Walber Rodrigues de Oliveira

Sistema de Visão Computacional para Reconhecimento de Modelos CAD Complexos

Recife

2018

Walber Rodrigues de Oliveira

Sistema de Visão Computacional para Reconhecimento de Modelos CAD Complexos

Trabalho apresentado ao Curso de Graduação em Ciência da Computação do Departamento de Informática da Universidade Federal de Pernambuco como requisito parcial para obtenção do grau de Bacharel em Ciência da Computação.

Universidade Federal de Pernambuco - UFPE

Centro de Informática

Graduação em Ciência da Computação

Orientador: Professor Carlos Alexandre Barros de Mello (CIn/UFPE)

Co-orientador: Bruno Medeiros de Oliveira (ISI-TICs)

Recife

2018

Agradecimentos

Agradeço inicialmente à minha família, estes são os alicerces de toda a minha jornada pela vida. Agradeço também a paciência e atenção despendidas por Carlos e Bruno durante toda a escrita deste documento e orientação durante a elaboração deste método.

Agradeço a Ju por estar comigo em todos os passos me mantendo são nos momentos mais difíceis e servindo como um leme para o meu futuro. Agradeço aos amigos de *sangue de barro* do RPG & Pizza: Aldineia, Noaldo, Thamires, Vitu e Thiego. Assim como aos meus amigos da Casa do Estudante: Izaildo, Daniel, Célio, Diogo e Felipão que estiveram comigo nos momentos mais difíceis da UFPE.

Agradeço à todos os meus amigos e colegas de curso, em especial aos componentes do Vanilo: Calegário, Danilo, Lapprand, Lasalvia, Milena e Thiago por participarem de quase todos os trabalhos dessa reta final de curso. Também aos que estão um pouco de antes: Jadson, Véras e César.

Agradeço ao ISI como um todo por ter me fornecido espaço, material e recursos para o desenvolvimento desta pesquisa, assim como toda a equipe do ISI que faz do trabalho quase uma diversão.

Resumo

A visão computacional cresceu bastante nas últimas décadas, juntamente com o poder de processamento dos computadores atuais. Neste âmbito, diversos sistemas de reconhecimento de objetos foram propostos. São observadas formas de casamento entre imagens de itens e da cena até abordagens que realizam o reconhecimento de instâncias utilizando-se de inteligencia artificial, passando também pelo casamento de modelos complexos como nuvens de pontos em outros sistemas de pontos. Porém, novos desafios surgem com a iminência da instauração da quarta revolução industrial, a chamada indústria 4.0. Nela, a produção deve ser automatizada com maior velocidade e eficácia possível. Porém, a remodelagem das linhas de produção nem sempre é viável financeiramente, portanto, métodos que possibilitem essa mudança com menor impacto são cruciais. No contexto industrial, a maior parte da produção é feita através de restritas especificações. Esta máxima se estende aos objetos que são produzidos e à própria linha de produção. Em meio aos estudos de casamento de objetos usando documentos de descrição, são comumente encontradas abordagens que realizam a busca de um item a partir da comparação entre a informação do modelo e características que são obtidas das imagens que se busca o objeto. Entretanto, existe uma carência de estudos relacionados às necessidades dos sistemas críticos das indústrias, ou seja, com uma verificação aprofundada da cena e do objeto. Com isto, neste estudo, é apresentado um método capaz de lidar com a acurácia necessária para esses sistemas, que utiliza de uma dupla análise das imagens de cena em regiões relevantes, para retirada do máximo de características adequadas possíveis. É feito ainda, um levantamento de métodos de casamento e reconhecimento de objetos apresentados na literatura, assim como de propostas de obtenção de características de imagens. São apresentados, por fim, experimentos com as variações do método proposto, demonstrando a sua viabilidade em diferentes contextos.

Palavras-chaves: Reconhecimento de Objetos. Detecção de Bordas. Visão computacional.

Abstract

The computer vision field has been growing in recent decades, along with the processing capability of computer systems. Thus, many object recognition systems were proposed. These approaches go from image matching of object and scene to machine learning based systems, passing through methods designed to match complex structure such as cloud points. However, new challenges have arisen, along with the fourth industrial revolution, the named Industry 4.0. Within its guidelines, the production has to be increased through automatization, aiming to conquer maximal speed and accuracy. Nevertheless, a complete remodeling process of production line may be too expensive. Thus, methods which aim to do this revolution with minimal impact on one's capital are essential. On the industry context, several parts of the process are guided by specific standards. This may also comprise the object and the production line itself. Among other works about matching of complex structures from descriptor files, it is common to see studies that perform the object tracking using a proximity evaluation between the describing files and features obtained from the image scene where the object is supposed to be found. However, these works have a minimal exploration of the needs of the industry; in other words, a complete understanding of both object and scene. Hereupon, this study presents a method which is capable of dealing with the needs of the industry such as accuracy with minimal impact on the production line, using a dual analysis in major regions building a whole scene understanding. This monography also provides a literature review of methods of tracking and item recognition along with image features detection. Experiments are provided, demonstrating the feasibility of the method in different contexts.

Keywords: Object recognition, Edge detection, Computer vision.

Lista de ilustrações

Figura 1 – Supressão dos não máximos 21
Figura 2 – Etapas 25
Figura 3 – Alterações DifFocusColor 27
Figura 4 – Refino de Arestas 29
Figura 5 – Escolha de Arestas Analisadas 30
Figura 6 – Comparativo entre Microquadrantes de mesmas coordenadas 31
Figura 7 – Montagem de Câmera 34
Figura 8 – Obtenção do Modelo Renderizado 35
Figura 9 – Modelo Para Casamento 35
Figura 10 – Resultado Casamento 36
Figura 11 – Detalhes do Casamento 37
Figura 12 – Filtragem Dupla 38

Lista de tabelas

Tabela 1 – Resultados 38

Lista de abreviaturas e siglas

SURF	*Speed Up Robust Features*
SIFT	*Scale Invariant Feature Transform*
GFT	*Good Features to Track*
CNN	*Convolutional Neural Network*
ROI	*Region of Interest*
M	Matiz
S	Saturação
B	Brilho
CAD	*Computer-Aid Design*
GPS	*Global Positioning System*
RAPiD	*Real-time Attitude and Position Detection*
C++	*C Plus Plus*
OpenGL	*Open Graphics Library*
OpenCV	*Open Source Computer Vision Library*

Sumário

1 INTRODUÇÃO 9
1.1 Motivação 9
1.2 Objetivos 10
1.3 Estrutura do Documento 11

2 ANÁLISE DA LITERATURA 12
2.1 Métodos baseados em Pontos-chave 12
2.2 Métodos baseados em Fluxo Óptico 14
2.3 Métodos baseados em *Templates* 15
2.4 Métodos baseados em Aprendizagem de Máquina 16
2.5 Métodos baseados em Arestas 18
2.6 Detecção de Bordas 20
2.6.1 Canny 20
2.6.2 DifFocusColor 22
2.6.3 Transformada de Hough 23

3 MÉTODO PROPOSTO 25
3.1 Pré-Processamento 25
3.2 Obtenção de Características de Imagens 27
3.3 Refino das Retas obtidas a partir da Transformada de Hough 28
3.4 Casamento 28
3.5 Análise de Microquadrantes 30
3.6 Inferência de Presença 32

4 EXPERIMENTOS 33
4.1 Definição de ambiente e Pré-processamento 33
4.2 Testes e Comparativos 35
4.3 Resultados 37

5 CONCLUSÃO 40
5.1 Trabalhos Futuros 41

REFERÊNCIAS 42

1 Introdução

Com a evolução da robótica, problemas de reconhecimento de objetos em imagens tornam-se capitais para a indústria. Diversas abordagens surgem para tentar sanar e melhorar esse processo. Dentre os métodos de reconhecimento de objetos abordados na literatura, podem-se citar métodos que realizam a localização a partir de imagens do objeto a ser detectado, métodos que realizam a localização de modelos que são um conjunto de estruturas complexas (geralmente, um conjunto de polígonos), e ainda abordagens que se destinam a realizar uma correspondência entre modelos representados por uma nuvem de pontos.

Na indústria, a utilização de modelos de referência é bastante difundida, por esses apresentarem com clareza a perfeição buscada para os itens produzidos e para o processo industrial. O entendimento do posicionamento e *status* de um objeto são desafios primordiais para a evolução da indústria (HERMANN; PENTEK; OTTO, 2016). O exame de objetos que foram moldados a partir de tais arquivos a partir de uma foto é o foco deste estudo. No desenvolver deste capítulo é introduzida a problemática de reconhecimento de objetos seguida de uma breve explanação das diretrizes deste estudo e, por fim, é apresentada a estrutura desta monografia.

1.1 Motivação

Como citado anteriormente, pode-se organizar as formas de realização de casamento entre objeto e cena descrevendo-se a cena e o objeto de diversas formas. Entre elas, objetos representados por uma imagem assim como a cena, o objeto representado por um modelo enquanto a cena é uma imagem, e também ambos sendo apresentados como um modelo. Dentre as variadas abordagens que visam realizar correspondência entre imagens, os métodos mais famosos utilizam da localização de pontos chave e criação de descritores obtidos de fotos reprodutivas do objeto a ser encontrado. Dentre estes, podemos citar os métodos SIFT(do inglês, *Scale Invariant Feature Transform*) (LOWE, 1999) e SURF (do inglês, *Speed Up Robust Features*) (BAY; TUYTELAARS; GOOL, 2006) sendo o segundo um melhoramento ao primeiro em termos de custo computacional.

De forma geral, ambos os métodos consistem na captura de pontos chave que não variam ou variam pouco em diversas imagens que contenham o objeto. A partir daí, uma série de vetores de características são obtidos destes pontos chave, denominados descritores. Em posse dos descritores, prossegue-se para a realização de uma correspondência entre os descritores já obtidos na imagem do objeto buscado e descritores que são obtidos de uma cena alvo, onde é pesquisado o item.

Existem ainda, a exemplo de (MIAN; BENNAMOUN; OWENS, 2006), métodos que utilizam de uma série de imagens de um objeto para construção de um modelo tridimensional e deste retirar a extração de descritores em três dimensões com o intuito de buscar uma correspondência com a cena. Para realizar a busca de modelos complexos compostos por polígonos, a exemplo dos modelo CAD (do inglês, *Computer Aided Design*), em cenas obtidas de câmeras no mundo real, alguns métodos (CHOI; CHRISTENSEN, 2010) (ARMSTRONG; ZISSERMAN, 1995) concentram-se na obtenção de características pronunciadas como arestas e círculos. A partir daí, é realizado o cálculo entre a distância destas características encontradas até o modelo projetado na cena em duas dimensões, buscando a melhor equivalência possível. Estas características são geralmente obtidas utilizando-se de uma série de filtros sobre a cena a qual se deseja buscar o objeto, sendo o algoritmo Canny para detecção de bordas (CANNY, 1986) associado à Transformada de Hough (PEDRINI; SCHWARTZ, 2008) os mais difundidos.

Outros métodos utilizam-se ainda de nuvens de pontos coletados por sensores de profundidade e disponibilizados em bases de dados como a encontrada em (ALDOMA et al., 2012) para realizar a correspondência com os objetos. Outros estudos abordam ainda o uso de métodos de reconhecimento de instâncias utilizando inteligência artificial, como demonstrado em (GOMEZ-DONOSO et al., 2017) (VIOLA; JONES, 2001) realizando a inferência de que tipo de item ou mesmo qual o item encontrado.

Dentre os diversos estudos, nota-se uma carência de estudos que sejam diretamente ligados a problemas apresentados hoje na indústria. A indústria demanda precisão em seu processo e bens projetados, um exemplo disto, é o processo de fabricação de gotejadores para irrigação. A largura do orifício de dispersão de água influencia diretamente nos gastos de água e na eficiência do sistema de irrigação (SOUZA et al., 2006). Um sistema de verificação da qualidade da produção dos bicos utilizando visão computacional pode ser aplicado minimizando perdas e aumentando a eficácia deste. Portanto, este trabalho apresenta um método de reconhecimento de arquivos de referência CAD viável no contexto da indústria 4.0. Ou seja, uma abordagem acurada e que cause um menor impacto no ambiente da linha de produção, evitando assim gastos com sua readequação e reorganização.

1.2 Objetivos

Este trabalho destina-se a apresentação de uma nova abordagem de reconhecimento de modelos CAD complexos em uma imagem RGB. Esta, obtida numa possível linha de produção industrial, indicando presença destes modelos e dos seus componentes.

São analisados diferentes métodos de procura de objetos em imagens utilizando-se ou não de arquivos de referência CAD. É analisada também a forma de obtenção de características pronunciadas de uma imagem. Com isto, são apresentados métodos de

obtenção de bordas em imagens, e a modificação de um deles. É compreendida, ainda, a estrutura de um modelo CAD, apresentando uma possível simplificação deste tipo de arquivo através da diminuição da quantidade de arestas e supressão de características oclusas.

