

LEONARDO GOMES

**ALINHAMENTO AUTOMÁTICO DE IMAGENS DE
PROFUNDIDADE PARA MODELAGEM 3D DE OBJETOS**

CURITIBA

2011

LEONARDO GOMES

**ALINHAMENTO AUTOMÁTICO DE IMAGENS DE
PROFUNDIDADE PARA MODELAGEM 3D DE OBJETOS**

Dissertação apresentada como requisito parcial à obtenção do grau de Mestre. Programa de Pós-Graduação em Informática, Setor de Ciências Exatas, Universidade Federal do Paraná.

Orientador: Prof. Dr. Luciano Silva

Co-orientadora: Profa. Dra. Olga Regina Pereira Bellon

CURITIBA

2011

Gomes, Leonardo

Alinhamento automático de imagens de profundidade para modelagem 3D de objetos / Leonardo Gomes. – Curitiba, 2011.
62 f. : il.; tab.

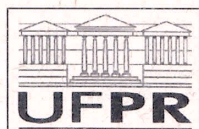
Dissertação (mestrado) – Universidade Federal do Paraná, Setor de Ciências Exatas, Programa de Pós-Graduação em Informática.

Orientador: Luciano Silva

Orientadora: Olga Regina Pereira Bellon

1. Imagem tridimensional - Simulação (Computadores). I. Silva, Luciano. II. Bellon, Olga Regina Pereira. III. Título.

CDD 006.693



Ministério da Educação
Universidade Federal do Paraná
Programa de Pós-Graduação em Informática

PARECER

Nós, abaixo assinados, membros da Banca Examinadora da defesa de Dissertação de Mestrado em Informática, do aluno Leonardo Gomes, avaliamos o trabalho intitulado, "ALINHAMENTO AUTOMÁTICO DE IMAGENS DE PROFUNDIDADE PARA MODELAGEM 3D EM OBJETOS", cuja defesa foi realizada no dia 25 de agosto de 2011, às 14:00 horas, no Departamento de Informática do Setor de Ciências Exatas da Universidade Federal do Paraná. Após a avaliação, decidimos pela aprovação do candidato.

Curitiba, 25 de agosto de 2011.

Prof. Dr. Luciano Silva
DINF/UFPR – Orientador

Prof.ª. Dra. Olga Regina Pereira Bellon
DINF/UFPR – Coorientadora

Prof. Dr. Marcelo Walter
UFRGS – Membro Externo

Prof. Dr. André Luiz Pires Guedes
DINF/UFPR – Membro Interno



AGRADECIMENTOS

Agradeço aos professores Olga Regina Pereira Bellon e Luciano Silva, não somente pela inestimável orientação mas também por me apresentarem à pesquisa e pelo grande apoio ao longo destes anos.

À CAPES que financiou este mestrado e apoia a pós-graduação no Brasil.

À UNESCO e ao IPHAN por acreditarem no trabalho de nossa equipe e nos proporcionar a gratificante oportunidade de ver nossa pesquisa aplicada na preservação de acervos importantes como as obras feitas pelo mestre Antônio Francisco Lisboa, O Aleijadinho.

À todos do grupo IMAGO pela amizade e grande disposição em ajudar.

Ao colega e amigo Maurício Pamplona Segundo, que trabalhou comigo na elaboração e desenvolvimento de um dos métodos aplicados neste trabalho.

Aos grandes amigos do grupo Hashi que estiveram ao meu lado desde o começo desta caminhada acadêmica e com certeza estarão ainda por muito tempo.

À minha namorada Giulianne pelo amor e companheirismo em todos os momentos.

E por fim agradeço minha família em especial meus pais que tanto se dedicaram por minha formação e me apoiaram por toda a vida.

SUMÁRIO

| | |
|--|------------|
| RESUMO | vi |
| ABSTRACT | vii |
| 1 INTRODUÇÃO | 1 |
| 1.1 <i>Pipeline</i> | 2 |
| 1.2 Motivação e objetivos | 8 |
| 1.3 Organização do trabalho | 10 |
| 2 MÉTODOS DE PRÉ-ALINHAMENTO | 11 |
| 2.1 4PCS | 12 |
| 2.2 Spin-image | 15 |
| 2.3 DARCES | 18 |
| 2.4 Alinhamento Baseado em <i>Principal Component Analysis</i> | 21 |
| 2.5 Método de Pingi <i>et al.</i> | 23 |
| 2.6 <i>Point Signature</i> | 25 |
| 2.7 Comparação | 26 |
| 3 IMPLEMENTAÇÕES | 28 |
| 3.1 Algoritmo de pré-alinhamento baseado em descritores SURF | 32 |
| 3.2 Definição de pares com sobreposição | 36 |
| 3.3 Abordagem proposta | 36 |
| 4 EXPERIMENTOS | 40 |
| 4.1 Avaliação do método de alinhamento de Pingi <i>et al.</i> | 46 |
| 4.2 Avaliação do método de alinhamento baseado em PCA | 48 |
| 4.3 Avaliação do método 4PCS | 49 |
| 4.4 Método desenvolvido | 50 |

| | |
|---------------------------------|-----------|
| | v |
| 4.5 Comparação geral | 52 |
| 5 CONCLUSÃO | 55 |
| 5.1 Trabalhos Futuros | 56 |
| BIBLIOGRAFIA | 58 |

RESUMO

No processo de modelagem 3D de objetos a partir de imagens, o pré-alinhamento é a etapa responsável por determinar uma aproximação da transformação rígida que alinha, em um mesmo sistema de coordenadas, duas superfícies vistas do objeto, representadas por imagens de profundidade (*range images*).

Na área de preservação digital de acervos naturais e culturais, foco deste trabalho, a fidelidade tanto da geometria quanto da textura do modelo 3D final é o principal requisito. Entretanto, investir em um sistema prático e ágil para modelagem 3D também é importante para viabilizar sua aplicação sobre acervos com diferentes características. Neste sentido, o estudo e implementação de métodos de pré-alinhamento de imagens de profundidade mais efetivos podem contribuir significativamente na construção de sistemas mais robustos e apropriados para a área.

Este trabalho apresenta um estudo sobre os métodos estado da arte de pré-alinhamento e compara-os com uma nova abordagem desenvolvida, mais efetiva para o problema. A abordagem proposta é baseada em pontos característicos, o que se mostrou bastante apropriado para o cenário avaliado. Dentre as principais contribuições do presente trabalho está a aplicação e substituição da etapa de pré-alinhamento manual utilizada no pipeline de modelagem 3D desenvolvido no grupo de pesquisa IMAGO.

ABSTRACT

In the process of 3D modeling from images, the rough registration step is responsible for finding an approximation of the rigid transformation that aligns, in the same coordinate system, two views of the object surface, represented by range images.

In the area of digital preservation of natural and cultural heritage, which is the focus of this work, the fidelity of both geometry and texture of the final 3D model is the main requirement. However, investing in a flexible and practical system for 3D modeling is also important to allow its application on objects with different characteristics. In this scenario, the study and implementation of more effective rough registration methods can contribute significantly in building more robust and appropriate systems for the area.

We present a study on the state-of-the-art methods for rough registration and compares them with a new approach more effective for the problem. The proposed approach is based on feature points, which seems to be well-suited for the evaluated scenario. Among the main contributions of this work is the application and replacing of the manual rough registration step used in the 3D reconstruction pipeline developed by the IMAGO research group.

