UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS DA COMPUTAÇÃO

Walber Rodrigues de Oliveira

Framework para Reconhecimento de Ferramental Industrial a Partir de Modelos Tridimensionais em Imagens Adquiridas de Câmera Monocular

Recife

2020

Walber Rodrigues de Oliveira

Framework para Reconhecimento de Ferramental Industrial a Partir de Modelos Tridimensionais em Imagens Adquiridas de Câmera Monocular

Trabalho apresentado ao Programa de Pós-graduação em Ciência da Computação do Centro de Informática da Universidade Federal de Pernambuco como requisito parcial para obtenção do grau de Mestre em Ciência da Computação.

Área de Concentração: Processamento de sinais e reconhecimento de padrões

Orientador: Carlos Alexandre Barros de Mello

Recife

2020

Catalogação na fonte
Bibliotecária Monick Raquel Silvestre da S. Portes, CRB4-1217

O48f Oliveira, Walber Rodrigues de
Framework para reconhecimento de ferramental industrial a partir de modelos tridimensionais em imagens adquiridas de câmera monocular / Walber Rodrigues de Oliveira. – 2020.
97 f.: il., fig., tab.

Orientador: Carlos Alexandre Barros de Mello.
Dissertação (Mestrado) – Universidade Federal de Pernambuco. CIn, Ciência da Computação, Recife, 2020.
Inclui referências e apêndice.

1. Processamento de sinais. 2. Reconhecimento de padrões. I. Mello, Carlos Alexandre Barros de (orientador). II. Título.

006.4 CDD (23. ed.) UFPE - CCEN 2021 - 21

Walber Rodrigues de Oliveira

“Framework para Reconhecimento de Ferramental Industrial a Partir de Modelos Tridimensionais em Imagens Adquiridas de Câmera Monocular”

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Mestre em Ciência da Computação.

Aprovado em: 22/12/2020.

BANCA EXAMINADORA

_____

Profa. Dra. Judith Kelner
Centro de Informática/ UFPE

_____

Prof. Dr. Marcelo Walter
Instituto de Informática / UFRGS

_____

Prof. Dr. Carlos Alexandre Barros de Mello
Centro de Informática / UFPE
(**Orientador**)

Dedico esta dissertação a toda minha família, amigos e professores que me deram o apoio necessário para chegar até aqui.

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 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 participantes do Sindicato, em especial a Deyvinho, João, David, Fernando e Rick, que entre uma reclamação sobre a árvore do prédio novo e alguma piada de continuidade do CUG me ajudaram também na produção da pesquisa. Agradeço aos amigos de sangue de barro do RPG & Pizza, aos meus amigos da Casa do Estudante Izaildo e Diogo. Agradeço a todos os meus amigos e colegas de CIn, em especial a Matheus e aos componentes do extinto Vanilo.

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, em especial a Bruno e Adriano que me apoiaram nos caminhos de elaboração da pesquisa e nos trâmites de aprovação e finalização do projeto.

RESUMO

A visão computacional tem como uma de suas áreas de atuação a busca de itens. O uso de tais técnicas possibilita o impulsionamento de diversos setores sociais e econômicos. A ferramentaria é o setor que produz os moldes para criação de estrutura básica de produtos; são criados moldes em forma de matrizes que podem vir a pesar até 20 toneladas e ocupar dezenas de metros quadrados. Cada matriz é formada de dezenas a centenas de componentes, muitos deles únicos. Na indústria automotiva estes moldes são concebidos em software de modelagem tridimensional e, através de um processo de manufatura aditiva, passam por diversas alterações até a construção do componente final. Após modelagem, a matriz é montada e utilizada em prensas hidráulicas que criam a estrutura metálica dos automóveis. Um passo fundamental consiste na verificação da montagem dos componentes antes de serem levados à prensa. A conferência é feita a olho nu e a única fonte de informação disponível é o modelo tridimensional. Falhas na estrutura podem causar danos estruturais e por a saúde dos colaboradores em risco. Com isto, esta dissertação apresenta um modelo de reconhecimento dos componentes de uma matriz de ferramentaria automotiva a partir do modelo tridimensional em imagem obtida a partir de câmera monocular. A solução proposta modela desde o padrão de aquisição de imagens até a geração de relatórios. Após a aquisição das imagens de cena, é inferida a posição global ideal da matriz utilizando geometria projetiva. Na próxima etapa, são renderizados os itens buscados numa imagem que se torna um modelo canônico da cena. Em paralelo a isto, a imagem de cena passa por um processo de extração de características, onde são utilizados algoritmos clássicos como o algoritmo de Canny, e soluções modernas baseadas em redes neurais de aprendizagem profunda (o HED - Holystically-nested Edge Network). A partir dos relatórios gerados é possível indicar problemas estruturais nos componentes e itens indesejados em cena. Além de demonstrar o percentual de casamento dos itens, a partir de uma análise comparativa entre a imagem de bordas da cena e o modelo canônico. A solução foi testada em uma indústria automotiva e obteve como resultados de casamento valores médios de 88% de acurácia e 91% de precisão além de conseguir identificar problemas estruturais nas ferramentas. A solução demonstra ser adequada para aumentar a eficiência do processo de conferência, trazendo melhorias de segurança e diminuindo os custos do processo.

Palavras-chaves: Reconhecimento de modelos tridimensionais especulares. Ferramentaria automotiva. Casamento de *template*. Visão computacional.

ABSTRACT

One of the areas of computer vision is object detection. The use of such algorithms allows improvements at diverse social and economic sectors. In the process of creating new automobiles, the basic structures of the products are manufactured under several processes. The tool construction sector produces such items that are created in the form of dies. Such dies can weigh up to 20 tons and occupy tens of square meters. Each die is made up of dozens of components, many of them unique. These molds are designed in a three-dimensional modeling software and, through an additive manufacturing process, undergo several changes until the construction of the final component. Once all the components are built, the die is assembled and goes to the stamping process. A fundamental step in this process is to verify the die components assembly before they are taken to the press. Usually, the conference is held with the naked eye and the only source of information available is the three-dimensional model. Failures in the structure can cause damage and put the health of employees at risk. Therefore, this study presents a algorithm to recognize the components of an automotive tooling die from the three-dimensional model in images obtained from a monocular camera. The proposed solution begin with the image acquisition and it has as last steps the generation of reports. After the acquisition of the scene image, the ideal global position of the matrix is inferred using projective geometry. In the following step, the searched items are rendered in an image that becomes a canonical model of the scene. simultaneously, the scene image goes through a process of feature extraction. Classical algorithms as Canny edge detection algorithm just as modern approaches based on deep learning (as HED - Holystically-nested Edge Network) were used. With these inputs, two reports are generated: the first one is capable of indicating scene problems, such as cracks and unwanted items. The second report demonstrates the matching percentage of the items, from a comparative analysis between the scene edge image and the canonical model. The solution was tested in an automotive industry and obtained as a result of matching the average values of 88% accuracy and 91% of precision in addition to being able to identify problems appear in the die. The algorithm shows to be suitable for increasing the efficiency of the conferencing process bringing improvements for security and decreasing the costs of the tooling process.

Keywords: Recognition of three-dimensional specular models. Automotive Tooling. Template matching. Computer vision.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

6DoF	*Six Degrees of Freedom*
BRISK	*Binary Robust invariant scalable keypoints*
BSDS 500	*Berkeley Segmentation Data Set and Benchmarks 500*
BSP	*Binary Space Tree*
CAD	*Computer-Aided Design*
DLT	*Direct Linear Transform*
DoG	*Difference of Gaussians*
EPnP	*Efficient Perspective-n-Point*
FAST	*Features from Accelerated Segment Test*
FREAK	*Fast Retina Keypoint*
GFT	*Good Features to Track*
GPU	*Graphic Process Unit*
HED	*Hollisticaly-nested Edge Detection*
IA	Inteligência Artificial
ILSVRC	*ImageNet Large-Scale Visual Recognition Challenge*
KLT	*Kanade Lucas Tomasi Feature Tracker*
MST	*Minimum Spanning Tree*
PnP	*Perspective-n-Point*
RANSAC	*Random Sample Concensus*
ReLU	*Rectified Linear Unit*
RPnP	*Robust Perspective-n-Point*
SIFT	*Scale Invariant Features*
STL	*Standard Transformation Language*
SURF	*Speeded-UP Robust Features*
SVD	*Singular Value Decomposition*

SVM	*Supported Vector Machine*
ToF	*Time of Flight*
TRL	*Technology readiness level*
UAV	*Unmaned Aerial Vehicle*
YOLO	*You Only Look at Once*

SUMÁRIO

1 INTRODUÇÃO

Com o avanço da tecnologia e o aumento da demanda industrial do mundo globalizado, indústrias de diversos setores buscam diferenciais competitivos sensíveis em todo seu modo de produção. Um exemplo disso, está presente na cadeia de produção automotiva, que precisa remodelar constantemente a aparência dos seus automóveis, buscando que estes tenham sempre um design renovado. Devido à grande competitividade do setor, diversos modelos precisam ser produzidos em um tempo cada vez mais curto. Isto faz-se fundamental para que seja possível alcançar o almejado diferencial de mercado atraindo a atenção da maior parte dos consumidores.

Para alcançar essa característica tão fundamental nos novos modelos, é necessário realizar constantes remodelagens da estrutura dos produtos. Para produtos com estrutura metálica, as estruturas são criadas a partir de chapas de aço que são estampadas, utilizando-se prensas hidráulicas. O molde presente na superfície da prensa, também chamado de matriz, é composto por uma série de ferramentas que são posicionadas de forma modular e estratégica, gerando a cada prensagem uma série de partes da estrutura do produto. Com isto, uma vez que as estruturas dos produtos estão em constante modificação, as matrizes precisam passar por constantes modificações e remodelagens de cada um dos componentes.

As remodelagens tornam cada vez mais difícil a utilização de moldes atuais já consolidados, gerando uma constante re-manufatura e re-invenção das principais ferramentas de estampagem destes componentes das matrizes. O principal processo de ajuste destes moldes é um processo similar ao de manufatura aditiva. É produzido um modelo tridimensional da matriz composta por diversas ferramentas. Uma vez criado o modelo tridimensional de cada ferramenta, estas são forjadas em metal. As ferramentas, que podem ter dimensões de ordem de grandeza das dezenas a centena de centímetros, são encaixadas na matriz transformando-se em um dos seus componentes. Ao final deste processo, é feita uma verificação de corretude, onde as peças devem seguir o guia formado pelo modelo tridimensional. Essa verificação é feita a olho nu, utilizando-se de um guia. Além disto, o processo de verificação é lento, custoso e a sua eficiência recai sobre a experiência e atenção do operador.

Uma vez conferida, a matriz é submetida à prensagem e uma peça piloto das estruturas dos produtos é criada. A partir do piloto, são analisados desgaste das chapas de metal, robustez das estruturas, suavidade das superfícies e gastos de matéria prima, entre outros problemas

e características associadas a cada produto. Para cada defeito presente na estrutura piloto gerada, modificações nas ferramentas são necessárias, gerando assim modificações pontuais em cada ferramenta. Os ajustes podem ser desde o desgaste de ferramentas, diminuição ou aumento da superfície de contato, rearranjo da matriz, etc. Cada ajuste é então anotado e as informações são repassadas para o engenheiro responsável pela criação do modelo tridimensional. Em posse destas informações, o engenheiro re-modela o modelo tridimensional com as correções e as repassa para a funilaria. A funilaria realiza a correção de cada ferramenta, seguindo o novo modelo. Uma vez corrigidas, as ferramentas voltam a ser montadas, compondo uma nova matriz e re-iniciando o processo.

Este processo de recriação do modelo e matriz, torna cada matriz única, além de acontecer diversas vezes até que uma matriz perfeita seja criada e o desenvolvimento em larga escala do novo produto possa ser iniciado. Falhas no processo de criação da matriz podem acontecer devido à um processo de verificação falho, que permita passar peças em locais indevidos, itens ausentes ou componentes esquecidos em cima da matriz. Por se tratarem de diversas ferramentas complexas, de extrema precisão, muitas delas únicas, mesmo os operadores mais experientes podem vir a fazer uma verificação falha. As falhas mais graves podem gerar danos incalculáveis tanto em termos financeiros como podem por em risco a saúde de operadores.

A Figura 1 apresenta a capa da revista Ferramental. Na imagem observa-se um exemplo de matriz de ferramentaria com seus componentes inclusos. É possível verificar que uma ferramenta base serve de suporte para os demais moldes. A Figura 2 exemplifica um modelo tridimensional componente de tais matrizes. Os componentes da imagem estão agrupados e a matriz apresenta as duas partes, superior e inferior que são ligadas aos seus respectivos locais numa prensa hidráulica. As matrizes ferramentais automobilísticas podem ocupar mais de $10m^2$ e serem formadas por centenas de componentes únicos.

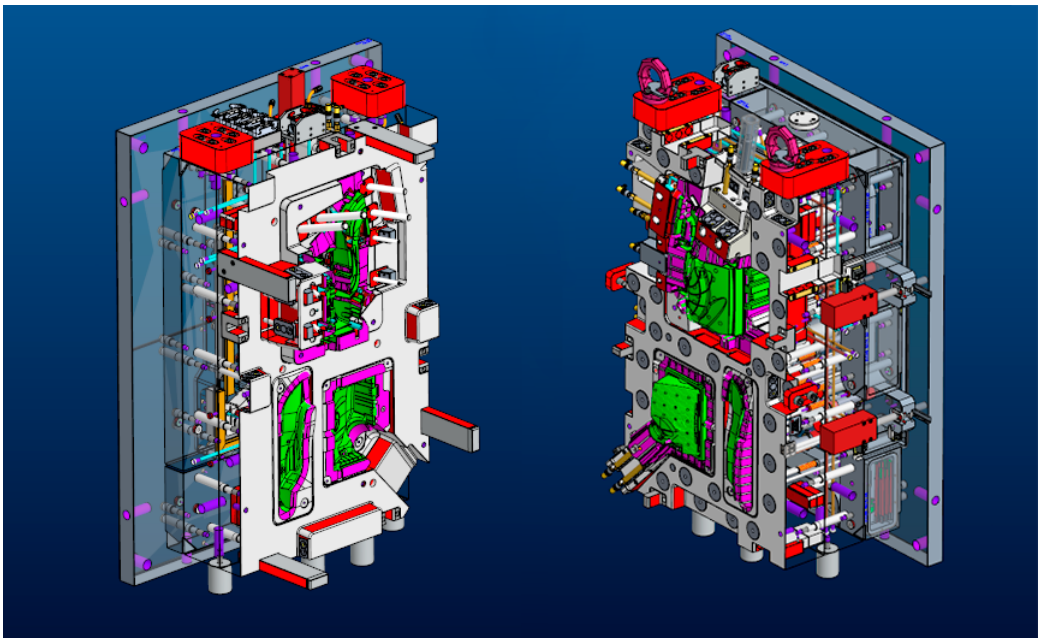
Sistemas de visão computacional já fazem parte de diversos processos da indústria. De modelos de verificação de qualidade à aplicações voltadas para segurança e saúde do trabalhador, é possível encontrar soluções que aumentam a eficiência industrial. O aumento do poder de processamento, viabilizou a aplicação de algoritmos cada vez mais rebuscados em ambientes operacionais plenos, aumentando ainda mais a robustez e eficiência de tais sistemas. Algoritmos de busca são capazes de detectar, reconhecer, rastrear objetos complexos com certo nível de eficiência em diferentes configurações e ambientes. Entretanto, estudos de busca voltados para ambientes fabris acabam por sofrer de diversos fatores como: desordem do plano de fundo, causada pelo constante trânsito de máquinas e pessoas; alto grau de especularidade

Figura 1 – Matriz de Ferramentaria presente na capa da Revista Ferramental. No detalhe, um molde de injeção de carcaça de farol.

Adaptada de (DIHLMANN, 2008).

dos objetos metálicos, que acabam por inibir o uso de algumas tecnologias e técnicas; além da necessidade de alta precisão, por ser um processo que é repetido com frequência e requer perfeição entre os componentes. Portanto, estudos eficientes em tais ambientes se faz importante para que possam ser evitadas falhas evitando por em risco à linha de produção e aos trabalhadores.

Figura 2 – Exemplo de arquivo tridimensional CAD de matriz de ferramentaria com componentes agrupados.



Adaptada de (FERRAMENTARIA, 2020).

1.1 OBJETIVOS

Como o processo fabril está em constante produção, qualquer solução proposta deve ter um menor impacto possível financeiro e procedural sobre a linha para obter uma maior chance de sucesso. Com isto, faz-se necessária a criação de uma solução capaz de realizar a verificação dos componentes de matrizes de forma automática utilizando-se dos processos já existentes dentro da própria indústria. Esta pesquisa apresenta uma solução que faz uso de técnicas de visão computacional para verificação da corretude da montagem das matrizes de prensas industriais. Os fatores de intervenção na linha de produção devem ser minimizados então o uso de equipamentos de baixo custo para aquisição de imagens e a exploração dos insumos já comuns ao cotidiano das indústrias deve ser preservado, como por exemplo, o uso dos modelos *Computer-Aided Design* (CAD) tridimensionais para guiar o processo de análise.

A solução apresentada deve culminar no impulsionamento da indústria, gerando modernização e automação dos seus processos, diminuindo riscos aos colaboradores, gastos de produção além de falhas do processo. Isto posto, indústrias que possuam acesso a soluções desta natureza podem lançar novos produtos com menor espaço de tempo no mercado, aumentando sua eficiência.

1.1.1 Ferramentaria automotiva

Uma das indústrias de transformação mais importantes, tanto no âmbito nacional quanto internacional, a indústria automotiva possui em sua fase de desenvolvimento de novos automóveis. Uma das etapas primordiais, pode durar anos desde a concepção dos modelos tridimensionais, até a prensagem do primeiro componente completo. Todas as etapas de construção das matrizes ferramentais estão dispostos na lista a seguir:

- Modelagem: nesta etapa, são criados moldes tridimensionais da matriz ferramental, onde cada componente é concebido, assim como é feito o arranjo ideal entre os componentes;
- Forja: a partir do modelo tridimensional, cada componente é forjado de forma independente. As peças que possuem menor contato entre si, podem ser construídas com moldes de barro/isopor e ferro fundido, enquanto peças que entram em contato com a chapa metálica tendem a ser produzidas em aço, através de um processo mais rebuscado.
- Montagem: uma vez que todas as peças foram forjadas, elas são agrupadas de acordo com o arranjo idealizado no modelo tridimensional. Esta etapa é realizada por um funcionário ferramenteiro de forma manual;
- Conferência: após realizar a montagem da matriz, é conferida a olho nu para verificar se todos os componentes se encontram nos locais corretos de acordo com o arranjo tridimensional;
- Prensagem: após conferidos, as matrizes são içadas e levadas para uma prensa hidráulica, onde uma chapa de metal é colocada entre as partes da matriz e o processo de prensagem ocorre;
- Analise de Resultado: a chapa de metal é submetida a um processo de conferência, onde é analisada a perfeição do processo, se houveram rachaduras, se foram criados chanfros incorretos, se o metal desgastou onde não deveria. Caso sejam identificadas imperfeições, o processo reinicia, seguindo novamente para a fase de modelagem tridimensional.

Sendo assim, o processo de construção dos moldes acaba sendo interativo, com intervenções humanas em todo o processo e iterativo, pois ocorre até que a chapa final de metal saia com as especificações corretas. Este processo pode se repetir de forma indefinida, e sempre que

os objetos são remodelados um novo modelo tridimensional único é criado para este objeto. Uma parte crucial do processo de construção das matrizes é a conferência. Por ser feita a olho nu, podem passar desapercebidos erros na montagem. Caso um erro ocorra, a segurança operacional é comprometida, uma vez que as peças fora de posição ao serem prensadas podem danificar a estrutura física da fábrica ou ainda por a vida dos colaboradores em risco.

O processo de conferência automatizado, diminui o risco de uma conferência errônea, entretanto uma intervenção dentro da fábrica requer um processo com baixo tempo e que não tenha grande impacto na linha de produção, evitando gastos. Dessa forma, este estudo tem como objetivo realizar a apresentação de um *framework* para o reconhecimento de uma matriz de ferramentaria automobilística composta de objetos tridimensionais complexos em imagens RGB obtidas a partir de câmeras monoculares. As imagens precisam ser obtidas em ambientes operacionais industriais, ou em simulações coerentes de tais características.

1.2 ESTRUTURA DO DOCUMENTO

Esta dissertação apresenta seu conteúdo disposto em seis capítulos. O Capítulo 2 inicia apresentando trabalhos de busca de objetos tridimensionais em imagens, utilizando de diferentes técnicas segundo o tipo de busca. Além disso, o capítulo elucida de forma breve estudos de detecção de características de imagens, modelos de estimativa do posicionamento do olho da câmera e um estudo de calibração.

O Capítulo 3 apresenta o modelo de reconhecimento de objetos tridimensionais complexos especulares a partir do modelo *Computer-Aided Design* (CAD). O capítulo descreve desde as preparações necessárias na cena e no modelo tridimensional, até a geração do relatório de casamento. Em seguida, no Capítulo 4, são apresentados os experimentos realizados. São apresentadas as três fases de experimentação realizada, desde a entrega da prova de conceito do projeto até sua fase de piloto, passando pela prototipação.

O Capítulo 5 apresenta, de forma condensada, os resultados colhidos durante as três fases de entrega. O capítulo discute ainda, as dificuldades e circunstâncias que levaram às devidas tomadas de decisão durante a execução do projeto.

O último capítulo apresenta as conclusões do projeto. São mencionados ainda, os trabalhos futuros incrementais visualizados a partir do estudo desenvolvido.

2 TRABALHOS RELACIONADOS

Um desafio comum dentro da computação, o reconhecimento de objetos é uma das muitas competências relacionadas com a visão computacional. Entretanto, faz-se relevante explicitar os diferentes conceitos que envolvem a busca por objetos. Detecção, rastreamento, reconhecimento e classificação, embora possam ter empregos similares para diversos contextos, apresentam definições diferentes para o problema de busca em visão computacional. Esses conceitos, nesta dissertação, seguem as definições dispostas aqui. Os algoritmos de detecção de objetos consistem dos estudos voltados para indicar se um determinado tipo de peça se encontra ou não em um local delimitado. Já os algoritmos voltados para o rastreamento de um determinado objeto focam em detectar a movimentação de um objeto e, possivelmente, seu posicionamento com 6 graus de liberdade em uma cena, ao longo do tempo. Estudos voltados para realizar classificação são responsáveis por afirmar se um determinado item pertence a uma certa classe de itens, enquanto que os modelos de reconhecimento apresentam foco em discriminar diferentes entidades de uma mesma classe.

Observa-se que diferentes algoritmos e estudos acabam por fazer uma amálgama desses conceitos, como é o caso do algoritmo apresentado em (REDMON; FARHADI, 2018). O estudo apresenta uma rede neural capaz de realizar a detecção e a classificação de diversos objetos em uma cena. Além disto, por apresentar um tempo de processamento baixo, é capaz de realizar o rastreamento num vídeo com algumas dezenas de quadros por segundo. Segundo apresentado em (Lepetit; Fua, 2005), os métodos que se utilizam apenas de características naturais da cena, tais quais arestas, em casamento com modelos tridimensionais acabam por acumular muitos erros para realização do processo de rastreio. Ademais, o estudo apresentado neste documento tem foco no reconhecimento de objetos complexos sem textura em imagens com plano de fundo complexo.

Ao longo deste capítulo será possível visualizar a revisão do estado da arte de métodos de busca de objetos tridimensionais. Além disto, serão explanadas as principais técnicas de extração de características de imagens utilizadas no estudo, assim como uma revisão de métodos para descoberta de projetividades responsáveis pelo posicionamento de câmera.

2.1 BUSCA POR OBJETOS TRIDIMENSIONAIS

Visando estabelecer melhor o problema de busca, este capítulo apresenta diferentes estudos cujas finalidades acabam por mesclar as competências de detecção, rastreamento, reconhecimento e classificação. O estudo sugerido por (Lepetit; Fua, 2005) elenca os métodos de busca de objetos tridimensionais em dois tipos. O primeiro agrupa os métodos baseados no casamento de arestas e o segundo classifica os métodos que se baseiam nas informações providas pelos pixels dentro das imagens dos objetos. Estes por sua vez, se dividem em fluxo óptico, casamento de *template* e no casamento de *keypoints*. Outros autores determinam ainda que a classificação dos métodos baseados em aresta é um subtópico dos métodos de casamento de *templates*. Há autores que defendem que o método de casamento de *templates* como sendo uma especificidade do método de fluxo óptico (Marks; Hershey; Movellan, 2010).

2.1.1 Métodos baseados em Arestas

Dentre as principais características de tais métodos, destaca-se o fato de terem baixo custo computacional combinado a sua facilidade de implementação. Uma outra grande vantagem destes métodos é que eles tendem a ser robustos a variações de iluminação, mesmo em equipamentos especulares, o que não é necessariamente verdade em métodos que consideram a porção interna dos objetos para casamento. Tais métodos podem ser classificados entre duas vertentes: os métodos que realizam a extração previamente das arestas e contornos da imagem da cena e os métodos que buscam a pose sem tal processamento (Lepetit; Fua, 2005).

Um dos principais métodos para busca, utilizando arestas, é o RAPiD apresentado em (HARRIS, 1993). Este estudo já passou por diversas melhorias (ARMSTRONG; ZISSERMAN, 1995) (DRUMMOND; CIPOLLA, 2002) (MIAN; BENNAMOUN; OWENS, 2006a) (CHOI; CHRISTENSEN, 2010). O RAPiD teve grande relevância por ter sido pioneiro ao apresentar o rastreamento em tempo real segundo apresentado em (Lepetit; Fua, 2005). Seu sistema de busca se baseia no uso de pontos de controle retirados do *wireframe* do objeto buscado que são comumente encontrados em imagens de bordas. Nesta seleção, são utilizados pontos de controle escolhidos a partir das arestas do modelo tridimensional. Para realizar a busca do objeto, é feita uma inicialização dos pontos de controle, que são projetados sobre o local que se espera encontrar o item buscado e então são casados a contornos da imagem alvo.

Após a inicialização, o rastreamento é realizado visando a predição dos pontos casados

entre os quadros da filmagem. Para cada quadro, é prevista a posição inicial e são calculadas a distância de cada ponto de controle até o contorno que havia sido casado. Esta busca é linear e realizada apenas nas direções perpendiculares à direção das arestas que geraram o ponto de controle. Calculada a distância entre os pontos de controle e as arestas mais próximas encontradas como casamento, é então calculada a nova posição do modelo. O estudo original apresentado em (HARRIS, 1993) considerava que apenas movimentações leves poderiam ser realizadas no posicionamento do objeto buscado e atualizava a posição apenas de forma linear. Para realizar a descoberta dos parâmetros que minimizam o vetor de erros é aplicado o método dos quadrados mínimos.