Em seguida, verificações do método apresentado são realizadas através de experimentos realizados com uma peça impressa em três dimensões. Além disto, são feitos comparativos que demonstram a viabilidade de utilidade do método. Por fim, é feita uma reflexão sobre o que é uma correspondência de qualidade adequada para o problema. Ou seja, uma correspondência entre modelos de referência projetados em duas dimensões em imagens que contenham um objeto semelhante ao procurado.

1.3 Estrutura do Documento

Este documento está disposto em cinco capítulos sendo os capítulos seguintes organizados da seguinte forma. No Capítulo 2, são analisadas cinco formas de realização de casamento e/ou reconhecimento de objetos em imagens. Em seguida, na subseção 2.6.1 são demonstradas duas formas de obtenção de bordas em imagens, seguido de uma explicação de como a Transformada de Hough pode ser usada para identificar segmentos de retas.

No Capítulo 3, é apresentado o mecanismo apresentado por esta monografia para realização do reconhecimento de modelos CAD. Dentro deste mesmo capítulo, na seção 3.2, é apresentada uma variação do filtro de detecção de bordas DifFocusColor, o DifFocusColor6m.

Já no Capítulo 4, é demonstrado a viabilidade do mecanismo proposto através de testes comparativos. Por fim, são feitas as considerações finais no Capítulo 5. Nele, são discutidos os resultados obtidos e abordados possíveis trabalhos futuros.

2 Análise da Literatura

A fim de realizar o reconhecimento de objetos em cenas, diversos métodos foram propostos. A topologia definida em CHOI; CHRISTENSEN diferencia os métodos como baseados em: detecção de pontos chave, arestas, fluxo óptico ou *templates*. Neste capítulo, fazemos uma breve revisão de alguns métodos selecionados acrescidos à uma breve explanação sobre reconhecimento de objetos utilizando Inteligência Computacional. Para realizar a procura de objetos tridimensionais (3D) ou instâncias em uma cena, faz-se necessária também a análise das possíveis formas de representação do objeto e da cena. É conveniente que um nível ideal de características seja obtido de ambos, objeto e cena, para que o casamento seja obtido de forma clara.

Em (GUO et al., 2014) são descritas as formas mais comuns deste tipo de representação, sendo estas: imagem de profundidade, nuvem de pontos ou ainda estrutura poligonal. Dentre os vários métodos propostos na literatura para representação da imagem de cena, temos o foco nos que buscam em imagem RGB única e portanto ao final do capítulo, apresentamos uma breve revisão sobre a detecção de primitivas a partir de uma imagem que passa por um processo de detecção de bordas.

2.1 Métodos baseados em Pontos-chave

Tipicamente, estudos de reconhecimento de objetos tridimensionais a partir de pontos-chave, consistem em três etapas: localizar pontos-chave, analisar localmente a superfície e suas proximidades e o casamento. Na primeira etapa, são coletados os bons candidatos a pontos-chave.

O ponto-chave é uma parcela da imagem do objeto que apresenta uma grande quantidade de características únicas e/ou complexas que variem pouco entre diversas imagens do mesmo objeto. Uma vez obtidos os pontos-chave, a vizinhança de cada um deles é analisada capturando informações como escala e posicionamento dos itens que os cercam. Em posse destas informações, há a concretização da segunda etapa onde são construídos descritores destas características. Por fim, cada descritor é buscado nas cenas que se deseja realizar o casamento e, de acordo com cada possível correspondência encontrada, são inferidas a presença e a posição do item na cena. Dois dos métodos mais difundidos para detecção de objetos a partir de pontos-chave são o *Scale Invariant Feature Transform* (SIFT) (LOWE, 1999) e o *Speeded Up Robust Features* (SURF) (BAY; TUYTELAARS; GOOL, 2006); ambos explanados a seguir.

O algoritmo apresentado por LOWE pode ser decomposto em quatro etapas: análise

espaço-escala, localização de pontos-chave, definição de orientação dos pontos-chave e descrição de pontos-chave. Para a primeira etapa, o autor propõe a representação em espaços e escalas que possa ser usada para identificar diferentes perspectivas do mesmo objeto. Para obtenção dos locais, a imagem passa por uma série de filtros gaussianos e em seguida um redimensionamento em metade do seu tamanho. Em (MESQUITA, 2017) estima-se o uso de três ou quatro octavos (sucessivos redimensionamentos) e 4 níveis de filtragem, por octavo, para a construção do espaço-escala. O espaço-escala é então representado numa pirâmide feita com a imagem de entrada contendo diferentes níveis de granularidade e escalas.

Construído o espaço-escala, a seleção de pontos estáveis se dá pelo uso de uma função que realize a diferença de Gaussianas entre níveis adjacentes no espaço escala. A partir daí, são identificados pontos de mínimo e de máximo locais. Estes são comparados às suas vizinhanças no espaço-escala; ou seja, aos oito pontos adjacentes, em duas dimensões, a eles na mesma imagem e às demais escalas acima e abaixo na pirâmide. Ele é um ponto selecionado apenas se permanecer como máximo ou mínimo após passar por todas as verificações. Lowe ainda afirma que o custo computacional das comparações é baixo, devido ao fato que a maioria dos pontos são descartados nas primeiras comparações.

Após esta fase, foram obtidos pontos que qualificaram-se minimamente como candidatos para pontos-chave. Porém, ainda mais refino faz-se necessário, pois, a qualidade e a quantidade dos pontos-chave adquiridos afeta diretamente no êxito do casamento alcançado. Portanto, São descartados os pontos através da adição de um limiar mínimo para os valores de diferença de gaussiana apresentados. Posteriormente, mais refinos são aplicados, retirando-se pontos cuja curvatura principal não seja adequada. Os pontos remanescentes deste processo são considerados verdadeiros pontos-chave.

Em posse dos pontos-chave, são realizadas mais análises. São observados os gradientes de todos os pontos que constam numa janela circular ao redor de um ponto-chave e o maior pico é assinalado como sua orientação. Por fim, para obtenção dos descritores, são re-calculados os valores dos gradientes da região adjacente e então uma função gaussiana é aplicada. É construído então um histograma de oito orientações com o valor de cada subregião constando o gradiente de cada subregião (MESQUITA, 2017).

O algoritmo SURF, apresentado em (BAY; TUYTELAARS; GOOL, 2006), é uma alternativa computacionalmente mais eficiente e que quase não apresenta ligeiras alterações de acurácia em relação ao SIFT, afirmam os autores. Seu baixo custo se dá pela utilização de filtros *box*. Devido à mudança no cálculo do filtro, as imagens não precisam ser sucessivamente reduzidas, o que também diminui a complexidade do método, uma vez que é poupado o custo dos redimensionamentos.

Além disso, no SURF, a definição da orientação do ponto-chave é realizada utilizando-se Wavelets de Haar baseadas na vizinhança do ponto de interesse. De forma

semelhante ao SIFT, após projetadas num espaço bidimensional a maior componente é considerada como sendo o vetor orientação dominante do ponto-chave. Contudo, os descritores são construídos em relação aos componentes x e y de forma separada, 128 ao todo, sendo 64 em cada direção (MESQUITA, 2017).

2.2 Métodos baseados em Fluxo Óptico

Enquanto métodos baseados em pontos-chave utilizam de imagens únicas para extração de características e realização do casamento, os métodos baseados em fluxo óptico lançam mão do acompanhamento da movimentação de pontos e regiões que pode ser observada entre diversos quadros de um vídeo. Abordagens do tipo são comumente usadas para realizar análise e rastreamento de objetos não-rígidos. Em (SENST et al., 2012) é citada a relevância destes métodos para reconhecer pessoas que se movem em multidões assim como monitoramento de atividades, já em (MARKS; HERSHEY; MOVELLAN, 2010) é apresentado um método que combina fluxo óptico e casamento de *templates* móveis e pode ser usado para realizar o rastreamento de movimento, textura e deformação como a presente em rostos que são reconhecidos por quadro em videos.

O desafio primordial desse tipo de abordagem é descobrir quais pontos são adequados para serem acompanhados. A acurácia e o custo computacional precisam se equilibrar para que o método seja adequado. Métodos computacionalmente eficientes como o *Kanade Lucas Tomasi Feature Tracker* (KLT) (TOMASI; KANADE, 1991) permanecem sendo vastamente utilizados. O KLT emprega a busca de características consideradas boas - *Good Features to Track* (GFT) - associado ao método de Lucas Kanade (LUCAS; KANADE, 1981) para acompanhamento da movimentação. Abordagens que combinam o KLT com a paralelização de placas de vídeo conseguem acompanhar cerca de dez mil pontos em tempo real, ou seja, mais que vinte e cinco quadros por segundo. (SENST et al., 2012)

A estratégia apresentada em (SENST et al., 2012) inova ao trazer uma forma diferente de análise da trajetória que não apenas utiliza-se da diferenciação de textura entre as imagens. São computadas afinidades temporal e espacial dos pontos entre os quadros e esses valores são armazenados como uma distância. Um grafo é construído com todas as possibilidades de movimentação e a distância calculada anteriormente é associada a cada uma das arestas de forma que quanto maior a distância, indicando menor afinidade, menor é o peso atribuído. Esse processo gera um grafo com pesos contendo todos os pontos conectados.

A partir do grafo gerado, é obtida outra estrutura, uma *Minimum Spanning Tree* (MST), um grafo sem ciclos que liga todos os nós e cuja soma do peso dos nós é o menor peso total possível. Como deseja-se identificar estruturas formadas por um conjunto de pontos que se move em conjunto, as arestas com menores pesos da MST vão sendo eliminadas;

estas ligações representam pontos de menor afinidade. O processo de eliminação se repete até que uma floresta seja criada a partir do grafo. Quando a floresta é criada, cada um dos componentes é considerado como um agrupamento. Cada agrupamento é considerado uma estrutura independente em movimento e são construídos mapas de densidade dos pontos e da malha. Por fim, é inferida finalmente a trajetória de cada ponto baseando-se no movimento da estrutura independente.

Em (MARKS; HERSHEY; MOVELLAN, 2010), é comentado que o método de detecção de fluxo óptico se aproxima do método de detecção de *template* uma vez que este é um caso especial daquele, onde há uma constante modificação do *template* que se busca na cena.

Mais detalhes sobre métodos de detecção baseado em *templates* são apresentados a seguir.

2.3 Métodos baseados em *Templates*

Métodos baseados em *template* são comumente utilizados para buscar objetos rígidos em cenas onde é feito um casamento entre uma imagem texturizada do objeto que se deseja buscar e uma imagem da cena. Os métodos mais simples podem ser representados por uma janela deslizante que procura a melhor correlação de iluminação entre as duas imagens. Com a evolução dos estudos, abordagens que apresentam o casamento a partir de modelos não rígidos já foram propostos, a exemplo das pesquisas apresentadas em (MARFIL et al., 2004) e em (MARKS; HERSHEY; MOVELLAN, 2010).

Para realizar o casamento, os métodos precisam descrever a imagem do objeto de tal forma a maximizar a chance de casamento sobre diferentes pontos de vista, contemplando cenas que apresentem os objetos parcialmente oclusos. Abordagens mais comuns que utilizam o valor de distribuição das cores nas imagens podem lidar com certa qualidade com detecção sob tais condições para objetos como mãos ou rostos. Além disso, podem ser estáveis também para problemas como profundidade e rotação. Há ainda técnicas estatísticas e outras que procuram histogramas de cores para realizar o casamento. Além destes, existem esquemas voltados para casamento com características apenas de formato e também com uma mescla de formato e cor do objeto. Entretanto, estes mecanismos não são sensíveis a mudanças do formato do objeto e/ou oclusão total do mesmo. (MARFIL et al., 2004)

Para sanar tal problema, surgem os processos que apresentam representação de um *template* dinâmico. O método proposto por MARFIL et al. utiliza de estruturas em forma de pirâmides. Este tipo de estrutura representa geralmente níveis diferentes de uma mesma imagem com qualidade e/ou informação reduzida. No algoritmo proposto por MARFIL et al. são utilizadas pirâmides contendo informações de Matiz (M), Saturação (S) e Brilho

(B), assim como as coordenadas da região analisada. É formada uma pirâmide de 4 níveis, onde são cruzadas informações dos valores de MSB associados a cada região com valores de limiar diferentes para cada nível. Assim, em cada nível, são descartadas as regiões que não apresentarem os valores de MSB dentro dos limiares. O processo de casamento começa com a definição de uma Região de Interesse (ROI - *Region of Interest*) atribuída de forma manual representando onde o objeto que está sendo acompanhado se encontra. Cada ROI tem sua pirâmide calculada e ela é comparada com a gerada para a imagem do objeto. As pirâmides da cena são comparadas com a pirâmide do objeto e, dois a dois, é medido um valor de distância entre níveis iguais das duas pirâmides; caso esse valor esteja abaixo do limiar pré-definido, o casamento é considerado positivo, poupando processamento posterior e tornando o método mais rápido.

Após encontrado o objeto, ainda são feitos ajustes para prosseguimento da análise dos próximos quadros. É construído um *bounding box* para o objeto encontrado e esse *bounding box* é expandido por uma borda. A análise para o quadro seguinte é feita inicialmente na região correspondente à borda. Isso poupa tempo de processamento para casos onde não há mudanças bruscas de posicionamento do objeto alvo entre quadros.