CAPÍTULO 1

INTRODUÇÃO

A modelagem 3D de objetos reais é um tema de estudo bastante ativo na área de Visão Computacional. Diversas aplicações têm sido propostas com o uso da modelagem 3D de objetos, alguns exemplos:

- Identificação biométrica [30];
- Geração de conteúdo 3D para aplicações multimídia como filmes ou jogos eletrônicos;
- Preservação digital [18, 20, 27, 35].

O processo de modelagem 3D de objetos possui várias etapas, cada uma com desafios em particular e diferentes técnicas para resolvê-las. Exemplos que apresentam um conjunto destas etapas, aqui chamado de *pipeline*, podem ser vistos em vários trabalhos na literatura [5, 9, 13, 15, 18, 35].

Este trabalho dá continuidade ao *pipeline* de modelagem 3D de objetos desenvolvido no grupo IMAGO¹, detalhado em Vrubel [35] que tem como foco a preservação digital de acervos culturais e naturais. Existem outros dois trabalhos que fazem uso deste *pipeline*, Andrade [2] que traz um modelo de aquisição de textura de alta resolução e Mendes [26] que apresentou um museu virtual para web permitindo uma maior divulgação e acesso aos acervos digitalizados.

A principal contribuição apresentada neste trabalho é a automação de uma das etapas mais trabalhosas do *pipeline*, o alinhamento. Esta etapa junto das demais que formam este *pipeline* são apresentadas na sequência.

¹<http://www.imago.ufpr.br>

1.1 *Pipeline*

Considerando somente a modelagem geométrica do objeto, que é a etapa em foco neste trabalho, o *pipeline* é composto por cinco etapas principais:

- Aquisição;
- Alinhamento;
- Alinhamento global;
- Integração;
- Geração do modelo 3D.

A etapa de aquisição é onde se obtém dados brutos referentes à geometria do objeto, onde as técnicas e sistemas utilizados variam muito para cada cenário. Dentre os sistemas óticos mais usados destacam-se a geometria a partir de vídeo [10], visão estéreo [37], silhuetas [22] e luz estruturada [21]. Existem também métodos que adotam equipamentos de ressonância magnética [34] e escaneamento *laser*. Uma explicação mais detalhada sobre os principais métodos de aquisição 3D pode ser visto na revisão feita por Pavlidis [27].

Dentro do *pipeline* adotado, a aquisição é feita com um equipamento de escaneamento *laser*, VIVID 910 da Konica Minolta ², que é capaz de capturar a informação geométrica do objeto com precisão de até 0,22 mm.

O equipamento é composto por um emissor *laser*, que realiza varreduras com um feixe *laser*, e uma câmera, que adquire imagens de onde o *laser* incide. O método se chama triangulação por causa do triângulo formado pela câmera, o emissor *laser* e o ponto onde feixe *laser* incide. A distância entre emissor e câmera é conhecida, o ângulo formado no emissor também, e o ângulo formado na câmera pode ser calculado pela posição em que o laser aparece projetado no campo de vista da câmera. Utilizando essas 3 informações é possível calcular a posição em que o *laser* incidiu.

²<http://www.konicaminolta.com/sensingusa/products/3d/non-contact/vivid910>

A imagem resultante deste processo é uma imagem com informações de profundidade em cada ponto, chamadas imagens de profundidade (*range images*).

As imagens de profundidade [4] obtidas em diversas vistas, são o resultado da etapa de aquisição, e são elas a serem alinhadas na etapa de alinhamento. A imagem de profundidade é uma das formas de se representar dados tridimensionais, mais utilizada por sensores de captura tridimensional. Essas imagens são representadas por uma matriz bidimensional, com amostras regulares da imagem, onde cada elemento dessa matriz, o *pixel*, tem um valor real associado à distância relativa naquele ponto entre o objeto e o sensor.

As imagens de profundidade são adquiridas a partir de uma determinada posição e angulação em relação ao objeto, o ponto de vista do sensor, por isso essas imagens de profundidade também são chamadas simplesmente de vistas neste trabalho. Um exemplo de imagem de profundidade é mostrado na figura 1.1.

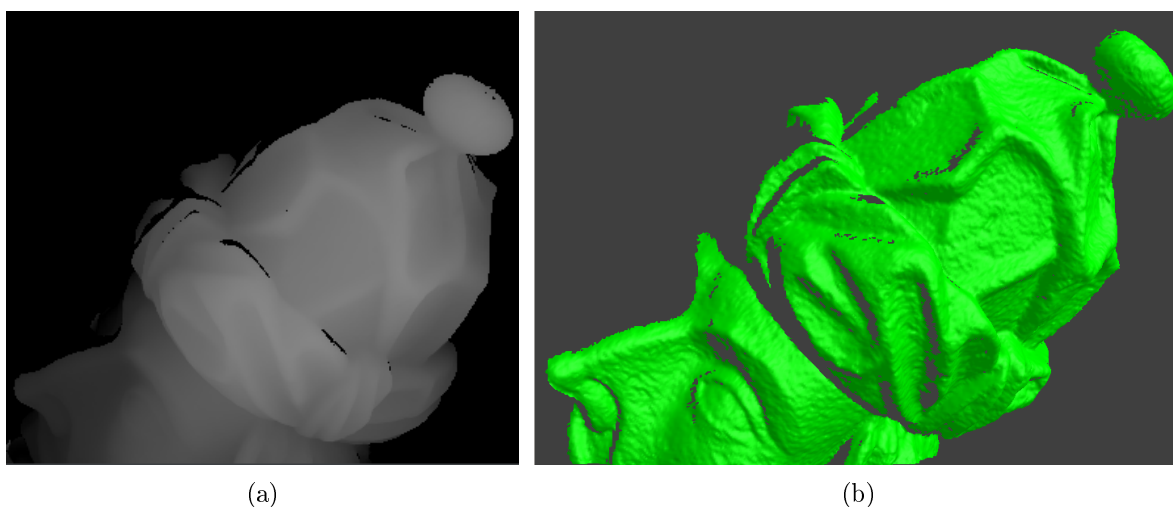


Figura 1.1: Exemplo de uma imagem de profundidade: em (a) é mostrada uma imagem em profundidade; quanto mais escuro for o ponto na imagem, maior é a distância daquele ponto ao sensor do *scanner* no momento da aquisição. Em (b) está a respectiva renderização 3D usando textura verde desta imagem de profundidade. Trata-se de uma das vistas adquiridas no processo de modelagem da estátua que representa o profeta Jonas, parte da obra "Os doze profetas" de Antônio Francisco Lisboa, mais conhecido como Aleijadinho.

O *scanner* calcula a informação de profundidade apenas das regiões que estão no seu campo de visão no momento da aquisição, conforme pode-se ver na figura 1.2, e por isso

somente uma imagem em profundidade não é suficiente para realizar a modelagem 3D integral do objeto desejado. Para isso várias vistas devem ser capturadas de forma que toda a superfície do objeto seja digitalizada.