Diversos estudos exploraram esta ideia apresentada pelo RAPiD. Dentre eles, inicialmente, os trabalhos buscaram melhorar pontos em aberto pelo modelo, como em (ARMSTRONG; ZISSERMAN, 1995). O estudo apresenta um modelo de busca similar, entretanto, ao invés de considerar todos os erros de forma independente, ele não realiza a busca dos pontos de controle nas arestas com erro acima de um determinado limiar. Este tratamento evita que o método tenha resultados muito distantes do esperado, evitando que o rastreio do objeto seja perdido e o erro seja acumulado entre quadros. Outra melhoria apontada pelos autores, é a utilização prévia do *Random Sample Concensus* (RANSAC) para eliminar valores fora da curva (*outliers*). Uma vez detectados os erros com maior confiança o processo de rastreio segue para o uso do método dos quadrados mínimos tal qual o apresentado por (HARRIS, 1993).

Uma evolução do modelo de busca, utilizando arestas, foi apresentado em (DRUMMOND; CIPOLLA, 2002). O estudo apresenta um modelo de busca de objetos similar ao RAPiD com melhorias em diversos pontos. Inicialmente, apenas as arestas visíveis são tratadas. Para descobrir quais arestas são visíveis, o modelo apresentado utiliza uma renderização do modelo para descobrir quais serão as características que estarão visíveis em cena. O estudo é focado na busca de objetos complexos, semi-articulados. A cena é capturada por câmeras posicionadas em servomecanismos e são apresentados resultados comparativos, utilizando múltiplas câmeras. Para realizar o rastreio em si, usa-se Álgebra de Lie para descoberta dos seis graus de liberdade do item buscado *Six Degrees of Freedom* (6DoF). Ou seja, o movimento tridimensional é decomposto e otimizado em seis geradores próprios do espaço de Lie.

Seguindo o modelo proposto em (DRUMMOND; CIPOLLA, 2002), outros métodos exploraram a busca utilizando a álgebra de Lie. Um exemplo destes é o apresentado em (CHOI; CHRISTENSEN, 2010). Neste modelo, é apresentado um estudo híbrido entre métodos baseados em *keypoints* e métodos baseados em arestas. Este estudo visa minimizar os problemas relaciona-

dos à perda do rastreio da peça, por oclusões completas ou uma movimentação muito rápida do objeto entre cenas. Nesses casos, foi utilizado o algoritmo *Speeded-UP Robust Features* (SURF) (BAY et al., 2008) para recuperação da posição inicial.

Outros diferenciais apresentados em (CHOI; CHRISTENSEN, 2010) estão na utilização de um pré-processamento do modelo para escolha das arestas relevantes para busca. Para descoberta das arestas é utilizada uma técnica de manter apenas as arestas formadas a partir da interseção de planos cuja angulação entre eles seja superior a 30º. Este tipo de processamento permite a criação de um modelo que mantém apenas arestas consideradas fortes. Estas são as arestas que os autores atribuem a maior chance de se apresentarem na cena. O tratamento de oclusão aplicado utiliza uma *Binary Space Tree* (BSP).

A metodologia híbrida recorre ao método baseado em *keypoints* sempre que um objeto é considerado se movimentar mais de 10 cm. Isto minimizou os problemas relacionados ao acúmulo de ruído entre quadros, comum entre métodos de rastreio utilizando arestas. Com isto, ele apresentou bons resultados em manter o rastreio do posicionamento global dos modelos buscados, mesmo que tenha apresentado uma quantidade de cálculos um pouco superior ao da utilização de marcadores fiduciais considerados o *ground truth*.

Para a utilização dos métodos baseados em *keypoints* faz-se necessário também um pré-processamento para a aquisição dos pontos chave que serão utilizados para realizar a busca global do objeto. Segundo os autores, é praticamente impossível manter a representação de todas as formas possíveis que o objeto pode se apresentar em cena. Portanto, o método propõe a armazenagem de uma certa quantidade de *keypoints* obtidos a partir da mesma face do objeto. Essas partes são obtidas e associadas a posições correspondentes no arquivo tridimensional de forma similar à apresentada por (Vacchetti; Lepetit; Fua, 2004). O casamento, utilizando *keypoints*, é explorado por outros tipos de estudos, alguns deles são apresentados na subseção 2.1.2.

Boa parte dos métodos centrados na busca de objetos tridimensionais se dedica a buscar correspondências na imagem de cena entre o modelo projetado e as arestas da cena. Um problema comum enfrentado por tais métodos é a grande quantidade de desordem de fundo, causada pela existência de itens diferentes do que são buscados. Este tipo de informação pode causar o casamento do modelo com falsos negativos e inviabilizar diversas técnicas. Uma outra relação que pode ser considerada comum é que itens buscados por modelos tridimensionais, tendem a ser construídos em coloração monocromática (SEO et al., 2014).

O algoritmo proposto em (SEO et al., 2014) apresenta um modelo de busca que utiliza da

informação desorganizada do plano de fundo para aumentar a eficiência do método. Ou seja, confiando que o objeto buscado é majoritariamente monocromático, enquanto no plano de fundo constam diversas texturas e outras informações. Para isto, o autor propõe a criação de uma estrutura de busca que realiza a extração das porções de pixels próximos às arestas projetadas. A estrutura condensa as informações tanto da peça buscada, quanto do fundo da cena. Com isto, são determinados os pontos que são as arestas, baseando-se na alteração mais brusca de textura considerando que os objetos construídos apresentam textura similar ao longo de sua extensão. Os resultados alcançados aumentam a eficiência dos métodos de busca, entretanto recai sobre a necessidade de ter um plano de fundo em grande desordem e que todo o objeto buscado tenha textura similar.

2.1.2 Métodos baseados em *keypoints*

Os métodos baseados em *keypoints* buscam realizar o casamento de características invariantes em diferentes quadros dos objetos. Idealmente, as informações coletadas devem ser invariáveis a modificações do objeto como escala e angulação. De forma geral, os métodos são baseados em três etapas: obtenção de *keypoints*, descrição das proximidades e, por fim, o casamento dos descritores.

As três etapas da construção destes métodos podem ser modificadas e otimizadas. Diversos estudos apresentam diferentes formas de realizar o casamento (LOWE, 2004), (BAY et al., 2008), (ALAHI; ORTIZ; VANDERGHEYNST, 2012), (CALONDER et al., 2010), (LEUTENEGGER; CHLI; SIEGWART, 2011), (Rublee et al., 2011), (MESQUITA, 2017); estes métodos buscam desde uma simplificação da escolha dos candidatos a *keypoints*, até os métodos que buscam a otimização do processo de casamento em si.

Um dos principais métodos é o *Scale Invariant Features* (SIFT) apresentado em (LOWE, 2004) que descreve um modelo em quatro etapas até a criação dos descritores. A primeira etapa é referente à construção do espaço-escala; a segunda etapa concerne à localização de *keypoints*; a terceira, à atribuição de orientação; e, por fim, uma quarta etapa de criação de um descritor da região de vizinhança. A primeira etapa do algoritmo descreve a criação do espaço-escala. Esta estrutura serve para descobrir características que não sejam sensíveis a redimensionamentos. Para a obtenção de tais características, a imagem passa por diversos filtros gaussianos e redimensionamentos que reduzem sua dimensão pela metade. Em (MESQUITA, 2017), são estimados o uso de três ou quatro octavos para a criação da estrutura. Cada

octavo, por sua vez, contém 4 camadas. O espaço-escala pode ser representado então, por uma pirâmide com a imagem de entrada contendo diferentes níveis de granularidade e escalas. Quanto mais alto o nível hierárquico na pirâmide, diminuem-se a quantidade de detalhes observados na imagem.

Para localizar pontos estáveis no espaço-escala, são calculadas a Diferença das Gaussianas *Difference of Gaussians* (DoG), utilizando a diminuição entre duas camadas adjacentes do espaço escala. Após isto, são calculados pontos de mínimo e máximo em cada imagem e então são realizadas comparações entre as vizinhanças dos pontos destacados, comparando-os por todo espaço-escala. Apenas os pontos que permaneçam de máximo em todo o espaço escala são mantidos; caso contrário, são descartados após a primeira comparação negativa. Os pontos sobressalentes passam ainda por mais comparativos, a fim de manter apenas os mais relevantes. São aplicados limiares para os valores resultantes da DoG e é analisada a curvatura dos pontos para verificar se são adequadas. Apenas os pontos sobressalentes são mantidos como *keypoints* adequados para etapa de descrição.

Para essa etapa, as regiões próximas aos *keypoints* são analisadas. os autores do *Scale Invariant Features* (SIFT) propõem a observação dos gradientes da região circular ao redor do *keypoint* com intuito de adquirir o maior pico e definir assim a orientação do *keypoint*. São recalculados os valores dos gradientes e uma função Gaussiana é novamente aplicada. Por fim, é construído um histograma de oito orientações com o valor de cada sub-região constando o gradiente de cada sub-região (MESQUITA, 2017).

Um outro método bastante difundido é o *Speeded-UP Robust Features* (SURF) (BAY et al., 2008). O estudo apresenta um resultado similar ao apresentado pelo SIFT, com alterações na construção do espaço escala e na definição da orientação dos *keypoints*. Para acelerar o processo de construção do espaço-escala são utilizados filtros *box* ao invés de redimensionar as imagens. Esta mudança permite uma aceleração do processo, com pequena perda de acurácia. Outra mudança apresentada é a utilização da *wavelet* de Haar na definição da orientação do ponto-chave (MESQUITA, 2017).

Com o avanço tecnológico, o uso de celulares *smartphones* se popularizou e a busca por algoritmos que sejam capazes de alcançar a mesma qualidade da busca apresentada pelo SIFT com menor custo computacional. O *Fast Retina Keypoint* (FREAK) (ALAHI; ORTIZ; VANDERGHEYNST, 2012) busca tal mérito, utilizando como inspiração a retina humana. Tal qual a retina humana, o descritor foca em porções centrais do descritor para terem maior peso de casamento.

Além disto, o processo de casamento do FREAK é feito de forma iterativa, onde, a cada nova iteração, o resultado do casamento é ajustado para melhores resultados. Para avaliar seu estudo, o autor compara o descritor proposto tanto com os descritores SIFT e do *Speeded-UP Robust Features* (SURF) quanto ao descritor binário *Binary Robust invariant scalable keypoints* (BRISK) (LEUTENEGGER; CHLI; SIEGWART, 2011). A avaliação demonstrou uma maior velocidade na realização do casamento.

Embora apresentem bons resultados para a busca de objetos, os métodos baseados em *keypoints* requisitam uma imagem texturizada de entrada dos itens a serem buscados. Este tipo de informação pode não estar disponível, ou limitar de alguma forma a busca dos itens, uma vez que os métodos tornam-se bastante sensíveis à especularidade dos objetos. Para os casos onde a informação geométrica do objeto está disponível, diferentes métodos baseados em *templates* foram propostos, conforme pode ser acompanhado na sub-seção a seguir.

2.1.3 Métodos baseados em *templates* e fluxo óptico

Alguns dos métodos já apresentados na subseção 2.1.1 podem ser considerados métodos baseados em *templates*. Entretanto, tais métodos podem apresentar diferentes métodos de busca. Inicialmente, os métodos mais simples baseados em *templates* consistiam numa janela deslizante que convolui sobre a imagem, buscando correspondentes de iluminação. Os métodos evoluíram, o uso de modelos não rígidos tridimensionais pôde ser viabilizado. Dois exemplos desta evolução estão presentes nos estudos apresentados por (Marfil et al., 2004) e (Marks; Hershey; Movellan, 2010).

O estudo apresentado em (Marfil et al., 2004) demonstra um sistema de busca que consiste de duas etapas. A primeira etapa que aborda o problema *bottom-up* que busca lidar com as modificações que o alvo pode sofrer devido às condições ambientais. A segunda etapa consiste numa abordagem *top-down* que busca lidar com a movimentação do objeto.

O processo total de busca é composto por quatro etapas, a etapa inicial consiste na construção de uma estrutura piramidal de segmentação da imagem de cena. Em um segundo momento, é feito o casamento do *template* buscado na cena. Com isto, caso o objeto tenha sido corretamente localizado, uma etapa de refinamento é aplicada e, por fim, o *template* é atualizado com a nova aparência do objeto. No modelo proposto em (Marfil et al., 2004), o *template* do objeto é representado pela estrutura piramidal de segmentação. Observa-se que este método é uma evolução dos métodos que exploravam uma janela deslizante com a

imagem buscada. Esta evolução do *template* permitiu um avanço no uso de *templates* não-rígidos, demonstrando como foram expandidas as fronteiras de busca destas técnicas.

Outros modelos de *template* foram propostos, buscando um aumento da eficiência no rastreio e busca de objetos não-rígidos. Em (Marks; Hershey; Movellan, 2010), é proposto o uso de um fluxo de Gaussianas que os autores chamam de *G-flow*. Nesse método, é realizado o rastreio de faces, considerando vértices de uma estrutura tridimensional ligados a informação de textura da região projetada do *template*. A partir dessa estrutura, o modelo atualiza-se de acordo com a movimentação das texturas mapeadas na estrutura. Para o acompanhamento dos pontos focais dentro do vídeo, são utilizados filtros de Kalman (WELCH; BISHOP, 1995). No geral, a técnica é híbrida entre os métodos de fluxo óptico e *templates*. Em (Marks; Hershey; Movellan, 2010), é afirmado ainda que os métodos clássicos de busca de fluxo óptico e de casamento de *templates* podem ser agrupados como casos particulares de um mesmo tipo de técnica de busca.

Os métodos de busca baseados em fluxo óptico utilizam da atualização constante de texturas no fluxo de quadros formado por um vídeo para realizar o rastreio e detecção de objetos. Comumente, estes métodos são utilizados para detectar e contar pessoas que se movem em multidões como citado em (Senst et al., 2012) ou um fluxo constante de objetos. Algoritmos já considerados clássicos, como é o exemplo do *Kanade Lucas Tomasi Feature Tracker* (KLT) (TOMASI; KANADE, 1991), conseguem realizar o rastreio de dezenas de milhares de pontos em tempo real. Para concluir tal tarefa, realizam o agrupamento de informações consideradas boas para rastreio *Good Features to Track* (GFT) e acompanham sua movimentação em video.

Em (Senst et al., 2012), é introduzido um modelo otimizado que armazena a trajetória possível dos pontos acompanhados em um grafo *Minimum Spanning Tree* (MST). São computadas afinidade temporal e espacial dos pontos e é coletada a distância entre eles. A partir da distância, a MST é criada e são retiradas as ligações cujo peso for menor. Uma vez retiradas, são criadas diversas florestas e esses agrupamentos são considerados como pontos pertencentes ao mesmo corpo em movimento. É construído então um mapa de densidade dos pontos e da malha e é calculada a trajetória de cada ponto de acordo com a estrutura.

Se por um lado existem métodos focados no uso das informações captadas a partir de um vídeo, outros métodos lançam mão de uma melhor coleta de informações da cena para realizar o rastreio dos itens. Um bom exemplo disso, são os métodos que utilizam modelos de busca em imagens de profundidade (RGB-D). Tais métodos podem realizar desde o casamento de modelos tridimensionais nas imagens de profundidade obtidas a partir de câmeras estéreo, até

o casamento de imagens RGB-D em nuvens de pontos. Alguns destes métodos são explorados adiante.

2.1.4 Métodos de casamento em imagens tridimensionais

O uso de imagens tridimensionais traz vantagens associadas ao fato das características morfológicas obtidas serem menos sensíveis a variações de iluminação, rotação e iluminação se comparadas a texturas. Associado a isto, a popularização de câmeras comerciais de baixo custo, tais como o Microsoft Kinect, tornou o estudo de tais métodos ainda mais pujante (Guo et al., 2014).

Uma característica clássica de algoritmos voltados para imagens tridimensionais, está no uso de *spin-images*, inicialmente, proposto em (Johnson; Hebert, 1999). Esse método foca no casamento de duas superfícies tridimensionais, buscando realizar seu alinhamento e verificação de corretude. Para realizar isso, o modelo propõe o uso de uma superfície construída a partir do modelo tridimensional composta por pontos com a orientação definida que são casados com os vértices de uma superfície similar construída a partir da imagem alvo. Segundo o algoritmo apresentado, ainda que formadas a partir de representações diferentes do mesmo objeto, elas têm comportamentos semelhantes, uma vez que a *spin-image* é baseada no formato do item.

Alguns estudos são extensões de algoritmos clássicos em imagens bidimensionais como é o caso do SIFT3D (Darom; Keller, 2012), *Speeded-UP Robust Features* (SURF)3D (KNOPP et al., 2010) e Harris 3D (SIPIRAN; BUSTOS, 2011). Estes estudos buscam expandir a compreensão dos métodos baseados em *keypoints* para o espaço tridimensional. Um outro método que se baseia nos conceitos de *keypoints* é o apresentado em (MIAN; BENNAMOUN; OWENS, 2006a). O autor propõe o uso de um tensor, uma estrutura tridimensional que mapeia a superfície do objeto, particiona-a em cubos para minimizar a dificuldade de casamento entre objeto e cena.

Embora apresentem algumas vantagens diante de métodos de reconhecimento de objetos, a obtenção de imagens tridimensionais pode apresentar problemas diante de cenas com alta especularidade. Diante das diversas variações possíveis para a cena onde se busca o objeto, métodos que possam generalizar a busca e manter boa acurácia são explorados. Algoritmos que usam aprendizagem de máquina tendem a ser bastante robustos a variações da cena. A seção a seguir explana, de forma breve, alguns destes métodos.

2.1.5 Métodos baseados em Aprendizagem de Máquina

Algoritmos de aprendizagem de máquina são uma sub-área da Inteligência Artificial (IA) e são vastamente utilizados na busca de objetos. De forma simplificada, para realizar o casamento entre objeto buscado e imagem de cena, tais algoritmos são treinados com uma série de imagens, contendo diversas possibilidades de como determinado objeto buscado pode aparecer na cena. A partir destes insumos, uma série de parâmetros são calibrados e com eles, é possível determinar se uma cena contém ou não determinado objeto.

Os primeiros métodos de detecção de objetos utilizavam máquinas de vetores de suporte *Supported Vector Machine* (SVM) como algoritmos a serem treinados. Um exemplo deste tipo de técnica pode ser observado em (Liebelt; Schmid; Schertler, 2008). Os autores apresentam um modelo de busca de objetos que além da detecção em cena, trata de informar o seu possível *bounding box*, contendo, assim, o informativo dos seus seis graus de liberdade. Visando realizar o casamento do objeto, o treinamento é realizado, utilizando uma renderização do objeto segundo diversos pontos de vista e uma simulação de diferentes condições de ambiente. A Figura 3 apresenta o modelo de treinamento apresentado em (Liebelt; Schmid; Schertler, 2008) , utilizando imagens renderizadas dos objetos buscados.

Figura 3 – Modelo de criação de base de dados sintética renderizada para treinamento de algoritmo de Inteligência Artificial (IA).

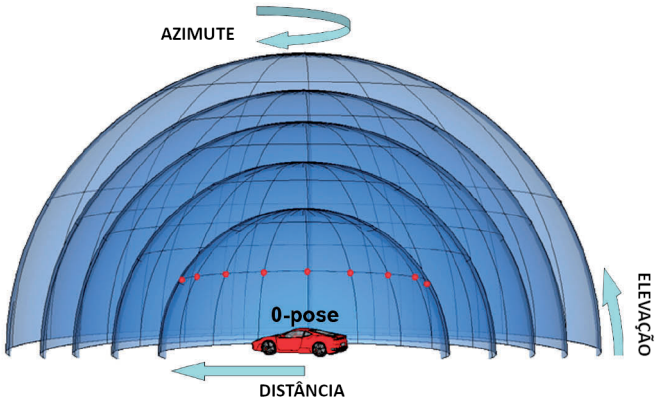


Imagem adaptada de (Liebelt; Schmid; Schertler, 2008).

Com a evolução dos métodos de aprendizagem de máquina, redes neurais convolucionais passaram a ser usadas para realizar o casamento de objetos em imagens. O uso de tais algoritmos permitiu ainda mais avanços no campo da detecção de objetos. Em (Gomez-Donoso et al., 2017), é apresentado um *framework* para detecção de objetos tridimensionais tal qual CAD em ambientes sintéticos. Para realizar este casamento, os autores propõem a criação de

um modelo de treinamento a partir de três secções do objeto tridimensional. O modelo de classificador completo, conta com três camadas GooLeNets com uma camada superior para aumentar a eficácia do método.

Os métodos ficaram ainda mais eficientes com a utilização de redes neurais convolucionais de aprendizagem profunda. O estudo apresentado em (SIMONYAN; ZISSERMAN, 2014) comenta que a viabilização de tais estudos se deu pelo aumento da capacidade computacional e principalmente das placas de vídeo *Graphic Process Unit* (GPU). Outro grande impulsionador dos modelos de busca, utilizando aprendizagem profunda, estão relacionados às competições. O *ImageNet Large-Scale Visual Recognition Challenge* (ILSVRC) é um dos maiores concursos e levou a grandes melhorias no uso das redes neurais convolucionais. A arquitetura de rede proposta VGG (SIMONYAN; ZISSERMAN, 2014) apresenta um algoritmo de treinamento que utiliza imagens coloridas de entrada com tamanho fixo de 224x224 pixels. Os *kernels* de convolução apresentam dimensões de 3x3 com intuito de descobrir uma possível orientação a partir do filtro, além de filtros de convolução 1x1 que são basicamente uma transformação linear da imagem. A arquitetura apresenta uma pilha de diferentes camadas com diversas profundidades e arquiteturas e, por fim, utiliza uma camada *softmax*. Todas as camadas são completamente interligadas. As camadas ocultas apresentam ainda uma retificação, utilizando o retificador *Rectified Linear Unit* (ReLU) e apenas uma delas apresenta normalização. Segundo os estudos apresentados, a camada de normalização não melhorou os resultados nos testes realizados, além de aumentar o consumo de memória e o tempo de reconhecimento. Como resultado do estudo, foi possível obter melhores resultados comparativos no desafio da ImageNet de 2014.

Um outro estudo bastante difundido do uso de redes neurais de aprendizagem profunda para busca de instâncias é o *You Only Look at Once* (YOLO) (REDMON; FARHADI, 2018). O estudo apresenta um modelo de detecção diferente das demais redes. Ao invés de analisar pequenas partes da imagem em busca dos objetos, a rede recebe a imagem completa para realizar a busca. A imagem é mapeada, utilizando diferentes *bounding boxes* e é atribuído um valor de erros para cada um deles. Baseado na quantidade de erro de cada *bounding box* e na recorrência de casamentos sobre a mesma região, surge um indicativo sobre a presença do objeto. A rede é treinada, utilizando a Darknet, e têm como entrada as imagens e os respectivos *bounding boxes* dos objetos a serem classificados. A extração de características é feita com uma rede neural convolucional de 53 camadas e, durante o processo, são aplicados diferentes processos de *data augmentation*.

Além de métodos puramente focados em construir classificadores utilizando redes neurais,

existem modelos voltados para aplicações reais que acabam por utilizar métodos híbridos para aumentar a robustez das aplicações. Um exemplo de aplicação híbrida é o apresentado em (BAYKARA et al., 2017). Os autores constroem uma aplicação capaz de realizar a detecção de objetos a partir de imagens coletadas de um veículo aéreo não triplado *Unmaned Aerial Vehicle* (UAV).

Para realizar a busca dos objetos, é construído um *framework* que, inicialmente, processa as imagens de entrada, corrigindo a distorção causada pelas lentes. Em um segundo momento, o método realiza um casamento de objetos, utilizando o método baseado em *keypoints*. Conforme os demais métodos de *keypoints*, realiza as três etapas de aquisição de *keypoints*, construção de descritores e casamento entre imagens, buscando a homografia adequada até o objeto alvo. Para a aquisição dos *keypoints*, no estudo (BAYKARA et al., 2017) utiliza-se o *Features from Accelerated Segment Test* (FAST) (VISWANATHAN, 2011), a fase de descrição é feita utilizando o FREAK e, por fim, a homografia é descoberta utilizando o *Random Sample Concensus* (RANSAC).

Após a detecção inicial, a aplicação foca em acompanhar a movimentação dos objetos no video através do acompanhamento de pixels que são alterados com menor variação entre quadros. De forma similar aos métodos de fluxo óptico, os objetos têm seu fluxo rastreado e para manter o rastreio dos objetos, é aplicado um filtro de Kalman.

A etapa final do *framework* proposto em (BAYKARA et al., 2017) consiste na classificação dos objetos rastreados. A classificação é feita entre veículos ou humanos. Esta etapa utiliza uma rede neural de aprendizagem profunda para classificação, a SqueezeNet (IANDOLA et al., 2016) com modelo pré treinado. Segundo os autores, o resultado é equivalente ao apresentado pela AlexNet (KRIZHEVSKY; SUTSKEVER; HINTON, 2017) com a vantagem de ter menor quantidade de parâmetros e ocupar menos memória.

O *framework* proposto em (BAYKARA et al., 2017) serve como exemplo da dificuldade enfrentada de busca de objetos em abientes reais. Para realizar algo robusto, diversas técnicas são aplicadas, envolvendo desde métodos de fluxo óptico até classificadores compostos por redes neurais de aprendizagem profunda. Tendo em vista tal complexidade, faz-se providente analisar formas de tratamento de imagem e extração de características úteis para realizar o casamento dos objetos tridimensionais. Alguns dos principais métodos são explanados na seção a seguir.

2.2 AQUISIÇÃO DE ARESTAS

Uma das principais atividades dentre os métodos de busca, consiste na coleta de informação da cena onde se espera encontrar o objeto. Um tratamento prévio das imagens, permite aumentar a eficiência dos métodos. Além disto, estes treinamentos fazem parte diretamente das etapas de alguns deles, como é o caso do tratamento demonstrado em (LOWE, 2004), que utiliza da DoG como um dos principais processos para construção dos seus *keypoints*.

Uma das informações mais importantes que podem ser extraídas da cena são as bordas. A partir da imagem de bordas é possível extrair informações importantes para o casamento de *templates* tais como retas ou outras primitivas. Este processamento viabiliza o casamento presente nos métodos de casamento a partir de aresta, principalmente os similares ao RAPID.

