sua perspectiva, como por exemplo, ser rotacionada, transladada, e ao mesmo tempo são sensíveis a oclusões, desordens ou



deformações no objeto (GRAUMAN e LEIBE, 2011; DE ARAÚJO, 2009; LISIN et al., 2005; DE ANDRADE, 2012).

Uma característica local, que é o foco deste trabalho, é descrita como um padrão dentro da imagem que difere de sua vizinhança, associada a mudança de uma ou várias propriedades da mesma, como intensidade, cor ou textura, podendo ser representadas por pontos de interesse, bordas ou trechos da imagem (TUYTELAARS e MIKOLAJCZYK, 2008).

Pontos de interesse são pontos que possuem uma posição bem definida dentro da imagem e cuja região a sua volta é rica em informação, tornando-o um ponto único, ou seja, que não se repete e distingue-se dos demais. Estes pontos, também conhecidos como pontos chaves, segundo (SCHMID et al., 1998), são comumente localizações na imagem onde o sinal muda em duas dimensões, ou ainda onde há a intersecção de duas bordas ou onde a própria borda da imagem muda subitamente.

Para (TUYTELAARS e MIKOLAJCZYK, 2008) características locais ideais devem ter as seguintes propriedades:

- *Repetição*: Dadas duas imagens do mesmo objeto ou cena, retiradas de condições diferentes de visualização, uma elevada percentagem das características detectadas na parte visível de ambas as imagens, devem ser encontradas em ambas as imagens.
- *Distinção / Informação*: Os padrões de intensidade subjacentes às características detectadas devem mostrar muita variação, de tal forma que as características possam ser distinguidas e correspondidas.
- *Local*: As características devem ser locais, de modo a reduzir a probabilidade de oclusão e permitir simples aproximações de modelo das deformações geométricas e fotométricas entre duas imagens obtidas sob diferentes condições de visualização.
- *Quantidade*: O número de características detectadas deve ser suficientemente grande, de tal modo que um número razoável de características possa ser detectado mesmo em pequenos objetos. No entanto, o número ótimo de características depende da aplicação. A densidade de características deve refletir o conteúdo de informação da imagem para fornecer uma representação compacta da imagem.

- *Precisão*: As características detectadas devem ser localizadas com precisão, tanto no local da imagem, quanto no que diz respeito à escala e forma.
- *Eficiência*: A detecção de características em uma nova imagem deve ser eficiente computacionalmente considerando a aplicação.

Ainda Segundo (TUYTELAARS e MIKOLAJCZYK, 2008) a repetição é a propriedade mais importante entre as citadas e pode ser alcançada de duas maneiras: invariância e robustez.

- *Invariância*: Quando são esperadas grandes deformações, a abordagem preferencial é modelar matematicamente essas deformações, se possível, e, em seguida, desenvolver métodos para detecção de características que não são afetados por essas transformações matemáticas.
- *Robustez*: No caso de pequenas deformações, que muitas vezes é suficiente para tornar os métodos de detecção de características menos sensível a tais deformações, isto é, a precisão da detecção pode diminuir, mas não tão drasticamente. Deformações típicas que são abordadas utilizando robustez são o ruído de imagem artefatos de compressão, borrão, etc.

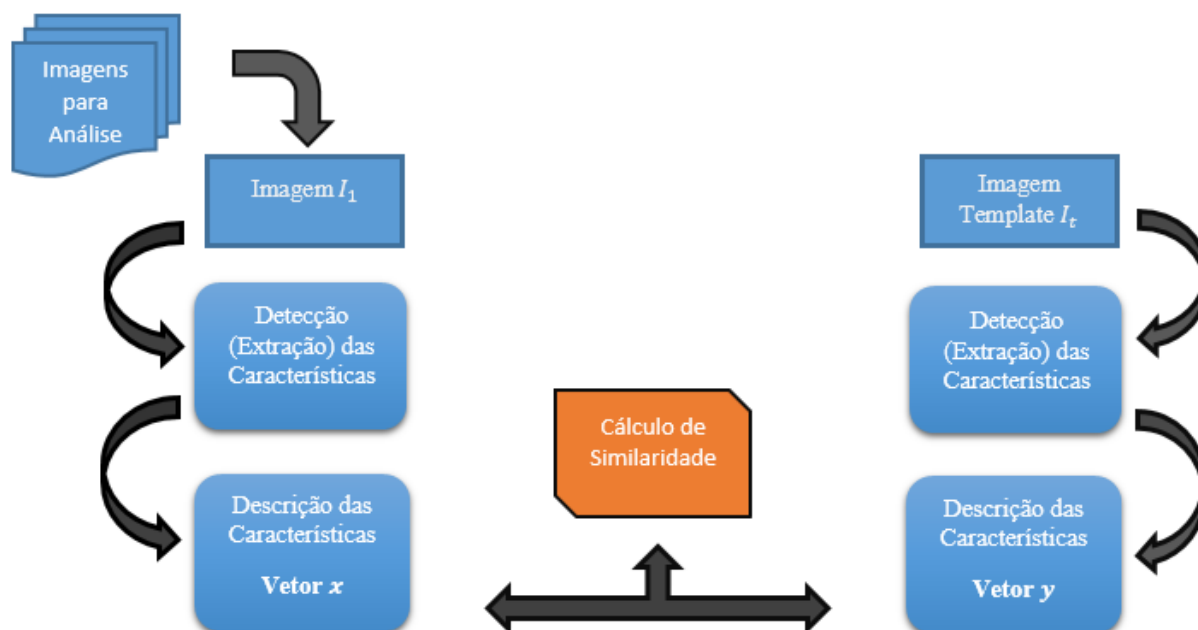
Tais características podem ser usadas para encontrar conjuntos esparsos de correspondências locais em diferentes imagens (SZELISKI, 2010) e portanto formam a base para as abordagens de reconhecimento de objetos específicos ou uma classe de objetos (GRAUMAN e LEIBE, 2011).

## 2.2 DESCRITORES DA IMAGEM

Após a detecção das características é preciso fazer a correspondência, ou seja, saber quais características detectadas correspondem em outras imagens (SZELISKI, 2010).

Mas para realizar a correspondência para classificação a partir do reconhecimento de um objeto (ou vários) é preciso descrever as características detectadas (extraídas) do conjunto de pixels (ou pontos de interesse), geralmente representadas por um conjunto de números (vetor) que representam as características do objeto, de modo que seja possível classificar uma imagem (NIXON e AGUADO, 2002). Um exemplo é o *Simple-Matching* que é demonstrado na Figura 3, na qual é comparado

através de uma medida de similaridade os descritores extraídos a partir das características detectadas de duas imagens, tendo como base uma Imagem *Template* que representa o objeto de interesse que se procura detectar nas imagens a serem analisadas.