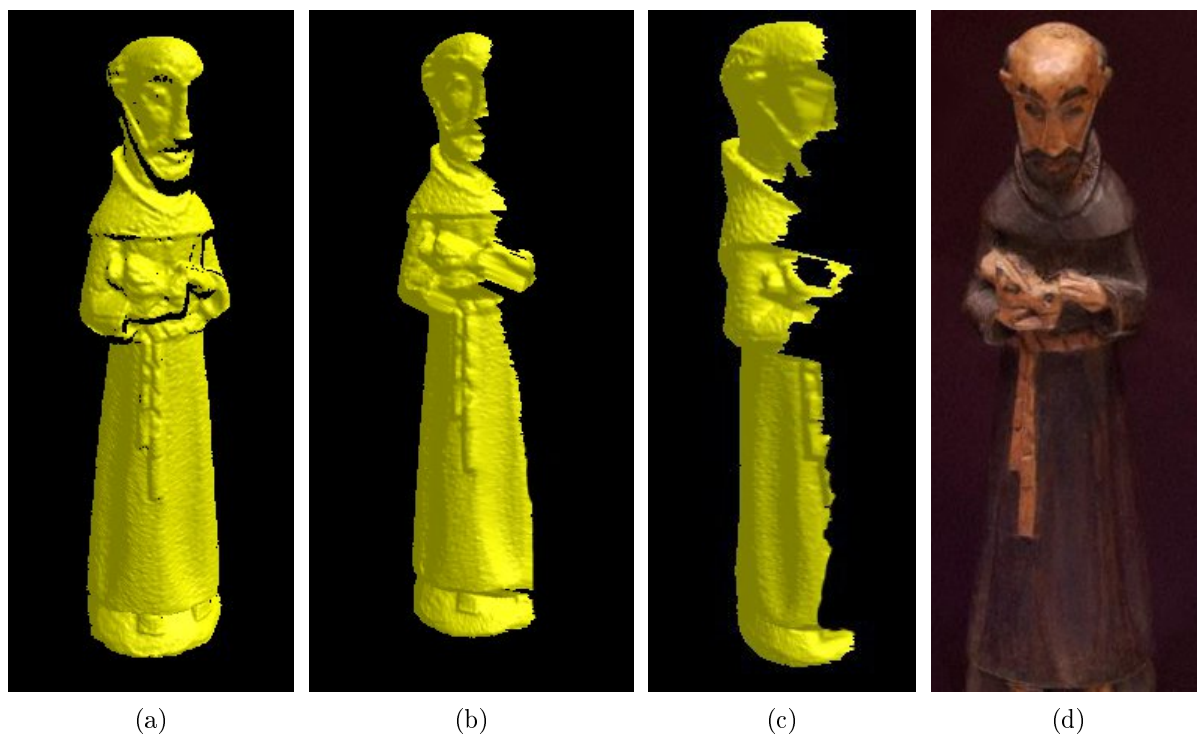


Figura 1.2: Exemplo de vistas de uma escultura de São Francisco de Assis: (a), (b) e (c) são imagens renderizadas com textura em amarelo, a partir de imagens de profundidade adquiridas de diferentes pontos de vista da escultura em madeira exibida na figura (d).

Como cada imagem de profundidade é capturada de um ponto de vista diferente do objeto, logo cada vista está em um sistema de coordenadas diferente, e o processo para colocá-las em um mesmo sistema de coordenadas global, ou seja alinhar as imagens, é o problema tratado na etapa de alinhamento. No *pipeline* apresentado por Vrubel [35], as vistas são agrupadas em pares que possuam superfície de sobreposição entre si, a organização desses pares é equivalente a um grafo conexo, onde as vistas são vértices, e os pares de vistas são arestas, e cada par de vistas nesse grafo é alinhado de forma que todos fiquem em um mesmo sistema de coordenadas.

Para alinhar cada par, assume-se M e S como sendo duas imagens de profundidade, adquiridas a partir de diferentes vistas de um mesmo objeto, mas com sobreposição entre

si, onde S está em um sistema de coordenadas global e M em um local. Considerando este cenário, existe uma transformação linear T , composta de uma translação e uma rotação, que minimiza a distância dos pontos correspondentes de M em relação a S no sistema de coordenadas global. Encontrar essa transformação é o objetivo da etapa de alinhamento [32].

O alinhamento propriamente dito é geralmente dividido em duas etapas principais: o pré-alinhamento (*rough registration*) e o alinhamento fino (*fine registration*). O pré-alinhamento é o foco principal deste trabalho, e consiste em adquirir uma estimativa inicial do alinhamento de duas superfícies com sobreposição quando não se tem nenhuma informação sobre o posicionamento relativo entre elas.

Em muitos sistemas de modelagem 3D, incluindo o sistema adotado no grupo IMAGO anteriormente a este trabalho, o pré-alinhamento é feito de forma manual [35]. Como esta é uma tarefa demorada e exaustiva [28], foram desenvolvidos vários métodos para automatizar a etapa de pré-alinhamento.

Este trabalho apresenta algumas das abordagens existentes na literatura para resolver o problema de pré-alinhamento e propõe um método que demonstrou, nos experimentos realizados, desempenho e eficiência superiores aos métodos estado da arte para o cenário deste trabalho.

Em geral os métodos de pré-alinhamento resultam em transformações que não possuem a precisão necessária para a modelagem 3D, por isso estas transformações são refinadas na etapa de alinhamento fino. Como os métodos de alinhamento fino costumam ser funções de minimização de erro sensíveis a mínimos locais, é importante encontrar uma boa estimativa inicial para não convergir para um alinhamento errado [32]. Um dos principais algoritmos utilizado para o alinhamento fino é o *Iterative Closest Point* (ICP) [6].

O ICP é um algoritmo iterativo que, a partir de duas malhas de pontos tridimensionais e uma estimativa inicial de alinhamento, gera correspondências por proximidade e calcula uma transformação que minimize a distância entre essas correspondências, esse processo

é repetido um número fixo de iterações ou até que algum limiar de erro seja atingido. As variações desse algoritmo [31] são muitas, baseadas nos algoritmos originais de Besl e McKay [6] que calcula correspondências entre pontos (*point-to-point correspondences*), e a de Chen e Medioni [9] que propõe minimizar a distância entre pontos e faces (*point-to-plane correspondences*).

No *pipeline* desenvolvido pelo grupo IMAGO a partir de uma estimativa inicial de alinhamento manual, uma variação do algoritmo ICP é aplicada [35]. Essa versão do ICP é dividida em duas fases: na primeira os parâmetros são ajustados de forma a enfatizar a convergência, visto que o ICP é vulnerável a mínimos locais; já na segunda etapa, o ICP é aplicado com parâmetros ajustados de forma a enfatizar a precisão, refinando com mais rigor o alinhamento obtido [35].

Na figura 1.3 um exemplo da diferença existente entre o alinhamento obtido na etapa de pré-alinhamento e alinhamento fino.