(RAMAN; AGGARWAL, 2009) contém um estudo de diferentes operadores de extração de bordas de imagens. Os autores nomeiam como extração de bordas o processo de identificação de descontinuidades abruptas numa imagem. Os métodos clássicos se baseiam em filtros bidimensionais que convoluem pela imagem, destacando ou inibindo características baseando-se no gradiente presente. Os extratores podem ser focados em obter características específicas como arestas com determinada angulação, geometrias específicas como orientação baseada em angulação, etc. Uma característica comum entre os extratores é que sua complexidade aumenta de acordo com a robustez à ruído. Por outro lado, filtros menos sensíveis a ruídos podem acabar por negligenciar arestas mais sutis, falhando ao reconhecer arestas reais. Outros problemas comuns podem ser relacionados a arestas que não estão relacionadas diretamente a alteração de gradientes na imagem. Portanto, a escolha do filtro mais adequado deve ser feita de acordo com as necessidades do problema a ser enfrentado. Os autores em (RAMAN; AGGARWAL, 2009) classificam os extratores de bordas em dois tipos: os baseados em gradientes e os Laplacianos. Os métodos baseados em gradiente buscam a variação entre máximo e mínimo na primeira derivada da imagem, enquanto os métodos Laplacianos buscam por pontos de zero na segunda derivada da imagem.

Um dos operadores mais difundidos, o operador de Sobel, consiste em um *kernel* de convolução 3x3. O operador indica máximos sempre que encontra grandes variações de gradiente nas duas direções, vertical e horizontal. Os *kernels* podem ser usados de forma separada numa imagem para detectar apenas variações em cada direção. Ao serem aplicados de forma combinada, apresentam o módulo do gradiente, conforme apresentado na Equação 2.1, em cada

ponto além de sua orientação apresentado na Equação 2.2.

$$| G |= \sqrt{Gx^2 + Gy^2} \tag{2.1}$$

$$\theta = arctan\left(\frac{Gx}{Gy}\right) - \frac{3\pi}{4} \tag{2.2}$$

Um exemplo de operador Laplaciano é o Laplaciano da Gaussiana (LoG). O Laplaciano de uma imagem destaca regiões com variação brusca de intensidade. Um processo comum para realizar a extração das arestas de forma mais precisa, uma vez que o Laplaciano é muito sensível a ruídos, é a utilização de uma suavização da imagem. Um exemplo de processo é o processamento com a Gaussiana da imagem, um filtro de suavização. A detecção Laplaciana de uma imagem pode ser representada pela Equação 2.3, onde I(x,y) representa a intensidade dos pixels.

$$L(x, y) = \frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2} \tag{2.3}$$

Para viabilizar os cálculos dos valores dos pixels que são discretos, são utilizados *kernels* de convolução 3x3. Como a detecção da Gaussiana também pode ser representada por um *kernel* de convolução, para diminuir a quantidade de operações na imagem alvo, o filtro pode ser composto pela convolução do Laplaciano sobre o *kernel* Gaussiano criando uma nova máscara de segmentação.

Um operador similar ao Sobel é o apresentado por Prewitt (LIPKIN, 1970). Com um *kernel* 3x3 e voltado para a descoberta de arestas horizontais e verticais. O estudo apresentado por (Costa; Mello; Santos, 2013) utiliza máscaras de Prewitt em seu processo de extração de arestas. O método é melhor explorado na subseção 2.2.2.

2.2.1 Canny

Canny apresentou um algoritmo de segmentação de bordas (CANNY, 1986) focado em cumprir todos os requisitos de uma boa detecção de bordas, ou seja: (i) não haver risco de uma borda não ser detectada; (ii) detectar as bordas com a localização correta; (iii) a computação das arestas acontece apenas uma vez evitando assim que ruídos remanescentes influenciem o encontro das bordas reais. Com isto, foi proposto um algoritmo composto de seis etapas.

Inicialmente o algoritmo processa a imagem alvo com um filtro Gaussiano para suavização da nitidez da imagem, diminuindo assim a quantidade de características. Este processamento define que quanto maior o *kernel* gaussiano, menor será a sensibilidade do filtro a ruídos, levando à diminuição das arestas falsas-positivas. Em um segundo momento, o algoritmo indica a verificação do gradiente da imagem. Para coletar tais informações, são utilizados os dois *kernels* de convolução propostos por Sobel.

O terceiro passo consiste em estabelecer a direção das arestas a partir do resultado dos dois *kernels*. A Equação 2.4 define o ângulo da aresta, entretanto, (RAMAN; AGGARWAL, 2009) alerta aos problemas de anulação do ângulo quando o ângulo de Gx for nula, pois serão assumidos o valor de 0 ou 90 graus.

$$\theta = tan^{-1}\left(\frac{Gy}{Gx}\right) \tag{2.4}$$

O quarto passo é relativo à construir a relação entre o ângulo total à uma direção e sentido. Os ângulos são então associados aos valores de 0º, 45º, 90º ou 135º. A partir daí, o quinto passo consiste na supressão de não máximos. Este processo realiza um estreitamento das arestas e manutenção apenas de pontos de máximo. Entretanto, após essa etapa ainda permanecem arestas sobressalentes que precisam ser extraídas. A partir disto, o algoritmo segue para mais uma etapa.

Nesta última etapa, são descartadas as arestas a partir de um modelo de duplo limiar. Seguindo o modelo, são mantidas as bordas que estejam acima do limiar. Além destas, são mantidas as bordas que tenham adjacência a bordas que estão acima do limiar, por serem consideradas componentes das demais. Entretanto, são descartadas as arestas que estejam abaixo do limiar inferior e também as que estejam entre os limiares, mas não tenham nenhum tipo de adjacência com uma aresta acima do limiar.

O duplo limiar pode tornar o filtro mais difícil de ser parametrizado. Uma alternativa de filtro de bordas robusto com menor quantidade de parâmetros é o (Costa; Mello; Santos, 2013) apresentado na subseção a seguir.

2.2.2 Diffocus

A abordagem proposta pelos autores do Diffocus, de forma similar ao (ALAHI; ORTIZ; VANDERGHEYNST, 2012) é centrada em conceitos oriundos da percepção humana. Para a

construção do filtro, são construídas etapas que simulam a percepção humana da diferença cromática e percepção da densidade de texturas. São consideradas de forma individual cada componente cromático da imagem e armazenados. Para o completo funcionamento do filtro, inicialmente, é realizado um procedimento para supressão da textura. São utilizados filtros morfológicos para supressão da textura. São realizadas operações de abertura e fechamento na imagem e seus resultados são complementados para evitar perdas de bordas. Após a supressão, é computado o valor da diferença cromática (dRdGdB). Em uma imagem RGB, os valores Δ que representam os valores de diferença cromática serão calculados seguindo a Equação 2.5, Equação 2.6 e Equação 2.7.

$$\Delta R = R - G + R - B \tag{2.5}$$

$$\Delta G = G - R + G - B \tag{2.6}$$

$$\Delta B = B - R + B - G \tag{2.7}$$

Cada um dos valores de Δ calculado é armazenado como camadas na imagem, formando uma imagem com seis camadas ao total. Com intuito de evitar que os valores dos deltas sejam negativos, os valores são elevados ao quadrado. A partir daí, as seis máscaras são suavizadas aplicando-se duas máscaras gaussianas com valores de σ e 2σ . Isto gera duas variações da imagem, uma com maior nitidez que a outra.

A partir disto, as duas imagens passam por um processo de detecção de bordas, utilizando máscaras de Prewitt apresentado em (LIPKIN, 1970) em três direções vertical, diagonal e horizontal. Uma vez extraídas as arestas, cada uma das imagens tem suas seis camadas combinadas em uma única camada considerando o maior valor para os pixels entre cada uma das cores. Após este processo, as duas imagens são unidas entre si através da multiplicação dos seus componentes e uma posterior normalização dentro do intervalo $[0, 1020]$.

Por fim, a imagem passa por um processo de estreitamento das arestas similar à supressão de não máximos ou erosão. Entretanto, em (Costa; Mello; Santos, 2013) é explicitado que isto pode causar uma retirada em excesso de arestas, por isso recomenda que seja utilizado apenas em situações específicas.

Visando aumentar a robustez de métodos de extração de arestas, autores propuseram o uso de algoritmos de aprendizagem de máquina para realizar a segmentação. O *Hollisticaly-*

nested Edge Detection (HED) (XIE; TU, 2015), uma rede neural que utiliza de técnicas de aprendizagem profunda é apresentada seguir.

2.2.3 HED

A heurística proposta em (XIE; TU, 2015) consiste num sistema completo de detecção de arestas que é capaz de aprender as principais características que devem ser segmentadas das imagens. A rede neural apresenta diferentes resultados aninhados como saída. É possível obter diferentes níveis de detalhes de suas camadas, o que permite o uso em diferentes contextos. A sua arquitetura apoia-se na apresentada por (SIMONYAN; ZISSERMAN, 2014) com a diferença que foi realizada a conexão entre camadas de convolução.

Para realizar o treinamento, é utilizado um conjunto de imagens e suas respectivas referências globais (*ground truth*). Por se tratar de um sistema que verifica a corretude a partir de imagens de saída, os autores propõem o uso de uma função de custo associada.

Os experimentos sobre a base de imagens *Berkeley Segmentation Data Set and Benchmarks 500* (BSDS 500) demonstraram o limiar de contorno fixo (ODS) mais alto entre os estudos comparados, $0,782$ se aproximando do *ground-truth* $0,8$. Comparativamente, o Canny obtém na mesma base de dados um ODS de $0,6$. Outro fator importante é que diante das outras redes neurais de aprendizagem profunda este algoritmo apresentou também os menores tempos de execução.

Outro ponto de destaque é que em testes realizados adicionando aleatoriamente mais 100 imagens ao conjunto de testes foram obtidos bons resultados. O testes foram realizados 5 vezes e foi possível aumentar o ODS até $0,797(\pm 0,003)$ se aproximando ainda mais dos valores apresentados como *ground truth*.

Elucidados os métodos extratores de características de cena, faz-se providente um estudo de como se dá a projeção e estimativa do posicionamento dos objetos tridimensionais em cena. Esta estimativa permite adquirir as informações necessárias para realizar o rastreio dos objetos, apresentando seu posicionamento com seis graus de liberdade. A seção a seguir apresenta uma revisão dos principais métodos de inferência de posicionamento de um objeto tridimensional em uma imagem bidimensional.

2.3 ESTIMATIVA DO POSICIONAMENTO DOS OBJETOS TRIDIMENSIONAIS

Uma das principais características dos métodos de busca é a identificação do local em que o objeto se encontra. Diferentes métodos foram propostos e realizam desde o rastreio através da descoberta da melhor homografia que correlacione os pontos, como é o caso do *Random Sample Concensus* (RANSAC) (FISCHLER; BOLLES, 1981), até a inferência do *Six Degrees of Freedom* (6DoF) dos objetos, técnica comum em métodos baseados em *templates*. Este estudo foca em métodos para a detecção de *templates* por serem comuns em métodos que realizam a busca de objetos tridimensionais.

A partir do posicionamento dos objetos no espaço, é possível prever colisões, assim como emular condições únicas que são necessárias ao funcionamento de diferentes sistemas. Uma das formas mais comuns da aquisição do posicionamento dos objetos consiste na descoberta do posicionamento do ponto focal da câmera que capturou a imagem. Esta descoberta permite a construção de um espaço de coordenadas centrado na câmera e consequentemente um mapeamento tridimensional do ambiente, permitindo a inferência do posicionamento dos itens.

Em aplicações reais, pode vir a ser providente o uso de marcadores fiduciais para aumento da acurácia na inferência do posicionamento. Ao final desta seção, é apresentado um método de aquisição de posicionamento fiducial rápido que demonstra ser bastante confiável para aplicações reais.

2.3.1 Métodos de Ponto e projeção

Os autores de (LEPETIT V., 2009) dividem os métodos *Perspective-n-Point* (PnP) entre iterativos e não iterativos. Os métodos iterativos tendem a ser mais precisos, entretanto, tendem a apresentar falhas de resultados quando encontram mínimos locais, além de terem um alto custo computacional. Para os métodos não-iterativos o problema da descoberta da projetividade responsável pelo posicionamento do ponto, geralmente envolve resolver as raízes de um polinômio de oitavo grau com apenas os coeficientes pares. Portanto, é possível descobrir a informação a partir de 4 soluções conhecidas. Entretanto, é comum que os métodos explorem o casamento utilizando mais pontos, para que seja possível lidar com ruídos possibilitando redundância entre os pontos.

Visando solucionar o problema, foram propostos métodos baseados em decomposição do valor singular de uma matriz *Singular Value Decomposition* (SVD). Entretanto, tais métodos

podiam ter complexidade $O(n^3)$. Para uma grande quantidade de valores de entrada foram propostos métodos que utilizam a Transformação Direta Linear *Direct Linear Transform* (DLT).

O algoritmo *Efficient Perspective-n-Point* (EPnP) proposto em (LEPETIT V., 2009) indica que para seu funcionamento é necessário, ao menos, o conhecimento de quatro duplas de pontos não coplanares. Cada dupla conterá os valores que atribuem o ponto tridimensional - em coordenadas de câmera - e os valores de posicionamento dos pixels da imagem da cena. Esta é a configuração mínima para funcionamento do método. Entretanto, como afirmado anteriormente, quanto maior a quantidade de duplas, mais preciso será o resultado.

A partir das correspondências, é solucionado o problema que consiste no núcleo da matriz (M) de tamanho $2n$ x 12 ou $2n$ x 9. De forma mais precisa, o resultado é a soma ponderada dos autovetores nulos da matriz. Dada a combinação linear correta, ou seja, as coordenadas de câmera para os pontos de controle, o problema se resume a resolver um sistema de equações quadráticas. O custo total assumido quando a quantidade de duplas for acima de 15 é basicamente o custo de computação da Matriz que cresce de forma linear. Isto posto, o método tem complexidade O(n). O método apresenta bons resultados nos experimentos que foram considerados desde ruído gaussiano na imagem, até a presença de ruídos impostos, como a correlação bidimensional de 25% dos pontos sendo feita de forma randômica (LEPETIT V., 2009). Uma relação importante também discutida em (LEPETIT V., 2009) é que, quando os pontos tendem a cobrir apenas uma pequena porção da imagem, a eficiência geral dos métodos é bastante impactada.

Um estudo mais recente foi proposto em (LI; XU; XIE, 2012). Assim como em (LEPETIT V., 2009) é proposto um modelo não-iterativo. Porém ele se diferencia ligeiramente dos demais estudos, por apresentar diferentes pesos para as correspondências. O estudo propõe a resolução em três etapas:

- Agrupar as duplas correspondentes em sub-conjuntos de três itens para adquirir polinômios de quarta ordem.
- Computar a soma dos quadrados das raízes dos polinômios para a construção de uma função de custo.
- Encontrar as raízes da derivada da função de custo produzida com intuito de encontrar os valores ótimos.

Os resultados apresentados pelo *Robust Perspective-n-Point* (RPnP) demonstram bastante robustez, com valores aproximados de eficiência dos demais métodos tanto em conjuntos onde não existe redundância (n<5) como quando existe (n>5). Uma outra forma de encontrar o posicionamento dos objetos se dá através do uso de marcadores fiduciais. Por serem um artefato introduzido artificialmente na cena, o uso de marcadores fiduciais permite diminuir a incerteza da localização dos pontos que é presente nos métodos PnP.

2.3.2 Descoberta de Posicionamento utilizando marcadores fiduciais

Dentre os estudos que apresentam o uso de marcadores fiduciais, um de baixo custo computacional e resultados eficientes é o ArUco (ROMERO-RAMIREZ; MUñOZ-SALINAS; MEDINA-CARNICER, 2018). Utilizando-se de métodos de binarização e extração de contornos, é possível realizar o reconhecimento rápido de marcadores quadrados impressos em papel.

O algoritmo de reconhecimento se dá em quatro etapas. A primeira etapa, consiste na aplicação sobre a imagem onde se busca o objeto de um limiar adaptativo. A partir de uma janela de tamanho m, o algoritmo zera os valores dos pixels que ficarem acima de um determinado limiar c. O algoritmo de binarização da primeira versão do ArUco é o apresentado em (OTSU, 1979). A segunda etapa é a extração de contornos da imagem. Como a maior parte dos contornos é irrelevante, é feito um processo de filtragem dos contornos, onde são descartados todos os contornos considerados pequenos demais. Em um segundo momento, os contornos são aproximados a polígonos e os que não fizerem parte de um polígono convexo bem definido, são descartados.

O terceiro passo do algoritmo realiza a extração dos marcadores. A partir da imagem binarizada, o polígono é dividido em uma matriz quadrada, onde as partes pretas são consideradas valores 0 e as brancas consideradas valores 1. Por ser uma matriz quadrada, são consideradas e armazenadas as quatro configurações possíveis de posicionamento. Ou seja, a matriz fonte e

O passo final de encontro dos marcadores é relativo ao refinamento de quinas a nível de subpixel. Para realizar a descoberta do exato posicionamento das quinas dos marcadores, são verificadas a angulação das retas que os compõem e, posteriormente, é computada sua interseção com uma precisão de sub-pixels. Os autores não recomendam este tipo de refino em lentes com alto valor de distorção tal quais olho-de-peixe.

A eficiência computacional do método é diretamente proporcional ao tamanho da imagem.

Imagens muito grandes apresentam muito mais arestas a serem refinadas e mais cálculos são necessários. Para acelerar a computação do método o autor propõe ainda a utilização de um método de limiar adaptativo global, e o redimensionamento da imagem para metade do seu tamanho. Estes incrementos tornaram o método quarenta vezes mais rápido se comparado ao método clássico do ArUco, que por sua vez ainda é superior aos demais métodos do estado da arte comparados.

Um dos fundamentos essenciais para a detecção de objetos é a utilização de câmeras calibradas. A calibração é necessária para poder aumentar a eficiência de diversos métodos, além de viabilizar o uso de técnicas de aumento de eficiência, como é o caso do (ROMERO-RAMIREZ; MUñOZ-SALINAS; MEDINA-CARNICER, 2018).

2.4 CONSIDERAÇÕES FINAIS

Neste capítulo, foram apresentados diversos estudos do estado da arte voltados para a busca de objetos tridimensionais em imagens bidimensionais. Dentre os estudos analisados, destacam-se os métodos baseados em *keypoints*, os métodos baseados em *templates*, métodos baseados em Inteligência artificial, além da vasta ocorrência de estudos voltados para o casamento de arestas extraídas dos modelos tridimensionais e da imagem alvo. Tais métodos derivam do RAPiD e apresentam uma busca aprimorada com possibilidade de rastreio do objeto em tempo real. Logo foram revisados e apresentados também, diversos estudos voltados para a extração de arestas de imagens RGB, destacando-se o método proposto por Canny para sistemas que necessitem de baixo custo computacional e o HED que conta com uma extração de características aprimorada, entretanto, por usar de uma rede neural de aprendizagem profunda, acaba sendo mais orientado a sistemas com alto poder computacional. Uma vez que diversos métodos de busca requerem inicialização da cena, foram revisados ainda, estudos que pudessem estimar o posicionamento de um objeto tridimensional da cena, a partir de pontos bidimensionais conhecidos. O método RPnP demonstra uma maior robustez, entretanto, os métodos clássicos como EPnP podem se demonstrar adequados, quando forem cumpridos os requisitos de maior eficiência. No capítulo a seguir, são cruzados diversos estudos para a criação de um *framework* de realizar o casamento entre objetos tridimensionais complexos em imagens RGB.

3 SISTEMA DE RECONHECIMENTO DE MODELOS TRIDIMENSIONAIS ESPECULARES

No contexto industrial, com o advento da manufatura aditiva, diversos desafios tecnológicos surgem. Dentre eles, é possível elencar o reconhecimento de conglomerados de peças, cuja única informação disponível é uma lista de modelos tridimensionais. Esses modelos contêm muitas peculiaridades que surgem devido ao fato do processo de construção manufatura aditiva alterar, de forma iterativa e interativa, o objeto da construção e, por consequência, todo o conglomerado a ser reconhecido. Visando resolver tal desafio, este capítulo apresenta uma abordagem que explora de métodos clássicos de busca de objetos tridimensionais, tais quais os propostos em (ARMSTRONG; ZISSERMAN, 1995), (DRUMMOND; CIPOLLA, 2002) (CHOI; CHRISTENSEN, 2010), aliado a métodos de processamento de imagens para aquisição de características da cena e técnicas de renderização para realizar o reconhecimento de objetos complexos sem textura.

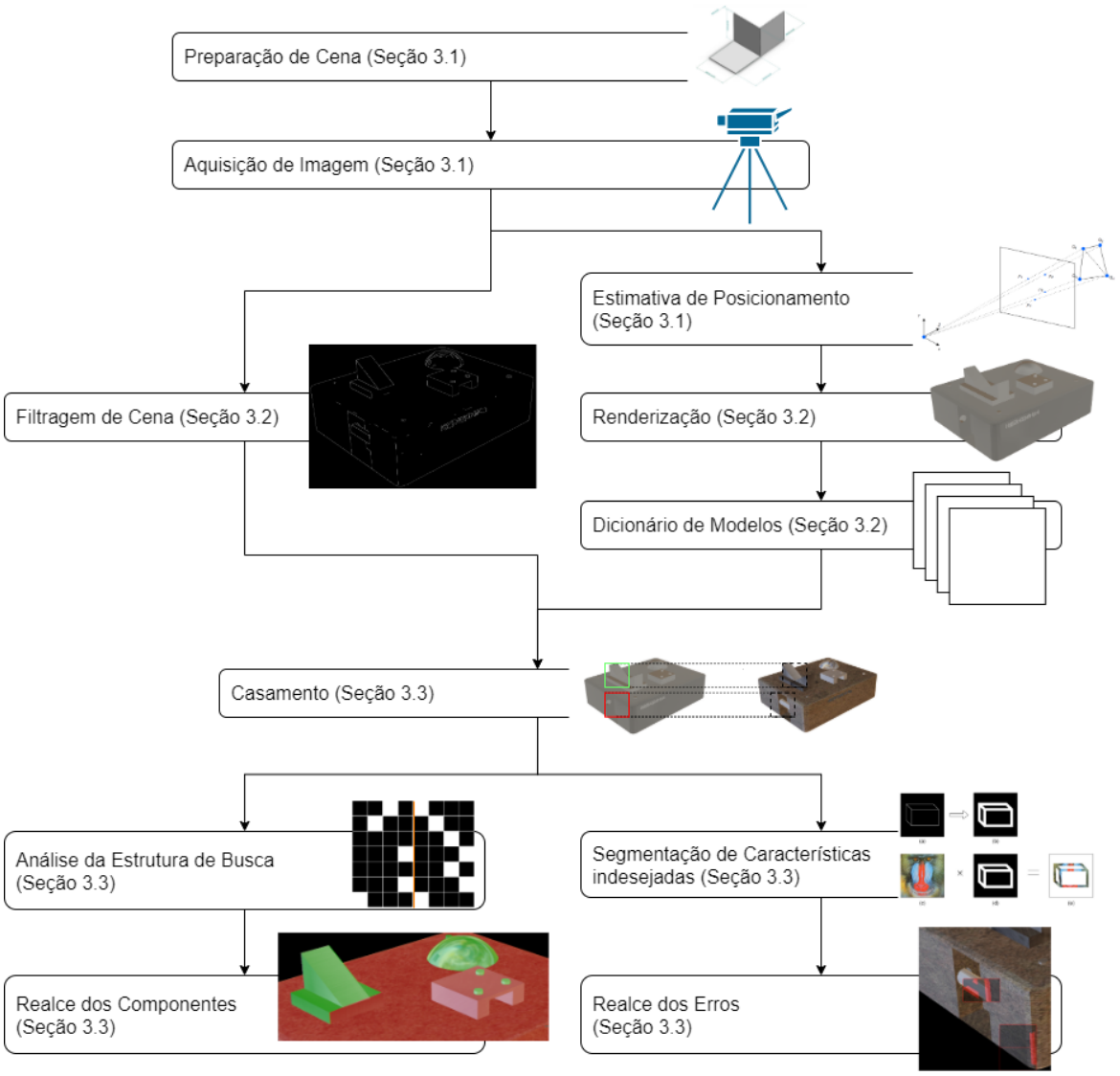
Visando facilitar a compreensão deste documento, algumas definições são listadas abaixo e esclarecem o uso de algumas palavras dentro do contexto deste trabalho:

- Matriz: molde que compõe a prensa de estamparia. São montadas por diversas ferramentas de acordo com o gabarito representado pelos modelos tridimensionais CAD;
- Ferramenta: peça que compõe uma matriz de estamparia. São componentes em metal com partes foscas e/ou especulares e apresentam diferentes funções no processo de estamparia;
- Conglomerado: lista na forma de arquivo digital, podendo conter de dezenas a centenas de modelos tridimensionais CAD com seus respectivos posicionamentos;
- Cena: ambiente real onde é buscado a representação física do conglomerado a ser representado por uma imagem retirada por uma câmera;
- Modelo: imagem renderizada do conglomerado de peças;

O processo de reconhecimento proposto busca: (i) analisar de forma aprofundada as características da cena, aqui representada através de uma imagem única retirada de uma das faces do conglomerado buscado; (ii) mitigar informações que tornam-se pouco importantes a depender do material componente do conglomerado de peças renderizado; (iii) detectar de forma mais precisa o posicionamento da câmera para uma melhor adequação do casamento; e

(iv) analisar a imagem, buscando defeitos e objetos inadequados na cena, além de medir qual a confiança quanto à presença de cada item tridimensional na cena. O conjunto de etapas, descritas de forma geral na Figura 4, busca viabilizar esses pontos, permitindo um casamento de modelos constituídos por centenas de peças em materiais comumente encontrados nas indústrias como metais e derivados.

Figura 4 – Fluxo geral do sistema proposto.



Fonte: O autor.

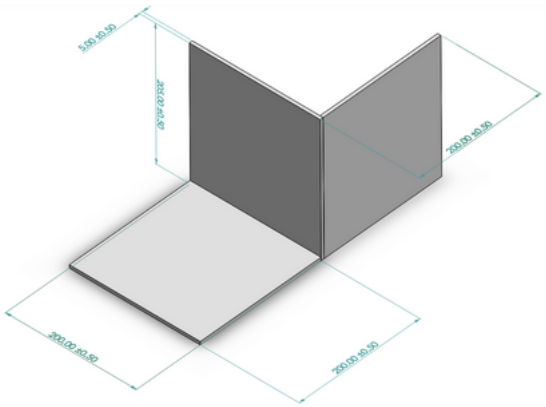
O fluxo proposto detalha desde o modelo de aquisição de imagens até a geração de relatórios visuais e textuais. O detalhamento das etapas é descrito nas seções a seguir.