2.4 Métodos baseados em Aprendizagem de Máquina

Vastamente utilizadas para problemas de classificação, a aprendizagem de máquina aparece como uma das formas mais promissoras de auxiliar o processo de reconhecimento de itens. Em (VIOLA; JONES, 2001) foi apresentado um estudo que utiliza de classificadores supervisionados para realizar o reconhecimento de rostos. Em (GOMEZ-DONOSO et al., 2017) é alcançada com grande eficiência a busca de arquivos de referência CAD em nuvens de pontos. Outros estudos utilizam a inteligência computacional como parte de seu processo. O (BAYKARA et al., 2017) utiliza uma rede neural convolucional (CNN - *Convolutional Neural Network*) durante parte do método de busca e reconhecimento de itens feito para Veículos Aéreos não Tripulados.

Três grandes contribuições foram feitas por VIOLA; JONES no seu estudo. Inicialmente é definido o conceito de imagem integral que consiste numa forma de reduzir o custo da análise de características de uma região retangular através da soma dos pixels à esquerda e acima do ponto de referência. Usando as imagens integrais, a soma de dois retângulos pode ser feita em quatro referências a *arrays*. A segunda contribuição consiste no uso do algoritmo AdaBoost para construir o classificador, permitindo um menor número de características importantes a ser selecionado. Por fim, o estudo ainda apresenta uma forma de combinar classificadores mais complexos para criação de uma cascata de forma a ter atenção focada em regiões mais promissoras da imagem, aumentando a velocidade de detecção. (VIOLA; JONES, 2001)

Para realizar a construção do classificador, o algoritmo recebe uma série de imagens de classificação positiva ou negativa e utiliza-se do AdaBoost para determinar quais as características que são buscadas e também no treinamento do classificador. A utilização do AdaBoost para procurar as melhores características consegue diminuir drasticamente a quantidade de características. Além disto, o processo é focado em distinguir retângulos que facilmente diferenciem imagens positivas das negativas.

O dispositivo proposto por VIOLA; JONES descreve ainda a construção de uma cascata de classificadores. Basicamente, ela consiste em uma árvore de decisão degenerada, onde a decisão afirmativa no primeiro classificador dispara a análise de um segundo classificador e assim sucessivamente, sendo interrompido imediatamente assim que um negativo é encontrado. Outros estudos foram desenvolvidos baseando-se no de VIOLA; JONES e, embora ele seja focado para a detecção de rostos, é possível realizar a busca de outros itens, baseando-se no mesmo processo.

Em (GOMEZ-DONOSO et al., 2017), é apresentado o LonchaNet, uma arquitetura para uso de redes neurais convolucionais na realização do casamento entre nuvens de pontos geralmente adquiridos a partir de imagens RGB-D e bases de dados sintéticos. Para realizar o casamento, são capturadas três imagens de todos os objetos na base de dados para criar a rede neural. Cada imagem corresponde a um dos eixos coordenados. A arquitetura prevista apresenta o uso de três GoogLeNets, uma para cada imagem juntas em uma camada superior à de classificação, com intuito de reutilizar o melhor de cada abordagem.

São utilizadas três fatias do objeto, estas representando cada uma um paralelepípedo cuja base é igual ao segmento de plano perpendicular a cada eixo e a altura é igual a 5% do tamanho do modelo. Para formar uma imagem, cada ponto é representado por uma área de 24x24 pixels, formando assim uma área coberta pelos pontos do objeto. As três imagens processadas em três GoogLeNets diferentes têm seus resultados passados para uma camada que conecta o resultado oriundo das três. Como cada uma das três ramificações do GoogLeNet é independente, a arquitetura proposta por GOMEZ-DONOSO et al. aproveita-se da velocidade de se trabalhar com imagens bidimensionais sem perder a informação tridimensional apresentada no modelo.

Além de abordagens como o método de VIOLA; JONES e GOMEZ-DONOSO et al., a aprendizagem de máquina surge como uma etapa essencial em alguns *frameworks* de detecção de objetos. Em (BAYKARA et al., 2017), uma rede neural convolucional é utilizada no processo de reconhecimento de itens. O processo descrito por BAYKARA et al. inicia-se com a aquisição de imagens, passando por um processo de retirada da distorção causada pelas lentes da câmera que é calibrada previamente. Em seguida, realiza um processo de detecção utilizando pontos-chave. São utilizados os algoritmos FAST (ROSTEN; DRUMMOND, 2006) para aquisição dos 200 melhores pontos-chave e FREAK

(do inglês, *Fast Retina Keypoint*) (ALAHI; ORTIZ; VANDERGHEYNST, 2012) para construção dos descritores, e em seguida é realizada uma computação da homografia entre dois quadros utilizando-se do RANSAC (do inglês, Random Sample Concensus) (FISCHLER; BOLLES, 1981).

Após a obtenção da homografia, é realizada uma verificação da quantidade de pixels que vão sendo alterados entre as imagens e, de forma semelhante a um método baseado em fluxo óptico, é realizado o mapeamento de regiões da imagem que se movem e, utilizando-se do filtro de Kalman, é realizado a identificação dos objetos que se movem.

Uma vez identificados, os objetos passam para um processo de reconhecimento, nesse estágio, é utilizada uma rede neural previamente treinada com vídeos obtidos da mesma altura que se imagina que a câmera estará posicionada. O *framework* retrata ainda uma etapa final de posicionamento utilizando-se de um Sistema de Posicionamento Global (GPS - do inglês, *Global Positioning System*). Observa-se que vários métodos são aplicados na construção do *framework* proposto por (BAYKARA et al., 2017); este é um exemplo da relevância de todas as técnicas abordadas nesta monografia.

2.5 Métodos baseados em Arestas

Estratégias que visam realizar o casamento através da detecção de arestas geralmente buscam focar na procura de contornos e bordas dos objetos. É uma alternativa que tende a ser mais barata computacionalmente e apresenta uma possibilidade de busca de objetos em imagens únicas. O RAPiD (do inglês, *Real-time Attitude and Position Detection*) é um estudo de busca baseado em contornos que veio a ser complementado por (ARMSTRONG; ZISSERMAN, 1995). Abordagens mais atuais como a proposta (CHOI; CHRISTENSEN, 2010) criam uma correlação com métodos baseados em pontos chave.

Neste tipo de abordagem, os objetos são representados por primitivas geométricas, geralmente arestas. Este conjunto de primitivas é projetado, considerando uma câmera já previamente calibrada, criando uma malha em duas dimensões de arestas. Com isto, é feita uma aproximação entre essas características e as arestas que constem na imagem de cena. Para realizar a detecção de arestas na imagem de cena, os métodos baseados no RAPiD utilizam-se de um processo de filtragem de imagens para detecção de bordas, seguido da obtenção dos segmentos de retas. Existem muitos estudos correlacionados com isso; este processo é melhor descrito na seção 2.6.

Para realizar a procura do objeto, o RAPiD infere qual a posição inicial do objeto utilizando-se do filtro de Kalman. A partir daí, para cada aresta projetada do modelo são computados i pontos. Para cada ponto i é calculada qual a aresta da cena apresenta a menor distância ortogonal à aresta que contém o ponto. Em posse do valor da distância, é calculada uma re-adequação da projeção do modelo e, ao final da verificação dos valores

para os i pontos, toda a projeção e posição do objeto é re-adequada baseando-se no filtro de Kalman.

O processo apresentado por ARMSTRONG; ZISSERMAN visa corrigir problemas apresentados pelo RAPiD. O método não apresenta tratamento para qualquer tipo de erro. Por exemplo, não importa quão grande seja a distância da aresta mais próxima ortogonal, seu valor sempre é computado para a re-adequação do modelo. Com isso, arestas que apresentem valores muito distantes do ideal causam uma distorção inadequada para o casamento. Além disto, não há nenhuma medida para obtenção dos i pontos, assim como nenhuma correlação entre os i pontos de controle: todos têm a mesma influência, mesmo que existam vários deles realizando o casamento da mesma aresta.

Com intuito de sanar estes problemas, o estudo de ARMSTRONG; ZISSERMAN apresenta uma abordagem que sempre define uma medida de quantidade mínima de pontos de controle para uma determinada aresta em análise forçando a redundância. Com intuito de diminuir os valores calculados em erro, utiliza-se do RANSAC (FISCHLER; BOLLES, 1981) para quais são as arestas provenientes pós filtro de detecção de bordas. Este processo diminui a quantidade de erros apresentados por arestas muito distantes, uma vez que remove grande quantidade de informações não tão relevantes para o problema.

Já o procedimento proposto por CHOI; CHRISTENSEN, embora seja também baseado no RAPiD, apresenta uma nova abordagem mista que envolve casamento com pontos chave e métodos de aproximação de arestas. Para realizar a busca, ele recebe uma imagem e um modelo: uma lista de triângulos em três dimensões (3D) que representam o objeto buscado. Inicialmente, são computadas as várias possíveis poses do objeto baseando-se nos possíveis pontos de vista do mesmo. Essa informação é salva de forma *off-line*. A depender da face buscada, apenas os triângulos que estão visíveis são armazenados. Este procedimento descarta arestas oclusas pelo próprio formato do objeto. Outro procedimento empregado é um processo que é realizado com intuito de simplificar os modelos 3D. Os triângulos adjacentes do modelo são analisados dois a dois descartando-se as arestas comuns a ambos, quando os dois triângulos apresentam angulação menor que 30 graus entre suas normais. Os segmentos de reta remanescentes são chamados de arestas fortes e têm uma maior propensão a serem encontrados na imagem de cena.

Para realizar o processamento da imagem da cena, são utilizados o detector de bordas Canny; já para estimar a posição do objeto, o método proposto por CHOI; CHRISTENSEN lança mão da utilização de uma busca por pontos chave, o SURF (BAY; TUYTELAARS; GOOL, 2006). Porém, apenas no primeiro quadro é realizado o casamento utilizando o SURF; nos quadros seguintes é feito um processo de cálculo de um vetor de erros de forma semelhante ao RAPiD: a aresta forte projetada do modelo é particionada em i pontos, e, em cada posição do vetor, é armazenada a distância entre o i-ésimo ponto da aresta e a linha correspondente mais próxima na cena. O vetor que pode ser construído

com os valores das distâncias é chamado de vetor de erro, a posição real do objeto é calculada por minimização do erro entre os quadros de um vídeo. Segundo os autores, essa abordagem consegue realizar a busca em tempo real. Ela mostra-se eficiente, pois, uma vez encontrado os objetos, a re-leitura do erro entre quadros se faz mais simples pois dificilmente o objeto se move abruptamente tornando a função de construção do vetor de erro mais custosa; e, caso o erro seja demasiadamente grande, o sistema basta ser re-iniciado utilizando-se o SURF.

2.6 Detecção de Bordas

Retas e/ou cônicas são essenciais para descrever objetos complexos a serem buscados, assim como arquivos de descrição CAD são constituídos por uma lista de triângulos e normais. Em cenários onde não é possível obter-se uma imagem prévia para classificação do objeto buscado, sendo impedido o uso de características como textura, faz-se necessário um método que aproxime a informação das imagens da cena (RGB) com primitivas matemáticas semelhantes aos modelos.

Logo, a detecção de bordas mostra-se como uma das mais robustas formas de processamento de imagem para realizar a busca de objetos sem textura. Além disto, uma vez filtrada, a imagem apresenta muito menos características para serem analisadas. Assim, este tipo de tratamento torna-se comumente usado não apenas em métodos voltados para casamento de arestas como também em métodos de casamento de *templates*.

A seguir, apresentamos um dos mais difundidos detectores de bordas, o método proposto por CANNY, seguido de uma breve explanação do método de COSTA; MELLO; SANTOS. Por fim, analisamos a técnica de utilização da Transformada de Hough para obtenção de retas.

2.6.1 Canny

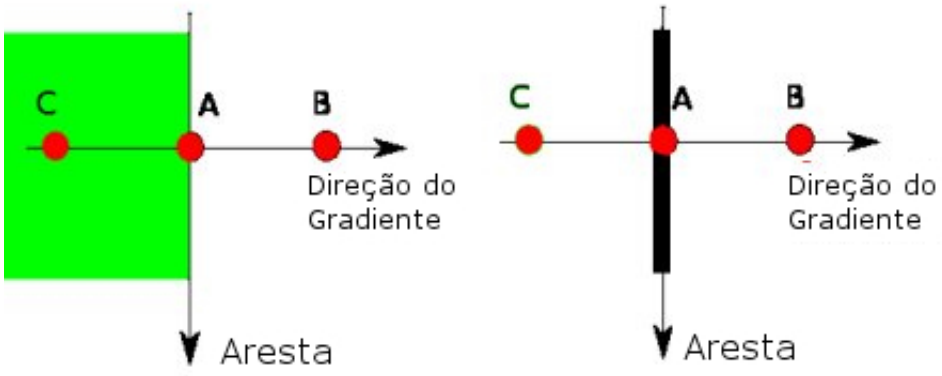
CANNY define que existem três requisitos para definição de um bom método de detecção de bordas:

1. Boa Detecção; não pode haver risco de uma borda real não ser detectada.
2. Boa Localização; as bordas encontradas devem ser o mais próximo possível da borda real.
3. Apenas uma única resposta para uma mesma aresta; um bom localizador de bordas deve evitar que ruídos remanescentes possam ser encontrados correlacionados a uma borda real.