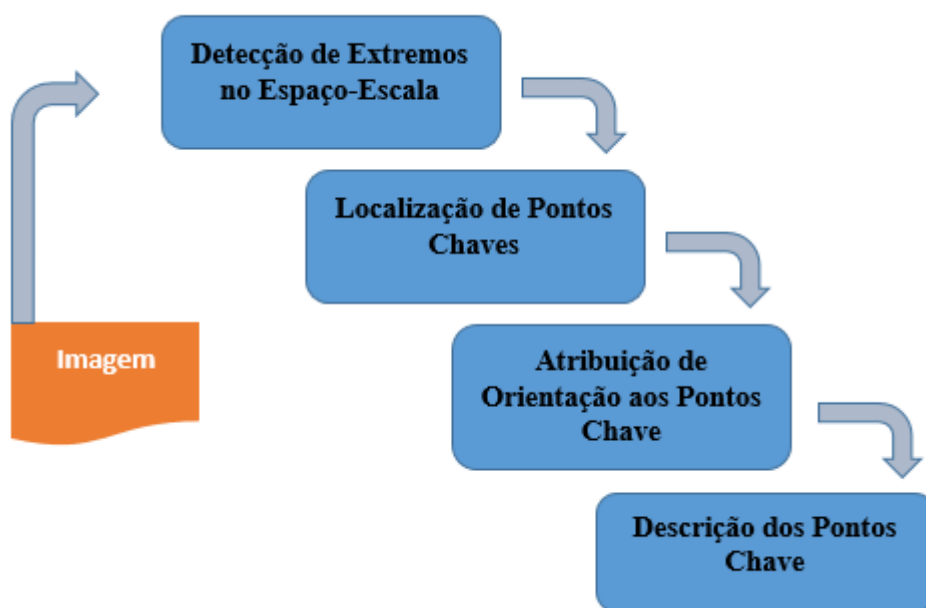
**Figura 3 - Exemplo de *Simple-Matching***  
 Fonte: Autoria Própria.

### 2.2.1 Scale Invariant Feature Transforms – SIFT

O algoritmo SIFT (*Scale Invariant Feature Transforms*) foi proposto por Lowe para a extração de características que fossem invariantes, de modo a possibilitar a correspondência de objetos dentro de cenas mesmo sobre diferentes pontos de perspectiva, após terem sofrido alguma transformação, seja na escala, rotação ou iluminação (LOWE, 2004). A característica de invariância tornou o algoritmo bastante reconhecido, tendo bastante aplicabilidade em diversos contextos de visão computacional e aprendizagem de máquina, como por exemplo, criação de imagens panorâmicas através da junção de várias imagens (BROWN e LOWE, 2007), mapeamento e navegação de robôs, identificação de digitais (PARK et al., 2008), reconhecimento de

objetos (LOWE, 1999; SUGA, 2008), reconhecimento de objetos 3D (FLITTON, 2010) dentre outros.

A obtenção de descritores é feita segundo Lowe (2004) em quatro etapas, conforme a Figura 4, e serão descritas brevemente a seguir.

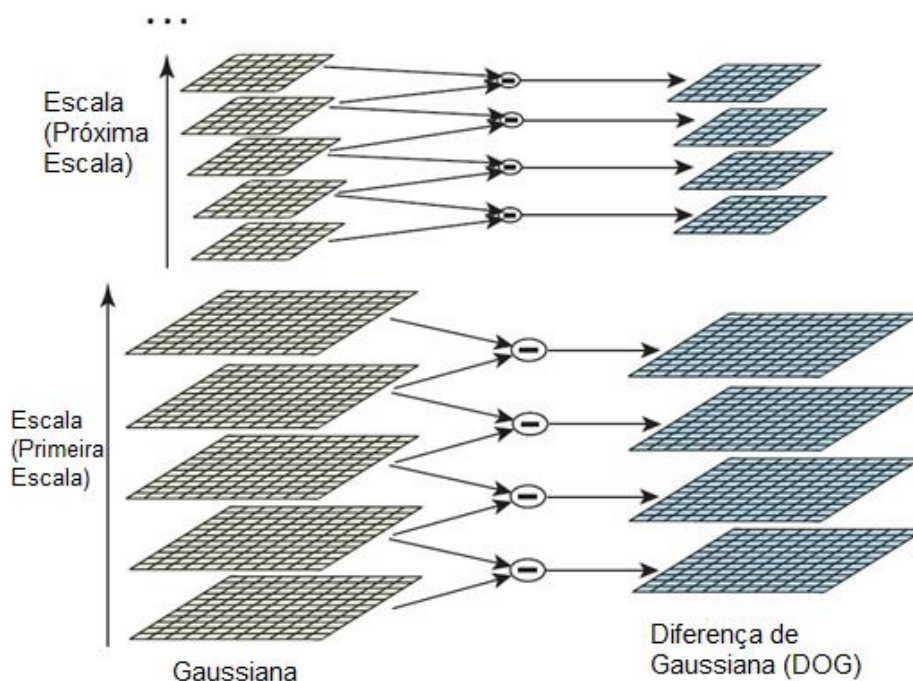


**Figura 4 - Etapas para obtenção do descritor SIFT.**  
**Fonte: Autoria Própria.**

A detecção de extremos no espaço-escala é realizada identificando potenciais candidatos a pontos de interesse, e de acordo com Lowe (2004) procurando extremos (valores locais máximos e mínimos), por meio da aplicação da função de diferença Gaussiana num espaço de escala de uma pirâmide de imagem, que é construída por meio das chamadas oitavas (*octaves*). A cada oitava a imagem inicial se torna a imagem cujo valor  $\sigma$  é o dobro do inicial, ao mesmo tempo a imagem tem sua escala reduzida pela metade.

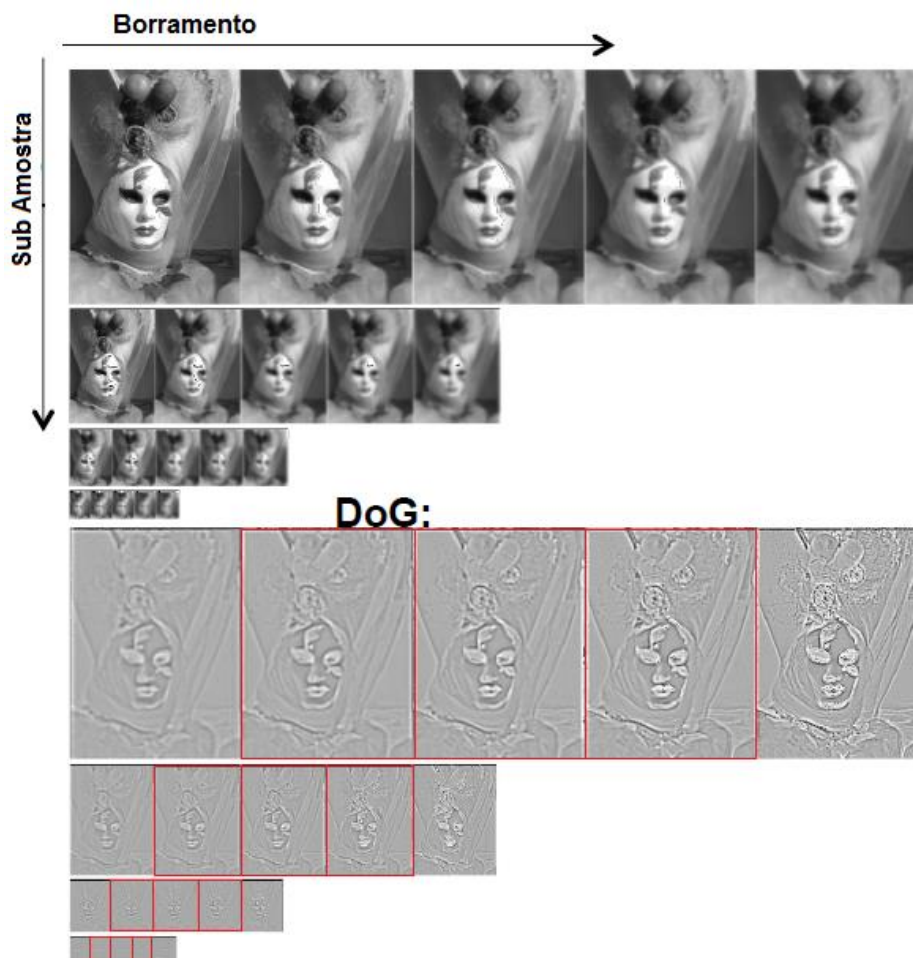
A aplicação da função de Diferença Gaussiana, como propôs (LOWE, 2004) consiste em aplicar Filtros Gaussianos que borram a imagem por meio de um fator  $\sigma$  (sigma) que varia em um intervalo na imagem e em seguida subtraí-la de outra com um valor  $\sigma$  diferente. Na Figura 5 é possível ver o esquema exemplificando a aplicação das diferenças gaussianas, a cada oitava (*octave*) é aplicado o Filtro Gaussiano com