3.1 CENA

A cena representa o ambiente no qual são buscadas as peças ou itens. No esquema proposto, a cena é preparada através de uma série de ancoragens, buscando minimizar os problemas de busca de itens e focando o algoritmo no reconhecimento do conglomerado. Como discutido anteriormente, é sabido que alguns algoritmos, principalmente os que utilizam modelos de aprendizagem de máquina, dedicam-se a mais de uma tarefa, a exemplo do estudo apresentado em (REDMON; FARHADI, 2018); entretanto, o foco da pesquisa aqui desenvolvida foi apenas o reconhecimento.

Com intuito de diminuir a incerteza proveniente da busca e da aquisição das imagens, o ambiente é preparado com marcadores rígidos, com formato tridimensional apresentado na Figura 5; em suas faces foram adicionados marcadores fiduciais dispostos com posicionamento conhecido. Para a aquisição das imagens, a câmera escolhida é de foco fixo (HARTLEY; ZISSERMAN, 2003), viabilizando o processo de calibração prévia. A calibração pode ser feita assim que o foco for alterado de forma visualmente adequada, tomando como base um pedestal fixo posicionado na cena a uma distância adequada para que, em uma única imagem, seja possível enquadrar toda a ferramenta buscada. A preparação da cena permite a garantia de que efeitos externos sejam minimizados, podendo focar no problema principal de reconhecer uma extensão maior do conglomerado de peças, assim como uma maior quantidade de detalhes.

Figura 5 – Exemplo de modelo de marcador rígido que serve de tela para os marcadores fiduciais que devem ser adicionados à cena.



Fonte: O autor.

3.1.1 Câmera

Para captura das imagens, uma câmera de foco fixo foi utilizada no modelo proposto neste trabalho. Este equipamento permite que seja feita uma calibração mais precisa, onde é possível adquirir valores intrínsecos tais quais os de distorção, a distância focal e o centro focal. Para realizar a calibração e aquisição dos valores da câmera, foi utilizado o método clássico de calibração, utilizando um *chessboard* conforme explicado no Apêndice A.

Uma vez que a câmera é posicionada e as imagens de referências para realizar a calibração são geradas, a imagem da cena que guia o processo de casamento é então adquirida. As imagens são armazenadas em formato *raw* para diminuir o nível de processamento sobre a imagem salva, assim como evitar o uso de algoritmos de interpolação e compressão.

3.1.2 Posicionamento esperado

Por se tratar de um método de casamento que utiliza o cruzamento de informações presentes em um modelo canônico *versus* as informações presentes na imagem de cena capturada, faz-se necessário uma estimativa inicial de onde as peças se encontram. Para essa estimativa inicial do posicionamento dos modelos na cena, foram utilizados marcadores fiduciais em um local conhecido da peça. Uma das entradas do programa desenvolvido é o cruzamento de qual o marcador rígido está presente em cada um das esquinas do item buscado. Esse cruzamento permite a descoberta de qual será o ponto de vista da peça que será renderizada.

O item componente do conglomerado tridimensional ao ser renderizado torna-se um modelo que guia todo o processo de identificação do que é buscado, para tal, é necessário ser feita uma projeção estimada de onde os itens estarão na cena. A precisão necessária da estimativa inicial é um fator crucial, pois interfere diretamente na eficiência final do casamento.

Com isto, para realizar a estimativa inicial de onde a ferramenta está no ambiente e, assim, poder renderizar a cena, foram utilizados marcadores fiduciais combinados, presentes nas faces dos marcadores rígidos impressos. Dentre os marcadores estudados, foram selecionados os disponíveis na biblioteca (ROMERO-RAMIREZ; MUñOZ-SALINAS; MEDINA-CARNICER, 2018), pois estes demonstraram boa integração e interoperabilidade com o software em desenvolvimento.

Dentre os diversos experimentos realizados, os estudos demonstraram que o uso isolado de um marcador fiducial, anexado a um local conhecido da cena, causava incongruências entre a renderização e a peça real por quase toda a extensão da imagem sintética. Para a identificação

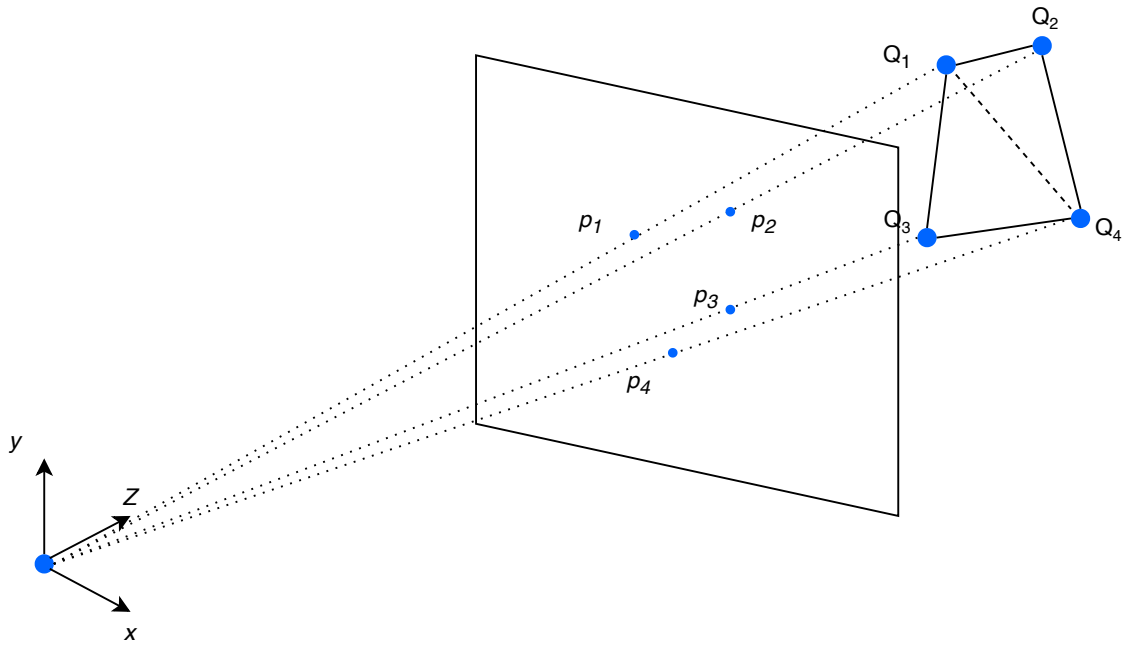
de conglomerados que tenham metros de extensão na cena, poucos graus de posicionamento errôneo do marcador em uma extremidade geram um distanciamento considerável entre a renderização e a imagem de cena; ou seja, vários centímetros ao longo das comparações feitas na extensão do conglomerado. Credita-se isto ao fato da ferramenta ter tamanho bastante superior ao do marcador na cena.

Para atenuar o desvio de posicionamento do marcador sem ter que necessariamente aumentar as dimensões do marcador impresso a ser anexado à ferramenta, são utilizados o posicionamento combinado de quatro marcadores e de métodos PnP (do inglês, *Perspective and Projection*) (HARTLEY; ZISSERMAN, 2003). Tais métodos permitem a projeção de um sistema de coordenadas projetivas do P3 (espaço projetivo tridimensional) na imagem da cena, a partir da estimação de posicionamento da câmera. Isto viabilizou uma projeção mais aproximada ao posicionamento real da ferramenta. Houve, portanto, uma diminuição do erro causado por desvios de ângulo devido à ligeira inclinação dos marcadores, detalhes imperceptíveis a olho nu, e consequente posicionamento da renderização quando adicionados os quatro marcadores posicionados ao longo da ferramenta na cena real.

Para computação dos métodos PnP, são necessários ao menos quatro pares de pontos conhecidos do espaço projetivo P2 e seus pontos correlacionados em coordenadas tridimensionais de mundo, vide Figura 7. Uma vez adquiridos os pares de pontos correspondentes, são utilizados métodos iterativos para determinar qual a projetividade que, partindo-se dos pontos tridimensionais, obtém-se os pontos bidimensionais descritos no par ou uma aproximação. Essa aproximação permitida, entra como um parâmetro do método iterativo e pode ser calculada de diversas formas, a exemplo do erro de reprojeção médio. Quanto menor o erro, melhor é a coerência da projetividade. Por se tratar de um algoritmo iterativo e sem um ponto de finalização bem definido, a otimização da função de erro de reprojeção pode apresentar resultados que são mínimos ou máximos locais ou globais, sendo os máximos globais os resultados mais precisos para o método. Entretanto, o método tende a apenas encontrar o máximo global quando não existe nenhum tipo de ruído entre as duplas de pontos encontrados (HARTLEY; ZISSERMAN, 2003).

Para a execução do método proposto nesta Dissertação, é necessário captar a localização em pixel dos centroides de cada marcador. Pois são captados o posicionamento dos centroides na imagem e sua informação é comparada às coordenadas tridimensionais conhecidas na cena para cada marcador. Em posse destas informações, lança-se mão do uso de métodos PnP para descoberta de qual o posicionamento da câmera na cena. Entretanto, como o processo de

Figura 6 – Perspectiva e Projeção.



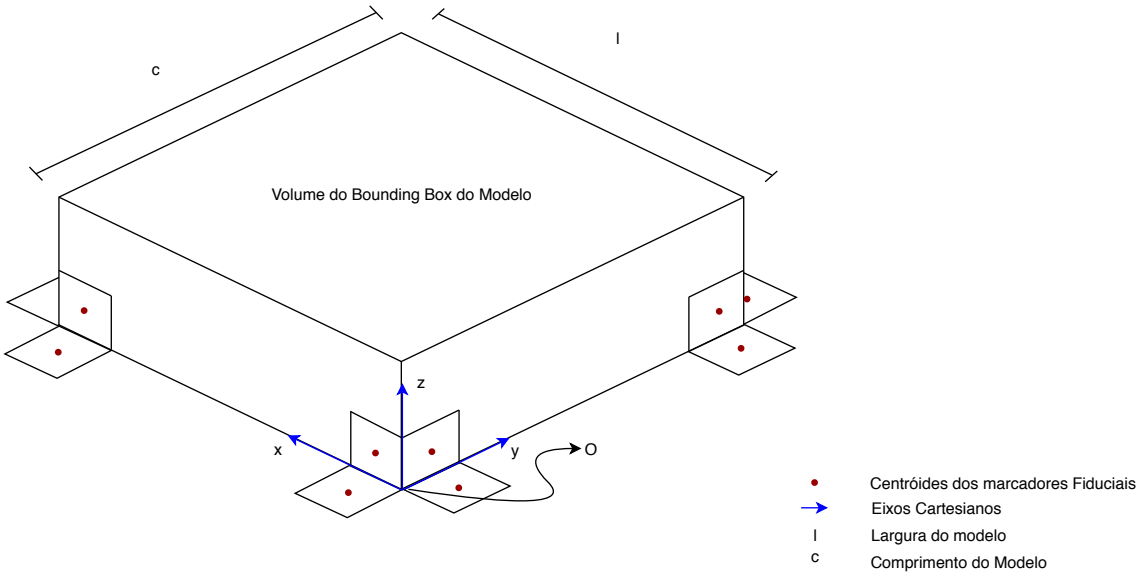
Fonte: O autor.

digitalização de uma imagem, desde a sua amostragem até a definição de em qual pixel se encontra o centroide do marcador, é composto por uma série de generalizações e incertezas, é considerado estatisticamente improvável, o alcance do máximo global durante a aplicação dos métodos iterativos de PnP. Métodos e formas de diminuir as incertezas e generalizações do processo de aquisição do centroides continuam sendo estudados e, aliados a estes, outros algoritmos iterativos e até mesmo diferentes funções de erro são utilizados para descobrir em tempo hábil qual a melhor projetividade.

Desta forma, foram considerados na solução três dos métodos PnP, sendo eles: EPnP (LEPETIT V., 2009), P3P (GAO et al., 2003) e o RPnP (LI; XU; XIE, 2012). Dentro do algoritmo proposto, são utilizadas quatro duplas (Q_i, p_i) de pontos, sendo eles $\{(Q_i, p_i)|Q_i \in P3, p_i \in P2\}$. Os pontos de mundo são obtidos do posicionamento esperado dos marcadores fiduciais na cena. Estes dados são cruzados com informações de largura e profundidade do objeto buscado; esses valores são aqui utilizados para determinar em qual coordenada tridimensional estará cada um dos centros dos marcadores rígidos confeccionados, considerando a origem em um local escolhido. A Figura 7 demonstra o processo de escolha dos pontos: uma estimativa dos centroides de cada marcador fiducial é a correspondência bidimensional do ponto; com isso, forma-se a dupla correspondente.

Em posse das duplas, é feita a estimativa de posicionamento da câmera, utilizando-se os

Figura 7 – Esquema tridimensional de distribuição dos marcadores e escolha dos pontos fiduciais.



Fonte: o autor.

três métodos de PnP. Os três valores retornados por cada um dos métodos são comparados. Para essa comparação, os valores tridimensionais relativos às coordenadas esperadas dos três marcadores são projetados, usando a nova câmera obtida. Com estes pontos, é calculada uma distância Euclidiana entre os centroides previamente obtidos dos marcadores e os reprojetados. O método que apresentar a menor média de distância ponto a ponto dos centroides é então o método considerado doravante no processo de reconhecimento.

Uma vez que os métodos PnP são métodos imprecisos, ainda é realizado mais um tipo de verificação de corretude. Baseando-se nas dimensões da imagem apenas, é considerada como saída útil dos métodos PnP, os valores que apresentarem distância Euclidiana média inferior a um limiar adequado definido de acordo com a aplicação.

3.2 MODELO

Para realizar o casamento de modelos CAD sem textura, uma série de algoritmos utiliza métodos que realizam a busca de pontos de correlação tridimensional *versus* bidimensional (CHOI; CHRISTENSEN, 2010), (MIAN; BENNAMOUN; OWENS, 2006b), (DRUMMOND; CIPOLLA, 2002), (ARMSTRONG; ZISSERMAN, 1995), (MIAN; BENNAMOUN; OWENS, 2006a). Para isto, são projetados os pontos tridimensionais na imagem da cena e comparadas as distâncias ao

longo de cada uma das arestas do modelo e sua correlação em uma imagem de bordas obtida da cena. Com isto, a principal informação de entrada da solução é uma lista de modelos tridimensionais CAD. Eles descrevem cada um dos componentes do conglomerado do ponto de vista geométrico, definindo seu formato e volume, assim como o posicionamento relativo entre eles.

Os arquivos CAD têm ainda uma vantagem em relação a outras formas de modelagem de sólidos: eles possuem uma relação fixa entre centímetros do mundo real e coordenadas, em uma relação direta de 1 para 1. Tal informação é usada como guia durante o processo de casamento.

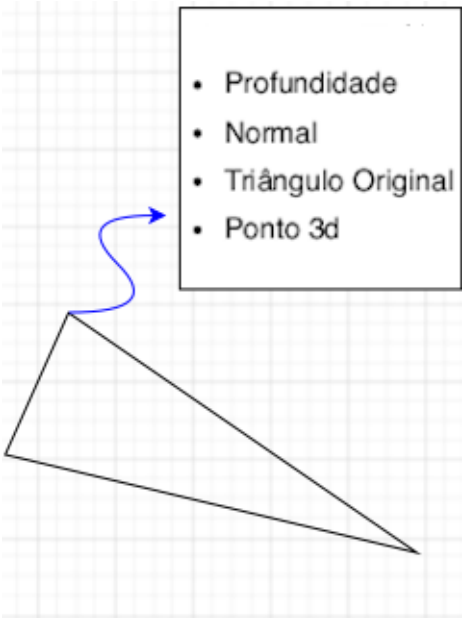
A principal etapa da busca dos itens numa imagem de cena, é a verificação da correlação entre o que se têm como modelo e as características apresentadas na imagem. Visando uma forma de assimilar as características que estarão presentes nas imagens da cena e os modelos tridimensionais, utiliza-se uma renderização para seleção de características dos itens buscados. A renderização permite suprimir ou evidenciar características a partir da exploração de diferentes modelos de iluminação, além de ser possível lidar com oclusão como por ser visto na subseção 3.2.3. Associada à renderização, é utilizado um filtro detector de bordas. Através da detecção de bordas da imagem do modelo renderizado é possível filtrar do arquivo CAD apenas as características que tendem a se destacar na imagem da cena, assim como suavizar as características menos pronunciadas.

3.2.1 Renderização

Durante a construção da renderização, podem ser utilizados algoritmos mais simples de *shading* como Phong (HUGHES et al., 2014) ou outros métodos de renderização mais elaborados como o *Ray Tracing* (WHITTED, 2005). Observa-se que o custo computacional da renderização é um dos fatores limitantes do tempo de execução da aplicação. Uma das principais características das renderizações, a iluminação é outro fator a ser observado. O uso de diversas fontes de luz pode tanto aproximar quanto distanciar da iluminação presente na cena. Por isso, foi utilizada apenas uma fonte de iluminação pontual com localização atrás do olho da câmera; os coeficientes de especularidade e rugosidade podem ser ajustados conforme o material buscado dos itens. Para realizar a oclusão dos itens da cena, é utilizado o algoritmo Zbuffer. Para o processo de rasterização foi utilizado o algoritmo de rasterização apresentado em (FOLEY et al., 1990), além deste, podem ser utilizadas técnicas de *anti-aliasing*.

Como uma forma de viabilizar etapas futuras do casamento, no momento que o Zbuffer é construído, pós-rasterização, é preenchida uma estrutura de dados, como pode ser observado na Figura 8. Nela constam informações importantes, tais como qual triângulo e de qual é o item do conglomerado cuja renderização contém cada pixel, qual o ponto tridimensional e qual a normal presente em um determinado pixel. Sendo que, os dois últimos são coletados baseando-se nas coordenadas baricêntricas do triângulo. Essas informações além de úteis para realizar o pixel *shader*, são insumo para os algoritmos de casamento e PnP. A sua coleta permite também otimizações futuras na solução, como descoberta de angulação entre faces de cada aresta, entre outras.

Figura 8 – Informação por pixel.



Fonte: o autor.

A partir da lista de modelos CAD tridimensionais da entrada, à medida que a rasterização vai sendo realizada e os pixels vão sendo calculados, um dicionário de modelos é construído. Cada peça que se acredita estar presente na cena, com ou sem oclusão, recebe um endereço específico nesta estrutura de dados. Nela, se mapeiam também informações úteis para outros algoritmos. Essa estrutura permite, posteriormente, o casamento peça à peça, além de viabilizar um destaque para cada peça e onde se espera encontrá-la. Uma vez que todas as peças do conglomerado são renderizadas em conjunto, uma imagem é gerada sendo esta um guia de como se espera que os modelos estejam arranjados na cena, consolidando esta imagem renderizada como um modelo canônico. Para realizar uma extração mais simplificada de informações do modelo canônico, a imagem é então submetida a um filtro detector de bordas. A

partir das bordas do modelo, é possível ter a quantidade mais adequada de informações para o casamento como pode ser visto na seção 3.3.

Os filtros de bordas segmentam da renderização, de forma geral, as arestas entre polígonos cuja a angulação seja mais acentuada. O estudo proposto por (CHOI; CHRISTENSEN, 2010), utiliza um modelo com arestas que ele considera fortes, onde os polígonos formam uma angulação entre si maior que 30 graus. No modelo aqui proposto, é poupada a etapa de descoberta de angulação entre todos os polígonos e são filtrados de forma simplificada apenas as arestas que se pressupõem destacar numa visualização aproximada da cena. Além das arestas presentes devido à angulação entre os polígonos componentes do arquivo CAD, outras arestas podem surgir por causa da própria reflexão especular. Regiões planas de itens especulares, apresentam arestas devido a iluminação. Com esta abordagem, é possível captar tais características, assim como neutralizá-las.

Durante os experimentos, foi possível observar a necessidade de boas escolhas de referencial para montar o pixel *shader*, assim como a quantidade de fontes de iluminação e o seu posicionamento. A iluminação pode gerar ou suprimir características importantes nos diversos itens do conglomerado, principalmente, os que têm maior capacidade especular. Portanto, os estudos levaram a consideração de diversos filtros e abordagens para supressão de características, como apresentado na seção a seguir.

3.2.2 Detecção de características do modelo

Como explicitado na subseção 3.2.1, a consolidação da imagem renderizada para se tornar modelo para o processo de casamento se dá após uma detecção de bordas. Essas informações são posteriormente comparadas com intuito de buscar discrepâncias e semelhanças, caracterizando o casamento entre a imagem do modelo e do mundo real. A detecção de bordas permite uma supressão de parte das informações, além de um refino de características que são consideradas úteis.

Dentro do escopo do projeto, foram estudados e comparados métodos de detecção de bordas e contornos como o (CANNY, 1986) e o (XIE; TU, 2015), além de métodos de limiarização de imagens, como o (OTSU, 1979). Os experimentos levaram ao uso de dois algoritmos detectores de bordas principais para a solução. Foram considerados fatores como tempo de execução e resultado para escolha dos algoritmos. Os métodos escolhidos para uso dentro da ferramenta foram o Canny e o HED (XIE; TU, 2015) (do inglês, *Holistic Edge Detector*),

mais robusto e baseado em Redes Neurais. A escolha de qual dos métodos é usado é uma das entradas da solução. Assim, o uso de dois algoritmos permite redundância no casamento, dando mais confiabilidade para gerar os relatórios de casamento.

Para imagens sintéticas como a de modelo o Canny apresenta boa aquisição de características e baixo tempo de execução. O uso do HED para refino do modelo, além do atraso, gera uma supressão de suavidades demasiada; acredita-se que pelo fato do modelo utilizado da rede neural ter sido treinado com imagens reais apenas.

3.2.3 Dicionário de modelos

Uma das características relevantes dos conglomerados é a grande quantidade de componentes. Algumas centenas de modelos tridimensionais únicos podem vir a compor uma única cena. Entretanto, a depender do ponto de vista que a cena é observada, alguns modelos podem vir a ficar completamente oclusos na cena e sua busca, uma vez que a face a ser observada já foi decidida pelos marcadores, conforme apresentado na subseção 3.1.2, torna-se dispensável.

Como descrito na subseção 3.2.1, após o carregamento de todos os itens do conglomerado, é gerado um mapa com a renderização de cada peça. O uso do *Z-buffer* permite uma fácil eliminação desta lista de peças completamente oclusas. Com isto, um dicionário de modelos que se almeja serem visíveis é criado, contemplando todas as informações que foram coletadas como modelo tridimensional gerador, pixels rasterizados, triângulos geradores da região rasterizada e suas normais.

Além dos modelos totalmente oclusos, modelos parcialmente oclusos também recebem um tratamento especial dentro do método proposto neste documento. Para cada imagem presente no dicionário de modelos renderizados é verificada a ordem de grandeza da quantidade de pixels rasterizados deste modelo único comparado com a quantidade total da renderização de toda a lista de modelos. Com isto, são descartados das etapas posteriores de casamento, os modelos que apresentarem uma quantidade de pixels muito inferior ao total da renderização.

Este tratamento evita comparativos desnecessários para o casamento das peças, além de diminuir a quantidade de falsos negativos, uma vez que, após a rasterização, pode ser possível que peças com apenas um pixel rasterizado sejam buscadas. Isso tornaria o relatório final descrito na subseção 3.3.4 confuso, além de causar uma indefinição do que está sendo buscado na cena. Com isto, tem-se o dicionário de modelos visíveis e relevantes consolidado para uso em etapas futuras como casamento e geração do relatório conforme explicado na

seção 3.3.

3.3 CASAMENTO

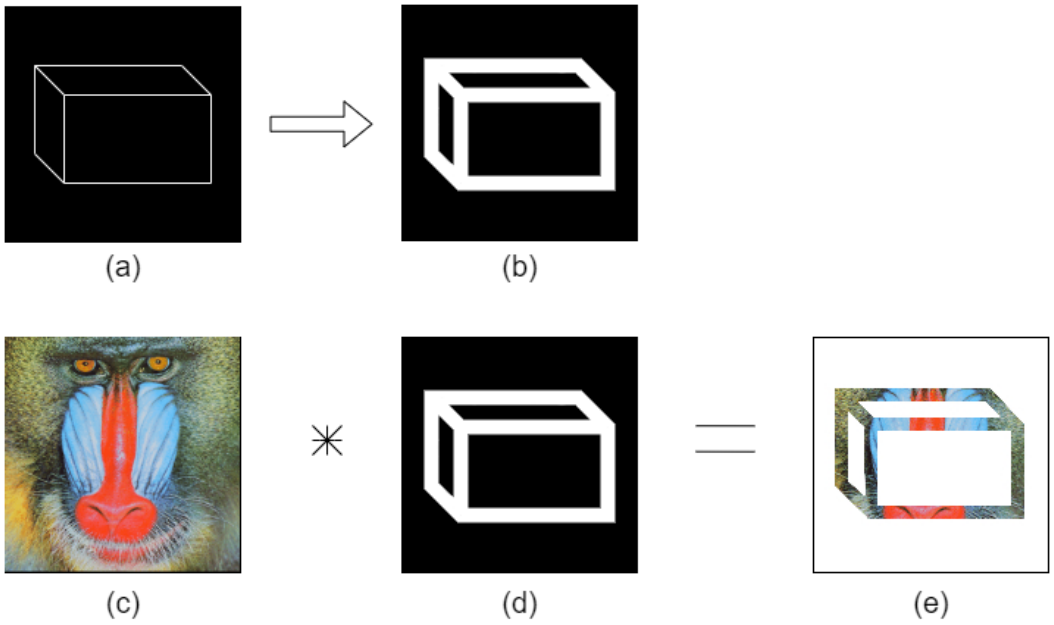
A etapa de casamento se divide em duas principais: uma busca por imperfeições e itens sobressalentes na cena em regiões longe das bordas; e um casamento próximo às regiões de borda, onde são buscadas correlações na imagem. Cada uma destas etapas gera um relatório de casamento que é representado por uma imagem. O primeiro relatório demonstra as regiões consideradas defeituosas em destaque. São considerados defeitos: rachaduras, objetos que estejam ocludindo a linha de visada até o conglomerado, etc. O segundo relatório prevê uma imagem que destaque as peças nas regiões esperadas e o nível de confiança sobre uma determinada peça estar sob este local.