Inicialmente, com intuito de diminuir os ruídos e falsos positivos, a imagem passa por uma filtro Gaussiano para redução de ruídos. Quanto maior o tamanho da máscara, menor a quantidade de características finas encontradas. Diminuída a quantidade de características, o processo segue para detecção dos gradientes da imagem. Estes vetores têm angulação perpendicular a uma borda, portanto, faz-se providente descobrir-los. Logo, CANNY propõe um processo de filtragem em quatro direções, vertical, horizontal e nas duas diagonais, para obter o gradiente das regiões. Cada região é mapeada para os ângulos $0°, 45°, 90°$ e $135°$. Caso o ângulo θ esteja no intervalo entre $[0°, 22°]$ ou entre $[157°, 180°]$ ele é mapeado para $0°$.

Em posse dos valores de angulação e intensidade dos gradientes da imagem, é realizado um processo de retirada de pixels que não venham a fazer parte de uma borda. Para o tal, é realizado um processo de supressão dos não máximos, onde, analisados todos os pixels, são considerados apenas os que apresentarem gradiente como máximo local. A Figura 1 representa o comparativo.

Figura 1 – Supressão dos não máximos



O Ponto A se encontra na Aresta, enquanto B e C se encontram no gradiente. O ponto A é comparado com os outros dois para verificar se ele é um mínimo local, caso seja, ele é considerado nos passos seguintes, caso não, é atribuído o valor nulo.

Do processo de supressão, restam apenas bordas mais estreitas. Contudo, após este processo ainda restam bordas ruins provavelmente remanescentes de variações de cores e ruídos. Assim, como tratamento posterior, é aplicada uma verificação dupla de limiar. São utilizados dois valores de limiar: um valor limiar de mínimo e um valor de máximo. Estes são usados na definição de quais bordas serão consideradas adequadas da seguinte forma:

- Permanecem as bordas cuja intensidade do gradiente esteja acima do limiar máximo;
- Permanecem também as bordas cujo valor esteja entre os limiares e tenham algum tipo de adjacência com um ponto de limiar acima do valor máximo, estas arestas são consideradas componentes das demais;

- São descartadas qualquer aresta que esteja abaixo do limiar mínimo;
- São descartadas as que tenham valores dentro dos dois limiares mas que não apresentem nenhum tipo de adjacência com pixels que apresentem valores de gradiente forte;

Para se considerar um pixel adjacente a outro, analisa-se se ele está dentro de uma região de 8-vizinhança em relação ao pixel analisado. Por fim, as bordas restantes são consideradas fortes. Embora o Canny apresente bons resultados diante de várias aplicações, pode apresentar grandes dificuldades em ser parametrizado empiricamente, logo, outros métodos precisaram ser desenvolvidos a exemplo do (COSTA; MELLO; SANTOS, 2013) que será explanado a seguir.

2.6.2 DifFocusColor

A abordagem apresentada por COSTA; MELLO; SANTOS apoia-se em conceitos oriundos da análise da percepção humana e utiliza-se de características como a diferenciação cromática e a percepção da densidade textural. O processo de consideração da diferença cromática é denominado dRdGdB. Para realizar tal ponderação, são consideradas de forma individual as intensidades de cada componente cromático de uma imagem. Os valores Δ obtidos serão a soma das diminuições dos valores do componente analisado pelos demais componentes cromáticos. Por exemplo, imagens que têm 3 componentes (RGB) terão valores ΔR ΔG e ΔB computados da seguinte forma:

- $\Delta R = R - G + R - B$
- $\Delta G = G - R + G - B$
- $\Delta B = B - R + B - G$

Cada ponderação efetuada é armazenada em um novo mapa cromático e, unido à informação completa das cores iniciais, é criada uma imagem com seis dimensões. O DifFocusColor pode ser descrito por completo nos cinco passos a seguir:

1. Inicialmente, é feito um processo de filtragem morfológico para supressão da textura. Neste passo, são realizadas operações de abertura e fechamento na imagem e seus resultados combinados para que haja a supressão sem perda de contornos. Após isso, é aplicado o mecanismo de obtenção de espaço cromático dRdGdB gerando uma imagem com seis dimensões.
2. As seis dimensões da imagem são suavizadas utilizando-se de duas máscaras Gaussianas. Esse processo gera duas imagens diferentes que têm também níveis de detalhes

diferentes, uma mais e outra menos detalhada. A imagem mais detalhada utiliza uma máscara com valores mais baixos de tamanho e σ para a obtenção da média local. Consequentemente, a imagem que apresentará menores detalhes sofre um desfoque maior, com valores de tamanho de máscara quadráticos em relação à máscara anterior e valor de σ dobrado.

3. Ambas as imagens com poucos ou muitos detalhes do passo anterior, passam por um processo de detecção de bordas usando as máscaras de Prewitt (LIPKIN, 1970) vertical, horizontal e diagonal.

4. Neste passo, ocorre a união das duas imagens de seis dimensões. Cada uma, individualmente, une suas seis dimensões em uma única, considerando sempre o maior valor para os pixels entre cada uma das cores. Para unir as duas imagens de uma dimensão em uma única imagem é realizada a multiplicação entre seus componentes e realizada a normalização dos valores dos pixels no intervalo $[0, 1020]$

5. No ultimo passo, a imagem unidimensional pode passar por um processo de estreitamento das arestas os autores evidenciam que esse passo pode eliminar arestas que já sejam estreitas o bastante por isso, seu uso pode ser moderado pela necessidade. O estreitamento é realizado através do uso de erosão morfológica, que pode causar alguma perda da informação.

2.6.3 Transformada de Hough

Propostas como DifFocus e o Canny podem ser utilizadas para tratar a imagem de tal forma que apenas características realmente pronunciadas sejam evidenciadas. Porém, para sistemas de busca de objetos faz-se providente também descobrir quais pixels correspondem a uma mesma curva. Em abordagens baseadas no RAPiD esta informação pode ser crucial para diminuir o tempo de processamento assim como melhorias de resultado através da análise mais aprofundada da cena, uma vez que pode haver uma melhor comparação entre as características encontradas nas curvas oriundas da imagem de cena e das presentes no modelo projetado.

Já utilizada em abordagens como (CHOI; CHRISTENSEN, 2010), a Transformada de Hough foi proposta em 1989 pelo próprio Hough, como comenta PEDRINI; SCHWARTZ. A transformada foi descrita inicialmente com intuito de descobrir retas e circunferências, mas pode ser estendida para quaisquer curvas. Embora seja bastante útil, a transformada requer um grande número de iterações o que pode ser inapropriado para diferentes tipos de aplicações (PEDRINI; SCHWARTZ, 2008). No estudo apresentado neste trabalho, é apresentada a forma de obtenção de retas utilizando a transformada, as demais alternativas demonstraram ser inviáveis para o escopo deste projeto.

Seja a Equação 2.1, a equação da reta em duas dimensões onde o valor de a representa o coeficiente angular da reta e b é o ponto que a mesma corta o eixo das ordenadas,

$$y = ax + b \tag{2.1}$$

fixados os valores de um ponto $P_1(x_1, y_1)$ existem infinitas retas passam por P_1, de acordo com a equação $y_1 = ax_1 + b$. Pondo-se em evidência os valores de a e b, a equação pode ser reescrita como a Equação 2.2

$$b = y - ax \tag{2.2}$$

O espaço criado por b e a é conhecido como espaço de parâmetros e todas as retas que passarem pelo ponto P_1 obedece à $b = -ax_1 + y_1$, assim, todos os pontos colineares numa imagem geram retas que se interceptam num mesmo ponto no espaço de parâmetros. Os pixels colineares podem representar uma mesma reta na imagem, logo, se interceptam no mesmo ponto no espaço paramétrico. Sendo assim, identificar muitas retas intersectadas no espaço paramétrico é equivalente a identificar candidatas a retas na imagem.

Mapeando-se os pixels para retas no espaço paramétrico e escolhendo os pontos de intersecção de várias retas é possível agrupa-los em segmentos de retas. Esta abordagem encontra problemas para casos em que a reta da imagem tenda a ser vertical, isso indica que os valores do coeficiente angular tendem ao infinito. Para solucionar tal problema, Hough propôs a utilização da equação polar da reta

$$\rho = x\cos\theta + y\sin\theta \tag{2.3}$$

em que ρ é a distância perpendicular da reta à origem $(0, 0)$ e θ é a angulação da reta com o eixo x. Ao invés de se utilizar um plano com os parâmetros (a, b) é utilizado um plano com parâmetros (ρ, θ) denominado espaço de Hough. Pontos colineares no plano (x, y) são curvas senoidais que se interceptam no espaço de Hough. O processo exige ainda a discretização do espaço de Hough. Observa-se que θ pode variar de 0 até 180, já o valor de ρ varia entre 0 e$\sqrt{m^2 + n^2}$ para uma imagem m x n(PEDRINI; SCHWARTZ, 2008).

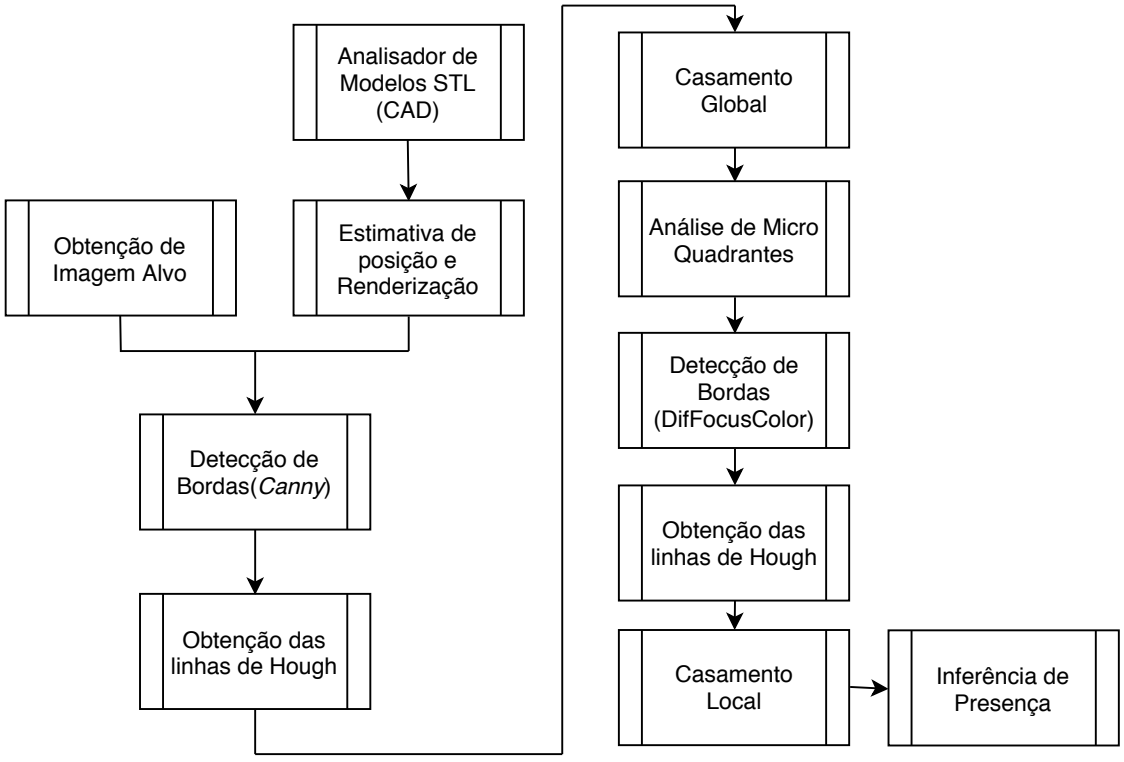
Mapeando-se os pixels não para o espaço paramétrico mas sim para o espaço de Hough, o espaço discretizado é analisado célula por célula e quanto mais pixels tiverem sido mapeados para determinada célula, maior a chance deles representarem uma reta. Nota-se ainda que a precisão é dada pela granularidade da divisão das células no espaço de Hough. Isso permite que mesmo que a detecção de bordas não seja perfeita, bons resultados podem ser encontrados, em compensação, aumenta consideravelmente o custo computacional de uma solução que o utilize.

3 Método Proposto

Este capítulo destina-se a explanação do método proposto nesta monografia o qual é baseado em detecção de bordas, ou seja, realiza o casamento através da comparação entre arestas de uma cena e de um modelo. A cena é uma imagem que pode conter ou não o objeto, ela representa o local onde se procura um ou mais itens; sendo assim, o modelo é uma lista de arquivos de referência técnica do(s) item(ns) buscado(s).