valor  $\sigma$ , em seguida é feita a subtração das imagens que resultam em Diferenças de Gaussiana, a próxima oitava inicia com a imagem que possui o dobro de  $\sigma$  inicial.



**Figura 5 - Exemplo da função *DoG* para cada oitava.**  
**Fonte: Lowe (2004).**

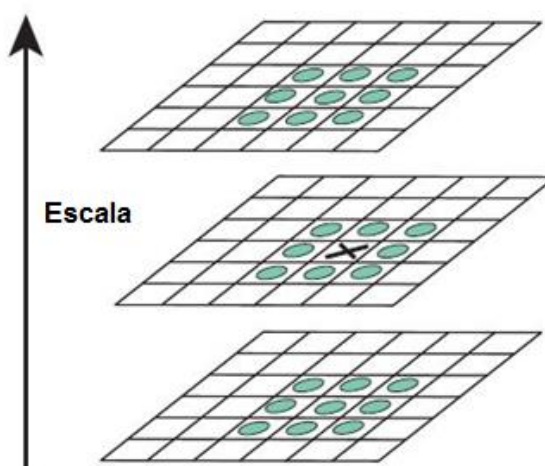
Após a aplicação da Diferença Gaussiana para cada oitava, é realizado a verificação para encontrar extremos. Essa busca é feita para cada intervalo de diferença gaussiana (excluindo a imagem com fator  $\sigma$  mínimo e máximo da oitava), como pode ser visto na Figura 6 (BELO, 2006).



**Figura 6 - Exemplo da aplicação DoG em cada oitava. A busca por extremos é realizada apenas nas imagens dentro do intervalo em vermelho.**

Fonte: Adaptado de LEUTENEGGER (2012)

Para se obter os possíveis candidatos a pontos chaves é feita a análise de cada pixel com seus 26 vizinhos nas diferentes escalas de sigma. A Figura 7 mostra o pixel X marcado, caso ele tenha valor maior ou menor que todos os seus vizinhos será um possível ponto de interesse (Lowe, 2004).

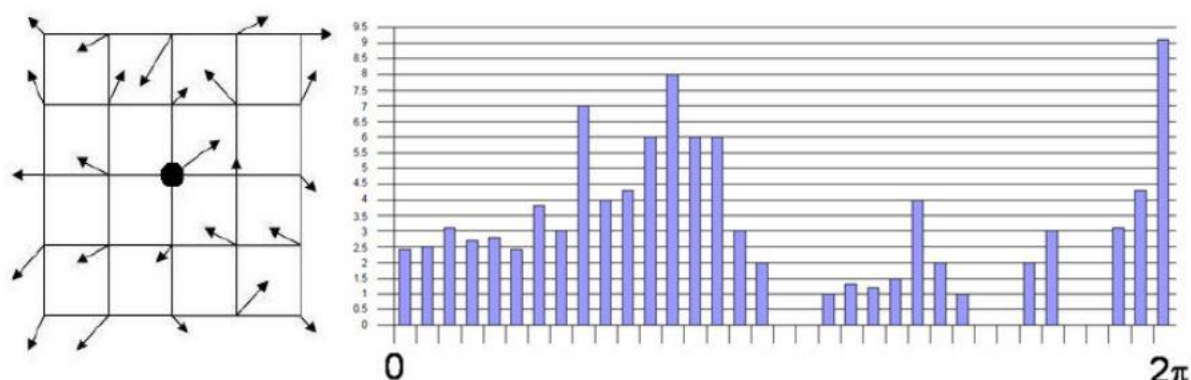


**Figura 7 - Detecção de extremos no espaço escala dos intervalos.**  
**Fonte: Lowe (2004)**

A localização dos candidatos a pontos chave utilizando a detecção de extremos gera muitos pontos que por sua vez são instáveis. Para determinar com maior precisão a localização e estabilidade dos pontos, Lowe (2004) propõe a utilização do cálculo de localização interpolada por meio da expansão quadrática da série de Taylor sobre a função da Diferença Gaussiana, e em seguida são filtrados os pontos por meio de um limiar de contraste, para descartar pontos com valores inferiores ao mesmo que é feita por meio do cálculo da expansão da série de Taylor de segunda ordem da Função de Diferença Gaussiana.

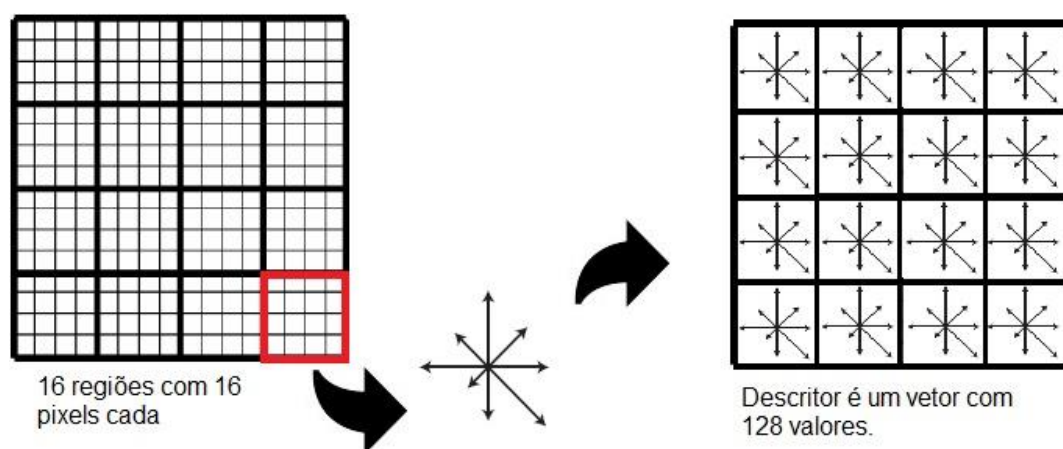
De modo a otimizar e selecionar pontos mais estáveis, é necessário remover os que possuem má localização mas que possuem forte resposta a arestas devido à alta resposta às mesmas da função DoG. Isto é feito utilizando o Hessiano (Matriz hessiana) com as derivadas parciais de segunda ordem da função DoG (Lowe, 2004).