3.3.1 Definição de região de análise

Para alcançar uma melhor definição do que é considerado região de borda ou centro da peça, foi utilizada uma abordagem similar à apresentada por (SEO et al., 2014). Ela determina uma região perpendicular de análise próxima às bordas do modelo para selecionar qual a região a ser segmentada e, posteriormente, usada na realização do casamento. Inicialmente, a renderização passa por um processo de filtragem e binarização como descrito na subseção 3.2.2. Essa imagem torna-se referência para um processo de segmentação da região de interesse na imagem da cena. Utilizando um método de busca de contornos no modelo, são agrupados todos os pontos de contorno em uma lista e posteriormente são expandidas as fronteiras de cada ponto da lista de contornos em um raio fixo para todos os pontos.

Essa máscara torna-se o modelo de segmentação de uma imagem de mesmo tamanho, em uma proporção 1 para 1. Com isto, ela é usada para segmentar as imagens retiradas da cena. Como a renderização está com o mesmo tamanho e posicionamento da peça a ser reconhecida na cena, são esperadas regiões de bordas bem definidas para todo o conglomerado. Um exemplo de como funciona o processo de criação da máscara e segmentação da imagem de cena pode ser visualizado na Figura 9.

Figura 9 – Criação de máscara e segmentação de uma imagem. (a) Imagem de contornos extraídos; (b) Máscara criada com pixels de contornos expandidos para 10 pixels; (c) Imagem alvo à ser segmentada; (d) Máscara a ser aplicada pixels pretos representam valor 0 e brancos 1; (e) Imagem segmentada resultante.



Fonte: o autor.

3.3.2 Detecção de incongruências fora da região de borda

Os modelos sem textura podem ser criados para definir peças sólidas ou mesmo moldes de outras peças que são construídas através de processos de manufatura aditiva, como é o exemplo das matrizes de ferramentaria próprias para estamparia de metais. Neste exemplo, ranhuras são um problema comum e que precisa ser evitado na construção dessas peças. Visando uma forma de identificação deste tipo de incongruência, além de uma forma rápida de eliminação de itens indesejados que podem constar em regiões distantes das bordas um relatório prévio é criado.

O relatório é constituído de uma imagem que apresenta possíveis defeitos presentes na cena real comparados à cena idealizada representada pela renderização, ou seja, rachaduras ou itens sobressalentes que não fazem parte da cena. Este relatório pode guiar o processo de montagem, levando à captura de mais imagens, até que seja considerada uma região livre de componentes indesejados.

Para construção do relatório, a imagem da cena é processada utilizando-se um algoritmo de detecção de bordas. Caso a imagem apresente muitos detalhes e nitidez demasiada, é aconselhável um pré-processamento com um filtro passa-baixa para suavização da imagem; este

procedimento pode aumentar a eficiência dos filtros de borda. Durante o processamento das imagens de cena, diferentes filtros demonstraram ser mais apropriados para determinados contextos de cena. Para o tratamento de cenas com grande quantidade de objetos metálicos que apresentam alta especularidade, o algoritmo (XIE; TU, 2015) demonstrou melhores resultados. O excesso de iluminação da cena, associada à especularidade do material dos itens buscados tem um melhor tratamento pela rede neural em comparação ao apresentado pelo Canny. Um dos principais fatores atribuídos a isto é a necessidade de definição dos limiares do filtro, como explicado na subseção 2.2.1.

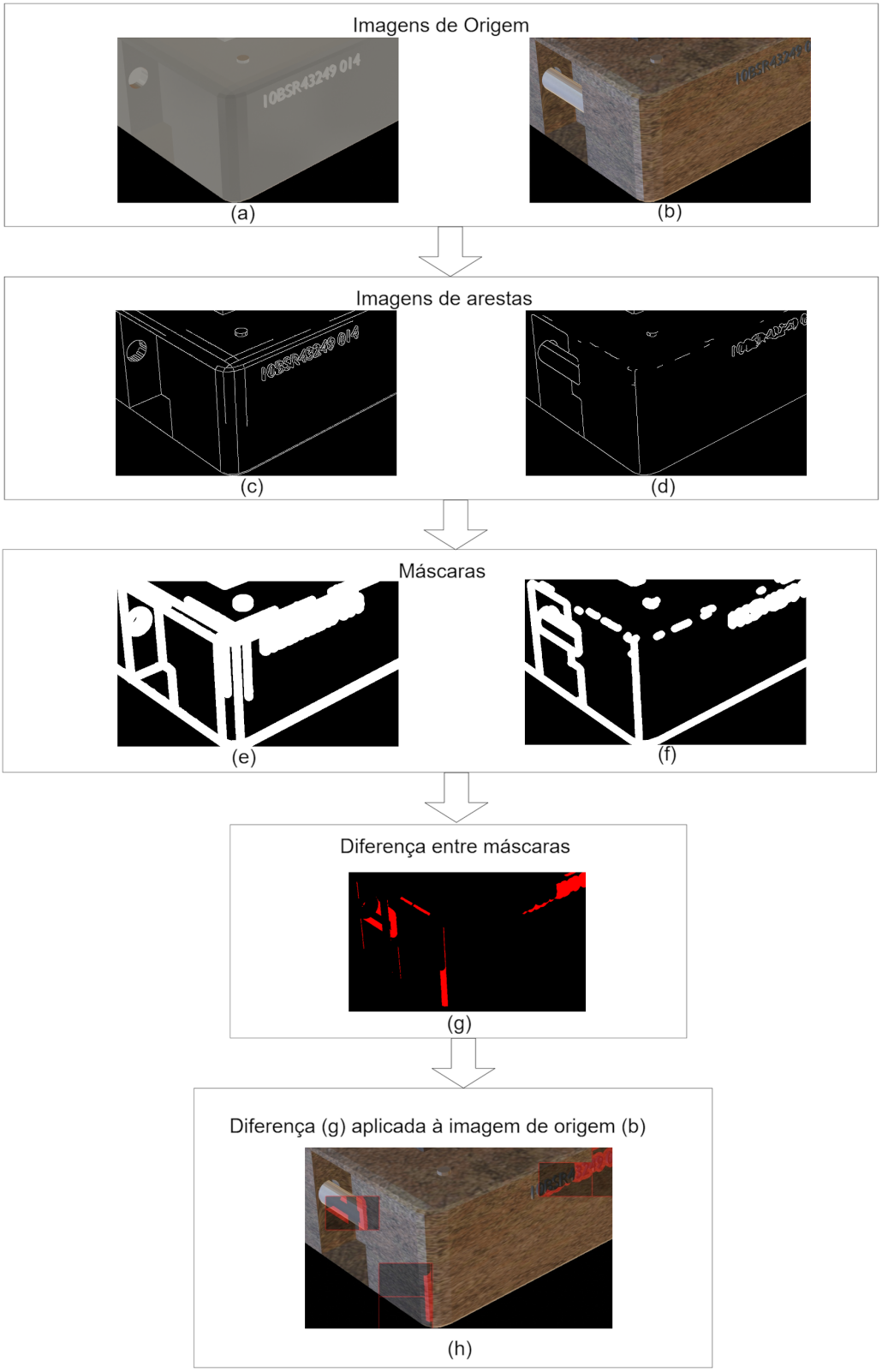
Em posse da imagem de bordas da cena, ela passa pelo mesmo processo de detecção de contornos e posterior criação de um círculo, com raio fixado por imagem, para cada ponto do contorno como descrito na subseção 3.3.1. Com isto, é possível destacar na imagem as bordas e saliências, como rachaduras, rasuras inesperadas, além dos objetos contidos que não façam parte da cena. Um exemplo deste processo pode ser observado na Figura 10. Com as duas imagens criadas, a máscara (Figura 10e) e a imagem das bordas da cena realçadas (Figura 10f), estas são então subtraídas, formando uma terceira imagem a partir da diferença (Figura 10g). Esta imagem resultante da subtração forma então um guia para realçar todas as incongruências da cena que estão longe das regiões de borda.

Como pode ser observado na Figura 10, visando demonstrar com maior clareza quais as regiões relevantes, um posterior tratamento é realizado. A imagem é dividida em quadrantes. Dentro de cada quadrante é verificada a quantidade de pixels diferentes de preto na imagem de diferença (Figura 10g). Caso a proporção entre pixels negros e coloridos apresente valor acima de um determinado limiar, esse quadrante é selecionado para ser destacado na figura final do relatório (Figura 10h), com as características que foram captadas na imagem de diferença das máscaras.

3.3.3 Análise de Proximidade as arestas do modelo

Para a análise da região próxima aos contornos das peças, é utilizado um método semelhante ao apresentado em (ARMSTRONG; ZISSERMAN, 1995) como (CHOI; CHRISTENSEN, 2010), (DRUMMOND; CIPOLLA, 2002), (SEO et al., 2014) entre outros, que consiste na busca linear a partir de pontos dispersos pelas arestas projetadas pelos contornos dos modelos, os pontos de controle. De forma mais específica, é criada uma região de busca similar à proposta por (SEO et al., 2014): neste método, é construída uma estrutura de busca que consiste em

Figura 10 – Exemplo de detecção de incongruências fora da região de borda.



Fonte: o autor.

uma imagem, cujas altura é a quantidade de pontos de controle e a sua largura corresponde a parcela de pixels da imagem de cena que se deseja analisar. A Figura 11 apresenta a retirada

de uma parcela de 3 pixels a partir de um ponto de controle central. No exemplo demonstrado pela imagem, caso fossem utilizados 100 pontos de controle a resolução da imagem seria de 100x7. Onde a coluna central de pixels da imagem seria correspondente aos pontos de controle e as colunas de valor mais baixo que a central representariam o que o autor do estudo (SEO et al., 2014) considera ser região externa da peça, de forma análoga, os pixels correspondentes às colunas de valores mais altos, seriam relativos à porção interna da peça buscada.

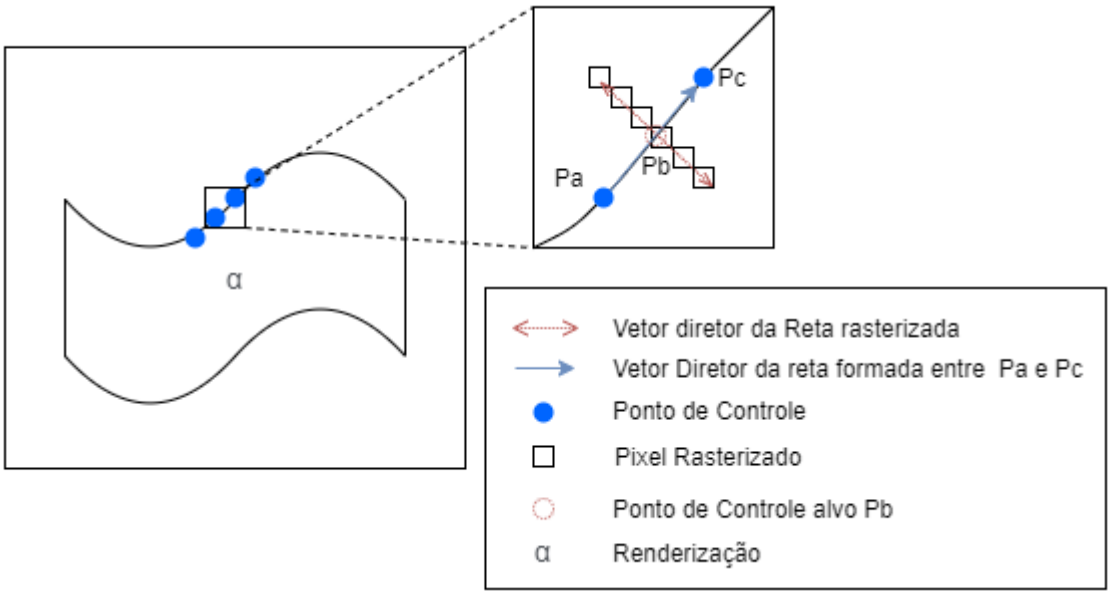
Almejando definir quais os pontos projetados que serão utilizados, os modelos são renderizados um a um, considerando apenas a sua parcela visível na cena, ou seja, já desconsideradas a parcela dos modelos que apresentaram oclusão parcial e as que apresentarem oclusão total como descrito na subseção 3.2.3. Para os conglomerados, notou-se que, para esta etapa de casamento, é mais adequado o uso de apenas iluminação ambiente para renderizar as peças com posterior extração apenas dos contornos. Este método permite um melhor aproveitamento para as cenas que apresentam muitos componentes e acelera o processo de casamento. O uso de bordas para realização do casamento pode levar a um falso negativo, devido à quantidade de características que cada item do conglomerado tem e a incapacidade dos detectores de borda conseguirem captar todas as nuances de cada imagem.

Para criação da estrutura de busca linear, foi utilizada uma alteração da heurística apresentada em (SEO et al., 2014). Para descobrir de forma facilitada quais os vetores diretores que guiarão o processo de extração dos pixels da cena. Após a retirada de todos os contornos da peça renderizada são adicionados em uma lista circular respeitando sua consecutividade. No exemplo apresentado na Figura 11, a consecutividade seria respeitada caso, o ponto Pa fosse ligado a Pb, Pb por sua vez seria ligado a Pa e Pc e este por sua vez, seria ligado a Pb e a um ponto externo ao detalhe.

Esta lista circular serve como um guia para o processo de criação da estrutura de busca, onde, para cada nó, são usadas as informações de dois nós em posições posteriores e anteriores para extração de um segmento de reta de onde é extraída a informação da direção que corresponde ao nó central buscado. No exemplo apresentado, é usado o Vetor diretor obtido pela diferença entre os potos Pa e Pc estes são encontrados acessando posições adjacentes ao contorno, uma vez que foi respeitada a consecutividade durante a criação. Isto possibilita uma suavização da aquisição da região que corresponde à fronteira, evitando que sejam adquiridos pontos diretamente retirados do próprio contorno do item buscado para a região de busca.

Uma vez criada a estrutura de busca, é possível criar uma imagem com os pixels extraídos da imagem de cena. Isto viabiliza a utilização de filtros e processamentos específicos para esse

Figura 11 – Seleção de pixels para armazenamento na estrutura de busca.



Fonte: o autor.

tipo de estrutura, como foi abordado em (SEO et al., 2014). Além disto, permite de forma simplificada a aquisição da distância dos centroides das regiões de saliência, computadas pós processamento da imagem de cena.

Para processamento da imagem de cena, foram utilizados dois métodos de filtragem principais como discutido no subseção 3.2.2. A estrutura de busca permite uma rápida aquisição de um vetor de erros a partir da imagem de bordas da cena. Para a construção, a partir do centro de cada linha da estrutura de busca, são verificados os dois sentidos interno e externo à procura de uma região de destaque e armazenada a menor distancia computada entre os sentidos. Se a estrutura de busca foi criada a partir de uma imagem de cena já binarizada, é possível buscar o primeiro pixel não negro a partir do centroide. Com isto, é então construído um *array* com as informações relativas a quais pixels e qual erro associado a cada pixel selecionado. Caso não sejam encontrados pixels em destaque ao longo das duas direções aquele pixel é marcado como distancia máxima e posteriormente indicará uma incongruência como explicado na subseção 3.3.4.

Para acelerar o processo de construção do *array*, são selecionados apenas uma fração dos pixels do contorno espaçados entre si com a mesma distância. A quantidade de pixels do contorno influencia na tomada de pontos, e consequentemente no tamanho do vetor de erro. Porém, aconselha-se a distancia mínima de 10 entre pixels escolhidos da estrutura. Pois, pontos escolhidos com um espaçamento menor podem aumentar muito a quantidade de pixels

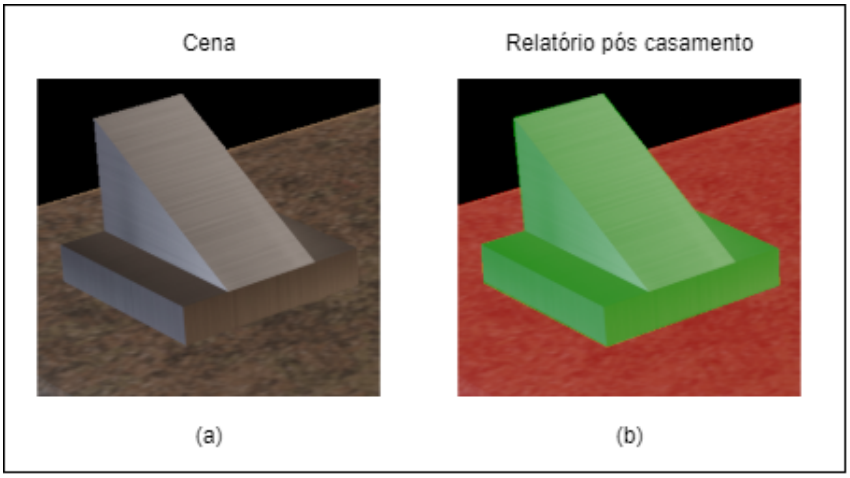
analisados, aumentando a complexidade do casamento.

3.3.4 Resultado do casamento

Em posse do vetor de erros dos contornos de cada peça, é possível então inferir um percentual de corretude de cada uma na cena, baseando-se no total da quantidade de cada um dos contornos que apresenta. Um relatório é escrito para cada modelo buscado com seu grau de precisão, informando a porcentagem de cada modelo, baseando-se na quantidade de pixels que apresentam distância mínima baseada no limiar da quantidade de pixels que foram usados para realizar a criação da estrutura de busca.

Em posse da porcentagem de cada um dos modelos é então renderizada uma imagem apenas com iluminação ambiente de cada modelo buscado, com uma variação de cores do vermelho (0%) até o verde (100%). É feita então uma sobreposição da renderização sobre a imagem cena, com um grau de transparência. Este relatório indica o grau de corretude de cada uma das peças de forma visual e permitindo correções de posicionamento de cada item. A Figura 12 exemplifica o relatório criado. A figura apresenta o mesmo recorte de uma imagem sintetizada com características similares ao do conglomerado real na forma de relatório. Observa-se que na imagem, apenas a ferramenta destacada em verde teve seu casamento considerado de sucesso, ao seu redor, a ferramenta que se apresenta em vermelho foi considerada fora de posição.

Figura 12 – Relatório de casamento de ferramentas do conglomerado. A peça (a) é considerada encontrada enquanto ao fundo, pode ser observada a peça (b) que serve de base de apoio para a peça (a) e é considerada não encontrada, devido a erros estruturais que nãos e destacam no recorte.



Fonte: o autor.

3.4 CONSIDERAÇÕES FINAIS

Ao longo deste capítulo foi apresentado um *framework* capaz de realizar o casamento de objetos tridimensionais complexos em imagens obtidas de uma câmera monocular. Tal estudo apresenta desde os passos iniciais de configuração da cena que se deseja obter a foto até a geração de relatórios de casamento com o percentual de confiança de cada item do conglomerado na cena, além de um relatório de incongruências fora das regiões de borda. Este último, é representado por uma imagem com destaques sobre os defeitos dos componentes, e é capaz de capturar ranhuras, imperfeições entre outros tipos de defeitos visuais. Para gerar tais relatórios, são apresentadas ainda, formas de uso de informações confiáveis, como marcadores fiduciais, para aquisição da posição inicial do conglomerado buscado. È demonstrado também a forma de criação dos modelos canônicos, obtidos a partir de renderizações dos arquivos tridimensionais de entrada e como realizar o comparativo desses modelos com as características obtidas da imagem de cena. Tal *framework* passou por experimentos descritos no capítulo a seguir.

4 EXPERIMENTOS

Durante a elaboração deste estudo, foram testadas diferentes configurações que pudessem auxiliar a busca do conglomerado, considerando as condições de câmera e cena disponíveis. O projeto foi desenvolvido em parceria com uma montadora de veículos automotivos com fábrica no Brasil e passou por diversas etapas até sua concepção. Em um momento inicial, foram testadas as configurações de busca de peças em metal a partir de imagens digitais. A evolução do projeto guiou a criação de um ambiente de testes, utilizando uma impressão tridimensional em escala reduzida de componentes reais de uma matriz de funilaria automotiva. Por fim, foram testados os algoritmos em ambiente fabril, utilizando imagens reais de uma matriz.

Visando proteger a propriedade intelectual da montadora, foram renderizadas imagens similares às da fábrica onde é possível observar o mesmo tipo de característica que gerou os resultados. Para estas renderizações das imagens de cena, foram modelados sólidos tridimensionais utilizando o software Blender (BLENDER.ORG, 2020). A iluminação da cena foi simulada utilizando uma imagem do ambiente fabril como plano de fundo, omitida nas imagens que constam neste documento. Esta foi uma Além disto, o material componente dos objetos foi trabalhado com texturas similares aos componentes reais encontrados na fábrica. Ainda que as imagens apresentadas ao longo deste capítulo apresentem simulações dos resultados apresentados nos relatórios originais, os dados compilados apresentados no Capítulo 5 se referem apenas à resultados reais, obtidos diretamente dos experimentos realizados.

Para todas as implementações de algoritmos presentes neste estudo, foi utilizada a linguagem de programação C++ disponível em (CPLUSPLUS, 2020), a menos que seja explicitado o contrário. Para a replicação dos algoritmos, lançou-se mão ainda do uso da biblioteca gratuita para visão computacional OpenCV apresentada em (OPENCV, 2020). Todos os testes foram realizados num computador com Sistema Operacional Windows 10, 16GB de memória RAM e processador Intel Core i7 2,8 GHz. Os algoritmos foram compilados utilizando-se a versão 15 do MSVC. As imagens de cena foram adquiridas a partir de uma câmera de foco fixo com abertura focal de 4mm. Este capítulo utiliza os termos específicos apresentados no Capítulo 3.

4.1 PROVA DE CONCEITO

As primeiras análises do problema levaram ao estudo de modelos de detecção de objetos em imagens através de uso de *keypoints*. O modelo apresentado por (LOWE, 2004) demonstrava boa acurácia, enquanto o estudo apresentado em (BAY et al., 2008) demonstrava acurácia similar e ainda maior eficiência em termos de custo computacional. Com isto, foi desenvolvido um sistema que utilizava o algoritmo de BAY et al. para a descoberta de itens em uma cena. O sistema foi construído, recebendo como entrada uma imagem dos itens a serem buscados e uma imagem de cena onde se pressupunha que o objeto estivesse. Com isto, era construído um relatório, onde constava o item buscado com sua projeção relativa.

Uma vez validada a busca de itens metálicos a partir de suas imagens, o estudo teve o foco voltado ao problema enfrentado dentro da montadora, como explicitado no seção 1.1. No problema descrito, é necessário verificar, em uma imagem RGB, a montagem e corretude de uma matriz de estamparia criada através da união de diversas ferramentas moldadas a partir de um modelo tridimensional.

Isto posto, foram então pesquisados estudos que permitissem o casamento entre modelos tridimensionais em imagens RGB. Visando aumentar o grau de complexidade da solução, assim como uma aproximação do desafio a ser enfrentado no ambiente fabril foi criado um ambiente simulado e reduzido, utilizando-se uma porção de componentes de uma matriz real. Para a simulação foi construída uma reprodução miniaturizada em impressora 3D de uma série de arquivos CAD pertencentes a um molde de uma ferramentaria. O material de construção utilizado foi polímero e a maquete representou uma redução de cerca de 8 vezes o tamanho real dos componentes.

4.2 RECONHECIMENTO DE CONGLOMERADO EM AMBIENTE SIMULADO POR MAQUETE

O levantamento bibliográfico levou à escolha do método apresentado por (CHOI; CHRISTENSEN, 2010) para realização do reconhecimento dos objetos. Como descrito na subseção 2.1.3, o método apresentado por CHOI; CHRISTENSEN realiza o casamento das intituladas *strong edges* com segmentos de reta obtidos a partir do processamento da imagem com o detector de bordas Canny e uma posterior aplicação da Transformada de Hough (DUDA; HART, 1972). O casamento das características serviu como guia para a realização do reconhecimento das peças.

Entretanto, o algoritmo apresentado para aquisição das arestas fortes apresentou elevado custo computacional na sua etapa de pré-processamento. Os modelos tridimensionais CAD selecionados para a impressão demandaram algumas dezenas de horas em pré-processamento. Este modelo com arestas fortes é ainda processado em tempo de execução através da utilização de um grafo de árvore BSP. A BSP é utilizada para tratamento de oclusão de arestas, a partir do ângulo de visada para a cena. Utilizando o modelo proposto em CHOI; CHRISTENSEN, apenas as arestas completamente visíveis são utilizadas para realização da detecção dos objetos.

O estudo apresentado por CHOI; CHRISTENSEN busca realizar a detecção do modelo CAD em vídeo, ou seja, informar a posição com 6 graus de liberdade de um modelo ao longe de diferentes quadros. Seu algoritmo é uma evolução do proposto em (DRUMMOND; CIPOLLA, 2002), com um diferencial que o SURF é utilizado para uma reinicialização rápida do posicionamento do objeto na cena para o caso de ocorrer uma grande perda da localização do objeto entre quadros do vídeo. Esse tipo de problema é comum de ocorrer por motivos de oclusão, rápida aceleração do objeto ou demais fatores externos na cena durante a filmagem.

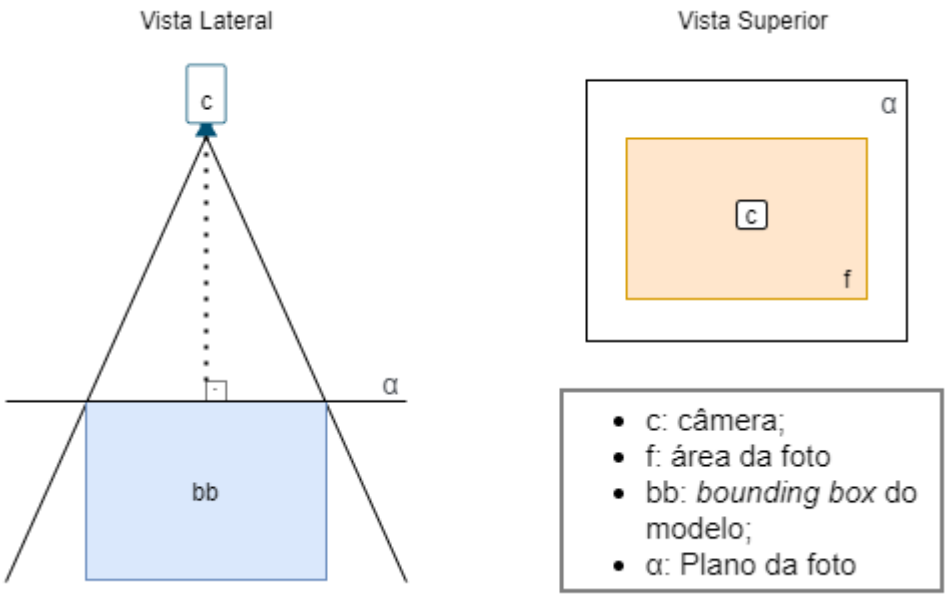
Como o foco desta pesquisa é o reconhecimento dos modelos CAD complexos a partir de uma única imagem RGB, a configuração de testes foi montada visando a retirada de uma imagem da maquete. Para captura desta imagem, a câmera foi posicionada com enquadramento exato do *bounding box* dos itens buscados, a seção 4.2 demonstra o modelo de posicionamento da câmera para retirada das imagens da cena. A câmera foi calibrada utilizando um *chessboard* como descrito na **??** e com erro de reprojeção médio mantido abaixo de 2.