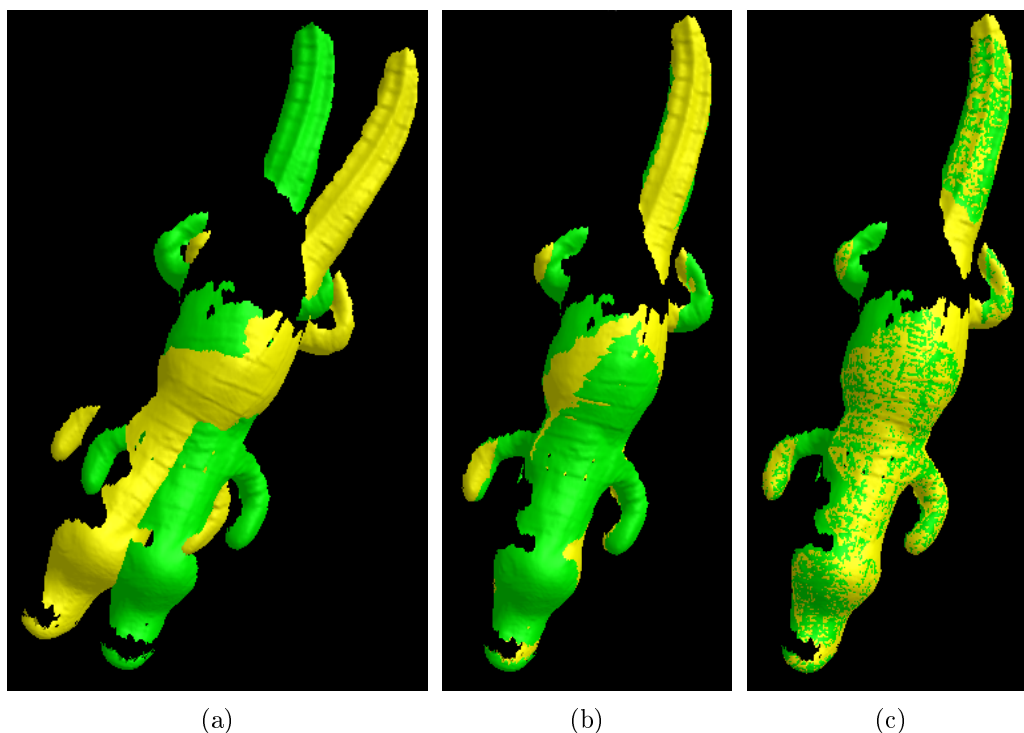


Figura 1.3: Alinhamento de duas imagens de profundidade de uma escultura indígena representando um jacaré, renderizadas em cores distintas: (a) mostra um par de imagens com a posição conforme as vistas foram obtidas; (b) o mesmo par de imagens após a etapa de pré-alinhamento e (c) o mesmo par de imagens após a etapa de alinhamento fino. Esta escultura está no Museu de Arqueologia e Etnologia (MAE) da UFPR.

Quando o alinhamento de um par de vistas é utilizado como referência de alinhamento global para o alinhamento de outros pares, o erro neste alinhamento é propagado para todos os que se baseiam nele de forma acumulativa. Com isso pode haver um acúmulo significativo deste erro de alinhamento nas últimas vistas alinhadas [29]. Para resolver este problema, após o alinhamento de todos os pares realiza-se um alinhamento global, que leva em conta todas as vistas e não somente um par delas, visando minimizar o acúmulo dos erros, distribuindo-o entre todos os pares [29].

Neste *pipeline*, o método adotado para o alinhamento global é o algoritmo de Pulli [29]. Este algoritmo gera pontos correspondentes chamados de virtuais na região de interseção de cada par de vistas, de forma que cada vista seja alinhada com estes pontos virtuais utilizando o método de mínimos quadrados, proposto por Horn [17].

Após o alinhamento global, vem a etapa de integração. Nesta etapa do *pipeline* o objetivo é gerar um único modelo 3D a partir de um conjunto de imagens de profundidade já alinhadas. No *pipeline* de Vrubel a integração é feita através de um método híbrido entre dois métodos. Primeiro é aplicado o método de Curless e Levoy [13] em conjunto com a técnica de *space carving* [13] o que gera uma representação volumétrica inicial. Em seguida aprimora-se essa representação volumétrica integrando-se somente o que está em consenso com o método de Wheeler *et al.* [36].

Se o objeto possuir concavidades muito profundas e regiões inacessíveis pelo feixe ou sensor do *scanner*, a aquisição não poderá ser feita adequadamente produzindo buracos no modelo final. No entanto neste *pipeline* é possível opcionalmente realizar o preenchimento dos buracos com informações estimadas utilizando o método de difusão desenvolvido por Davis *et al.* [14].

Com o modelo integrado e sem buracos, a malha poligonal final é gerada através do método *Marching Cubes* [23], obtendo-se uma malha de triângulos a partir da interpolação de superfícies em uma grade de cubos. Considerando que o foco do trabalho é a etapa de alinhamento, maiores detalhes das demais etapas do *pipeline* podem ser obtidas em Vrubel [35].

1.2 Motivação e objetivos

Neste trabalho são propostas contribuições para duas limitações existentes no *pipeline* desenvolvido pelo grupo IMAGO [35], visando simplificar significativamente as primeiras etapas.

No pipeline original [35], regiões que não tenham sido devidamente digitalizadas só poderiam ser identificadas após a integração das vistas. Como a etapa da integração depende do alinhamento e ambas originalmente eram processos lentos, tinha-se o resultado visual do que foi digitalizado apenas em um outro momento, que poderia inclusive ser outro dia. E tratando-se de acervos naturais e culturais nem sempre é viável ter acesso

prolongado aos objetos de forma que se possa digitalizar uma segunda vez para fazer a aquisição de regiões que não foram digitalizadas, além de ser muito trabalhoso.

Não ter uma forma prática de visualizar o que foi digitalizado também favorece que o operador do *scanner* adquira vistas redundantes, que não trazem contribuição para o processo e ainda atrapalham por aumentarem a quantidade de dados para serem processados e armazenados.

Para esta limitação é proposto neste trabalho uma forma de visualizar todas as vistas alinhadas compondo o que foi capturado até o momento durante a etapa de aquisição.

A segunda limitação identificada é a ausência de uma abordagem automática para o pré-alinhamento, este processo até então no pipeline [35] era feito de forma manual. Cabia ao usuário definir pares de vistas com sobreposição para todas as vistas e gerar o pré-alinhamento utilizando comandos do mouse em uma interface gráfica para rotacionar e transladar as vistas até ficarem aproximadamente alinhadas. Este processo é especialmente trabalhoso para objetos de grande porte, como as esculturas barrocas que são apresentadas do capítulo 4, nas figuras 4.5 até 4.9. Elas possuem perto de 2 metros de altura e foram necessárias entre 100 e 170 vistas para cada.

Para esta limitação a contribuição foi o desenvolvimento de uma nova abordagem de pré-alinhamento automático, integrado ao *pipeline*.

Visando tornar mais ágil e simplificado o processo de modelagem 3D do *pipeline*, este trabalho tem por objetivos:

- Realizar um estudo sobre as abordagens para pré-alinhamento.
- Implementar um método de pré-alinhamento automático integrado ao *pipeline* [35] desenvolvido no grupo IMAGO.
- Implementar uma ferramenta que facilite a identificação de regiões não digitalizadas em tempo real, durante o processo de aquisição de vistas do objeto.

1.3 Organização do trabalho

Este trabalho está organizado em 5 capítulos, sendo o primeiro a introdução, o segundo uma revisão sobre os algoritmos de pré-alinhamento estado da arte, o terceiro traz em detalhes a abordagem proposta para o pré-alinhamento e as demais ferramentas implementadas. O capítulo 4 mostra experimentos sobre as implementações realizadas seguido da conclusão contendo também trabalhos futuros no capítulo 5.