Na abordagem descrita neste trabalho, é usado um modelo refinado (posteriormente referido apenas como modelo) que é um conjunto de arestas obtidas a partir de um arquivo CAD(Computer-Aided Design). Os processos aqui descritos vão desde a obtenção do modelo refinado, até a inferência de presença deste na imagem de cena. Ao todo ele consiste em 11 etapas que são enunciadas em ordem na Figura 2.

Figura 2 – Etapas



3.1 Pré-Processamento

Obtidas as fotos da cena de forma *offline* essas são usadas para alimentar o programa juntamente com uma lista de arquivos de referência CAD para a obtenção do modelo. A imagem de cena segue para detecção de bordas, assim como métodos baseados no RAPID,

ou seja, primeiro passa por uma filtragem para detecção de bordas e posteriormente é extraída uma lista de arestas utilizando-se da Transformada de Hough. Essa lista de arestas é utilizada para realização do casamento como descrito na seção 3.4.

A lista de arquivos de referência fornece as informações do formato do objeto da lista, uma nuvem de triângulos e suas respectivas normais, e o posicionamento de cada um dos seus componentes, isto é, a posição de cada um deles num plano cartesiano de três dimensões. Uma vez que os objetos apresentados para serem buscados na lista podem estar sobrepostos entre si, quando a cena for construída com todos eles, além de cada um conter características oclusas a partir do ponto de vista, faz-se necessária um aprimoramento da informação que consta nos arquivos CAD.

Assim, com intuito de realizar tal refino, a seguir são apresentadas duas abordagens. Em ambas as abordagens são guardadas as informações relativas ao posicionamento de cada componente em relação ao sistema de coordenadas. Essa informação é usada como um guia para inferência de posição na seção 3.6.

A primeira abordagem baseia-se no método apresentado em (CHOI; CHRISTENSEN, 2010). Os autores demonstraram um processo de refino do *layout* do modelo CAD para achar arestas que podem ser consideradas fortes. Para tal, realiza-se um estudo entre todos os triângulos representativos do modelo, sendo escolhidas, para um modelo aprimorado, apenas as arestas que apresentem triângulos com angulação maior que 30 graus entre si. As arestas provenientes deste refino são consideradas características fortes, ou seja, com grande propensão a serem encontradas na foto do objeto, sendo então usadas posteriormente para realização do casamento. Este método, porém, mostrou-se bastante custoso devido à alta complexidade computacional, além de apresentar bordas remanescentes que desaparecem a depender da iluminação do local.

Assim, como forma de aproximar-se ainda mais dos componentes reais da cena, e tornar o processo mais rápido, a segunda abordagem apresentada consiste na utilização de um modelo obtido a partir de uma renderização com projeção ortogonal e pontos de iluminação estratégicos. Uma vez renderizados em projeção os modelos CAD, a imagem é salva e passa pelo processo de detecção de bordas; ou seja, sendo submetida a uma filtragem, para esta etapa foi escolhido o filtro Canny devido a sua menor complexidade computacional, e, posteriormente, obtenção de arestas utilizando-se da Transformada de Hough. Essas linhas obtidas constituem a nova lista de arestas representativa do modelo e são usadas nas partes posteriores do método de casamento. Este processo cria mais bordas onde consta a iluminação e facilmente elimina as arestas oclusas e mais suavizadas.

3.2 Obtenção de Características de Imagens

Como descrito na Figura 2 a estratégia utiliza-se inicialmente do filtro de Canny tanto no pré-processamento das imagens provenientes da câmera, como na obtenção do modelo. Além do filtro Canny, foi utilizada uma variação do filtro DifFocusColor (COSTA; MELLO; SANTOS, 2013) como uma abordagem extra para a obtenção das bordas de uma imagem.

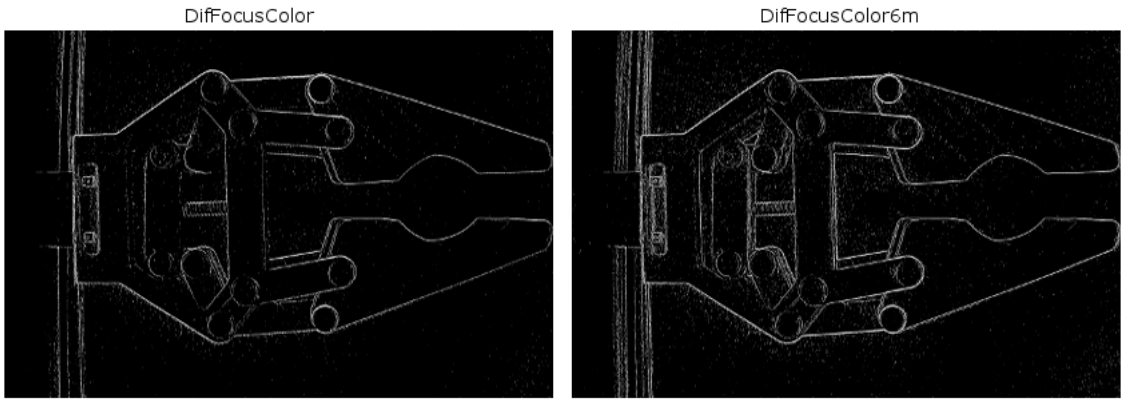
As alterações ao filtro consistem em duas novas máscaras adicionadas ao passo de detecção de bordas. O filtro utiliza-se de máscaras que convoluem com a imagem para fins de detecção de bordas através da combinação de imagens geradas. Assim, além das quatro máscaras descritas, são utilizadas as máscaras vertical inversa(a) e horizontal inversa(b):

$$a = \begin{pmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{pmatrix} b = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{pmatrix}$$

Essas duas máscaras adicionais refinaram o resultado encontrado pelo filtro, como pode ser visto na Figura 3. A alteração aqui apresentada, denominou-se como DifFocusColor6m.

Além da adição das máscaras, um método de limiarização foi adicionado, para uso do filtro em associação com a Transformada de Hough. A imagem de saída do filtro é truncada entre [0, 255]; todos os pixels acima ou iguais a determinado limiar ζ recebem o valor de 255.

Figura 3 – Alterações DifFocusColor



Nota-se uma maior quantidade de bordas detectadas na parte central da imagem que é gerada utilizando-se do filtro DifFocusColor6m.

Uma vez que a imagem passou pela filtragem, é utilizada a Transformada de Hough para detecção dos segmentos de reta da imagem. Este método de detecção de retas usando os filtros e a Transformada de Hough pode mostrar-se bastante complexo de ser

parametrizado empiricamente. Assim, apenas para a imagem oriunda da renderização do modelo é aplicado um processo de refino das linhas de Hough descrito a seguir.

3.3 Refino das Retas obtidas a partir da Transformada de Hough

Embora a transformada de Hough apresente bons resultados para características gerais, mostrou-se conveniente a criação de uma forma de reduzir ainda mais segmentos que podem ser considerados ruídos, ou seja, eliminação de segmentos de reta pequenos e/ou sobrepostos, assim como a expansão máxima dos segmentos que apresentassem um certo tipo de adjacência, como descrito a seguir.

São analisadas de forma exaustiva as arestas entre si, de forma que, sejam dados um limiar ϵ de distância, e um de angulação θ e seja uma aresta representada por dois pontos em duas dimensões que indicam a localização do segmento de reta na imagem que o gerou. As arestas são alteradas da seguinte forma:

1. Cada aresta analisada que tenha seus dois pontos com distância menor que η até uma mesma aresta, é descartada a menor aresta.

2. Caso uma aresta tenha um ponto com distância menor que η até a outra e a diferença de ângulo entre elas seja menor que θ, é adicionada uma nova aresta ao modelo correspondente aos pontos mais externos das duas arestas menores que são em seguida descartadas.

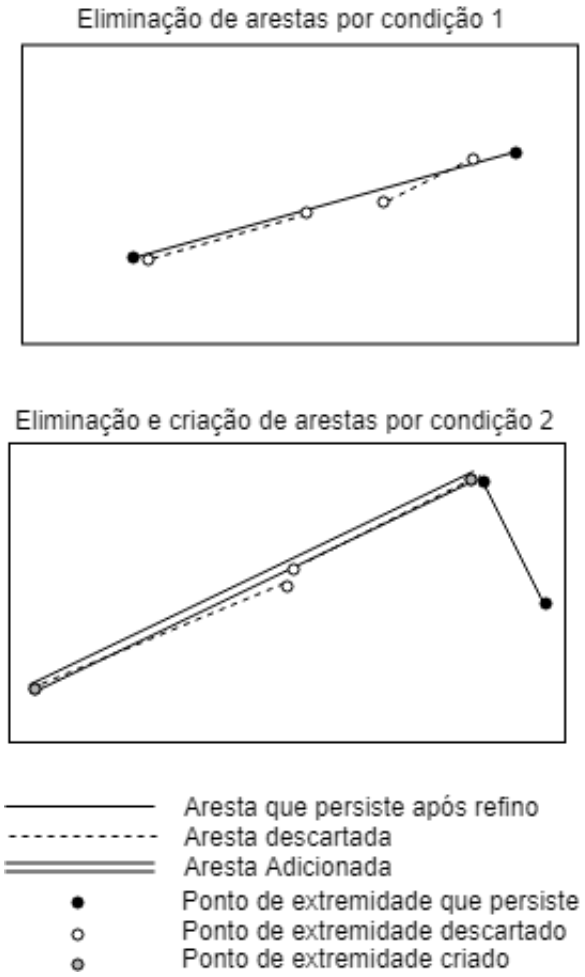
Este processo está representado na Figura 4 e se repete enquanto for possível diminuir a quantidade total de arestas. Como resultado, o modelo refinado apresenta arestas de maior comprimento e uma quantidade menor de arestas. Trata-se de um processo custoso, mas, como pode ser feito durante um pré-processamento, seu impacto sobre a usabilidade final no método é reduzido.

3.4 Casamento

Obtidas as arestas oriundas do modelo e da cena, é iniciada a análise de proximidade entre as características das duas fontes. Em métodos como o RAPID que realizam o casamento e procura de objetos baseados em modelos é checada a distância euclidiana ortogonal a pontos da aresta de modelo até as arestas de cena mais próximas (DRUMMOND; CIPOLLA, 2002).

Com intuito de diminuir o espaço de busca, o método aqui apresentado realiza a abordagem apresentada na Figura 5. Consideram-se dois valores de limiar: de distância α e de angulação β. Cada aresta do modelo é inserida em um *bonding box*. Caso este tenha

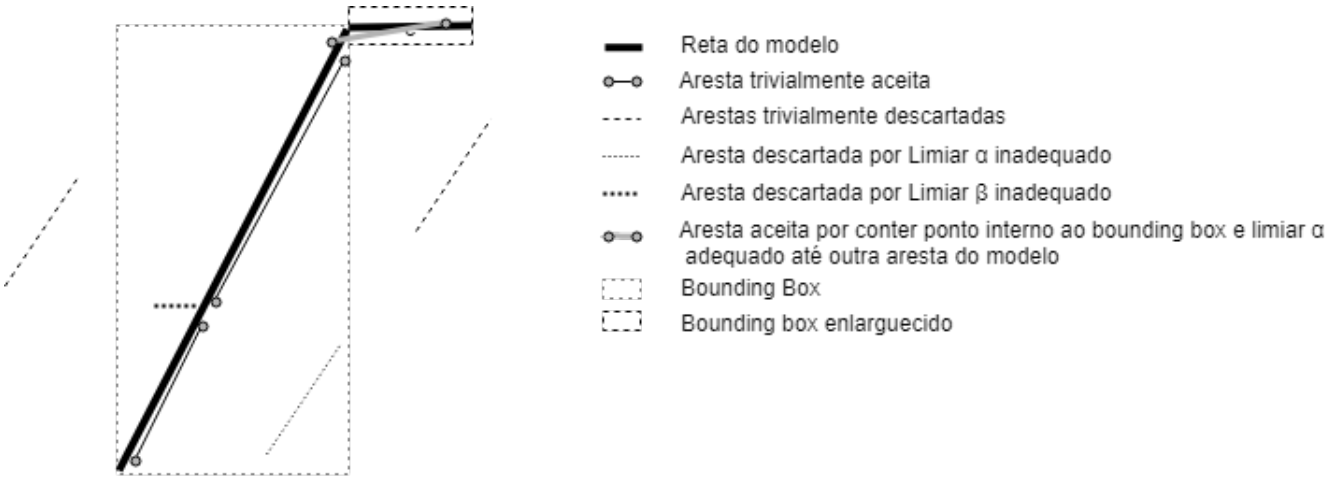
Figura 4 – Refino de Arestas



largura ou altura menores que α é criado um outro *bounding box* com dimensões expandidas por esse mesmo valor; caso contrário, apenas o próprio *bounding box* é considerado. A aresta de cena é considerada candidata à análise apenas caso esteja com ao menos um ponto dentro deste *bounding box*.

A análise prossegue para cada aresta considerada útil no passo anterior. Caso os dois pontos da aresta de cena estejam internos ao *bounding box*, isto indica que a aresta de cena está completamente inclusa dentro dele; assim, são medidas a angulação e a distância relativa ao modelo e, comparando-se estes valores com α e β, a aresta é então considerada uma correspondência ou não. Se apenas um dos pontos da aresta está dentro do *bounding box* é contabilizada a distância do ponto sobressalente até as outras arestas do modelo e, caso seja encontrada alguma aresta de modelo próxima o bastante deste ponto, é comparada sua angulação à β. Se esta aresta de cena encontre correspondência durante estas comparações é então armazenada para os passos seguintes, do contrário é descartada.