O refino dos modelos tridimensionais, associados ao uso da árvore BSP demonstrou resultados parcialmente satisfatórios. O algoritmo não demonstrou tratamento particular para arestas parcialmente oclusas. Visando acelerar o processo de aquisição das características buscadas na cena, foi criado então um algoritmo para a extração de um modelo a partir da renderização do conglomerado. Neste estágio, o modelo renderizado é processado, utilizando a transformada de Hough e os segmentos de reta são armazenados em uma lista para cada modelo renderizado do conglomerado.

4.2.1 Métodos baseados em *templates*

Para realizar o casamento das características, seguiu-se o padrão de métodos similares ao (ARMSTRONG; ZISSERMAN, 1995), também apresentado por CHOI; CHRISTENSEN. Para tais métodos, é verificada a distância Euclidiana em pixels entre pontos de controle obtidos a partir

Figura 13 – Posicionamento da câmera para retirada de fotos da maquete.



Fonte: o autor.

da projeção das arestas do modelo CAD até os pontos das arestas da imagem da cena que são resultados da busca linear dos segmentos de reta próximos, como explicado na subseção 2.1.3. O vetor com todos os erros é então armazenado e utilizado associado a métodos de PnP para técnicas de rastreio. Como o método proposto tem como foco o reconhecimento, foram armazenados os segmentos de reta próximos obtidos a partir da transformada de Hough.

Em posse das arestas próximas à projeção do modelo, foi criada uma imagem onde constavam apenas as retas próximas para verificação da corretude do modelo. Foi então observado que o método levava a captura de muitas arestas que eram falso positivos, ou seja, ranhuras na peça impressa capturada pela imagem da cena ou algum outro tipo de imperfeição na imagem captado pelo detector de bordas. Em determinados locais, era possível visualizar as arestas criadas pelo fio utilizado pela impressora 3D para construir a peça.

A variação de parâmetros de entrada da transformada de Hough, assim como do Canny, gera menos ou mais resultados falso positivos. Dentre os segmentos de reta visualizados como falso positivos, pôde-se verificar a presença de segmentos de reta com coeficiente angular com diferença de dezenas de graus do apresentado pela aresta projetada. Visando uma maior confiabilidade do casamento, as arestas que apresentem valores tão altos são desconsideradas.

A partir dos segmentos de reta resultantes desta etapa de casamento, a imagem de cena é então dividida em retângulos, baseando-se no *bounding box* do modelo projetado. Para cada

retângulo mapeado, foi então verificada a presença de arestas adequadas em seu interior. Em caso afirmativo, o retângulo é então sinalizado como um positivo para realização do casamento. Baseando-se na quantidade total de retângulos com casamento adequado, acima de 80%, a peça é considerara presente na região da cena.

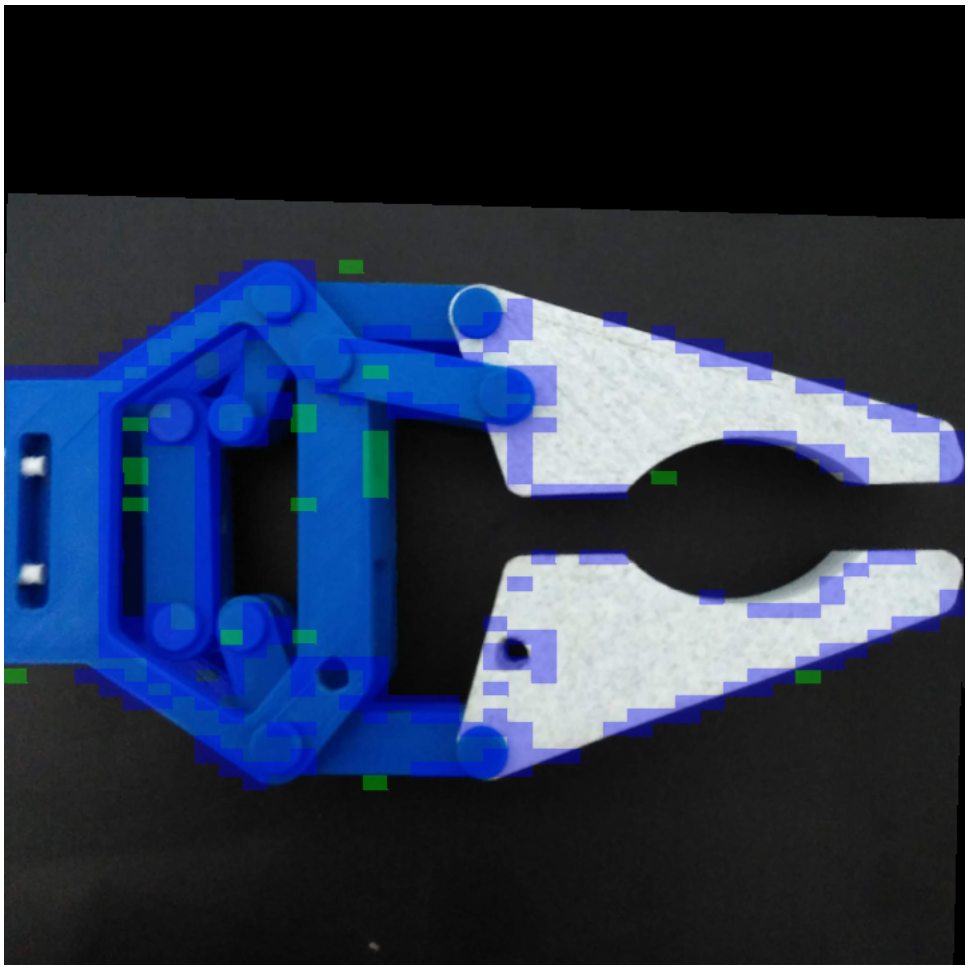
Os experimentos apresentaram bons resultados para o ambiente de testes montado, entretanto, tornou-se evidente a ineficácia para realização do casamento devido à dificuldade de parametrização adequada do algoritmo de bordas e da Transformada de Hough. Diferentes regiões da imagem requeriam diferentes parametrizações do filtro, gerando assim incongruências para os resultados esperados. Como forma de contornar tal problema, fez-se providente a verificação dupla de cada retângulo, seguindo diferentes limiares para a realização da filtragem usando o mesmo algoritmo, ou utilizando filtros de bordas diferentes e considerados mais robustos. Para o processo de filtragem, foi implementado o algoritmo detector de bordas apresentado em (Costa; Mello; Santos, 2013). Os resultados do algoritmo para um modelo tridimensional genérico, podem ser visualizados na Figura 14. Na figura, observa-se um destaque em verde para os quadrantes que só foram reconhecidos após o uso do filtro DifFocus. Como resultado, foi possível identificar a presença das ferramentas da maquete, além de indicar regiões com modelos tridimensionais ausentes.

Visando incrementos aos algoritmos de casamento similares ao (ARMSTRONG; ZISSERMAN, 1995) e buscando uma forma de tornar o sistema mais robusto, foram estudados algoritmos com base em inteligência artificial que pudessem ser usados paralelamente aos algoritmos já desenvolvidos, aumentando a eficácia do sistema.

4.2.2 Casamento utilizando técnicas de Inteligência Artificial

Os estudos mais comuns de busca de itens, utilizando visão computacional, são voltados para atividades como detecção e classificação como explicado no subseção 2.1.5. A maioria dos modelos que utiliza como entrada modelos tridimensionais, tende a fazer o casamento diretamente com a nuvens de pontos ou com imagens RGB-D. Com isto, foram considerados os estudos que pudessem realizar atividades de detecção e classificação ou reconhecimento diretamente em imagens RGB. Como apresentado na subseção 2.1.5, diversos estudos são voltados para o uso de redes neurais convolucionais para realizar tais atividades, como é o exemplo da rede YOLO v3 apresentada em (REDMON; FARHADI, 2018). Uma vez que as redes neurais apresentam bons modelos de detecção a partir de um treinamento apurado, foi validada

Figura 14 – Relatório de reconhecimento de garra impressa em impressora tridimensional.



Fonte: o autor.

a ideia que era possível realizar a busca dos objetos sem textura, treinando-se a rede apenas com imagens renderizadas das ferramentas.

Para realizar o treinamento da rede neural, foram utilizadas imagens renderizadas do conglomerado com diferentes variações de iluminação, diferentes fontes de luz e diferentes configurações de posicionamento da câmera. Foram geradas renderizações com o vetor direcional do foco da câmera voltado para a peça e com diferentes configurações de posicionamento da câmera. De forma similar ao proposto em (Liebelt; Schmid; Schertler, 2008) variando o posicionamento do olho da câmera de forma longitudinal e paralela ao redor da peça, possibilitando, assim, a captura de diferentes visões para a renderização. Associado a isto, foram utilizadas imagens capturadas aleatoriamente do ambiente onde a maquete poderia estar contida. Cerca de 20 imagens serviram como diferentes planos de fundo para cada uma das renderizações. Também foram utilizadas técnicas de *data augmentation* para variar a dimensão da renderização, rotações em torno do próprio eixo da renderização além de oclusão parcial. Em posse

desta base de imagens foi realizado o treinamento do algoritmo de REDMON; FARHADI.

Após o treinamento, foram realizados testes de detecção da maquete diretamente em vídeo, utilizando-se uma câmera de foco automático e resolução 1080p. O algoritmo apresentou bons resultados detectando a maquete com cerca de 70% de confiança em cerca de 0,5 quadros por segundo. Com os resultados alcançados foi possível fazer uma validação cruzada do casamento obtido por uma ou mais imagens entre o método de *template* e de Inteligência Artificial (IA), o Yolo.

Com os resultados combinados dos algoritmos de reconhecimento, utilizando arestas e a rede neural treinada com as imagens renderizadas, os experimentos evoluíram para execução diretamente em fábrica, com imagens reais retiradas diretamente do setor de ferramentaria automotiva de uma fábrica montadora de carros.

4.3 AMBIENTE FABRIL

Visto que os métodos testados e desenvolvidos em ambientes simulados apresentaram resultados satisfatórios, os algoritmos foram então submetidos a um ambiente de testes real. Os testes foram realizados em maio de 2019 em uma montadora de veículos automotores. Para realização dos testes, foram utilizados uma matriz de estamparia em estado de finalização com cerca de 72 ferramentas componentes. A matriz apresenta uma base cujas dimensões são 4,50 m de comprimento e 2,30 m de largura. Os componentes todos são metálicos, alguns apresentando parte de sua superfície polida, consequentemente especular.

Para viabilização do uso dos algoritmos desenvolvidos na etapa de testes em maquete, foi necessário realizar ajustes relacionados à configuração da montagem da cena. O modelo esperado de utilização da câmera conforme descrito na seção 4.2 mostrou-se inviável, devido à impossibilidade de aquisição aérea das imagens. Com isto, foram pesquisados modelos de viabilização de posicionamento com maior liberdade para o posicionamento da câmera.

4.3.1 Montagem de cena

O uso de marcadores fiduciais para aquisição do posicionamento do olho da câmera foi fundamental. Entre os estudos levantados, o trabalho apresentado em (ROMERO-RAMIREZ; MUñOZ-SALINAS; MEDINA-CARNICER, 2018) apresentou fácil integração com a solução em desenvolvimento. Os primeiros testes consistiram no uso de um marcador único posicionado em

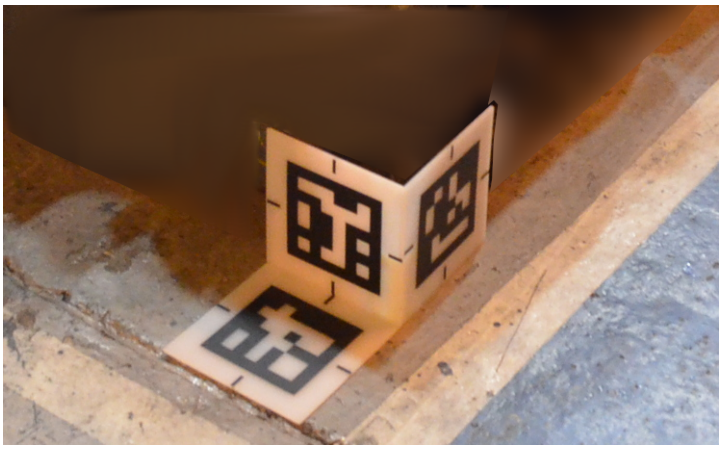
uma região de esquina, visando que o centro do marcador pudesse coincidir com o limite do *bounding box* da matriz. O posicionamento do marcador em local pré-determinado permitiu a inicialização da cena de outros pontos de vista, viabilizando o posicionamento da câmera em um ponto de vista mais adequado. Como o posicionamento do marcador é conhecido, antes de realizar a renderização do modelo, é possível assumir o marcador como o centro do sistema tridimensional e posteriormente realizar as transformações necessárias para alinhamento adequado entre modelo projetado e a matriz real da cena. Contudo, o uso de um único marcador posicionado na esquina gerou problemas relacionados à angulação presente no marcador. Ainda que o marcador estivesse ligeiramente desalinhado com a matriz, isto gerou incongruência entre a projeção do modelo renderizado e da imagem da cena. Outro problema comum durante os testes aconteceu devido ao marcador ter sido impresso em material frágil. Isto impedia o marcador de ficar ereto, resistindo a intempéries do ambiente. Com intuito de sanar tais problemas, foram utilizados múltiplos marcadores rígidos, como descrito na subseção 3.1.2.

Por conseguinte, foram construídos marcadores rígidos que pudessem estar visíveis, sem causar oclusão de forma danosa para o casamento. Os marcadores foram construídos em acrílico fosco, diminuindo a interferência da iluminação ambiente que pode causar problemas durante o processo de reconhecimento do marcador.

Embora os testes fossem conduzidos visando um reconhecimento em foto única, o projeto do marcador possibilita que, ao serem usados quatro marcadores posicionados ao redor da matriz, condizentes com o seu *bounding box*, fosse possível sempre visualizar ao menos quatro marcadores fiduciais impressos nas faces dos marcadores rígidos, possibilitando o uso de métodos PnP. Os marcadores foram construídos representando um quadrado com 20 cm de lado e 5 mm de espessura. Os marcadores fiduciais foram posicionados de forma que os centroides do marcador rígido e fiducial impressos casassem. Além disto, há uma margem no marcador para facilitar o seu reconhecimento. A Figura 15 representa o modelo de disposição dos marcadores na cena da fábrica. Aspirando aumentar a quantidade de informações confiáveis na imagem da cena, foram adicionados ainda quatro segmentos de reta nas laterais dos marcadores. Os segmentos são componentes de duas retas mediatrizes concorrentes que cruzam o centro do marcador. Estes segmentos permitem uma verificação da eficiência da configuração da cena e algoritmos, através da verificação de sua projeção.

Como a matriz completa compreende uma região muito vasta, cerca de $9m^2$, a câmera utilizada foi posicionada de forma que pudesse enquadrar boa parte dos componentes, além dos marcadores posicionados no solo. A Figura 16 demonstra o posicionamento completo para

Figura 15 – Marcador posicionado em cena.



Fonte: o autor.

uma matriz dentro de uma fábrica, nela é possível observar a relação de tamanho das matrizes no ambiente, além de analisar a incidência de iluminação disforme e trânsito constante de pessoas. As matrizes estão dispostas em um armazém de *Try-out* que contêm fluxo constante de máquinas pesadas e pessoas. Uma vez fixado o foco, a câmera é submetida então a um processo de calibração.

Figura 16 – Matrizes posicionadas na região de *Try-out* da ferramentaria.

Fonte: adaptada de (GASPEC, 2009)

Após calibração, a câmera pode retornar ao pedestal e as imagens de cena podem então

ser capturadas. Durante os experimentos, foram utilizadas imagens com resolução 4096 x 2304 pixels, em formato *raw*, como explicado na subseção 3.1.1. Este formato foi adotado visando a diminuição da interferência que pode vir a ser causada por algoritmos de compressão de imagens.

4.3.2 Ferramentas e Modelo

Como exemplificado no Capítulo 1, o processo fabril para estamparia metálica, passa por modificações iterativas nas ferramentas que compõem uma matriz. Cada modificação necessária nas ferramentas é guiada pela adequação a um modelo tridimensional. No referido exemplo da montadora de automóveis, são modelos CAD que guiam o processo. Uma matriz completa pode conter de dezenas a centenas de componentes.

A Figura 17 apresenta diferentes características que podem ser encontradas nas matrizes de ferramentaria e que se repetiam também nas imagens obtidas na fábrica automotiva. A região mais externa da matriz, onde ocorre o contato com a chapa de metal que vem a se tornar a peça de funilaria após a prensagem, tende a ser polida e apresentar um alto grau de especularidade, como pode ser observado na Figura 17a. As regiões que não apresentam contato direto com a placa de metal podem ou não apresentar tal característica. As regiões e peças mais externas que servem de base de apoio para as demais peças são moldadas a partir de moldes de barro ou isopor e, portanto, podem apresentar imperfeições em sua superfície assim como (Figura 17b). A estrutura ainda conta com a presença de pistões que são responsáveis por suspender outras ferramentas. Durante o processo de prensagem, é comum também existirem ferramentas que se deslocam em movimentos calculados, com intuitos específicos, como cortar a chapa de metal, ou realizar alguma dobra. Em tais regiões, é comum a presença de lubrificantes, que podem variar entre secos, como o grafite, ou pastosos, como graxa; a Figura 17c demonstra um componente com depósito de tal produto. Os lubrificantes acabam por trazer novas caraterísticas visuais para as peças, como alterações de textura ou variações de suas cores. As ferramentas apresentam formas variadas além de diferentes tipos de simetria. Outra característica possível dentro do conglomerado é a simetria bilateral entre alguns de seus componentes. Neste tipo de configuração, duas ferramentas podem apresentar características semelhantes, mas em posicionamento espelhado a partir de um plano que bissecte o conglomerado.

Como os algoritmos para reconhecimento do conglomerado já levantados baseiam-se em

Figura 17 – Características das ferramentas componentes do conglomerado. (a) Exemplo de especularidade dos componentes que ficam em contato com a chapa de metal durante a prensagem; (b) Ranhuras na parte mais externa da ferramenta, feita para suportar outros componentes; (c) Os círculos negros são depósitos de grafite, usado como lubrificante seco no processo.

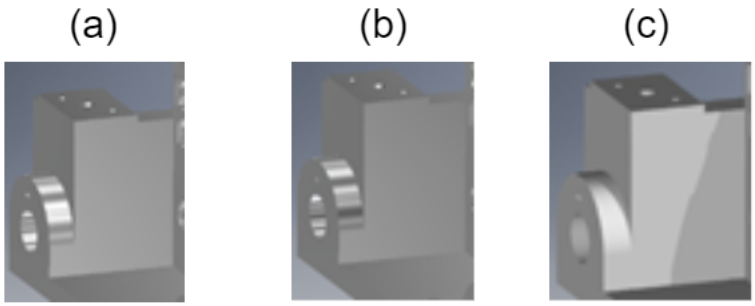
(a) (b) (c)

Fonte: o autor.

casamentos de *templates*, a escolha do *pixel shader* para a renderização dos modelos se faz pertinente. Os insumos principais para a renderização são arquivos *Standard Transformation Language* (STL) convertidos a partir de um modelo CAD proprietário. Em posse da lista de arquivos STL, as peças podem então ser renderizadas. Como explicado na subseção 3.2.2, diferentes algoritmos de iluminação podem gerar diferentes características nas ferramentas renderizadas. A Figura 18 demonstra a mesma região com três diferentes modelos de iluminação em uma renderização de uma das peças do conglomerado. Na imagem, pode-se observar que arestas são geradas a partir da especularidade considerada no material em todos os exemplos. Fatores como a quantidade e posicionamento dos pontos focais de luz influenciam na quantidade de arestas, como pode ser observado na região esquerda central das imagens Figura 18a e Figura 18b. Na renderização utilizando o algoritmo proposto por WHITTED demonstrada na Figura 18c pode-se observar uma maior quantidade de características que tendem a se apresentar na cena, como sombra e arestas mais detalhadas. Este tipo de informação causa variações nos casamentos das peças com as imagens de cena. No modelo ideal buscado, é possível encontrar uma renderização exata da cena, com uma relação direta entre a textura e iluminação da peça da cena com a renderização.

Em detrimento da construção de uma renderização com características mais aproximadas das reais, os algoritmos de renderização podem apresentar um elevado custo computacional. Os esquemas de renderização mais custosos, como é o exemplo do (WHITTED, 2005), demonstraram melhor desenvolvimento de características. Entretanto, o tempo de montagem de uma cena completa requer um elevado poder computacional associado a uma alta otimização para viabilizar seu uso. Enquanto os modelos de iluminação com diversos pontos de luz, aumenta-

Figura 18 – A influência da iluminação sobre a mesma região de uma ferramenta renderizada:(a) Um único ponto de luz; (b) Dois pontos de luz; (c) RayTracing.



Fonte: o autor.

ram a complexidade da escolha de qual o referencial de posicionamento. Logo, foi adotado um algoritmo de iluminação menos custoso para realização dos demais experimentos, conforme o proposto na subseção 3.2.1.

Geradas as imagens, é possível então submetê-las a um filtro de bordas, criando assim as arestas que são buscadas na imagem de cena. O detector de bordas de Canny (CANNY, 1986) demonstrou bons resultados para essa etapa devido a padronização da coloração na renderização, permitindo que o ajuste dos limiares para uma região da renderização fosse aproximado do ideal para toda ela. A Figura 19 exemplifica a aplicação do filtro em diferentes condições de iluminação todas com os mesmos valores de limiar definidos na entrada do algoritmo. Observa-se, na Figura 19c, uma maior quantidade de arestas encontradas. Na Figura 19a, destacam-se ainda uma maior quantidade de arestas na região superior destacada que na Figura 19b. Como ambas as cenas foram renderizadas com o algoritmo de Phong, apenas o posicionamento dos pontos de luz causou tal variação. Portanto, a adição de múltiplas fonte de luz além de adicionar complexidade ao algoritmo, aumentar a quantidade de entradas necessárias, ainda pode suprimir características consideradas fundamentais para o casamento. Um outro avanço que foi aplicado ao modelo é o fato de não precisar usar mais as linhas de Hough (DUDA; HART, 1972) para identificação das retas como havia sido proposto ainda em fase de testes em maquete na seção 4.2. Isto permitiu uma expansão do modelo de casamento, além de uma diminuição dos processos aplicados sobre o método. Permitindo assim, a diminuição da necessidade de seleção das variáveis da Transformada de Hough, como explicado a seguir.

Figura 19 – Resultados do Canny sobre mesma região de uma das ferramenta. (a) Fonte de Luz única; (b) Fonte de Luz dupla; (c) *RayTracing*.

Fonte: o autor.

4.3.3 Casamento

Para os testes relacionados ao casamento completo da matriz com os modelos tridimensionais de entrada, foram utilizados imagens obtidas da cena conforme explicitado na subseção 4.3.1. As imagens adquiridas foram submetidas a dois tipo de casamento, testes usando o casamento de *template* evoluídos a partir do modelo proposto na seção 4.2, além dos testes utilizando a rede neural apresentada em (REDMON; FARHADI, 2018) que apresentou bons resultados nos testes preliminares.

Para os testes realizados com o uso da rede Neural, foram utilizadas as renderizações do modelo conforme o apresentado na subseção 4.2.2. Foi escolhida uma parcela de ferramentas que compunham o conglomerado e que estavam parcialmente oclusas na cena. Utilizando-se imagens da fábrica como plano de fundo e as mesmas técnicas de *data augmentation*. A arquitetura treinada não conseguiu realizar a detecção dos modelos. Tentando encontrar uma forma de melhorar os resultados, ainda foram experimentados a criação de diferentes arquiteturas a partir de técnicas como o uso de imagens de bordas da renderização para realizar o casamento em uma imagem de bordas da cena e variação de cores para cada componente renderizados. Entretanto, a detecção apenas ocorria com 25% de margem de confiança, o que causava diversos falsos positivos na cena.

Em posse dos resultados negativos do uso da Rede Neural, o projeto evoluiu para o uso isolado do modelo de detecção de *templates*. O processo de reconhecimento utilizando *templates* proposto inicia-se desde a descoberta do posicionamento inicial da câmera, até o módulo de busca de correspondências entre a imagem da cena e o modelo. Com intuito de descobrir onde a câmera se encontra, foi montada a cena segundo a subseção 4.3.1. Conforme detalhado na

subseção 3.1.2, foram utilizados três métodos de PnP para a descoberta de onde se encontra o foco da câmera. O método de avaliação dos três algoritmos foi o uso da distância euclidiana entre os centroides dos marcadores e a reprojeção dos pontos a partir do esquema de câmera reportado pelo método. Os resultados comparativos dos módulos podem ser visualizados na Tabela 1.