Atribuição de orientação aos pontos chave é realizada para obter a invariância à rotação, Lowe, (2004) propôs o cálculo de orientações e magnitudes do gradiente para cada ponto chave, com base nas diferenças dos pixels pertencentes a imagens presentes nos intervalos do espaço-escala. Um histograma é montado representando todas as orientações entre 0 e  $2\pi$  e cada ponto vizinho ao ponto chave é adicionado ao histograma. O pico do histograma servirá para definir a orientação do ponto chave, mas além dele, picos que possuem ao menos 80% o seu valor, também serão considerados, o que poderá levar o ponto chave a ter mais de uma orientação (Lowe, 2004).



**Figura 8 - Histograma dos gradientes.**  
**Fonte: Adaptado de Gonzáles & Meggiolaro (2010)**

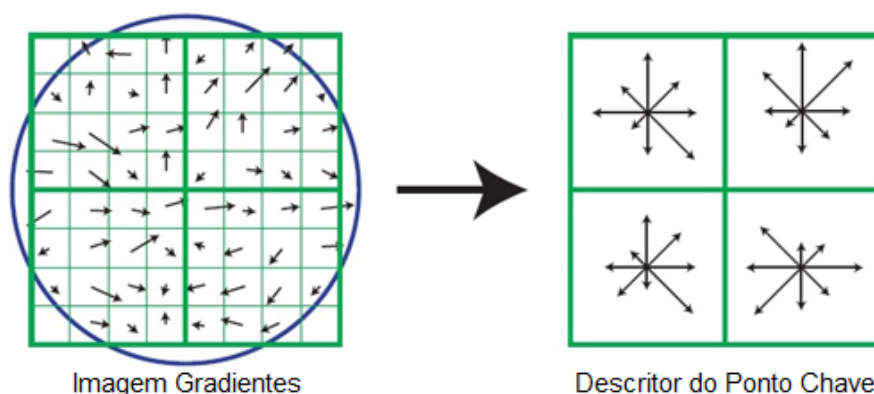
A construção do descritor é feita considerando uma região de tamanho 4 ao redor do ponto chave dividida em sub-regiões 4x4. Para cada sub-região é calculado um histograma considerando 8 orientações, o que leva a criação de um vetor de tamanho 128 que descreve o ponto chave, computado a partir das magnitudes e orientações dos gradientes (Lowe, 2004).



**Figura 9 - Exemplo de Histograma para cada uma das 16 regiões.**  
**Fonte: Adaptado de Belo (2006)**

A função gaussiana é aplicada na região a volta do ponto chave para dar peso a magnitude do gradiente de cada ponto na vizinhança para evitar mudanças súbitas na posição e dar menos ênfases a gradientes que estão longe do ponto chave (GONZÁLES e MEGGIOLARO, 2010). Na Figura 10 é demonstrado a gaussiana como um círculo azul em volta da região.





**Figura 10 - Região de aplicação da Gaussiana para enfatizar os pontos vizinhos.**  
**Fonte: Lowe (2004)**

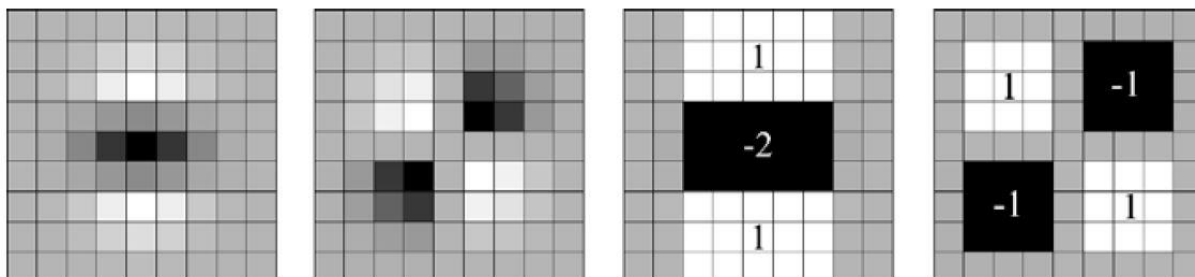
### 2.2.2 Speeded-Up Robust Features – SURF

O SURF (Speeded-Up Robust Features) foi proposto em 2006 por (BAY et al., 2008), e tem uma abordagem muito similar ao SIFT, porém, encontra os pontos de interesse e os descreve computacionalmente mais rápido (BUENO, 2011).

Para fazer a detecção de pontos de interesse, diferentemente do SIFT que usa *Difference of Gaussian*, o SURF realiza esse processo por meio de filtros Haar que tem formato de caixa e se baseia no determinante da Matriz Hessiana computando as derivadas de segunda ordem da Gaussiana, utilizando integrais de imagens que permitem a rápida computação dos filtros (BAY et al., 2008), para obter a localização e escala do ponto chave, as localizações são interpoladas e procura-se descartar pontos de baixo contraste ou localizados em arestas, similarmente ao que é feito no SIFT.

O espaço escala também é analisado iterativamente aumentando a escala do filtro e reduzindo o tamanho da imagem (BARBOSA, 2014).

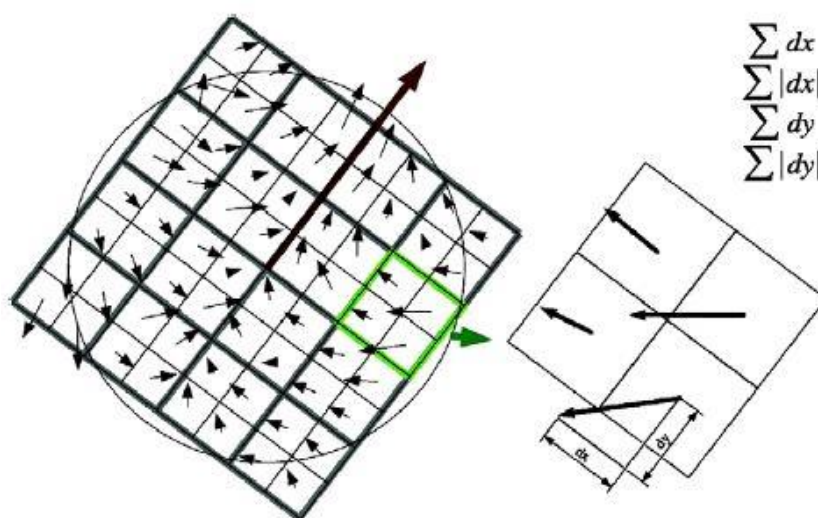
Para fazer a atribuição, Bay et al. (2008) propôs a realização de atribuição de orientação utilizando as respostas da aplicação de wavelet na direção horizontal e vertical para a vizinhança do ponto.



**Figura 11 - Aplicação das derivadas de segunda ordem da Gaussiana e aplicação do Filtro Caixa.**

Fonte: BAY et al. (2008).

O descritor é computado com base na construção de um histograma de orientação dos gradientes dos pontos presentes nas 16 regiões ao redor do ponto de interesse. Para cada sub-região é analisada a soma das respostas dos filtros e o módulo das respostas nas direções vertical e horizontal (Figura 12), ou seja, 4 valores, uma para cada direção é computado, para cada sub-região, resultando em um vetor de 64 dimensões (BAY et al., 2008).



**Figura 12 - Descritor SURF.**

Fonte: BAY et al. (2008).

### 2.2.3 Fast Retina Keypoint – FREAK

Descritores binários são uma alternativa para aplicações que requerem desempenho em termos de tempo computacional. O FREAK é um descritor binário inspirado no sistema de retina humana apresentando por (ALAHÍ, ORTIZ e VANDERGHEY, 2012), sendo uma extensão dos conceitos básicos de outro descritor chamado BRIEF (SCHMIDT, KRAFT, *et al.*, 2012).