Figura 5 – Escolha de Arestas Analisadas



3.5 Análise de Microquadrantes

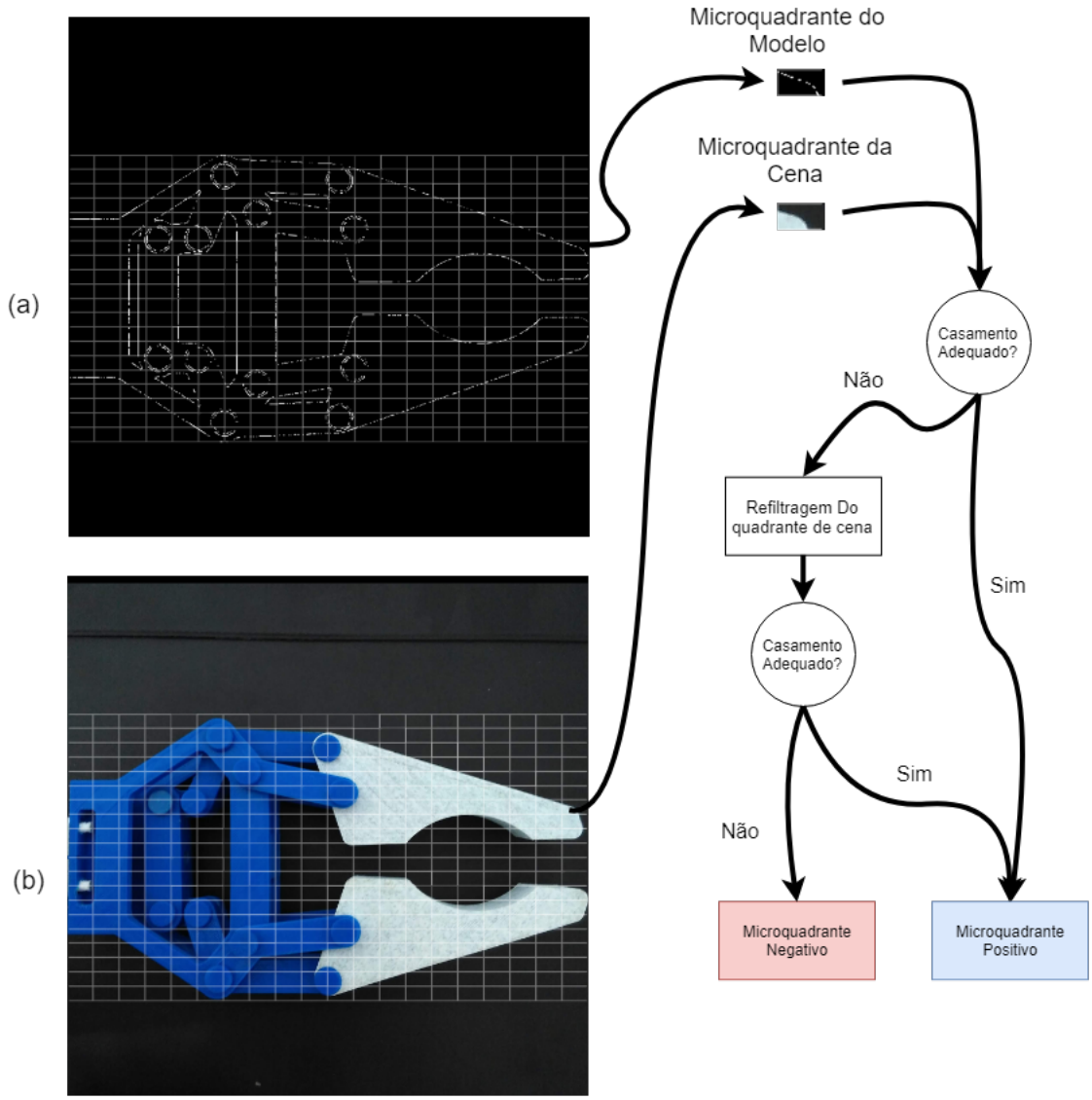
Como forma de avaliação de presença dos componentes do modelo na imagem, lança-se mão do conhecimento da posição que o componente deve aparecer na imagem de cena. Uma vez que o modelo é projetado em duas dimensões um *bounding box* que contenha todas as peças buscadas é construído. O mesmo é então fracionado em estruturas menores designadas microquadrantes. Logo, os microquadrantes são estruturas compostas por: duas coordenadas, representando cada uma um ponto da diagonal do retângulo da região que deseja-se mapear; um conjunto de arestas que estão presentes dentro do microquadrante mapeado. Um mapa de microquadrantes é uma lista contendo todos os microquadrantes de uma determinada área de uma imagem.

Nesta etapa, são preenchidos dois mapas de microquadrantes. O primeiro mapa representa a área do modelo projetado em duas dimensões; cada microquadrante contém as arestas do modelo internas a ele. Além deste, outro mapa é criado, o mesmo é limitado pela área onde o modelo tem que aparecer na cena, aqui todo microquadrante contém linhas obtidas através da Transformada de Hough e que passaram pela etapa de casamento descrita na seção 3.4. Linhas que ultrapassem os limites de determinado microquadrante passam por uma etapa de corte em duas dimensões com o método de Cohen-Sutherland (NEWMAN; SPROULL, 1979).

De cada mapa é extraída a informação, microquadrante por microquadrante, de qual o tamanho das arestas que estão presentes no mesmo. Então, caso haja arestas de modelo em determinado microquadrante o tamanho de suas arestas é comparado ao tamanho das arestas no microquadrante de mesmas coordenadas do mapa de arestas de cena. Um exemplo comparativo pode ser visto na Figura 6. Baseando-se num limiar de comparação γ, cada microquadrante é analisado quando existem arestas de modelo dentro

do micro quadrante e a análise pode ser positiva ou negativa. A análise positiva indica que o tamanho das arestas de cena encontrado é próximo ou maior que γ, do contrário, a análise é negativa.

Figura 6 – Comparativo entre Microquadrantes de mesmas coordenadas



Os microquadrantes que apresentam análise negativa, nesta etapa, são então submetidos a um novo processo de filtragem. Como descrito na seção 3.2 a imagem de cena é novamente filtrada, sendo a escolha de qual filtro feita de forma empírica baseando-se nas características da cena que se deseja analisar. São feitas modificações à filtragem inicial: apenas a porção relativa ao microquadrante é filtrada com o novo filtro ou com o mesmo filtro anterior e nova parametrização. Caso sejam encontradas arestas que realizem o casamento como descrito na seção 3.4 e seu tamanho seja adequado segundo o valor de γ, o microquadrante tem análise final positiva.

3.6 Inferência de Presença

Uma vez que todos os microquadrantes foram duplamente analisados, é estimada agora a presença ou não do modelo como um todo ou dos componentes na cena. Para analisar os componentes de forma individual é usada a informação remanescente ao refino, tal qual descrito na seção 3.1, um grupo de arquivos CAD representa o modelo tridimensional, sendo cada componente da peça respresentado por um arquivo deste grupo, é possível rastrear qual a posição trdimensional, assim como bidimensional, relativa de cada componente do grupo de entrada de forma individual. Como almeja-se encontrar o modelo e possíveis imperfeições, foram produzidos dois tipos de saída para designar a presença de um objeto. A primeira baseia-se na quantidade total de microquadrantes com análise positiva dividido pela quantidade total de microquadrantes do modelo não vazios. Caso esta análise apresente mais de 70% de microquadrantes corretamente encontrados, a peça é tida como positivamente encontrada.

Uma segunda abordagem tenta identificar a maior quantidade de estruturas presentes na peça e, para tal, classifica os microquadrantes que forem mais externos ao componente como microquadrantes de borda e os internos a estes, como microquadrantes de estruturas centrais. Isso permite uma análise mais detalhada de alguma possível imperfeição no modelo como um furo ausente ou mesmo uma peça com estruturas internas semelhantes desalinhadas ao modelo.

Observa-se que é possível expandir um pouco mais a análise: caso os microquadrantes mais externos ao topo e à base da projeção da peça forem encontrados, é possível inferir uma pequena translação desta na largura da cena, pois translações mais bruscas depreciariam a análise do topo e base. Analogamente, caso apenas os externos laterais sejam encontrados, a peça tende a estar levemente transladada em relação à altura da cena. Ainda assim, a procura da posição que a peça se encontra na cena, foge do escopo deste estudo.

4 Experimentos

Neste capítulo, são apresentados experimentos do método proposto. Foram variados valores de entrada de um sistema que implementa o método. O sistema foi completamente desenvolvido, utilizando-se da linguagem de programação $C++$ (CPLUSPLUS, 2018). Esta demonstrou ser a mais adequada para a produção do programa devido à sua familiaridade com as bibliotecas que foram utilizadas no projeto, assim como sua possibilidade de otimização avançada dos processos envolvidos.

Das variações dos valores de entrada, os mais adequados limiares foram definidos de forma empírica. Em posse dos limiares, foram definidos dois valores comparativos de quantidade de microquadrantes, realizando um mapeamento de vinte microquadantes por dimensão da projeção e quarenta microquadrantes por dimensão, totalizando, respectivamente, 406.400 áreas a serem duplamente filtradas. Três fotos representativas de cenários foram escolhidas para demonstração do método.

4.1 Definição de ambiente e Pré-processamento

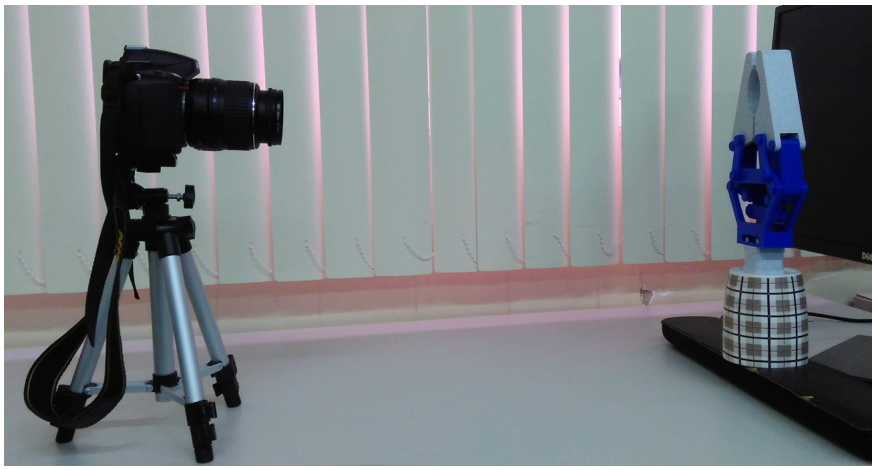
Para realizar o reconhecimento do objeto, o sistema precisa ser alimentado com fotos do ambiente que se deseja procurar o item e também uma descrição do objeto. Para o ambiente, foram obtidas imagens com câmera fixa e com foco, zoom e posicionamento considerados ideais, ou seja, ao obter-se uma imagem onde ocorre um casamento perfeito, os objetos aparecem completos e já com seu posicionamento e sua orientação corretos. Assim, a câmera deve ser colocada direcionada à esteira ou plataforma onde os objetos devem ser identificados. Uma foto da montagem para realização dos experimentos pode ser observada na Figura 7. Foi utilizada uma câmera de foco fixo com abertura 4mm, com fotos tiradas em resolução de 4 *megapixels*.

Após a obtenção das fotos da câmera, elas são consideradas imagens de cena e são repassadas como entrada para o sistema. Para representar o modelo buscado, foi impresso em três dimensões o modelo de uma pinça de robô (WASU, 2018)[1].

Como descrito no Capítulo 3, a lista de arquivos de descrição CAD é utilizada para gerar os modelos que serão buscados. A renderização do modelo foi feita, utilizando-se o ambiente de desenvolvimento livre OpenGL (do inglês, *Open Grafics Library*) juntamente com uma biblioteca com funções voltada para renderização de gráficos bidimensionais e tridimensionais. Seu uso foi escolhido por apresentar boa qualidade e desempenho, utilizando otimizações da GPU (OPENGL, 2018). O modelo é projetado utilizando-se uma

[1] A impressora utilizada foi uma XYZ35 RecifeMecatron com camadas de $0,02$mm e material Pla.

Figura 7 – Montagem de Câmera



projeção ortogonal, os planos limítrofes do volume de projeção são baseados no *bounding box* criado pela união dos aquivos de referência CAD. Como discutido na seção 3.1, para enfatizar arestas salientes assim como suprimir as bordas mais suaves, é utilizada uma iluminação ambiente, aliada a um foco de luz próximo a uma das extremidades do volume de projeção. Por fim, uma imagem do modelo renderizado é capturada do *buffer* da tela de exibição e então submetida à detecção de bordas.