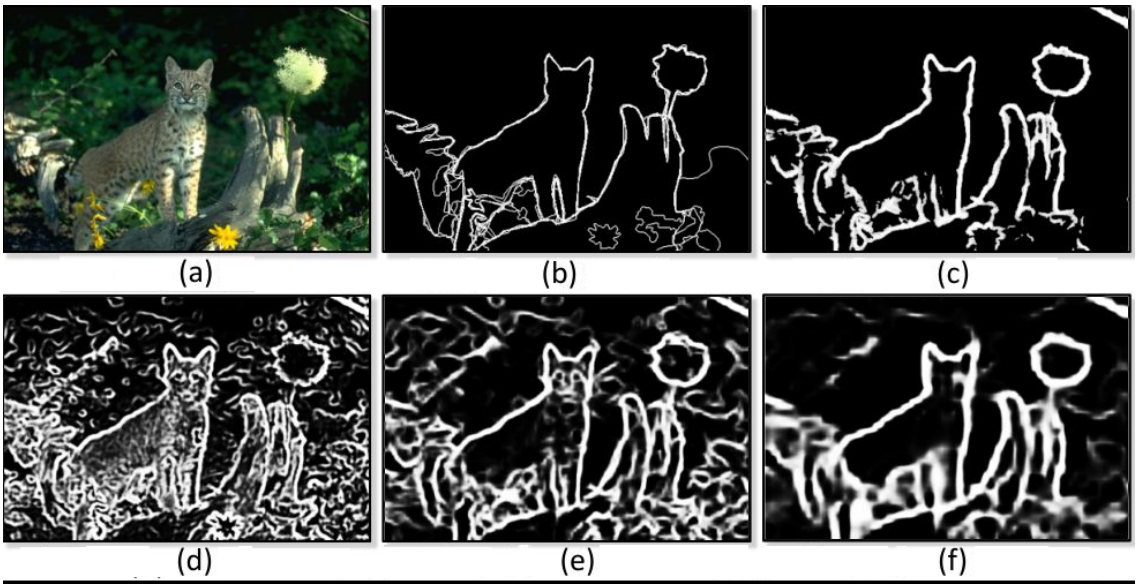
Para o processo de casamento em si foram considerados os ajustes apresentados em seção 3.3. Isto permitiu a criação de dois tipos de relatórios: um de inconformidades nas regiões longe das bordas; outro de casamento, onde constam os resultados de congruência entre a cena e cada ferramenta.

Um outro ganho do processo de segmentação da região de busca similar ao apresentado por (SEO et al., 2014) é que permitiu o uso da informação próxima as arestas de forma simplificada. No modelo prévio apresentado em seção 4.2 de divisão das regiões de busca por quadrantes, ocorria o fato do limiar do quadrante coincidir com uma região de borda. Isto poderia gerar um falso negativo, devido a aresta da cena se encontrar ligeiramente acima da representação do quadrante.

Isto posto, foi utilizado o modelo de reconhecimento do conglomerado proposto neste documento no Capítulo 3. Uma das etapas fundamentais para o reconhecimento da imagem da cena está no uso dos filtros de bordas. O uso de filtros para detecção e comparação das características demonstrou bons resultados em ambiente de testes. Entretanto, a parametrização utilizada nos filtros de bordas demonstrou uma dificuldade ainda maior para o cenário real. A especularidade associada aos demais fatores exemplificados na subseção 4.3.2 aumenta a dificuldade de uso de apenas dois parâmetros para toda a região da imagem. Com isto, foram estudados filtros que pudessem apresentar resultados mais robustos baseados no uso de redes neurais. Dentre os levantados, o (XIE; TU, 2015) foi escolhido. Como saída do processo de detecção de características do filtro, podem ser utilizadas diferentes camadas, que apresentam diferentes filtragens de características. A Figura 20 demonstra as camadas apresentadas para uma mesma imagem de análise a Figura 20a. As camadas mais baixas, como a Figura 20d e Figura 20e tendem a apresentar uma alta quantidade de características que podem ser mal entendidas no processo de casamento. Contudo, as camadas mais altas tendem a suprimir resultados de forma muito incisiva, observa-se que muitas características da imagem estão suprimidas mesmo na camada combinada. Com isto, tais características estão presentes no modelo podem aparecer em destaque na cena.

A partir das análises da saída do algoritmo, foi escolhida a terceira camada (Figura 20e) por

Figura 20 – Detalhe da aplicação do filtro de bordas HED. (a) Imagem original; (b) *Ground truth*; (c) Saída final combinada; (d) Saída camada lateral 2; (e) Saída Lateral 3; (f) Saída lateral 4.

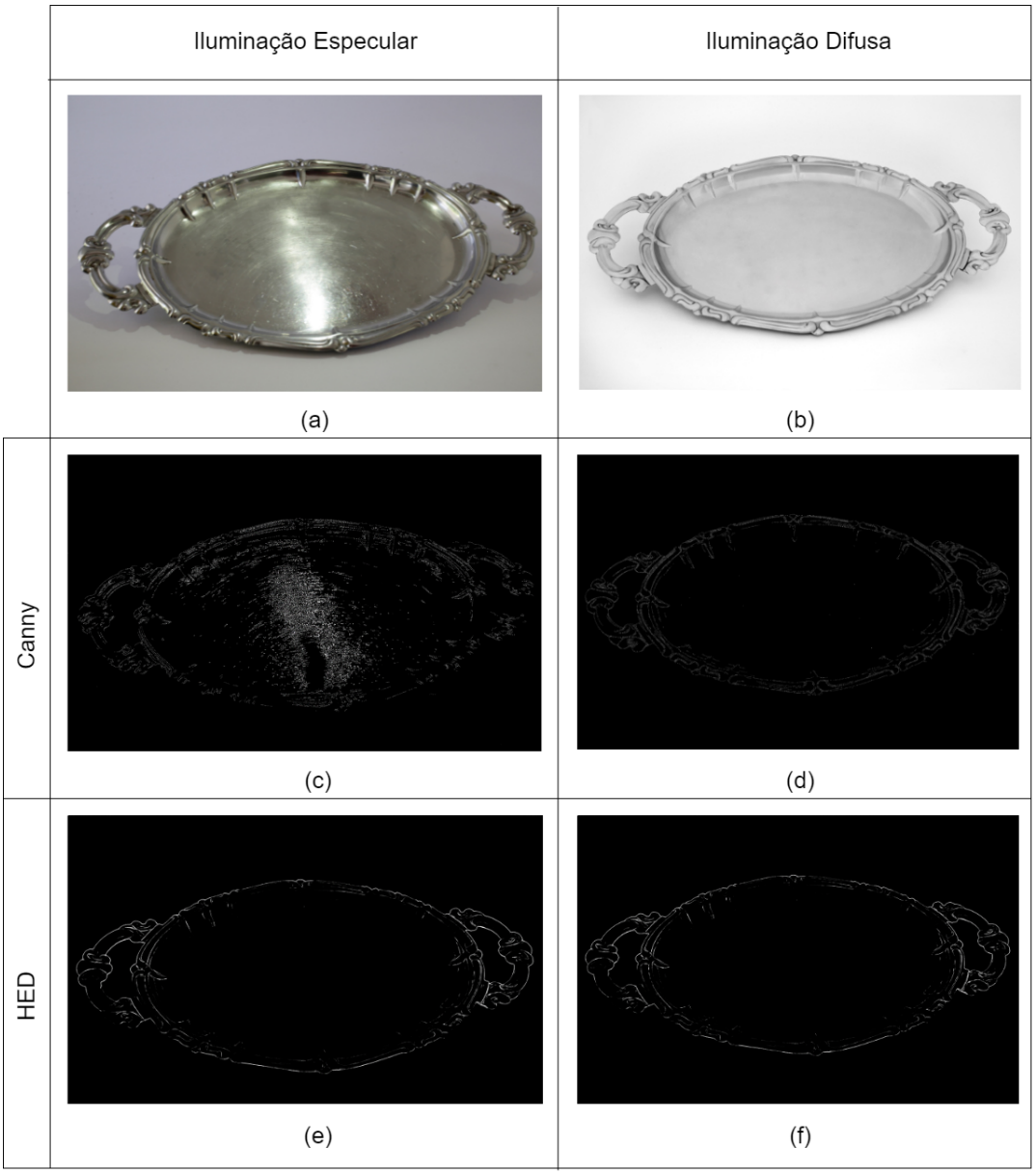


Fonte: adaptada de (XIE, 2015).

expor melhor supressão de características causadas pela iluminação, sem apagar características presentes nos modelos que são renderizados. Um posterior tratamento a imagem de saída da camada é dado, através da aplicação do algoritmo de supressão de não máximos apresentado em (CANNY, 1986). O uso deste algoritmo acrescentou precisão às bordas detectadas através de seu estreitamento. Isto evitou falsos positivos no processo de detecção das características longe das bordas proposto na subseção 3.3.2. Além do HED, foi utilizado também o algoritmo de Canny para gerar os relatórios e o resultado comparativo pode ser acompanhado na Figura 21. Essa figura apresenta dois cenários de teste. No primeiro, estão apresentados os resultados do uso dos algoritmos de borda sobre uma bandeja metálica cuja foto foi retirada sem nenhum tipo de tratamento de iluminação. Na Figura 21c e na Figura 21d, apresenta-se o resultado da filtragem de bordas da imagem com o algoritmo de Canny para as imagens da Figura 21a e da Figura 21b. A iluminação causa grande influência no resultado do filtro cuja região central detecta diversas bordas, entretanto, na imagem com iluminação difusa, o filtro apresenta bom resultado, detectando caraterísticas que seriam pertinentes no modelo proposto em subseção 3.2.2. Como pode ser observado na Figura 21e e na Figura 21f, o HED apresenta boa detecção de características em ambos os cenários, com imagens resultantes similares.

Com as imagens de bordas, puderam ser gerados os relatórios das regiões fora da região de borda e das regiões próximas. A região de borda foi considerada com uma margem de 10 pixels centrada nas bordas do modelo. Alguns resultados do relatório de incongruências podem ser

Figura 21 – Comparativo entre Canny e HED.

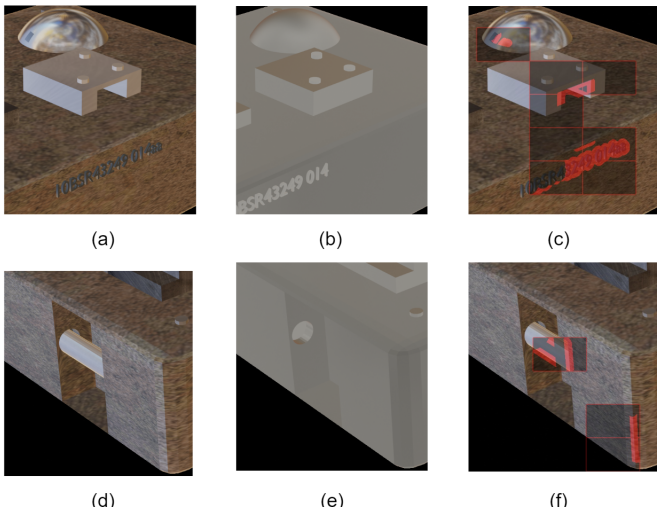


Fonte: o autor.

visualizados na Figura 22. A figura mostra regiões incongruentes entre o modelo renderizado e a cena. Em vermelho, é possível visualizar itens que não compõem a cena buscada. No primeiro cenário (Figura 22a, b e c), é possível visualizar incongruências na altura da ferramenta. Dois pedaços de metal foram soldados a parte superior para corrigir sua altura. Em consequência disto, toda a parte superior do relatório, apresenta destaque para a posição das arestas criadas pela parte rebaixada da ferramenta. Outro fator de destaque são os textos escritos na parte lateral e o posicionamento dos dois furos centrais que, na Figura 22b, podem ser visualizados em uma porção mais central do item, o que difere do ambiente real descrito na Figura 22a com

furos mais espaçados. Além destes, observa-se a presença do parafuso utilizado para adicionar a tampa metálica que cobre parte da ferramenta e não está presente no modelo, assim como o grafite, usado como lubrificante seco na ferramenta esquerda superior. No segundo destaque apresentado na figura (Figura 22d, e e f), pode-se ver outra região da mesma ferramenta analisada. Nestas imagens, é possível observar a presença de um item sobressalente na parte inferior da cena. Este item é uma ferramenta utilizada para erguer as matrizes e deveria ter sido retirada após o uso. A partir do relatório, pode-se realizar uma nova adequação da matriz, corrigindo os defeitos encontrados como itens sobressalentes, ou fora de posição. Este tipo de procedimento auxilia na composição da cena e diminui a chance de acidentes ocasionados por tais problemas.

Figura 22 – Relatório de incongruências em regiões fora das bordas.

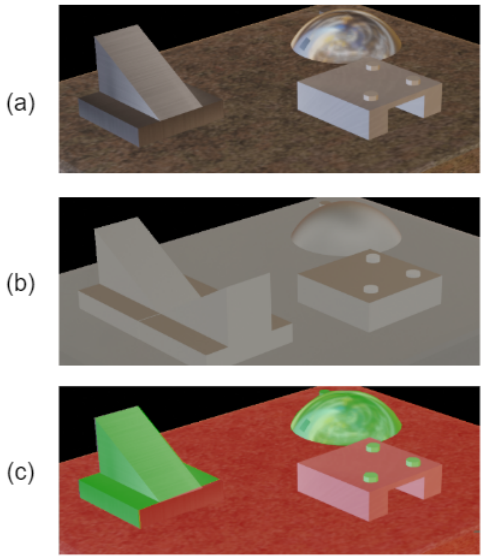


Fonte: o autor.

Para afirmar que uma peça se encontra na cena, foi utilizado o padrão de 70% de correspondências, conforme apresentado na seção 3.3. A Figura 23 mostra uma simulação de resultados apresentados em imagens simuladas da fábrica pelo casamento de algumas ferramentas que se apresentavam a cerca de 2,5 m do foco da câmera. Nas três imagens presentes na figura, observa-se a imagem alvo da cena (Figura 23a), o modelo renderizado considerado estado ideal da cena (Figura 23b) e o resultado do casamento (Figura 23c). Destaca-se que a cena apresenta-se em um estado de completude anterior em relação ao modelo, ou seja, mais peças ainda precisam ser adicionadas à cena. Isto pode ser observado comparando-se a peça sobressalente que pode ser visualizada na Figura 23b e na Figura 23a onde consta uma ferramenta ausente. O relatório apresentado na Figura 23c destaca a incongruência da cena,

mantendo a região em vermelho. As ferramentas destacadas em verde foram reconhecidas com uma porcentagem de 95%. A base, ou seja, a ferramenta maior que comporta todas as demais do conglomerado, além da peça central que suporta os três parafusos, não foram encontradas. A base em especial nos experimentos da fábrica não foi encontrada em nenhum dos testes realizados. Enquanto a peça central que apresenta-se como suporte para os três parafusos encontrados, também apresenta incongruências de formato e não é considerada encontrada, esta foi uma das situações encontradas na fábrica que a solução descrita neste trabalho foi capaz de identificar.

Figura 23 – Relatório de casamento.



Fonte: o autor.

4.4 CONSIDERAÇÕES FINAIS

Neste capítulo foram explicitados os experimentos realizados para a conclusão deste trabalho. O capítulo descreve todas as etapas de produção do *framework* desde a etapa de Prova de conceito até a conclusão do piloto, com experimentação em fábrica. São demonstrados o ambiente que foram utilizados para validar cada etapa, além de quais algoritmos se demonstraram eficazes para cada situação. Para a produção do piloto, pôde-se ter acesso ao ambiente fabril real e puderam ser verificados os diversos desafios relacionados à incongruências das peças, assim como problemas relacionados à alta reflexividade dos materiais que compõe o ferramental. Ainda na fase de piloto, foi possível verificar a validade dos métodos de PnP para aquisição do

posicionamento inicial dos itens buscados, assim como dos algoritmos de extração de características, Outro ponto importante coberto nos experimentos foi a experimentação dos métodos de aprendizagem de máquina utilizados para a busca de objetos tridimensionais, Através de um modelo de treinamento utilizando imagens sintéticas, que demonstraram bons resultados na fase de protótipo, entretanto não se adequaram ao ambiente fabril se demonstrando ineficientes na fase de piloto. Os resultados compilados de todas as fases são apresentados no capítulo a seguir.

5 RESULTADOS E DISCUSSÕES

Ao longo deste capítulo, são discutidos os resultados encontrados. O projeto foi desenvolvido ao longo de um ano dentro do Instituto Senai de Inovação para Tecnologias da Informação e Comunicação em parceria com a General Motors do Brasil. A solução foi amadurecida, seguindo o nível de prontidão tecnológica *Technology readiness level* (TRL) desenvolvido pela NASA (HÉDER, 2017). O projeto fez parte de um programa que levou inovação para dentro da ferramentaria da fábrica automotiva. As seções a seguir descrevem os resultados das etapas de prototipação e implantação do projeto piloto.

5.1 PROTOTIPAÇÃO

Com a evolução do projeto, obteve-se acesso às primeiras informações do problema real, o modelo tridimensional de uma matriz. O modelo é composto por centenas de componentes, e o formato dos componentes é variado, como apresentado na subseção 4.3.2. Outro fator importante que tornou-se conhecido, foi que apenas imagens RGB poderiam ser retiradas da cena. Não deveriam ser realizados vídeos ou câmeras que pudessem captar imagens RGB-D.

Uma vez que apenas imagens RGB únicas poderiam ser usadas, foram descartados a utilização inicial de métodos de fluxo óptico e métodos de casamento utilizando imagens de profundidade. Um outro fator importante é que havia o conhecimento do material que era componente da matriz. Por serem no geral formados por aço, imaginou-se que os itens poderiam apresentar alta especularidade, como foi descoberto posteriormente e pode ser observado na Figura 17. A especularidade do material, poderia vir a trazer problemas para câmeras que capturam imagens RGB-D principalmente as baseadas em tempo de voo *Time of Flight* (ToF).

Ainda sem informações mais importantes sobre o ambiente real, como imagens do ambiente onde o objeto seria buscado decidiu-se pela construção da maquete. Conforme comentado na seção 4.2 foi utilizado um subconjunto de componentes que fossem aglutinados e com características consideradas relevantes, como partes curvilíneas, convexas, grande quantidade de quinas, etc. Este subconjunto foi selecionado devido às características da impressora 3D utilizada. A impressora, só permitia a impressão de peças com dimensões de até 20cm x 20cm, o que tornaria inviável a impressão de todos os componentes da matriz, tendo em vista que alguns dos componentes teriam espessura menor que o fio polímero de impressão utilizado.

Em posse da impressão e do modelo tridimensional, foram implementados métodos de rastreio que pudessem viabilizar o casamento dos objetos. Os métodos baseados no RAPID demonstraram bastante atinência ao problema e, com isso, foi escolhido o modelo apresentado em (CHOI; CHRISTENSEN, 2010). A utilização das arestas fortes demonstrava uma boa seleção de características do modelo pois as arestas que eram mantidas após este pré-processamento se assimilavam com as arestas obtidas da imagem de bordas da foto retirada da maquete mas requisitavam um vasto tempo de pré-processamento.

Além disto, o tratamento de oclusão, utilizando a BSP mostrou-se deficitário por não apresentar um tratamento claro para arestas que permaneciam parcialmente oclusas. Com isto, a solução evoluiu para o uso de uma renderização do objeto como modelo, de forma similar a proposta em (ARMSTRONG; ZISSERMAN, 1995). Este tratamento acelerou a obtenção de um modelo já com oclusão e ainda possibilitou uma aproximação das caraterísticas reais encontradas na cena, devido à possibilidade de alteração da iluminação.

A imagem de cena, uma foto retirada da maquete foi tratada com um filtro de bordas e posterior aplicação da transformada de Hough. O conjunto de métodos apresentou bons resultados, com o condicionante da necessidade de re-parametrização constante. As imagens apresentavam características similares às do modelo, entretanto, alterações de iluminação ou foco, guiavam a necessidade de uma nova parametrização do filtro e da Transformada. Com isto, foi implementado o filtro de bordas DifFocus que apresentou bons resultados no ambiente da maquete, assim foi aplicado o modelo de filtragem dupla de sub-regiões apresentado na seção 4.2.

Com a filtragem dupla, os resultados de micro-quadrantes corretamente encontrados melhoraram em até 8%, a depender do tamanho que eram fixados os micro-quadrantes. O método de reconhecimento da maquete apresentou bons resultados. Para realizar a descoberta das peças, é necessário parametrizar o tamanho e consequentemente a quantidade de micro-quadrantes, esta informação pode deixar o algoritmo mais ou menos custoso, além disto, os quadrantes podiam recair sobre regiões muito longe das arestas, que não necessariamente era esperado obter informações úteis, com isso, uma forma de centralizar a análise apenas na região projetada das arestas do modelo tornaria o método mais ágil e robusto. Um outro ponto negativo é que o método ainda requeria que a câmera estivesse em posição fixa e conhecida conforme descrito na seção 4.2.

A segunda vertente de reconhecimento baseada em métodos de inteligência artificial foi concebida com intuito de aumentar a robustez do sistema. Dentre os métodos estudados,

mostraram-se promissores os resultados apresentados pela YOLO, entretanto, o modelo de treinamento utilizando nas imagens RGB inviabilizaria a utilização da rede. Com isto, foi criada a base de treinamento conforme apresentada na seção 4.2 que coincide com o modelo proposto em (Liebelt; Schmid; Schertler, 2008). O casamento da técnica de treinamento com a base sintética aliado a detecção apresentada pela rede de aprendizagem profunda YOLO demonstrou resultados satisfatórios de casamento. Foi possível reconhecer os objetos do modelo em ambiente laboratorial com 70% de confiança com velocidade de execução de 5 à 10 quadros por segundo. Isto viabilizou a dupla detecção do objeto na imagem, uma realizada pelos algoritmos de visão computacional usando *templates* e o método baseado em IA.

Os resultados coletados na fase de prototipação, guiaram o projeto à evolução para uma fase piloto. Nesta etapa, foi possível adentrar até a fabrica e coletar imagens reais dos objetos. A seção a seguir explica os resultados obtidos.

5.2 RESULTADOS DA FASE DE PILOTO

Com imagens reais das ferramentas, foi possível corroborar informações previamente discutidas como a presença de alta especularidade em diversos itens. Entretanto, surgiram novos desafios. O ambiente onde as matrizes estão localizadas impede o posicionamento da câmera conforme proposto na fase de protótipo. Como dispostas em um ambiente com fluxo constante de pessoas e máquinas, a iluminação variava constantemente nos objetos causando ainda maior variação. Outro fator importante observado é a diferença básica das peças construídas em moldes de barro ou isopor, elas apresentavam ranhuras e diferenciações dos modelos que não eram mapeadas. Além disso, características como pinturas e depósitos de lubrificantes. As diferenciações básicas do modelo, limitariam a utilização de métodos baseados em *keypoints* mesmo que fossem obtidas imagens RGB texturizadas dos objetos.

Para a produção do piloto foram executadas duas atividades iniciais, o treinamento do classificador presente no YOLO com os modelos tridimensionais relativos à ele e a adequação do posicionamento da câmera para possibilitar a verificação da matriz utilizando o método de *template*. O treinamento inicial da rede neural se deu de forma similar à realizada durante a fase de prototipação. Entretanto, os resultados alcançados não foram satisfatórios. Não foi possível encontrar nenhuma das peças treinadas, mesmo com baixo índice de confiança. Nenhum dos treinamentos mostrou-se adequado. Além do método de treinamento utilizado durante a prototipação, foram utilizadas variações de textura para renderização, treinamento

com imagens de bordas para realizar o casamento na imagem de bordas da cena e nenhum dos classificadores apresentou resultado positivo. Acredita-se, que as imagens não apresentaram características relevantes o bastante para realizar o treinamento do classificador com eficiência o bastante para realizar a detecção das peças complexas.

A alternativa apresentada para calcular posicionamento da câmera mostrou-se adequada em casos onde a imagem apresenta baixo erro entre o posicionamento dos marcadores e o modelo estimado. A Tabela 1 demonstra os valores do erro de reprojeção absoluto médio encontrado ao aplicar os 4 métodos durante os experimentos. Além de apresentar os melhores valores médios, o RPnP ainda se apresenta como o método mais estável. Os resultados são oriundos de 6 imagens de cena diferentes, que foram retiradas considerando duas linhas de visada retiradas de quinas opostas da matrizes, além de 2 tipos de matrizes, uma superior e uma inferior.

O EPnP também demonstrou resultados acurados, entretanto, como não são utilizadas duplas com alto grau de redundância, o algoritmo apresentou uma maior variância dos resultados. Além disto, os valores das coordenadas buscadas apresentam uma considerável proximidade de distribuição dos pontos - 2 pontos dos 4 considerados apresentam uma distância em coordenadas inferior a 15 - sendo este um dos experimentos reconhecidos como mais problemáticos para o EPnP, os métodos se comportaram conforme o esperado.

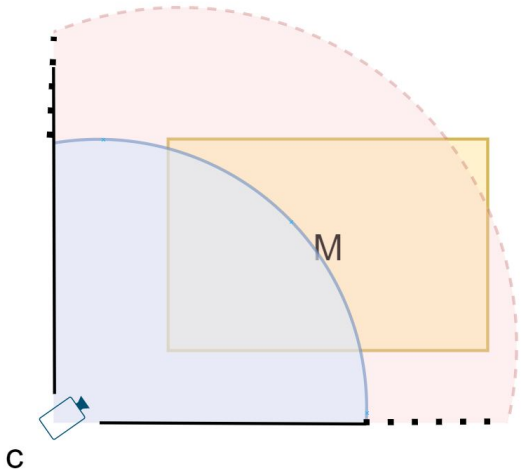
Tabela 1 – Valores do erro absoluto entre a localização dos centroides estimados utilizando o ArUco e sua projeção relativa após a aplicação do método PnP

	P3P	**EPnP**	**RPnP**	**DLS**
Média Aritmética	35,47	22,0	**17,44**	20,0
Desvio Padrão	12,31	17,28	**5,42**	14,26

Embora o posicionamento seja adequado para as peças até certa distância da câmera, os itens que se encontram a aproximadamente mais que 4 metros de distância da câmera sofrem distorção projetiva a ponto de não corresponderem com a imagem de forma adequada. A Figura 24 representa a amplitude da área adequada para realização do casamento. A região apresentada em azul destaca-se como a região onde a renderização apresenta fidedignidade à imagem de cena. A área hachurada em vermelho pode apresentar resultados de falso negativo ou positivo e não é considerada na produção do relatório descrito. Durante os experimentos cerca de 95% dos casamentos falsos positivos encontrados estão presentes na área em

vermelho.

Figura 24 – Região onde os itens são projetados em conformidade com a imagem da cena. Na figura C representa a câmera e M a matriz ferramental.



Fonte: o autor.

A extração de características na fase de piloto se deu através da utilização dos dois filtros de bordas Canny e HED. Os valores do Canny foram ajustados de forma empírica para o ambiente fabril, com valores que pudessem lidar com acurácia de resultado no foco da câmera. O HED apresentou resultados superiores ao lidar com a região completa da imagem, isto aumentou a eficiência geral do algoritmo de verificação de incongruências na região longe da borda descrito na subseção 3.3.2.