Por meio de descritores binários é possível utilizar a distância de Hamming para fazer a correspondências entre duas regiões, já que as mesmas são descritas por string binárias. Os descritores binários utilizam a comparação de intensidade das imagens, e são compostos geralmente de uma região com uma amostragem de pares. Considerando uma região em torno do ponto chave para descreve-la como uma string binária, é escolhido um número de pares nessa região. Então, sobre todos os pares é comparado o valor da intensidade do primeiro ponto no par com o valor da intensidade do segundo ponto do par. Se o primeiro valor é maior que o segundo é escrito '1' na cadeia, caso contrário '0' (ALAHÍ, ORTIZ e VANDERGHEY, 2012).

Existem muitos padrões de amostragem para se comparar as intensidades entre pares de pixels. Nesta abordagem é utilizada o padrão de amostra inspirado na biologia da retina do olho (Figura 13), onde sobre cada ponto é aplicado um filtro Gaussiano para se tornar menos sensível a ruído. O valor do filtro varia conforme o valor do ponto de modo a simular a natureza da retina o qual possui mais células ganglionares na região Fovea (Figura 13) (ALAHÍ, ORTIZ e VANDERGHEY, 2012), cada região denominada região receptiva, e sobre o qual é aplicado um filtro com valor variado.

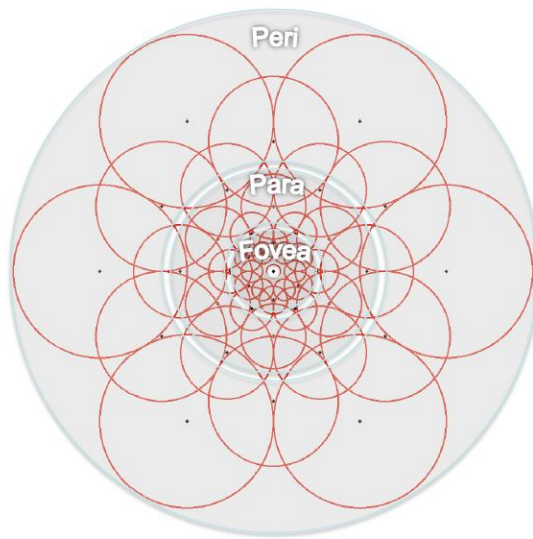
O descritor é construído por meio da comparação das diferenças de intensidades entre os pares com um limiar, sendo ela dada por:

$$T(P_a) = \begin{cases} 1, & \text{se } (I(P_i) - I(P_j)) > 0 \\ 0, & \text{caso contrário,} \end{cases} \quad (3)$$

A string binária formada pela sequência de bits é dada por:

$$F = \sum_{0 \leq a < N} 2^a T(P_a) \quad (4)$$

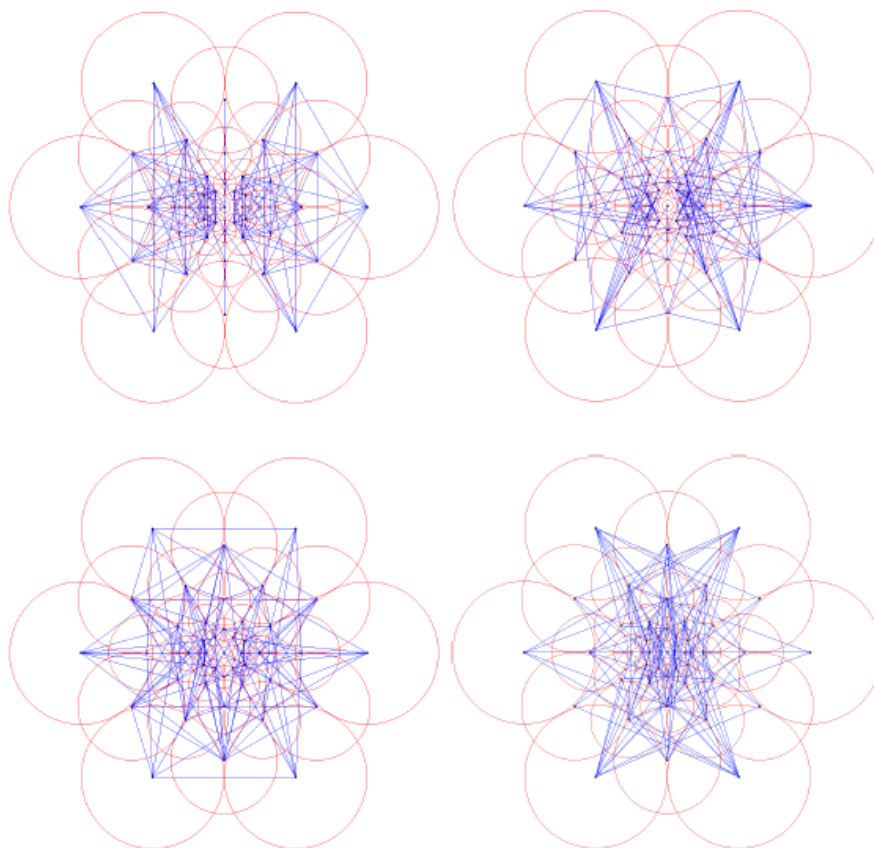
Onde  $P_a$  é um par  $(P_i, P_j)$  com  $i, j \in \mathbb{N}$  com  $i \neq j$ ; e  $N$  é o tamanho desejado do descritor (ALAHÍ, ORTIZ e VANDERGHEY, 2012) (KRIZAJ, STRUC, *et al.*, 2014).



**Figura 13 - Padrão de amostragem FREAK.**  
**Fonte: ALAHÍ, ORTIZ e VANDERGHEY, (2012)**

Há a possibilidade de milhares de pares com algumas dezenas de regiões receptivas, o que leva a um grande descritor, e muitos desses pares podem não ser úteis para descrever a imagem. Para contornar esse problema, os autores utilizam um algoritmo com a estratégia de aprendizagem de pares baseada em outro descritor binário chamado ORB (RUBLEE, RABAUDE, *et al.*, 2011). Os melhores pares são aprendidos a partir de um conjunto de treinamento (ALAHÍ, ORTIZ e VANDERGHEY, 2012).

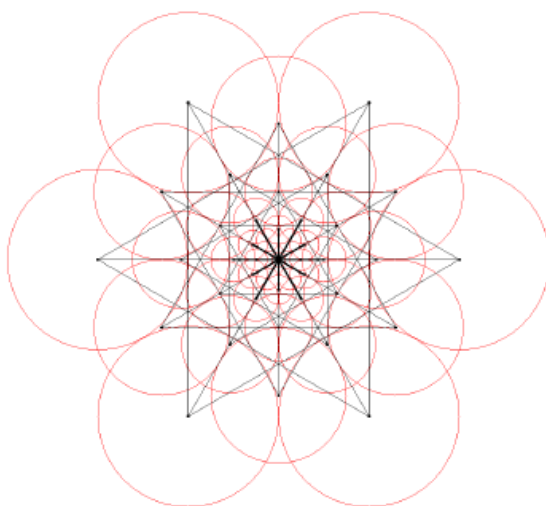
A forma final do descritor possui 512 pares os quais são testados pela Equação (4), para formar a sequência de bits. Os 512 pares são então divididos em quatro *clusters*, como pode ser observado pela Figura 14, onde é possível visualizar uma estrutura simétrica devido a orientação do padrão ao longo do gradiente global (ALAHÍ, ORTIZ e VANDERGHEY, 2012).