Como vários processos de filtragem e manipulação de imagens são necessários, foi utilizada a biblioteca de licença livre OpenCV (do inglês, *Open Source Computer Vision Library*) (OPENCV, 2018). No processamento da imagem do modelo obtida no passo anterior, foi utilizado o detector de bordas Canny, e, como explicado na subseção 2.6.1 são necessários dois valores de limiar para escolha das bordas. Foram escolhidos então, de forma empírica os valores de limiar inferior e superior respectivamente iguais à 25 e 50. A Figura 8 exemplifica a obtenção do modelo de arestas (b) a partir da imagem renderizada (a).

Uma vez obtida a imagem proveniente do tratamento do borda, é feita a obtenção de segmentos de retas para formarem o novo modelo, utilizando a Transformada de Hough. A discretização do espaço de Hough, explicado na subseção 2.6.3, é feita variando-se ρ pixel por pixel e θ em graus. Foi considerado o número mínimo de intersecções que uma reta candidata no Espaço de Hough precisaria ter iguais a cinco. Ainda foram considerados número mínimo de pixels para que um segmento fosse considerado, assim como uma distância mínima para fazer a ligação entre pixels e considerá-los uma aresta, sendo esses valores respectivamente iguais à 1 e 5, todos os valores foram obtidos empiricamente. Um exemplo da obtenção de segmentos a partir da imagem filtrada é apresentado na Figura 9.

Figura 8 – Obtenção do Modelo Renderizado

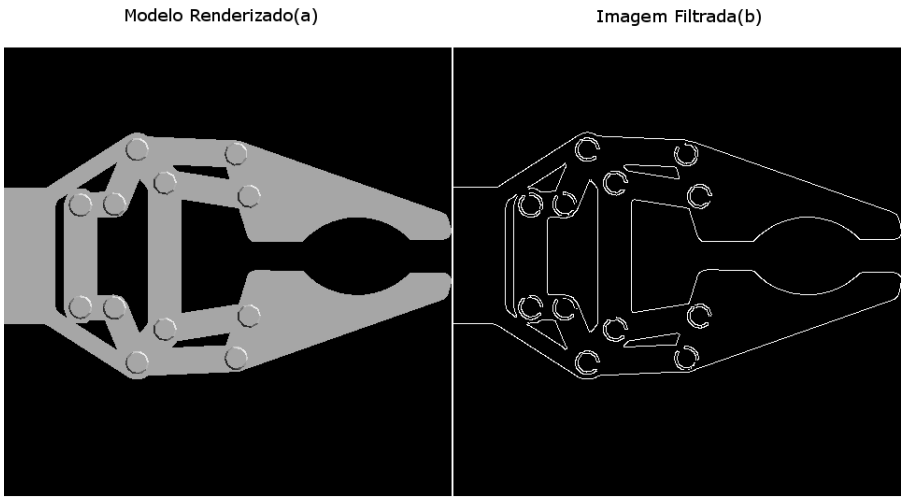
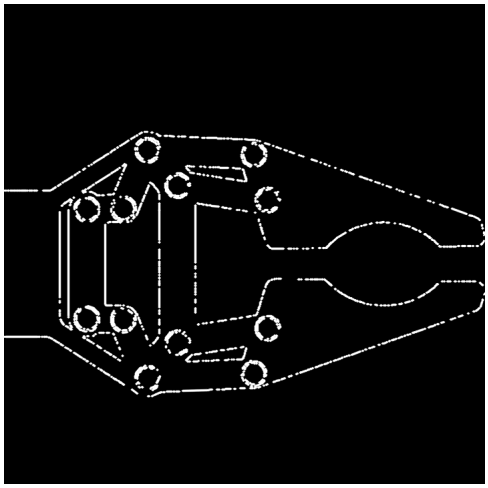


Figura 9 – Modelo Para Casamento

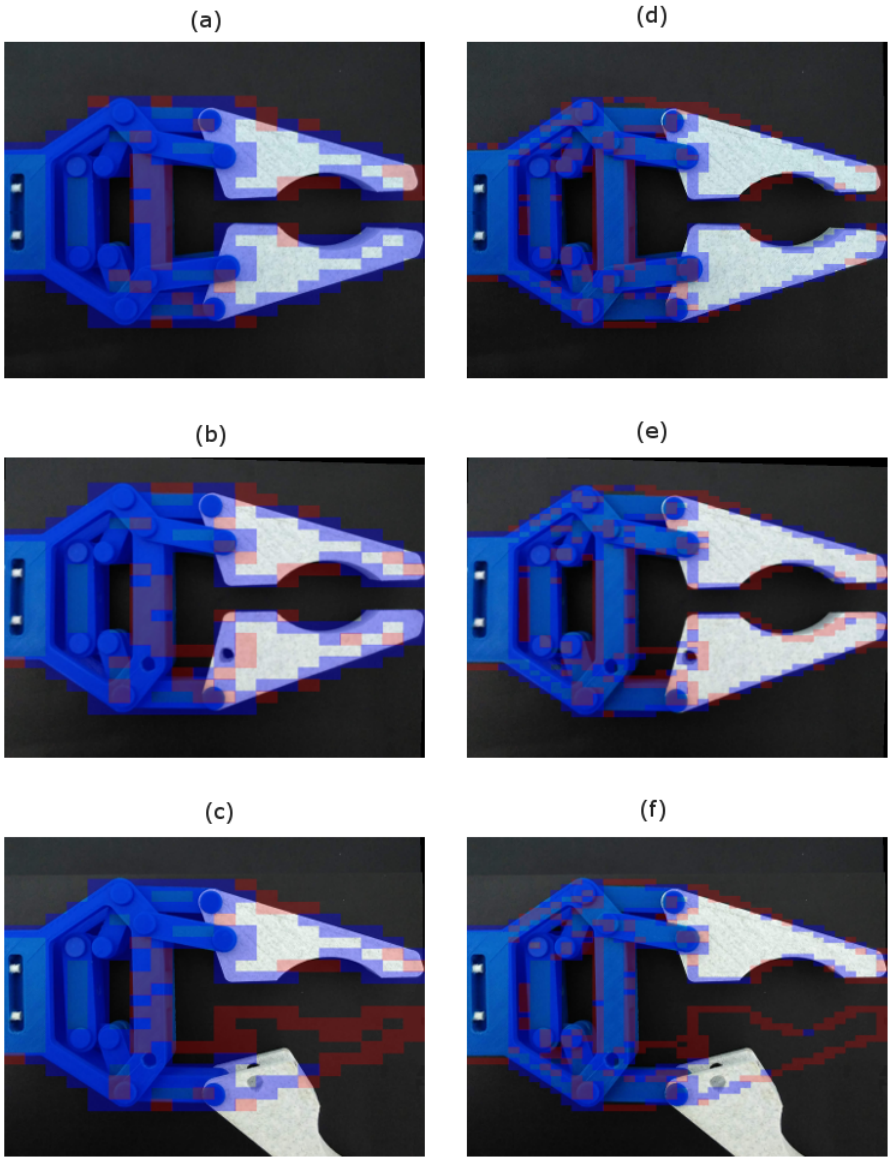


4.2 Testes e Comparativos

Como o método proposto baseia-se na análise de microquadrantes duplamente filtrados, foram realizados comparativos entre o uso de diferentes filtros, assim como diferentes tamanhos dos retângulos relativos aos microquadrantes. As imagens na Figura 10 apresentam os comparativos de diferentes ambientes, apontando os microquadrantes que são considerados positivos, ou seja, que realizaram casamento, em azul e, analogamente, os considerados negativos em vermelho. O objeto com todos os componentes em posição correta é apresentada nas imagens (a) e (d), com um pequeno componente ausente nas imagens (b) e (e) e, por fim, com o mesmo componente ausente e uma peça maior deslocada como pode ser observado em (c) e (f). A figura apresenta dois mapeamentos: um feito com 400 microquadrantes relativos à coluna esquerda da Figura 10 e outro com 6.400

microquadrantes relativos à coluna da direita da Figura 10.

Figura 10 – Resultado Casamento

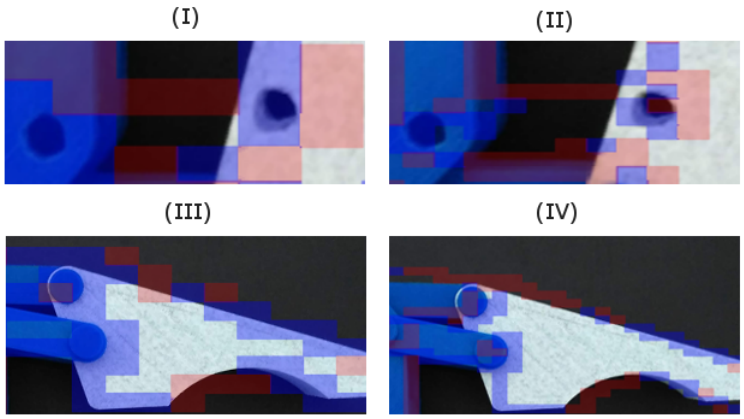


Figuras (a), (b) e (c) representam a tentativa de casamento com microquadrantes distribuídos com 20 horizontais e 20 verticais $(20x20)$. Analogamente, nas figuras (d), (e) e (f), distribuídos $(40x40)$. Microquadrantes positivos sombreados em azul e os negativos em vermelho.

Nas ilustrações (b) e (e) da Figura 10. onde uma peça central está ausente, nota-se uma maior precisão do resultado negativo para o componente ausente quando os microquadrantes são menores (e). A Figura 11, nas ilustrações (I) e (II), demonstra em detalhes as regiões onde ocorre a falta do componente em cada um dos casos. São observados ainda os destaques avermelhados na região superior das imagens (a) e (d) da

Figura 10, estes são resultado do deslocamento das peças que formam a pinça naquela região. O detalhe do deslocamento pode ser observado nas imagens (III) e (IV) da Figura 11.

Figura 11 – Detalhes do Casamento



Figuras (I) e (II) apresentam os detalhes da ausência do componente. Já as imagens (III) e (IV) detalham um deslocamento das peças que sustentam as garras

Para realizar a observação do efeito de filtragem dupla da imagem de cena, a Figura 12 representa em verde os microquadrantes que são identificados após a utilização de uma segunda filtragem. Ao todo, após a segunda filtragem, foram identificados 18 novos microquadrantes positivos, existindo anteriormente 93 (destacados em azul).

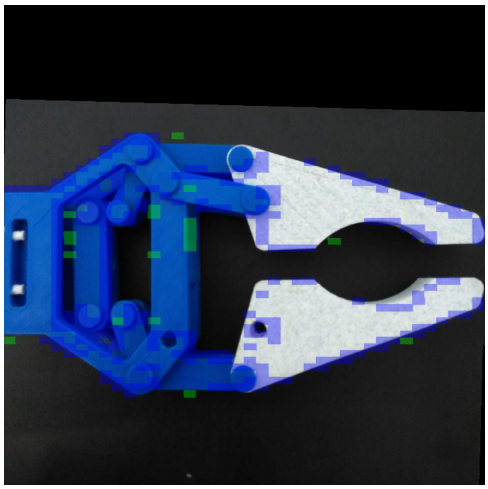
4.3 Resultados

Para a demonstração dos resultados dos comparativos, estes foram organizados na Tabela 1. A tabela organiza os resultados entre comparativos realizados utilizando-se 20 e 40 microquadrantes por dimensão. São demonstrados também os valores em relação aos microquadrantes identificados na primeira e na segunda filtragem. Além disto, a tabela apresenta o total de microquadrantes identificados para o modelo completo.

O processo de filtragem dupla demonstra melhorias sutis em termos de porcentagem total, com pico de 4,1% em relação ao total de microquadrantes positivos encontrado no mesmo experimento. Espera-se uma melhora substancial neste valor quando as condições da cena sejam mais arbitrárias, como um fundo colorido, diferentes materiais e texturas e iluminação nos componentes do item.

As porcentagens de acerto do item como um todo decaem pouco em relação à quantidade total de microquadrantes analisados. A diferença máxima percentual entre

Figura 12 – Filtragem Dupla



Em azul os microquadrantes obtidos do casamento feito com o filtro Canny. Destacados em verde os microquadrantes cujo casamento só é obtido após a realização do segundo processo utilizando-se do filtro DifFocusColor6m

imagens é de aproximadamente 13% (imagens (a) e (c)) como apresentado na Tabela 1. Isso demonstra que o reconhecimento geral da peça (união de todos os componentes), pode ser robusto em relação a oclusão e aumenta a importância da dupla filtragem, uma vez que cada microquadrante mostra-se essencial no entendimento da cena.

Tabela 1 – Resultados

Imagem (tipo)*	**Microquadrantes** (Dimensão)	**Totais **** (não-nulos)	**Positivos** (1ª filtragem)	**Positivos** (2ª filtragem)	**Positivos** (total)
(a)	20x20	195	161 (82,56%)	5 (2,56%)	166 (85,1%)
(b)	20x20	195	147 (75,4%)	7 (3,6%)	154 (79%)
(c)	20x20	195	137 (70,2%)	4 (2%)	141 (72,2%)
(d)	40x40	481	308 (64,1%)	17 (3,5%)	325 (67,6%)
(e)	40x40	481	277 (57,6%)	20 (4,1%)	297 (61,7%)
(f)	40x40	481	272 (56,5%)	10 (2%)	282 (58,6%)

*As imagens destes experimentos encontram-se na Figura 10. ** Refere-se à quantidade de microquadrantes não nulos formados a partir do modelo.