Em detrimento do bom rendimento apresentado pelo HED, o seu tempo de execução que gira em torno de 2 minutos e 30 segundos - cerca de 100 vezes o tempo médio de execução da criação do relatório utilizando o Canny, o que impacta na eficiência geral da solução. Ainda assim, o relatório de busca por incongruências na região fora de borda demonstrou o resultado esperado. Como descrito na Figura 22 pôde-se observar incongruências na matriz que não deveriam estar presentes segundo o processo adequado de montagem esperado. Tanto a região superior da Figura 22c que apresenta incongruência na altura da peça base da matriz - a base real apresenta cerca de 2 cm de altura a menos do que deveria ter segundo o modelo CAD - quanto a Figura 22f demonstra a existência da ferramenta de suporte da matriz que não deve estar presente durante o processo de prensagem.

Ao longo dos experimentos foram adicionados também itens que compunham o ambiente para verificar a viabilidade do algoritmo. Foi possível encontrar dois parafusos com cerca de 5 centímetros de comprimento e arruelas com 2 centímetros de diâmetros ambos compostos

de metal que foram dispostos ao longo das peças. Além destes, foi testada a oclusão com uma luva de proteção de algodão. Também foi possível verificar o posicionamento da luva quando depositada em uma região distante das bordas dos objetos e a oclusão causada por ela inviabilizou o casamento. Ou seja, um exemplo de um casamento verdadeiramente negativo. A Tabela 2 apresenta como foram classificados cada um dos possíveis casamentos que puderam ser observados.

Tabela 2 – Classificação do casamento

Classificação	**Descrição**
Verdadeiro positivo	Item ou conglomerado de itens considerado encontrado e que consta na posição correta na imagem de cena
Verdadeiro negativo	Não foi possível encontrar o item com o algoritmo e ele não se apresenta em cena ou encontra defeitos estruturais
Falso positivo	Objeto foi considerado em cena, entretanto, ele não se encontra, ou possui defeitos estruturais
Falso negativo	Objeto não foi considerado em cena, entretanto, ele está em seu posicionamento correto com sua estrutura correta

Seguindo este modelo de classificação proposto, foi possível verificar a acurácia do algoritmo implementado. A partir do relatório de casamento, foram classificadas uma a uma as peças visíveis de acordo com a tabela. Os resultados gerais estão dispostos na Tabela 3. Para a obtenção destes resultados, foram analisados ferramenta por ferramenta de 6 casamentos. Os casamentos foram obtidos de diferentes configurações de posicionamento da câmera, levando em consideração duas quinas opostas além de dois tipos de matrizes, uma superior e uma inferior. A partir dos resultados manuais, a tabela foi construída. Na tabela é possível observar a média aritmética dos resultados apresentados do casamento geral das peças além do desvio padrão encontrado.

Tabela 3 – Resultados gerais de casamento

	Acurácia	**Cobertura**	**Precisão**	**F1 (Média Harmônica)**
Média Aritmética	0,778	0,711	0,658	0,661
Desvio Padrão	0,07	0,268	0,076	0,163

Foram realizados seis testes com ângulo de visualização de duas quinas opostas da matriz. Os experimentos consideraram a adição de itens sobressalentes à cena, causando oclusão de partes do conglomerado. Além destes, foi realizada a retirada de itens da lista de entrada do modelo para verificar como o método se comportava. Também foram retirados itens da cena sem sua retirada da lista de entrada dos modelos.

De forma geral, o método apresentou a acurácia adequada, ou seja, identificou de forma adequada tanto peças que estavam corretamente posicionadas ou formadas em cena, como apresentou resultado de casamento negativo para que não se apresentavam em cena. Também foi relatado uma boa cobertura dos valores, uma vez que os objetos corretamente posicionados foram encontrados na maior parte dos casos. Entretanto, os experimentos demonstraram que a consideração geral dos objetos apresentou uma precisão baixa. A consequência disto foi o impacto na média harmônica observada na tabela.

Atribuiu-se tal comportamento ao fato que ao considerar a cena de forma geral, surgem casamentos falsamente positivos ocasionados pela deformação projetiva dos objetos. Com isto, os resultados foram revisitados, considerando apenas os objetos que se encontravam no raio adequado, centralizado no olho da câmera. A Tabela 4 apresenta os valores obtidos.

Tabela 4 – Resultados de casamento em região de baixo índice de deformação projetiva

	Acurácia	**Cobertura**	**Precisão**	**F1 (Média Harmônica)**
Média Aritmética	0,88	0,691	0,917	0,744
Desvio Padrão	0,06	0,222	0,204	0,143

Os valores apresentados na segunda análise demonstram um aumento significativo em acurácia e precisão. Este aumento se dá pela maior coerência entre o modelo projetado e o verificado em cena. O aumento na acurácia e precisão ocasionou também no aumento da média harmônica entre a cobertura e precisão. Embora a cobertura tenha apresentado resultados ligeiramente inferiores, o novo valor recai sobre o intervalo formado a partir do valor da média composto com o desvio padrão.

Isto posto, os resultados apresentados no relatório de proximidade apresentaram boa acurácia para as peças analisadas tanto na presença, quanto na ausência dos itens que se encontravam até 4m do olho da câmera. A Figura 12 é resultado de um dos casos de teste, e demonstra o casamento correto de uma ferramenta de corte componente da matriz. Ao seu lado deveria ser possível visualizar mais uma ferramenta de corte, similar a ela. Este é um

exemplo de como o sistema é capaz de acusar a presença e ausência das ferramenta de forma correta.

Os testes guiaram também a percepção de informações que eram até então desconhecidas pelos desenvolvedores. Em nenhum dos testes a ferramenta base foi considerada encontrada, devido à grande quantidade de incongruências construtivas que a mesma apresenta. Os especialistas presentes indicaram que é comum a base apresentar as incongruências e que ela não era um fator determinante da qualidade do processo por servir apenas como apoio para as demais.

Um dos fatores importantes que puderam ser observados a partir dos resultados é a necessidade de aquisição de diversas imagens da matriz ferramental, devido a amplitude da mesma. Já era esperado que fosse necessária a construção de mais de um relatório para realização do casamento completo. Tendo em vista a necessidade de verificação de itens que permanecem oclusos por outras ferramentas para a linha de visada da câmera. No planejamento inicial, esperava-se ser necessária aquisição de duas imagens da matriz retiradas entre duas quinas não adjacentes. Entretanto, os experimentos demonstraram a necessidade da aquisição de pelo menos quatro imagens, uma por quina para maior confiabilidade do casamento realizado pelo sistema.

Todo o processo, até gerar os relatórios de casamento, tem um tempo aproximado de dez minutos. Este tempo compreende desde a etapa inicial de posicionamento dos marcadores rígidos, até o processamento final e geração dos dois relatórios, Este processo, conta ainda com o posicionamento do tripé, calibração do foco da câmera, aquisição das imagens de cena e para calibração da câmera e processamento geral dentro do *framework*. O processamento do *framework* realiza a calibração da câmera e o processamento geral das informações com geração de renderizações, aquisição de características, casamento e geração das duas imagens de relatório.

5.3 CONSIDERAÇÕES FINAIS

O capítulo 5 possui a descrição dos resultados obtidos, foram levantados dados de todas as etapas de protótipo e piloto do *framework*. Os resultados obtidos na fase de protótipo, demonstraram bons resultados na realização do casamento utilizando comparativos de arestas entre um modelo canônico renderizado e as arestas obtidas a partir da filtragem de bordas da cena. A fase de protótipo ainda guiou o estudo para a utilização da rede neural Yolo

para detecção do ferramental. Entretanto, durante a fase de piloto, as características reais do ambiente fabril, levaram a adaptações do framework, com a utilização de modelos de PnP para a descoberta do posicionamento da câmera. Dentre estes métodos, houve o destaque para o RPnP que demonstrou maior robustez e menores índices de erro quadrático médio.

Para validar o casamento, foram realizados 6 experimentos com 2 bases diferentes e 2 configurações de câmera, com imagens obtidas a partir de duas quinas opostas. Os resultados compilados estão apresentados com todos os itens que são esperados estar presentes na cena e com destaque para os casamentos de objetos até uma distância próxima ao foco da câmera. Nesta segunda configuração os resultados demonstraram maior valores médios de acurácia, cobertura e precisão. O capítulo a seguir, apresenta a conclusão geral do estudo, demonstrando também melhorias e trabalhos futuros.

6 CONCLUSÕES

Este trabalho propôs um método de reconhecimento de instâncias tridimensionais especulares em uma imagem RGB obtida a partir de uma câmera monocular. O objetivo do trabalho foi propor métodos capazes de realizar esta tarefa dentro do contexto de uma indústria automotiva. O projeto alcançou o nível de maturidade de inovação de piloto, ou seja, protótipo funcional validado em ambiente operacional.

O algoritmo apresentado busca minimizar as incertezas do processo para gerar um modelo de classificação acurado e preciso. Para realizar tais méritos utiliza uma lista de modelos tridimensionais dos objetos buscados e uma imagem da cena real. O sistema tem como saída dois relatórios, um que aponta defeitos estruturais nos componentes da cena e um que verifica se os itens estão em seu lugar correto. A solução preza por um mínimo impacto dentro do ambiente fabril, minimizando problemas próprios do processo industrial otimizado com a proposta de intervenção sutil. Ao longo deste capítulo são demonstradas as contribuições e possibilidades já mapeadas de trabalhos futuros.

6.1 CONTRIBUIÇÕES

O método proposto realiza o casamento a partir de diferentes etapas. Como resultado, foi possível atingir o objetivo inicial de criar um sistema que realiza o reconhecimento de componentes de uma matriz automotiva a partir de uma imagem obtida da região de *try-out*. Até onde foi possível levantar durante o estudo é proposto um algoritmo inédito para a tarefa apresentada cujas principais contribuições alcançadas estão dispostas na lista a seguir:

- Através da minimização das incertezas, como a utilização de marcadores para adquirir o posicionamento inicial dos objetos, além de uma verificação dupla da cena, buscando imperfeições e artefatos coerentes, foi possível criar um algoritmo de reconhecimento de objetos especulares completamente adaptado ao problema de verificação de componentes de uma matriz automotiva com uma precisão média de 91%.

- O algoritmo proposto não requer a criação de modelo prévio utilizando imagens reais do objeto e realiza o casamento de objetos complexos em ambiente com plano de fundo com alto grau de desordem, utilizando-se apenas do modelo tridimensional como entrada.

- Foi possível testar e validar algoritmos extratores de características de cena conceituados como o Canny e o HED em um problema real, demonstrando a sua amplitude de utilização e eficiência tecnológica.
- Foi possível demonstrar as limitações do treinamento utilizando bases de imagens sintética renderizada de algoritmos conceituados de detecção e classificação de itens em problemas reais.
- O algoritmo apresentado demonstra ser uma ferramenta adequada para a verificação de itens, sendo um processo adequado a aumentar a eficiência e segurança da região de ferramentaria automotiva.

6.2 DIFICULDADES ENCONTRADAS

Para alcançar as contribuições supracitadas, foi necessário evoluir o algoritmo através de diversas etapas seguindo o padrão de amadurecimento iterativo da solução para o desenvolvimento de inovação para a indústria. Dentre as dificuldades encontradas, destaca-se a ineficiência parcial do ambiente de prototipação criado. Atribui-se a isto, a complexidade e quantidade dos componentes associados ao alto grau de especularidade apresentada pelos objetos em metal.

Os resultados apresentados na fase de protótipo pela detecção utilizando a rede neural Yolo guiaram a um falso entendimento do problema. A reformulação da base de treinamento seguindo diferentes indicadores e tentativas, geraram um atraso na conclusão do projeto.

Outro problema enfrentado foi o alto grau de especularidade dos materiais, associado a um ambiente com iluminação mal distribuída do ambiente fabril. Esta configuração gera diversas informações que podem ser consideradas falsas pelos algoritmos extratores de bordas. Além da iluminação especular, a presença de pinturas variadas, como marcações a giz ou tinta, além da presença dos depósitos de grafite em determinados componentes. Para sanar tais problemas foi necessário realizar diversas calibrações empíricas dos parâmetros do filtro de bordas Canny até ser encontrado um valor ideal.

Com relação à inferência inicial do posicionamento dos itens, a calibração de câmera clássica associada aos métodos de PnP demonstrou não ser o bastante para adequar uma imagem renderizada a imagem real. Isto limitou a aplicação do método e acabou por diminuir a eficiência geral com a apresentação de falsos positivos, como pode ser visto no Capítulo 5.

Além disto, também faz necessário aumentar a quantidade de relatórios a serem gerados para verificação completa da matriz.

Um outro fator impactante no resultado apresentado foi a dificuldade de realizar os experimentos enquanto o ambiente estava completamente operacional. As matrizes precisavam ser constantemente realocadas dentro do *try-out* além do curto período de tempo que foi possível ter acesso ao ambiente. Isto diminuiu a quantidade de experimentos que puderam ser realizados.

6.3 TRABALHOS FUTUROS

Durante os testes do piloto em ambiente operacional foi possível coletar dados necessários para realizar melhorias no funcionamento mais adequado do sistema. Dentre os levantamentos, a diferenciação entre os diferentes tipos de problemas reconhecidos no relatório de incongruências longe da região de borda. Como alguns dos erros encontrados são recorrentes, à exemplo de regiões amassadas ou ranhuras, acredita-se ser viável construir um classificador de tais problemas e adicioná-lo ao sistema.

Uma outra melhoria a ser realizada identificada, consiste na criação de um relatório que case diferentes ângulos visualização. Este tipo de validação permitirá o cruzamento dos dados, aumentando a eficiência geral do sistema.

Uma outra melhoria visualizada é no aumento da eficiência do sistema. A utilização de programação paralela, preferencialmente utilizando GPU do sistema, deve trazer boa otimização do tempo de execução do sistema. Espera-se que ao realizar a otimização do sistema, isso viabilize a adição de métodos de rastreio que podem vir a contribuir com a minimização do impacto de uso da ferramenta em fábrica.

Outros futuros incrementos estão na identificação de objetos que possam vir a aparecer em mais de uma matriz. Tal recorrência pode viabilizar a utilização de classificadores de inteligência artificial. Esta ação também aumentará a robustez do sistema, permitindo a análise mais aprofundada utilizando arestas apenas das peças únicas.

Além das melhorias previstas, faz-se providente aumentar o número de experimentos para diminuir a incerteza associada ao método. De forma adicional a outras matrizes automobilísticas e em ferramentas de outros segmentos industriais.

REFERÊNCIAS

ALAHI, A.; ORTIZ, R.; VANDERGHEYNST, P. Freak: Fast retina keypoint. In: . [S.l.: s.n.], 2012. p. 510–517.

ARMSTRONG, M.; ZISSERMAN, A. Robust object tracking. In: *Asian Conference on Computer Vision*. [S.l.: s.n.], 1995. I, p. 58–61.

BAY, H.; ESS, A.; TUYTELAARS, T.; GOOL], L. V. Speeded-up robust features (surf). *Computer Vision and Image Understanding*, v. 110, n. 3, p. 346 – 359, 2008. ISSN 1077-3142. Similarity Matching in Computer Vision and Multimedia. Disponível em: <http://www.sciencedirect.com/science/article/pii/S1077314207001555>.

BAYKARA, H. C.; BıYıK, E.; GüL, G.; ONURAL, D.; ÖZTüRK, A. S. 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.

BLENDER.ORG. *About*. 2020. Disponível em: <https://www.blender.org/about/>.

CALONDER, M.; LEPETIT, V.; STRECHA, C.; FUA, P. Brief: Binary robust independent elementary features. In: . [S.l.: s.n.], 2010. v. 6314, p. 778–792.

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.

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.

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.

CPLUSPLUS. *Overview*. 2020. Disponível em: <http://www.cplusplus.com/info/>.

Darom, T.; Keller, Y. Scale-invariant features for 3-d mesh models. *IEEE Transactions on Image Processing*, v. 21, n. 5, p. 2758–2769, 2012.

DIHLMANN, C. Ferramental - revista brasileira da industria de ferramentarias. gravo, 2008. ISSN 1981-240X. Disponível em: <www.revistaferramental.com.br>.

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.

DUDA, R. O.; HART, P. E. Use of the hough transformation to detect lines and curves in pictures. *Commun. ACM*, Association for Computing Machinery, New York, NY, USA, v. 15, n. 1, p. 11–15, jan. 1972. ISSN 0001-0782. Disponível em: <https://doi.org/10.1145/361237.361242>.

FERRAMENTARIA, R. *Projetos*. 2020. Disponível em: <https://fribeiro.com.br/fabricacao-de-moldes-plasticos/projetos/>.

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>.

FOLEY, J. D.; DAM, A. van; FEINER, S. K.; HUGHES, J. F. *Computer Graphics: Principles and Practice (2nd Ed.)*. USA: Addison-Wesley Longman Publishing Co., Inc., 1990. ISBN 0201121107.

GAO, X.-S.; HOU, X.; TANG, J.; CHENG, H.-F. Complete solution classification for the perspective-three-point problem. *IEEE Trans. Pattern Anal. Mach. Intell.*, v. 25, p. 930–943, 2003.

GASPEC, F. *Ferramentaria Gaspec*. 2009. Disponível em: <http://www.gaspec.com.br/ingles/equipamentos/equip02.html>.

Gomez-Donoso, F.; Garcia-Garcia, A.; Garcia-Rodriguez, J.; Orts-Escolano, S.; Cazorla, M. 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.

Guo, Y.; Bennamoun, M.; Sohel, F.; Lu, M.; Wan, J. 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, 2014.

HARRIS, C. Tracking with rigid models. In: _____. *Active Vision*. Cambridge, MA, USA: MIT Press, 1993. p. 59–73. ISBN 0262023512.

HARTLEY, R.; ZISSERMAN, A. *Multiple View Geometry in Computer Vision*. 2. ed. USA: Cambridge University Press, 2003. ISBN 0521540518.

HÉDER, M. From nasa to eu: the evolution of the trl scale in public sector innovation. *The Innovation Journal*, v. 22, p. 1, 2017.

HUGHES, J.; DAM, A. van; FOLEY, J.; MCGUIRE, M.; FEINER, S.; SKLAR, D. *Computer Graphics: Principles and Practice*. Addison-Wesley, 2014. (The systems programming series). ISBN 9780321399526. Disponível em: <https://books.google.com.br/books?id=OVpsAQAAQBAJ>.

IANDOLA, F. N.; HAN, S.; MOSKEWICZ, M. W.; ASHRAF, K.; DALLY, W. J.; KEUTZER, K. *SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size*. 2016.

Johnson, A. E.; Hebert, M. Using spin images for efficient object recognition in cluttered 3d scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 21, n. 5, p. 433–449, 1999.

KNOPP, J.; PRASAD, M.; WILLEMS, G.; TIMOFTE, R.; GOOL, L. V. Hough transform and 3d surf for robust three dimensional classification. In: DANIILIDIS, K.; MARAGOS, P.; PARAGIOS, N. (Ed.). *Computer Vision – ECCV 2010*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010. p. 589–602. ISBN 978-3-642-15567-3.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. *Commun. ACM*, Association for Computing Machinery, New York, NY, USA, v. 60, n. 6, p. 84–90, maio 2017. ISSN 0001-0782. Disponível em: <https://doi.org/10.1145/3065386>.

Lepetit, V.; Fua, P. *Monocular Model-Based 3D Tracking of Rigid Objects: A Survey*. [S.l.: s.n.], 2005.

LEPETIT V., M.-N. F. . F. P. Epnp: An accurate o(n) solution to the pnp problem. *Computer Vision, International Journal on*, v. 81, n. 155, 2009.

LEUTENEGGER, S.; CHLI, M.; SIEGWART, R. Brisk: Binary robust invariant scalable keypoints. In: . [S.l.: s.n.], 2011. p. 2548–2555.

LI, S.; XU, C.; XIE, M. A robust o (n) solution to the perspective-n-point problem. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, v. 34, n. 7, p. 1444–1450, 2012.

Liebelt, J.; Schmid, C.; Schertler, K. Viewpoint-independent object class detection using 3d feature maps. In: *2008 IEEE Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2008. p. 1–8.

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

LOWE, D. G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, Kluwer Academic Publishers, Hingham, MA, USA, v. 60, n. 2, p. 91–110, nov. 2004. ISSN 0920-5691. Disponível em: <http://dx.doi.org/10.1023/B:VISI.0000029664.99615.94>.

Marfil, R.; Bandera, A.; Rodriguez, J. A.; Sandoval, F. 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.

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, 2010.

MESQUITA, R. G. de. *Reconhecimento de Instâncias Guiado Por Algoritmos de Atenção Visual*. Tese (Doutorado) — UNIVERSIDADE FEDERAL DE PERNAMBUCO, Recife, 2 2017.

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.

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.

OPENCV. *Overview*. 2020. Disponível em: <https://opencv.org/>.

OTSU, N. A Threshold Selection Method from Gray-level Histograms. *IEEE Transactions on Systems, Man and Cybernetics*, v. 9, n. 1, p. 62–66, 1979. Disponível em: <http://dx.doi.org/10.1109/TSMC.1979.4310076>.

RAMAN, M.; AGGARWAL, H. Study and comparison of various image edge detection techniques. *International Journal of Image Processing*, v. 3, 03 2009.

REDMON, J.; FARHADI, A. Yolov3: An incremental improvement. *arXiv*, 2018.

ROMERO-RAMIREZ, F.; MUñOZ-SALINAS, R.; MEDINA-CARNICER, R. Speeded up detection of squared fiducial markers. *Image and Vision Computing*, v. 76, 06 2018.

Rublee, E.; Rabaud, V.; Konolige, K.; Bradski, G. Orb: An efficient alternative to sift or surf. In: *2011 International Conference on Computer Vision*. [S.l.: s.n.], 2011. p. 2564–2571.

Senst, T.; Evangelio, R. H.; Keller, I.; Sikora, T. 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.

SEO, B.-K.; PARK, H.; PARK, J.-I.; HINTERSTOISSER, S.; ILIC, S. Optimal local searching for fast and robust textureless 3d object tracking in highly cluttered backgrounds. *IEEE Transactions on Visualization and Computer Graphics*, v. 20, p. 99–110, 01 2014.

SIMONYAN, K.; ZISSERMAN, A. *Very Deep Convolutional Networks for Large-Scale Image Recognition*. 2014.

SIPIRAN, I.; BUSTOS, B. Harris 3d: A robust extension of the harris operator for interest point detection on 3d meshes. *The Visual Computer*, v. 27, p. 963–976, 11 2011.

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

Vacchetti, L.; Lepetit, V.; Fua, P. Stable real-time 3d tracking using online and offline information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 26, n. 10, p. 1385–1391, 2004.

VISWANATHAN, D. Features from accelerated segment test (fast). In: . [S.l.: s.n.], 2011.

WELCH, G.; BISHOP, G. *An Introduction to the Kalman Filter*. USA, 1995.

WHITTED, T. An improved illumination model for shaded display. In: *ACM SIGGRAPH 2005 Courses*. New York, NY, USA: Association for Computing Machinery, 2005. (SIGGRAPH '05), p. 4–es. ISBN 9781450378338. Disponível em: <https://doi.org/10.1145/1198555.1198743>.

XIE, S. *Holistically-Nested Edge Detection*. 2015. Disponível em: <https://github.com/s9xie/hed>.

XIE, S.; TU, Z. Holistically-nested edge detection. In: *Proceedings of IEEE International Conference on Computer Vision*. [S.l.: s.n.], 2015.

APÊNDICE A – CALIBRAÇÃO DE CÂMERAS

Uma das principais necessidades da computação visual tridimensional consiste na calibração das câmeras. A partir da calibração, é possível extrair informações importantes das imagens bidimensionais. É uma área transversal, abordada tanto na visão computacional quanto na fotogrametria. As técnicas podem ser divididas em dois tipos: calibração fotogramétrica e auto-calibração Zhang (2000).

O processo de calibração utilizando fotogrametria se baseia no uso de um objeto tridimensional com características conhecidas. Geralmente, os objetos utilizados são compostos por dois ou três planos ortogonais entre si. O processo que realiza a auto-calibração não utiliza de objetos, mas sim da câmera que se move por uma cena estática. Caso as imagens tenham sido retiradas com a mesma câmera com parâmetros internos fixos, bastam três imagens de diferentes posições para conseguir recuperar a parametrização extrínseca a partir da triangulação. A simplicidade do método impacta nos resultados, que são potencialmente não-confiáveis. Existem ainda técnicas que utilizam da verificação de pontos de fuga para posições ortogonais, além da calibração, utilizando rotação Zhang (2000).

O modelo de câmera mais comumente utilizado para os estudos envolvendo objetos tridimensionais é o modelo *pinhole*. Uma relação importante a ser observada, é de como os pontos são projetados no plano projetivo bidimensional da imagem, a partir deste modelo de câmera. A Equação 7.1 demonstra como um ponto em coordenadas de mundo (M) é projetado no plano projetivo bidimensional P2 (m). Destaca-se que *M* e *m* são acrescidos de uma coordenada com valor 1 em seu final, por estarem no plano visível do espaço projetivo e não no infinito.

$$m = A[Rt]M$$

Por sua vez, *A* representa a matriz de valores intrínsecos da câmera dados por:

$$A = \begin{pmatrix} \alpha & \gamma & ux \\ 0 & \beta & uy \\ 0 & 0 & 1 \end{pmatrix}$$

Onde os valores de *u* representam os eixos da imagem, α e β representam o as coordenadas do centro e γ representa o fator de escala. Os valores representados pela matriz3x4 [Rt] representam a matriz de valores extrínsecos, onde *t* representa a localização global do olho da câmera e R o valor da matriz de rotação 3x3 que pode ser encontrada a partir da utilização do método de Rodrigues. Foge ao escopo deste trabalho aprofundar-se no estudo do espaço projetivo. Em (Foley et al. 1990) é apresentado um vasto material sobre o plano projetivo e suas características. O procedimento de calibração de câmera proposto em Zhang (2000) consiste, em resolvera Equação 1 através do uso de homografias entre os valores conhecidos encontrados na imagem. De forma resumida, o método é decomposto em seis etapas:

- Imprimir um padrão conhecido e adicioná-lo a uma superfície planar.
- Capturar algumas imagens de diferentes posições.

- Detectar o posicionamento das características conhecidas que estão sendo buscadas.
- Elucidar os parâmetros intrínsecos da câmera.
- Estimar os coeficientes radiais de distorção utilizando o método dos quadrados mínimos.
- Otimizar a estimativa minimizando os parâmetros.

Os experimentos realizados utilizaram de um papel impresso com um padrão quadriculado com bordas brancas, que permite uma fácil aquisição dos pontos de interesse. O método proposto por Zhang tornou-se bastante popular, sendo inserido dentro de bibliotecas abertas de visão computacional.