CAPÍTULO 2

MÉTODOS DE PRÉ-ALINHAMENTO

A etapa de alinhamento consiste em colocar todas as vistas em um mesmo sistema de coordenadas, ou seja descobrir a transformação geométrica mais adequada para posicioná-las. Existem métodos que trabalham em todas as vistas simultaneamente como Blais [7] que adota o método de otimização *Simulated Annealing*. Outra técnica foi apresentada por Masuda [24,25], ele utiliza um algoritmo robusto a dados discrepantes e se baseia no casamento de um descritor chamado *Signed Distance Fields*. E há também o método de Silva *et al.* [33] que utiliza algoritmos genéticos e propõe uma nova medida de erro chamada *Surface Interpenetration Measure*. Porém, devido à quantidade de dados utilizada usualmente no cenário deste trabalho estes métodos apresentam custo computacional proibitivo no que diz respeito a memória e processamento [35].

Uma alternativa de menor custo computacional aos métodos que trabalham com o alinhamento de todas as vistas simultaneamente seria trabalhar com elas aos pares. As vistas são organizadas em pares de forma análoga a um grafo conexo e com a exigência de que haja sobreposição entre os pares, e com isso é possível tornar o alinhamento de todas as vistas diretamente ou transitivamente referentes ao sistema de coordenadas de uma mesma vista.

Sobre o alinhamento de pares de vistas o algoritmo *Iterative Closest Point* (ICP) [6] é um método muito conceituado e amplamente adotado com sucesso em diversos trabalhos de preservação digital [20,31,35]. Porém o método ICP, como algoritmo de otimização, é sensível a mínimos locais, pois ele converge para o alinhamento mais próximo que minimize a distância entre dois conjuntos de pontos. Portanto para garantir um resultado adequado, é necessário que exista uma boa estimativa inicial de alinhamento. E os métodos de pré-alinhamento são os responsáveis por encontrar essa estimativa inicial para a transformação

que alinha pares de vistas com sobreposição.

A maioria das técnicas de pré-alinhamento buscam correspondências entre características de duas imagens, sejam elas pontos, faces, curvaturas, entre outras características. E uma vez identificadas estas características, é gerada uma transformação que minimiza a distância entre essas correspondências. As estratégias para encontrar correspondências variam muito, algumas técnicas procuram por características especiais que possam ser casadas em ambas as imagens, por exemplo *Spin-images* [19] e *Point Signature* [11]. Outras inferem correspondências procurando por conjuntos de pontos com alguma propriedade euclidiana ou algébrica em especial, como o DARCES [8] e o 4PCS [1]. Uma exceção é o método de alinhamento baseado em *Principal Component Analysis* [12] que alinha pares de vistas com sobreposição sem utilizar correspondências e sim com o alinhamento dos centros de massa e os eixos principais destas vistas. A seguir são apresentados os algoritmos estado da arte estudados neste trabalho.

2.1 4PCS

O 4PCS [1] é um método de pré-alinhamento desenvolvido recentemente, mas que já é adotado em vários centros de pesquisa de visão computacional. Ele propõe ser rápido e ao mesmo tempo robusto a ruídos. Sua técnica consiste em inicialmente selecionar um conjunto base de 4 pontos na primeira imagem, sendo 3 pontos aleatórios e um quarto ponto aproximadamente coplanar aos outros 3. Em seguida propriedades invariantes a movimentos rígidos são calculadas para esses pontos, e na segunda imagem procura-se um conjunto de 4 pontos com as mesmas propriedades. A busca de correspondências fica então restrita somente aos conjuntos que possuam as mesmas propriedades dos 4 pontos da primeira imagem, e o alinhamento é gerado para o conjunto de melhor correspondência. Essa estratégia torna o método ágil por restringir o espaço de busca de candidatos somente aos pontos aproximadamente coplanares de mesma propriedade da base. Entretanto sua maior limitação está em pares de vistas que possuam pouca região de sobreposição, abaixo de 50%, como será mostrado no capítulo 4.

Para uma melhor compreensão do método 4PCS considere a, b, c, d sendo 4 pontos coplanares na primeira imagem e e o ponto de intersecção entre as retas formadas por ab e cd . Duas medidas $r1$ e $r2$ respectivamente as proporções do comprimento de ae em relação a ab e ce em relação a cd são calculados como mostram as equações 2.1 e 2.2:

$$r1 = \frac{\|a - e\|}{\|a - b\|} \quad (2.1)$$

$$r2 = \frac{\|c - e\|}{\|c - d\|} \quad (2.2)$$

Na segunda imagem todos os pontos são combinados em pares e cada par tem seus dois pontos intermediários calculados relativos a $r1$ e $r2$ como mostram as equações 2.3 e 2.4:

$$e1 = q1 + r1(q2 - q1) \quad (2.3)$$

$$e2 = q1 + r2(q2 - q1) \quad (2.4)$$

Nas equações 2.3 e 2.4, $q1$ e $q2$ formam o par de pontos analisados na segunda imagem, $e1$ e $e2$ são pontos intermediários que possuem a mesma proporção de distância que e em relação a ab e cd , respectivamente.

Caso a distância euclidiana entre o ponto $e1$ de um par de pontos na segunda imagem e $e2$ de outro par de pontos na mesma imagem esteja abaixo de um pequeno limiar, provavelmente os dois pares são correspondentes da segunda imagem aos pontos a, b, c, d da primeira.

A figura 2.1 ilustra toda essa situação, para todos os pares de pontos na segunda imagem, na figura 2.1 (c), $e1$ e $e2$, são calculados conforme as equações 2.3 e 2.4, nas cores cinza e amarelo respectivamente. O $e1$ de ab e o $e2$ de cd devem coincidir no

ponto e independente de qualquer movimento de translação ou rotação que se faça no conjunto de pontos da primeira imagem. Portanto os pontos $q1$, $q4$, $q5$, $q3$ da segunda imagem que tem $e1$ e $e2$ aproximadamente na mesma posição, são fortes candidatos a serem correspondentes de a , b , c , d .

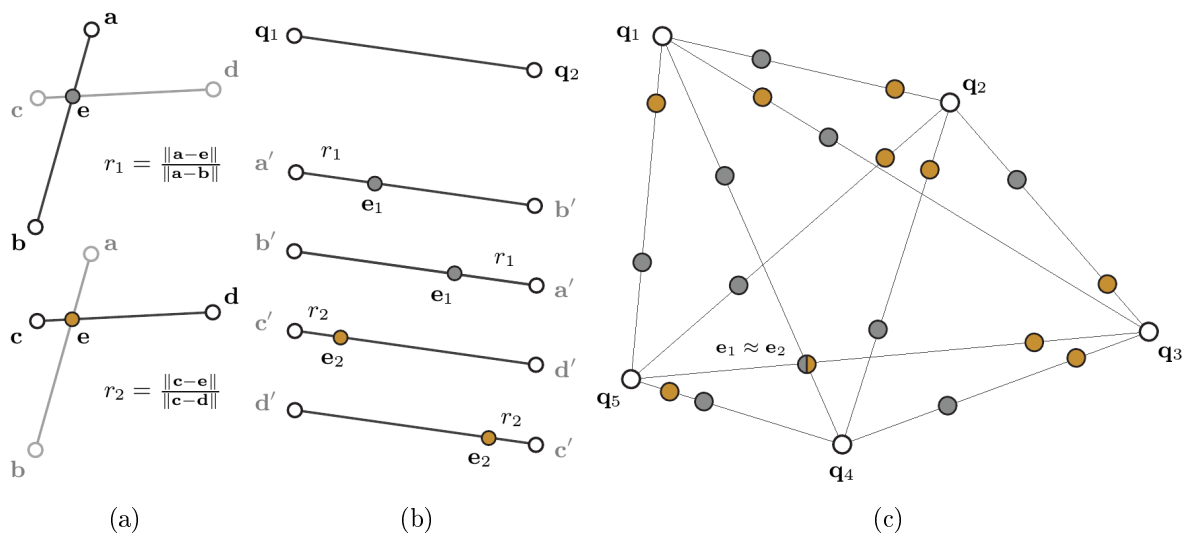


Figura 2.1: Exemplicação do método 4PCS, extraída do artigo [1]. Na figura (a) pode-se ver em cinza o ponto intermediário entre a e b , e em amarelo o ponto intermediário entre c e d . Na figura (b) é exemplificado o cálculo de pontos intermediários candidatos $e1$ e $e2$ para os pares de pontos de outra imagem de profundidade. Na figura (c) nota-se que o pontos intermediários candidatos $e1$ e $e2$ dos pontos $q5$, $q3$, $q4$, $q1$ estão localizados aproximadamente na mesma posição o que faz desses pontos correspondentes a a , b , c , d da primeira vista.