**Figura 14 – Os 512 pares selecionados divididos em quatro grupos (128 por grupo).  
Fonte: (ALAH, ORTIZ e VANDERGHEY, 2012)**

Para acelerar ainda mais a correspondência o FREAK usa uma abordagem em cascata: quando é feita a correspondência de dois descritores, primeiro é comparado apenas os primeiros 16 bytes. Se a distância é menor do que um limiar, a comparação continua com os próximos 16 bytes. Como resultado, mais de 90% dos candidatos são descartados com os primeiros 16 bytes.

Para estimar a rotação do ponto chave são somados os gradientes locais dos pares selecionados. A orientação é estimada então com base nos 45 pares selecionados que estão dispostos simetricamente com relação ao centro do padrão de amostragem (Figura 15).



**Figura 15 - Pares selecionados para computar a orientação.**  
**Fonte: (ALAHÍ, ORTIZ e VANDERGHEY, 2012).**

### 3 RECONHECIMENTO DE PADRÕES EM IMAGENS

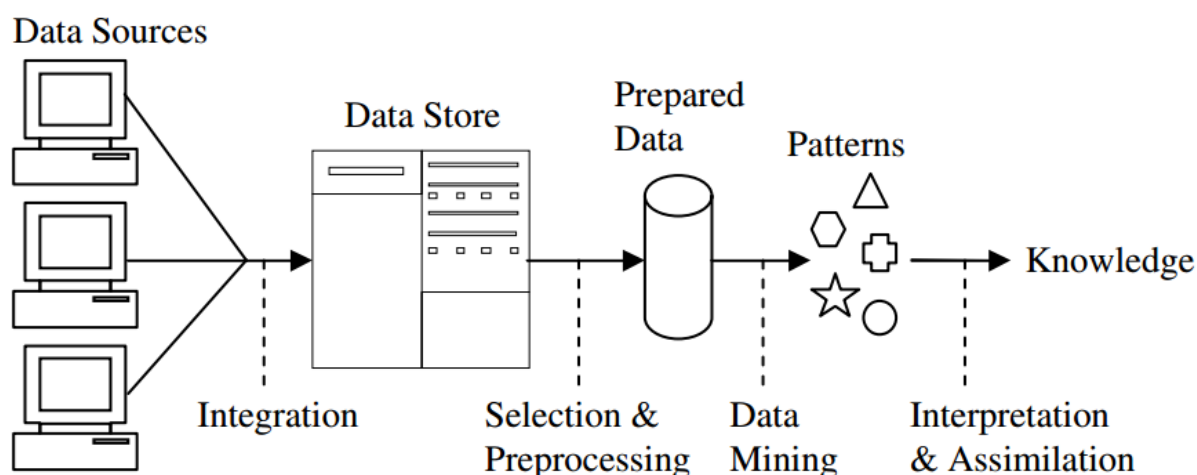
O Reconhecimento de Padrões pode ser entendido como uma das etapas do processo de descoberta de conhecimento (*Knowledge Discovery Process*, KDP), que pode ser definido como um processo não trivial de extração de informação desconhecida e útil dos dados (CIOS, PEDRYCZ, *et al.*, 1998), ou também como ampla área de estudo que tem a ver com tomadas de decisões automáticas (STARCK e MURTAGH, 2006) ou ainda uma área de estudo científico que tem como escopo a classificação de objetos em um número de categorias ou classes (THEODORIDIS e KOUTROUMBAS, 2006).

O modelo de descoberta de conhecimento de Fayyad *et al.* apud (CIOS, PEDRYCZ, *et al.*, 1998) é constituído de nove passos, definidos como:

1. *Desenvolver e compreender o domínio da aplicação*: Esta etapa inclui aprender todo conhecimento relevante previamente e também as metas do usuário do conhecimento descoberto.
2. *Criação de um conjunto de dados de destino*: Aqui são selecionados um subconjunto de variáveis (atributos) e pontos de dados (exemplos) que serão utilizados para executar tarefas de descoberta. Esta etapa inclui geralmente consultar os dados existentes para selecionar o subconjunto desejado.
3. *A limpeza de dados e pré-processamento*: Esta etapa consiste na remoção de outliers, lidar com o ruído e valores em falta nos dados, e respondendo por informações sequência de tempo e das alterações conhecidas.
4. *Redução dos dados e projeção*: Esta etapa consiste em encontrar atributos úteis, aplicando redução dimensão e métodos de transformação, e encontrando representações invariantes dos dados.
5. *Escolher a tarefa de mineração de dados*: Nesta etapa são conciliados os objetivos definidos no passo 1 com um método de Mineração de Dados em particular, como a classificação, regressão, clustering, etc.
6. *A escolha do algoritmo de mineração de dados*: São selecionados métodos para busca de padrões nos dados e ocorre a decisão de quais os modelos e parâmetros dos métodos utilizados podem ser apropriados.

7. *A mineração de dados*: Esta etapa gera padrões em uma forma de representação especial, tais como regras de classificação, árvores de decisão, modelos de regressão, tendências, etc.
8. *Interpretar padrões*: Aqui o analista visualiza os padrões extraídos e modelos, e a também visualiza os dados baseados nos modelos extraídos.
9. *Consolidação conhecimento descoberto*: A etapa final consiste em incorporar o conhecimento descoberto no sistema de performance, e documentar e reportar às partes interessadas. Esta etapa também pode incluir a verificação e resolução de conflitos potenciais com o conhecimento que se acreditava anteriormente.

É possível ver um resumo gráfico por meio da Figura 16 da sequência de passos do processo de descoberta de conhecimento.



**Figura 16 - Processo de Descoberta de Conhecimento.**  
**Fonte: BRAMER (2007)**

De modo geral, os seres humanos são bons em reconhecer padrões, geralmente é fácil distinguir o som da voz humana do som de um violino ou ainda diferenciar caracteres, por exemplo (DOUGHERTY, 2013).

Todos os dias, reconhecemos rostos em torno de nós, porém, fazemos isso de maneira inconsciente e por não conseguir explicar essa experiência, tem-se a dificuldade de escrever um programa de computador para que faça o mesmo. Por exemplo, cada face humana tem um padrão composto de combinações de estruturas particulares (olhos, nariz, boca etc.) localizadas em posições definidas na mesma. Pela análise



de amostra de imagens de faces, um programa deve ser capaz de extrair um padrão específico para um rosto e identifica-la ou reconhece-la como sendo um rosto, isso seria reconhecimento de padrões (DOUGHERTY, 2013).

Em imagens, um padrão pode ser definido como um conjunto de descritores, que são as características extraídas da imagem (GONZALEZ e WOODS, 2002).

### 3.1 RECONHECIMENTO DE OBJETOS

O mundo real é constituído de um amontoado de objetos, eles podem aparecer um na frente do outro e em diferentes perspectivas. A versão mais desafiadora do reconhecimento é o reconhecimento de categoria genérica (ou classe), que pode envolver reconhecer instâncias de classes extremamente variadas, tais como animais ou móveis (SZELISKI, 2010).