A utilização de quadrantes menores, embora aumente o custo computacional consideravelmente, uma vez que requisita uma maior quantidade de análises e possivelmente refiltragens, apresenta-se como uma viabilidade para sistemas críticos em que a perfeição dos itens, assim como sua organização, é buscada. É possível notar leves distorções do modelo, assim como desalinhamento de peças, como no demonstrado na Figura 11. Para considerar um modelo completo em casamento, definiu-se que um quantitativo mínimo de 70% de microquadrantes encontrados são necessários. Assim nenhuma das três imagens

relativas aos experimentos com dimensão $40x40$ ($(d), (e)$ e (f)) teria realizado casamento isso demonstra como o sistema lida com a precisão, se aumentada a quantidade de microquadrantes para análise. Como os detalhes das imagens (II) e (IV) da Figura 11 demonstram regiões que são corretamente não encontradas.

Problemas que necessitem de maior ou menor acurácia devem manipular a quantidade de microquadrantes para maximizar o ganho na análise. O uso de uma maior quantidade de microquadrantes permite que pequenas discrepâncias de formato da peça sejam encontradas. Por outro lado, o método ainda faz-se viável num contexto onde é necessário encontrar os modelos com maior velocidade e menor análise dos detalhes da peça, pois, definindo-se uma menor quantidade de microquadrantes, é possível encontrar a peça mais rapidamente e com diferentes condições de cena, lidando até mesmo com problemas como oclusão.

5 Conclusão

Com o advento da indústria 4.0 e da modernização das linhas de produção, cada vez mais, a avaliação e acompanhamento dos processos de forma automatizada é necessária. Uma das competências fundamentais para acompanhamento de processos é a visão computacional. No âmbito da indústria, processos de montagem, armazenamento, e produção requerem uma grande precisão e o redimensionamento e redesenho do processo produtivo pode ter um custo exorbitantemente alto dentro desses padrões. Com isso, soluções precisas e que consigam se adequar ao contexto atual da indústria se fazem necessários.

Esta monografia apresentou um sistema de reconhecimento de arquivos de descrição de objetos a partir de uma única imagem RGB. Uma vez que esses arquivos de descrição não contêm informações como textura, faz-se providente a extração da maior quantidade de características possível da imagem da cena. Para resolver esses problemas, é apresentada uma nova forma de obtenção de características de imagens, assim como um sistema de reconhecimento que é capaz de ajustar o nível de precisão à necessidade de cada problema contemplando o contexto que cada objeto buscado pode estar inserido.

Através da manipulação dos filtros de entrada, assim como da definição do tamanho dos microquadrantes, é possível definir a precisão e encontrar erros ínfimos de montagem e formato das peças avistadas assim como componentes. Por utilizar apenas de uma imagem RGB, o sistema tem impacto mínimo na reorganização da linha industrial, fazendo com o que os custos atribuídos ao seu uso, sejam diminuídos.

Porém, cada ambiente tem organização específica como, por exemplo, sua própria iluminação e disposição. Assim como cada peça componente do modelo pode ter características menores a serem buscadas, mas que são cruciais ao funcionamento do sistema. Sendo a calibração a padronização e definição dos melhores valores de limiar e entrada e filtros para realizar o casamento.

A calibração mostrou-se uma das grandes dificuldades para os experimentos do método proposto. Inicialmente, durante a obtenção das imagens de cena, onde foram necessárias várias fotos até que a câmera fosse posicionada na mesma posição padronizada pelo arquivo CAD. Após isso, diversos valores de limiar, assim como combinações entre diferentes filtros para o primeiro e segundo processo de filtragem, foram utilizados. A necessidade de interferência manual para escolha dos filtros e limiares, assim como a dificuldade de organização da câmera para obtenção da imagens de cena, foram as maiores dificuldades encontradas para realização do estudo apresentado nesta monografia.

5.1 Trabalhos Futuros

Para que o método seja ainda mais eficaz e preciso, a utilização de tamanhos dinâmicos para os microquadrantes demonstra ser um grande incremento ao mecanismo. Basear-se num mapa de densidade de características do modelo que possibilite criar de forma proporcional maiores ou menores microquadrantes parece ser uma abordagem promissora para se diminuir a quantidade de avaliações assim como apontar ainda mais incongruências.

Como o único resultado, em relação ao casamento, apresentado pelo método proposto é a existência ou não de um casamento, estender o estudo para utilização de métodos de busca numa cena como o RANSAC faz-se bastante providente. Certamente, surgem novos desafios como, por exemplo, a projeção dos modelos de forma adequada a depender de cada casamento possível. Tratar problemas como este também fazem parte desse estudo futuro.

É necessário ainda, realizar o comparativo com outros estudos de casamento de arquivos de descrição tridimensionais. Comparar a precisão no casamento, em relação a quantidade de componentes corretamente encontrados da peça buscada, assim como a perfeição destes. Além disto, uma análise da velocidade em que eles são executados sob mesmas condições de obtenção dos arquivos de descrição e de imagens de cena. A partir dos resultados obtidos, descobrir também formas de engrandecimento mútuo entre as abordagens.

Referências

ALAHI, A.; ORTIZ, R.; VANDERGHEYNST, P. Freak: Fast retina keypoint. *IEEE Conference on Computer Vision and Pattern Recognition*, Ieee, New York, p. 8, 2012. CVPR 2012 Open Source Award Winner. Citado na página 18.

ALDOMA, A. et al. Tutorial: Point cloud library: Three-dimensional object recognition and 6 dof pose estimation. *IEEE Robotics Automation Magazine*, v. 19, n. 3, p. 80–91, Sept 2012. ISSN 1070-9932. Citado na página 10.

ARMSTRONG, M.; ZISSERMAN, A. Robust object tracking. In: *Asian Conference on Computer Vision*. [S.l.: s.n.], 1995. I, p. 58–61. Citado 3 vezes nas páginas 10, 18 e 19.

BAY, H.; TUYTELAARS, T.; GOOL, L. V. Surf: Speeded up robust features. In: *In ECCV*. [S.l.: s.n.], 2006. p. 404–417. Citado 4 vezes nas páginas 9, 12, 13 e 19.

BAYKARA, H. C. et al. Real-time detection, tracking and classification of multiple moving objects in uav videos. In: *2017 IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI)*. [S.l.: s.n.], 2017. p. 945–950. Citado 3 vezes nas páginas 16, 17 e 18.

CANNY, J. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8, n. 6, p. 679–698, Nov 1986. ISSN 0162-8828. Citado 3 vezes nas páginas 10, 20 e 21.

CHOI, C.; CHRISTENSEN, H. I. Real-time 3d model-based tracking using edge and keypoint features for robotic manipulation. In: *2010 IEEE International Conference on Robotics and Automation*. [S.l.: s.n.], 2010. p. 4048–4055. ISSN 1050-4729. Citado 6 vezes nas páginas 10, 12, 18, 19, 23 e 26.

COSTA, D. C.; MELLO, C. A. B.; SANTOS, T. J. d. Boundary detection based on chromatic color difference and morphological texture suppression. In: *2013 IEEE International Conference on Systems, Man, and Cybernetics*. [S.l.: s.n.], 2013. p. 4305–4310. ISSN 1062-922X. Citado 3 vezes nas páginas 20, 22 e 27.

CPLUSPLUS. *Overview*. 2018. Disponível em: <http://www.cplusplus.com/info/>. Citado na página 33.

DRUMMOND, T.; CIPOLLA, R. Real-time visual tracking of complex structures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 24, n. 7, p. 932–946, Jul 2002. ISSN 0162-8828. Citado na página 28.

FISCHLER, M. A.; BOLLES, R. C. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, ACM, New York, NY, USA, v. 24, n. 6, p. 381–395, jun. 1981. ISSN 0001-0782. Disponível em: <http://doi.acm.org/10.1145/358669.358692>. Citado 2 vezes nas páginas 18 e 19.

GOMEZ-DONOSO, F. et al. Lonchanet: A sliced-based cnn architecture for real-time 3d object recognition. In: *2017 International Joint Conference on Neural Networks (IJCNN)*. [S.l.: s.n.], 2017. p. 412–418. Citado 3 vezes nas páginas 10, 16 e 17.

GUO, Y. et al. 3d object recognition in cluttered scenes with local surface features: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 36, n. 11, p. 2270–2287, Nov 2014. ISSN 0162-8828. Citado na página 12.

HERMANN, M.; PENTEK, T.; OTTO, B. Design principles for industrie 4.0 scenarios. In: *2016 49th Hawaii International Conference on System Sciences (HICSS)*. [S.l.: s.n.], 2016. p. 3928–3937. ISSN 1530-1605. Citado na página 9.

LIPKIN, B. *Picture Processing and Psychopictorics*. Elsevier Science, 1970. ISBN 9780323146852. Disponível em: <https://books.google.com.br/books?id=vp-w_pC9JBAC>. Citado na página 23.

LOWE, D. G. Object recognition from local scale-invariant features. In: *Proceedings of the Seventh IEEE International Conference on Computer Vision*. [S.l.: s.n.], 1999. v. 2, p. 1150–1157 vol.2. Citado 2 vezes nas páginas 9 e 12.

LUCAS, B. D.; KANADE, T. An iterative image registration technique with an application to stereo vision. In: *Proceedings of the 7th International Joint Conference on Artificial Intelligence - Volume 2*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1981. (IJCAI'81), p. 674–679. Disponível em: <http://dl.acm.org/citation.cfm?id=1623264.1623280>. Citado na página 14.

MARFIL, R. et al. Real-time template-based tracking of non-rigid objects using bounded irregular pyramids. In: *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (IEEE Cat. No.04CH37566)*. [S.l.: s.n.], 2004. v. 1, p. 301–306 vol.1. Citado na página 15.

MARKS, T. K.; HERSHEY, J. R.; MOVELLAN, J. R. Tracking motion, deformation, and texture using conditionally gaussian processes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 32, n. 2, p. 348–363, Feb 2010. ISSN 0162-8828. Citado 2 vezes nas páginas 14 e 15.

MESQUITA, R. G. de. *Reconhecimento de Instâncias Guiado Por Algoritmos de Atenção Visual*. Tese (Doutorado) — UNIVERSIDADE FEDERAL DE PERNAMBUCO, Recife, 2 2017. Citado 2 vezes nas páginas 13 e 14.

MIAN, A. S.; BENNAMOUN, M.; OWENS, R. Three-dimensional model-based object recognition and segmentation in cluttered scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 28, n. 10, p. 1584–1601, Oct 2006. ISSN 0162-8828. Citado na página 10.

NEWMAN, W. M.; SPROULL, R. F. (Ed.). *Principles of Interactive Computer Graphics (2Nd Ed.)*. New York, NY, USA: McGraw-Hill, Inc., 1979. 124 p. ISBN 0-07-046338-7. Citado na página 30.

OPENCV. *Overview*. 2018. Disponível em: <https://opencv.org/>. Citado na página 34.

OPENGL. *Overview*. 2018. Disponível em: <https://www.opengl.org/>. Citado na página 33.

PEDRINI, H.; SCHWARTZ, W. *Análise de imagens digitais: princípios, algoritmos e aplicações*. THOMSON PIONEIRA, 2008. ISBN 9788522105953. Disponível em: <https://books.google.com.br/books?id=13KAPgAACAAJ>. Citado 3 vezes nas páginas 10, 23 e 24.

ROSTEN, E.; DRUMMOND, T. Machine learning for high-speed corner detection. In: LEONARDIS, A.; BISCHOF, H.; PINZ, A. (Ed.). *Computer Vision – ECCV 2006*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006. p. 430–443. ISBN 978-3-540-33833-8. Citado na página 17.

SENST, T. et al. Clustering motion for real-time optical flow based tracking. In: *2012 IEEE Ninth International Conference on Advanced Video and Signal-Based Surveillance*. [S.l.: s.n.], 2012. p. 410–415. Citado na página 14.

SOUZA, L. O. C. d. et al. Avalição de sistemas de irrigação por gotejamento, utilizados na cafeicultura. *Revista Brasileira de Engenharia Agricola e Ambiental*, scielo, v. 10, p. 541 – 548, 09 2006. ISSN 1415-4366. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1415-43662006000300002&nrm=iso>. Citado na página 10.

TOMASI, C.; KANADE, T. *Detection and Tracking of Point Features*. [S.l.], 1991. Citado na página 14.

VIOLA, P.; JONES, M. Rapid object detection using a boosted cascade of simple features. In: *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*. [S.l.: s.n.], 2001. v. 1, p. I–511–I–518 vol.1. ISSN 1063-6919. Citado 3 vezes nas páginas 10, 16 e 17.

WASU, S. *Robot Gripper*. 2018. Disponível em: <https://grabcad.com/library/robot-gripper-16>. Citado na página 33.