Para todos os conjuntos de 4 pontos com essa característica são escolhidos aqueles que geram a melhor transformação pelo critério *Largest Common Point set* (LCP), ou seja, pelo maior conjunto de pontos em comum.

Um conceito importante sobre o 4PCS é o tamanho da base formada pelos 4 pontos aproximadamente coplanares escolhidos da primeira imagem. Quanto maior a distância entre os pontos da base, maior será a precisão do alinhamento obtido e mais robusto ele se torna frente a ruídos [1]. Isso ocorre pelo fato de pequenas imprecisões na posição dos pontos se tornarem proporcionalmente maiores quando relacionadas a bases de tamanho menor. Bases pequenas também fazem o algoritmo se tornar mais lento pois no caso geral aumentam a quantidade de combinações existentes que devem ser avaliadas. Mas ao

mesmo tempo quanto maior a base, maior a chance de um dos pontos não fazer parte da região de sobreposição entre as imagens, por isso o tamanho da base é calculado de acordo com uma estimativa do tamanho da região de sobreposição entre o par de vistas. Caso essa não seja indicada, o algoritmo é aplicado diversas vezes para diferentes tamanhos decrescentes de base. E ao final o melhor resultado é considerado como o melhor pré-alinhamento.

Essas medidas tornam o método robusto a ruídos e imprecisões, quando a área de sobreposição for suficientemente grande, o que permite ao 4PCS ser aplicado diretamente sobre dados brutos, sem necessidade de tratamento prévio.

2.2 Spin-image

Spin-image [19] é o nome dado a um descritor invariante a movimentos rígidos que é associado a um ponto orientado de uma superfície. Esse descritor foi inicialmente desenvolvido para o reconhecimento de imagens 2D, mas também é adotado como ferramenta para algoritmos de alinhamento.

O método é simples, *Spin-images* são geradas para pontos de ambas as imagens, e então comparadas entre si. Aquelas que forem mais semelhantes são consideradas correspondentes, e a partir destas correspondências são geradas as transformações, sendo que as que resultarem no melhor alinhamento são adotadas como pré-alinhamento.

O descritor *Spin-image* pode ser entendido como uma imagem bidimensional que descreve um ponto p em uma superfície. A partir de p , um plano tangente e o vetor normal são determinados. Para cada ponto dentro de uma vizinhança de p , a distância α relativa ao vetor normal de p , e a distância β relativa ao plano tangente de p são calculadas conforme as equações 2.5 e 2.6.

$$\alpha = \sqrt{\|x - p\|^2 - (n(x - p))^2} \quad (2.5)$$

$$\beta = n(x - p) \quad (2.6)$$

Nas equações 2.5 e 2.6 p é o ponto referencial da *Spin-image*, n é a sua normal e x é o conjunto de pontos vizinhos adotados para gerar a *Spin-image*.

Uma imagem com *pixels* de valor zero é criada e as distâncias calculadas são adotadas para definir a posição do ponto na imagem. α é discretizado para um valor inteiro no eixo X e β para o eixo Y . O número na célula é usado para identificar a quantidade de pontos que pertencem àquela região. Esse processo pode ser melhor entendido com a figura 2.2.

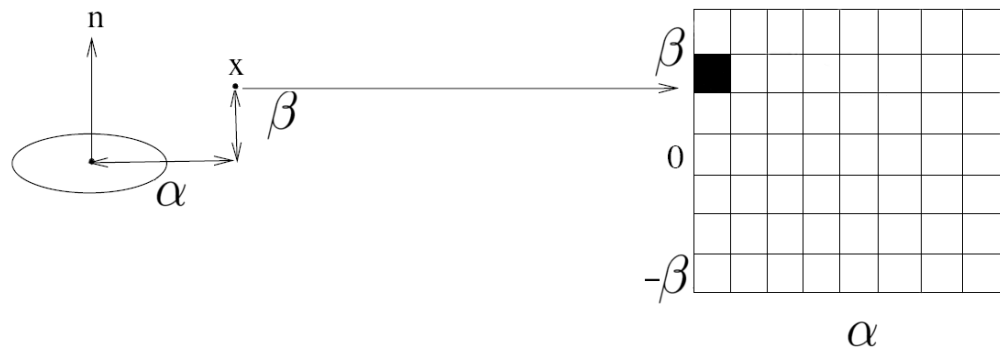


Figura 2.2: Representação de como é criado o descritor *Spin-image*. À esquerda pode-se observar um ponto x que possui uma distância α até a normal do ponto descrito, e uma distância β ao plano tangente do mesmo. A direita pode-se observar o ponto de coordenadas α e β na *Spin-image* tem seu valor acrescido em 1 (bloco preto).

A figura 2.3 mostram duas vistas, renderizadas em cores diferentes de um mesmo objeto, onde é possível perceber a semelhança das *Spin-images* entre pontos correspondentes em vistas distintas.

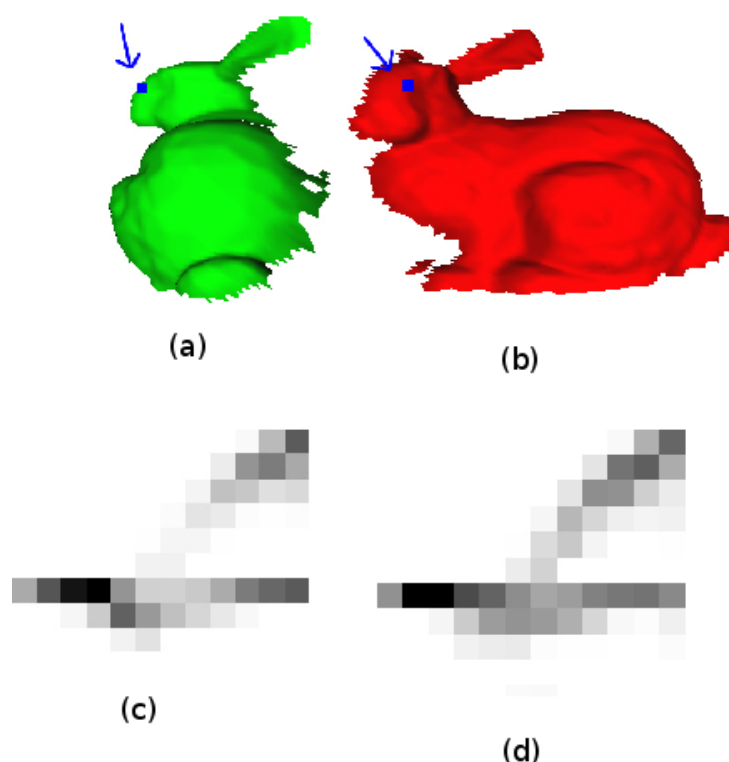


Figura 2.3: Representação de duas *Spin-images* de dois pontos correspondentes em duas imagens de profundidade diferentes de um mesmo objeto: (a) e (b) são as representações dos pontos, (c) e (d) são as *Spin-images* destes pontos respectivamente. Imagem retirada da tese de doutorado de Johnson [19].