## 3.2 APRENDIZAGEM DE MÁQUINA

### 3.2.1 Aprendizagem Supervisionada

### 3.2.2 Aprendizagem Não Supervisionada

### 3.2.3 Classificação

### 3.2.4 Agrupamento

## 3.3 BAG OF WORDS

## 3.4 AVALIAÇÃO DA CLASSIFICAÇÃO

### 3.4.1 Medidas

### 3.4.2 Validação Cruzada

### 3.4.3 Teste de Hipótese

## 4 MATERIAL E MÉTODOS

Este Capítulo apresenta a base para o entendimento acerca da construção do modulo para a classificação de imagens, fornecendo uma visualização geral do projeto e a breve explicação das ferramentas utilizadas em seu desenvolvimento.

### 4.1 FERRAMENTAS

Inicialmente pretende-se utilizar a Linguagem de Programação JAVA versão SE 8, para desenvolver o componente levando em consideração a utilização dos conceitos do Paradigma de Orientação a Objetos.

#### 4.1.1 Pré-Processamento

Para a tarefa de detecção características e descrição foi escolhido a biblioteca de visão computacional OpenCV, originalmente desenvolvida pela Intel. Ela possui código fonte aberto para uso acadêmico e comercial, e foi escrita em C e C++ contando com mais de 500 funções que podem ser utilizadas em aplicações de diversas áreas tais como: medicina, segurança e inspeção de produtos em fabricas (BRADSKI e KAEHLER, 2008).

A versão escolhida para realização do projeto foi a 2.4.9, nela estão presentes todos os algoritmos responsáveis por extração e descrição de características que serão utilizados no trabalho.

#### 4.1.2 Aprendizagem e Classificação

Para as tarefas de aprendizagem e classificação pretende-se utilizar a *Waikato Environment for Knowledge Analysis* (WEKA) em sua versão de desenvolvedor 3.7.x, utilizando como uma biblioteca para a linguagem Java. A WEKA provém um conjunto de algoritmos de aprendizagem de máquina e ferramentas para pré-processamento de dados, que são muito utilizados em tarefas de *Data Mining*.

#### 4.1.3 Base de Dados

Pretende-se utilizar bases disponíveis de forma livre para os testes de validação do modelo. A base inicialmente escolhida<sup>1</sup> está presente de forma livre e contém conjuntos de imagens com várias transformações (perspectiva, escala, rotação), que servirá como teste para a acurácia do modelo cujas entradas sejam imagens que estejam nessas condições.

### 4.2 ANÁLISE E PROJETO

Para facilitar e organizar a construção do componente de classificação de imagens, será utilizada a linguagem de modelagem UML (*Unified Modelling Language*) por meio do *software Astah* versão *Community*<sup>2</sup>. O mesmo possui diversas ferramentas para modelagem gráfica, os quais são utilizados na área de Engenharia de *Software* para documentação.

A seguir são especificados alguns dos artefatos para documentação e construção do componente.

---

<sup>1</sup> <http://www.robots.ox.ac.uk/~vgg/research/affine/>

<sup>2</sup> <http://astah.net/editions/community>

#### 4.2.1 Requisitos do Sistema

Pretende-se estabelecer requisitos funcionais que descrevam as funcionalidades que o componente deverá desempenhar, os quais haverão de ser especificadas, para documentação, testes e validação. Também pretende-se estabelecer requisitos não funcionais, referentes ao desempenho do componente.

#### 4.2.2 Diagrama de Classes

A modelagem do diagrama de classes servirá para dar uma visão geral da estrutura e relação entre as classes que serão implementadas. O intuito é fornecer uma base para o entendimento de todo o projeto e de como ele será construído.

#### 4.2.3 Diagrama de Componentes

Com o diagrama de componentes pretende-se apresentar uma visão da estrutura física e especificar os componentes das quais o modulo de classificação de imagem seja dependente.

#### 4.2.4 Diagrama de Atividades

Dar uma visão geral do fluxo de atividades desempenhadas pelo modulo ao longo das tarefas executadas.

### 4.3 IMPLEMENTAÇÃO E VALIDAÇÃO

Inicialmente foi detectado um problema, onde os extratores de características não retiram um número fixo de características, e esse número de pontos variável é um problema no que diz respeito a utilização da API Weka que necessita de um conjunto de dados com número de atributos fixo. Para solucionar este problema foi utilizado o modelo *Bag of Words*.

Portanto, juntamente com a combinação dos algoritmos de descrição de características e algoritmos de classificação, pretende-se avaliar este modelo.

A avaliação ocorrerá utilizando medidas a serem definidas para avaliação do desempenho e eficiência, observando-se bases de dados disponíveis de forma livre, a serem definidas.

Após a conclusão do processo de avaliação, pretende-se também criar uma base de conhecimento contendo a representação de diversos objetos a serem detectados nas imagens que chegarem de entrada para o componente.

## 5 RESULTADOS ESPERADOS

Espera-se com o desenvolvimento deste trabalho, analisar os descritores de imagens presentes na literatura assim como os algoritmos de aprendizagem de máquina; possibilitando uma avaliação da combinação de descritores e classificadores de modo a encontrar um modelo eficiente para integrar ao componente de classificação que pretende-se desenvolver.

O que se espera é um modulo funcional, que seja capaz de classificar imagens com base em objetos identificados automaticamente por meio de algoritmos inteligentes baseadas em técnicas de aprendizado de máquina e processamento de imagens.







## REFERÊNCIAS

- ACHARYA, T.; RAY, A. K. **Image processing: principles and applications**. [S.l.]: John Wiley & Sons, 2005.
- ALAHÍ, ; ORTIZ, ; VANDERGHEY, P. Freak: Fast retina keypoint. **Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on**, p. 510--517, 2012.
- ALVES, G. T. M.; GATTASS, M.; CARVALHO, P. C. P. **Um Estudo das Técnicas de Obtenção de Forma a partir de Estéreo e Luz Estruturada para Engenharia**. Rio de Janeiro: Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro, 2005. 88p p.
- BALLARD, D. H.; BROWN, C. M. **Computer Vision**. Englewood Cliffs, : Prentice Hall, 1982.
- BARBOSA,. **Efficient Database Image Search**. Faculdade de Engenharia da Universidade do Porto. Porto, p. 18. 2014.
- BAY, H. et al. Speeded-up robust features (SURF). **Computer vision and image understanding**, v. 110, n. 3, p. 346--359, 2008.
- BELO, F. A. W. **Desenvolvimento de Algoritmos de Exploração e Mapeamento Visual**. Rio de Janeiro: Dissertação de Mestrado - Departamento de Engenharia Elétrica, 2006. 276 p.
- BEZDEK, J. C. et al. **Fuzzy Models and Algorithms for Pattern Recognition and Image Processing**. [S.l.]: Springer, v. 4, 2005.
- BOSCH, A.; MUÑOZ, X.; MARTÍ, R. Which is the best way to organize/classify images by content? **Image and vision computing**, 25, n. 6, 2007. 778-791.
- BOURIDANE, A. **Imaging for Forensics and Security: From Theory to Practice**. [S.l.]: Springer, v. 106, 2009.
- BRADSKI, G.; KAEHLER, A. **Learning OpenCV: Computer vision with the OpenCV library**. [S.l.]: O'Reilly Media, Inc., 2008.
- BRAMER, M. **Principles of Data Mining**. [S.l.]: Springer, v. 180, 2007.
- BROWN, M.; LOWE, D. G. Automatic panoramic image stitching using invariant features. **International journal of computer vision**, 74, n. 1, 2007. 59-73.
- BUENO, L. M. **Análise de descritores locais de imagens no contexto de detecção de semi-réplicas**. Campinas: Universidade Estadual de Campinas, 2011.