O algoritmo de alinhamento por *Spin-image*, começa com a geração de *Spin-images* para todos os pontos da primeira imagem, em seguida elas são geradas para um grupo de pontos dispersos, definidos aleatoriamente na segunda imagem, usualmente é utilizada uma amostragem de 10% dos pontos.

A etapa seguinte do algoritmo é a de correspondência. Para cada *Spin-image* da segunda imagem, a diferença é calculada em relação a todas as *Spin-image* na primeira imagem. As que tiverem menor diferença são consideradas possíveis correspondentes, onde a diferença é calculada através da soma das diferenças *pixel a pixel* da *Spin-image*.

Para garantir que sejam escolhidas as melhores correspondências, uma etapa de filtragem é realizada sobre as correspondências candidatas. A média e o desvio padrão das diferenças de todas as *Spin-images* são calculadas, e 3 vezes o desvio padrão é adotado como limiar para escolha das correspondências mais precisas [19].

Um grupo de correspondências é utilizado para gerar uma transformação e o erro dela é calculado através da soma das distâncias entre pontos correspondentes. Caso o erro esteja acima de um limiar pré-determinado, uma nova tentativa de agrupamento é realizada, até que não existam mais possibilidades de agrupamento ou que se encontre uma transformação válida.

Um ponto importante no método *Spin-image* é a resolução das imagens, pois é vital que a distribuição dos pontos em ambas as imagens seja uniforme para produzir *Spin-images* consistentes. Uma proposta de solução para garantir a uniformidade dos pontos pode ser vista em Johnson [19], onde é proposta a geração de novos pontos através da interpolação de pontos entre os triângulos, até que se tenham duas malhas de pontos uniformes. *Spin-images* é tido como um método bastante robusto por testar diferentes agrupamentos ao gerar alinhamento, uma limitação do método é a sensibilidade a falhas em imagens que possuam simetrias, ou regiões repetidas. A simetria é de fato um dos grandes desafios na construção de métodos de pré-alinhamento e uma limitação para a maioria dos métodos de pré-alinhamento existentes.

A abordagem proposta valida as correspondências geradas em um teste que cruza as candidatas a correspondência entre si, o que torna a abordagem menos sensível a regiões repetidas e simétricas. A abordagem proposta é vista em detalhes no capítulo 3.

2.3 DARCES

O método *Data-Aligned Rigidity Constrained Exhaustive Search* (DARCES) [8] baseado em *Random Sample Consensus* (RANSAC) [16] trabalha com a idéia de busca por correspondências através da força bruta. Na primeira imagem um conjunto de 3 pontos é selecionado aleatoriamente, com a condição de que estes formem um triângulo aproximadamente equilátero. Na segunda imagem é realizada uma busca por pontos que formem um triângulo equivalente ao da primeira imagem. A busca possui uma natureza exaustiva, garantindo que se uma resposta existir ela será encontrada, utilizando-se somente a

informação da geometria da nuvem de pontos do objeto.

Considere que o triângulo aproximadamente equilátero construído na primeira imagem contenha os pontos $s1$, $s2$, $s3$, como mostrado na figura 2.4.

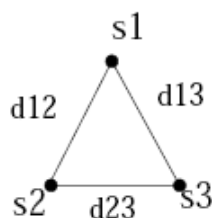


Figura 2.4: Pontos na primeira imagem, formando um triângulo aproximadamente equilátero.

O objetivo é encontrar pontos equivalente na segunda imagem, $m1$, $m2$, $m3$, correspondentes a $s1$, $s2$, $s3$ respectivamente. Como não se pode fazer nenhuma inferência sobre a posição de $m1$, assume-se que um ponto qualquer escolhido aleatoriamente na segunda imagem seja $m1$. Todos os pontos da segunda imagem são possíveis candidatos para $m1$, portanto eles são todos empilhados em uma ordem aleatória e avaliados um a um.

Para realizar a busca por $m2$ assume-se como candidatos todos os pontos que estejam de $m1$ a uma distância igual a que existe entre $s1$ e $s2$, visto que a distância é invariante sob transformações rígidas. A busca é realizada dentro de uma esfera com centro em $m1$ e raio equivalente à distância da reta $s1s2$, como mostrado na figura 2.5. Todos esses candidatos a $m2$ também são empilhados para serem avaliados.

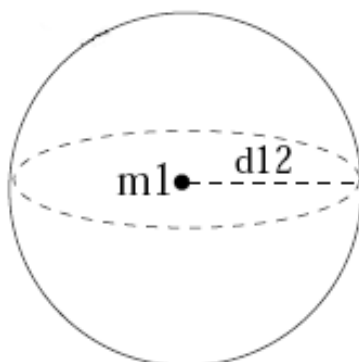


Figura 2.5: Esfera onde é realizada a busca por pontos candidatos a $m2$.

Para realizar a busca por $m3$ assume-se como candidatos os pontos que estejam a uma distância da reta $m1m2$ igual a distância de $s3$ e a reta $s1s2$. Essa busca é realizada na interseção das duas esferas de centro em $m1$ e $m2$ e raio que seja a distância de $s1s3$ que é a mesma $s2s3$ visto que se trata de um equilátero, como na figura 2.6. Os candidatos a $m3$ também são empilhados.

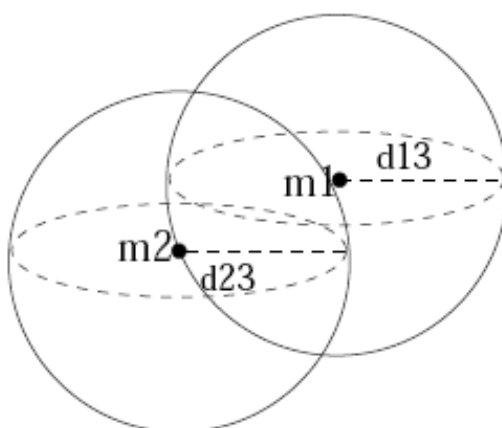


Figura 2.6: Interseção das esfera onde os pontos candidatos a $m3$ são procurados.

Para cada conjunto candidato a $m1$, $m2$, $m3$ uma transformação é calculada e o erro dessa transformação é avaliado. O erro neste algoritmo é calculado através da soma da distância dos pontos da primeira imagem com relação ao ponto que esteja mais próximo na segunda imagem. Por questões de otimização, apenas um conjunto esparsos dos pontos

da primeira imagem é avaliado, aproximadamente 10% dos pontos. Caso esteja dentro de um limiar aceitável essa transformação é aceita como a correta, caso contrário continua-se avaliando novas transformações candidatas. Caso nenhuma transformação candidata seja considerada válida, os pontos $s1$, $s2$, $s3$ escolhidos não fazem parte da região de sobreposição das duas imagens. Consequentemente novos pontos $s1$, $s2$, $s3$ são escolhidos e o algoritmo é repetido novamente.