CARVALHO, I. A. D. **Classificação de imagens de pornografia e pornografia infantil utilizando recuperação de imagens baseada em conteúdo**. Brasília: Universidade de Brasília, 2012.

CIOS, K. J. et al. **Data Mining: A Knowledge Discovery Approach**. [S.l.]: Springer, 1998.

CIPOLLA, R.; BATTIATO, S.; FARINELLA, G. M. **Machine Learning for Computer Vision**. [S.l.]: Springer Publishing Company, Incorporated, 2012.

CONCI, A.; AZEVEDO, E.; LETA, F. R. **Computação Gráfica: Teoria e Prática**. Rio de Janeiro: Elsevier, v. 2, 2008. 407 p.

DE ANDRADE, F. D. S. P. **Combinação de descritores locais e globais para recuperação de imagens e vídeos por conteúdo**. Biblioteca Digital da Unicamp. [S.l.]. 2012.

DE ARAÚJO, S. A. **Casamento de padrões em imagens digitais livre de segmentação e invariante sob transformações de similaridade**. Universidade de São Paulo. [S.l.]. 2009.

DE LA CALLEJA, J.; FUENTES, O. Automated classification of galaxy images. **Knowledge-Based Intelligent Information and Engineering Systems**, p. 411-418, 2004.

DESELAERS, T. Features for image retrieval. **Rheinisch-Westfälische Technische Hochschule, Technical Report, Aachen**, 2003.

DOUGHERTY, G. **Digital image processing for medical applications**. [S.l.]: Cambridge University Press, 2009.

DOUGHERTY, G. **Pattern Recognition and Classification: An Introduction**. Camarillo, CA, USA: Springer, 2013.

DOUGHERTY, G. **Pattern Recognition and Classification: An Introduction**. Camarillo: Springer, 2013.

FLITTON, ; BRECKON, T. P.; MEGHERBI, N. A 3D extension to cortex like mechanisms for 3D object class recognition. **Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on**, p. 3634--3641, 2012.

FLITTON, G. T.; BRECKON, T. P.; BOUALLAGU, N. M. Object Recognition using 3D SIFT in Complex CT Volumes. **BMVC**, p. 1-12, 2010.

GONZÁLES, G. L. G.; MEGGIOLARO, M. A. **Aplicação da Técnica SIFT para Determinação de Campos de Deformações de Materiais usando Visão Computacional**. PUC-Rio. [S.l.], p. 109p. 2010.

GONZALEZ, R. C.; WOODS, R. E. **Digital Image Processing**. 2. ed. [S.l.]: Prentice Hall, 2002.

GRAUMAN, K.; LEIBE, B. **Visual object recognition**. [S.l.]: Morgan & Claypool Publishers, 2011.

LAGAREIRO, A. C. Fotografias tipo “selfie”: Uma forma de expressão ou a manifestação de um comportamento preocupante? Disponível em: <[http://www.pucsp.br/nppi/coluna\\_eletronica/2014/selfie-janeiro-2014.pdf](http://www.pucsp.br/nppi/coluna_eletronica/2014/selfie-janeiro-2014.pdf)>. Acesso em: 17 set 2014.

LECUN, Y.; HUANG, F. J.; BOTTOU, L. Learning methods for generic object recognition with invariance to pose and lighting. In: **Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on**, 2, 2014. II-97.

LEUTENEGGER, S. **Image Keypoint Detection, Description, and Matching**. AI Robots Summer School. [S.l.]. 2012.

LIBERMAN, F. **Classificação de imagens digitais por textura usando redes Neurais**. Rio Grande do Sul: Universidade Federal do Rio Grande do Sul. Instituto de Informática. Curso de Pós-Graduação em Ciência da Computação, 1997.

LISIN, D. A. et al. Combining local and global image features for object class recognition. **Computer Vision and Pattern Recognition-Workshops, 2005. CVPR Workshops. IEEE Computer Society Conference on**, 2005. 47-47.

LOWE, D. G. Object recognition from local scale-invariant features. **Computer vision, 1999. The proceedings of the seventh IEEE international conference on**, v. 2, p. 1150--1157, 1999.

LOWE, D. G. Distinctive image features from scale-invariant keypoints. **International journal of computer vision**, v. 60, n. 2, p. 91--110, 2004.

NIXON, M. S.; AGUADO, A. S. **Feature Extraction and Image Processing**. 1. ed. [S.l.]: Newnes, 2002.

PARK, U.; PANKANTI, ; JAIN, A. Fingerprint verification using SIFT features. **SPIE Defense and Security Symposium**, p. 69440K-69440K, 2008.

PAULA FILHO, P. L. **Processamento Digital de Imagens**. Notas de Aula. 2014.

PAULA FILHO, P. L. D. **Reconhecimento de espécies florestais através de imagens macroscópicas**. Tese (Doutorado) – Universidade Federal do Paraná. Curitiba. 2013.

PEDRINI, H.; SCHWARTZ, W. R. **Análise de Imagens Digitais: Princípios, Algoritmos e Aplicações**. São Paulo: Thompson, 2008.

PENATTI, O. A. B. **Estudo comparativo de descritores para recuperação de imagens por conteúdo na web**. Biblioteca Digital da Unicamp. [S.l.]. 2009.

POLAKOWSKI, W. E. et al. Computer-aided breast cancer detection and diagnosis of masses using difference of Gaussians and derivative-based feature saliency. **Medical Imaging, IEEE Transactions on**, v. 16, n. 6, p. 811-819, 1997.

RAOUI, Y.; BOUYAKHF, E. H. A. D. M.; REGRAGUI, F. Global and local image descriptors for Content Based Image Retrieval and object recognition. **Applied Mathematical Sciences**, 5, 2011. 2109-2136.

ROMANO, R. A.; ARAGON, C. R.; DING, C. Supernova recognition using support vector machines. **Machine Learning and Applications, 2006. ICMLA'06. 5th International Conference on**, p. 77-82, 2006.

SCHMID, C.; ROGER, M.; BAUCKHAGE, C. Comparing and evaluating interest points. **Computer Vision, 1998. Sixth International Conference on**, p. 230-235, 1998.

SCHMIDT, A. et al. The Comparison of Point Feature Detectors and Descriptors in the Context of Robot Navigation. **CLAWAR Workshop on Perception for Mobile Robot Autonomy (PEMRA'12), Poznan,(CD-ROM)**, 2012.

STARCK, J.-L.; MURTAGH, F. **Astronomical Image and Data Analysis**. 2. ed. [S.l.]: Springer, 2006.

SZELISKI, R. **Computer Vision: Algorithms and Applications**. [S.l.]: Springer, 2010.

TAGLIAFERRI, R. et al. Astronomical object recognition by means of neural networks. **Neural Nets WIRN VIETRI-98**, p. 169-178, 1999.

THEODORIDIS, S.; KOUTROUMBAS, K. **Pattern Recognition**. 3. ed. San Diego: Elsevier, 2006. 837 p.

TIAN, D. P. A Review on Image Feature Extraction and Representation Techniques. **International Journal of Multimedia and Ubiquitous Engineering**, v. 8, n. 4, p. 385-396, 2013.

TUYTELAARS, T.; MIKOLAJCZYK, K. Local invariant feature detectors: a survey. **Foundations and Trends in Computer Graphics and Vision**, 3, 2008. 177-280.