A natureza de busca exaustiva do algoritmo possui um custo muito elevado para alinhar vistas com grande número de pontos, além de gerar muitos casamentos ambíguos. Definir um triângulo base adequado é um problema, pois quanto maior for o triângulo equilátero formado pelos pontos $s1$, $s2$ e $s3$, mais robusto a ruídos será a base, porém isso fará aumentar as chances dos pontos $m1$, $m2$ e $m3$ correspondentes não existirem na segunda imagem. Além disso, pode-se considerar um aumento considerável do espaço de busca, visto que a busca pelos pontos $m1$, $m2$ e $m3$ é feita dentro de uma área relativa à distância entre os pontos da base. E caso contrário, apesar de ser mais eficiente, quanto menor for o triângulo formado pelos pontos $s1$, $s2$, $s3$, mais sensível a ruído e mais candidatos $m1$, $m2$ e $m3$ serão encontrados na segunda imagem. Uma solução mais adequada pode ser definida em função do erro médio para que o menor triângulo seja adotado sem que os ruídos comprometam demais o método.

DARCES é um método considerado robusto, mas pouco eficiente para um grande número de pontos, servindo melhor em aplicações onde o tempo não é requisito.

2.4 Alinhamento Baseado em *Principal Component Analysis*

O método de alinhamento baseado em *Principal Component Analysis* (PCA) [12] é um dos métodos de pré-alinhamentos mais rápidos existentes. O par de imagens em profundidade tem seus centros de massa e eixos principais calculados. A translação que alinha os centros de massa junto com a rotação que alinha os eixos principais são a transformação resposta para o par de imagens.

O centro de massa de um conjunto de pontos é a média das coordenadas x , y , z desse conjunto. E o eixo principal é calculado através dos auto-vetores da matriz de covariância como mostra a equação 2.7.

$$Cov = \frac{1}{N} \sum_{i=0}^{N-1} (p_i - \bar{p})(p_i - \bar{p})^T \quad (2.7)$$

Na equação 2.7, N é o número de pontos, \bar{p} é o centro de massa, p_i é o ponto corrente da imagem que está sendo avaliado.

Para cada imagem aplica-se uma decomposição para obter a direção U do eixo principal conforme a equação 2.8.

$$Cov = UDU^T \quad (2.8)$$

A decomposição resulta em U que é a direção do eixo principal da imagem e D que é a matriz de escala.

A rotação é definida pelo produto dos auto-vetores dos eixos principais das duas imagens conforme mostra a equação 2.9. E a translação se consegue alinhando os dois centros de massa das imagens conforme a equação 2.10.

$$R = U_1 U_2^{-1} \quad (2.9)$$

Onde U_1 é a direção do eixo principal da primeira vista e U_2 a direção do eixo da segunda vista.

$$t = \mu_2 - R\mu_1 \quad (2.10)$$

Onde μ_2 é o centro de massa da segunda vista, μ_1 é o centro de massa da primeira e R é a rotação entre as vistas.

Os eixos principais indicam a direção da maior variabilidade dos pontos. Os centros de massa e os eixos principais variam pouco caso o conjunto de pontos também varie pouco. Por essa razão, alinhar os eixos principais e os centros de massa resulta em uma melhor aproximação do alinhamento de duas vistas dependendo do tamanho da região de sobreposição.

Caso as imagens possuam uma região de sobreposição suficientemente grande, em geral acima de 70%, os eixos principais de ambas devem ser semelhantes, e ao calcular uma transformação que alinhe esses eixos, conseqüentemente irá alinhar as duas imagens.

Este método, embora seja bastante rápido, tem sérias restrições de uso: precisa de uma grande área de sobreposição, é sensível a ruído e pode gerar transformações totalmente incorretas quando se trabalha com objetos que possuem simetrias.

2.5 Método de Pingi *et al.*

O método proposto por Pingi *et al.* [28] não é feito para um cenário genérico, ele foi desenvolvido especialmente para o pré-alinhamento de um grande conjunto de imagens de profundidade. O método parte do pressuposto que a aquisição das imagens foi feita seguindo um padrão de orientação. Este padrão pode ser circular, vertical ou horizontal, que é um cenário comum para a digitalização de objetos de grande porte, como estátuas ou murais onde se utiliza um *scanner 3D* fixo sobre um tripé móvel.

Semelhante ao *Spin-image* descrito anteriormente na seção 2.2, prováveis correspondências são encontradas com o uso de um descritor próprio, que é aplicado sobre um conjunto de pontos da primeira vista e comparado contra todos os pontos da segunda.

Diferente do *Spin-image*, que utiliza um conjunto esparsos de pontos aleatórios de uma vista para comparar com os pontos da outra, neste método há critérios para essa seleção. A vizinhança de cada ponto no mapa de profundidade é utilizada para o cálculo da variância da orientação das normais. Caso essa variância seja muito baixa, significa que o ponto está em uma região muito lisa, portanto não serve como boa correspondência. E caso

a variância seja muito elevada, apesar de ser um ponto bastante característico, a probabilidade deste ponto não existir na outra vista por estar em uma região de reentrâncias é alta. Por isso só são selecionados aqueles pontos que possuam um valor intermediário entre o limiar superior e inferior de variância. Esta situação é mostrada na figura 2.7.

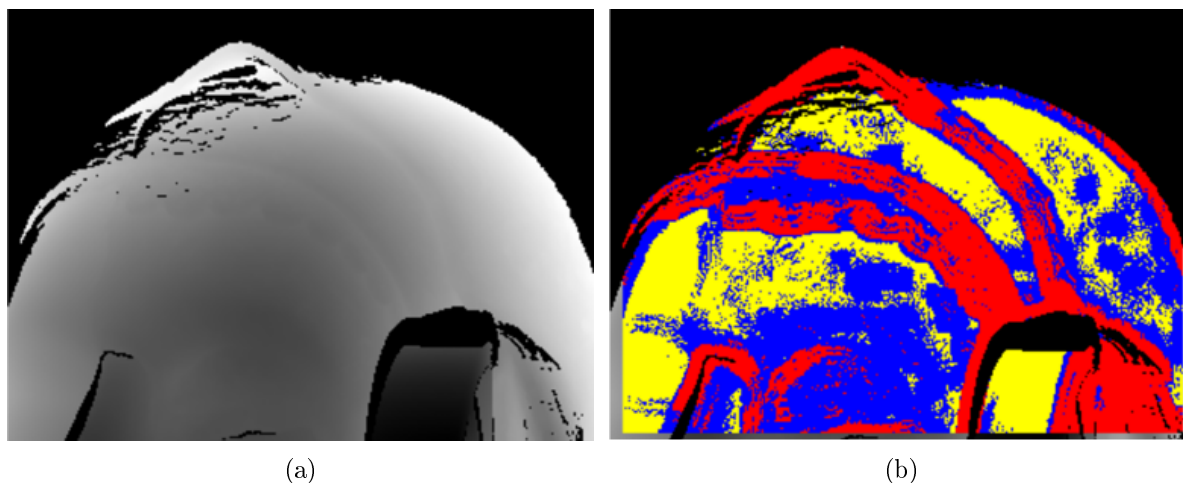


Figura 2.7: Exemplo de imagem de variância usadas no método de Pingi *et al.*: (a) imagem de profundidade e (b) imagem de variância onde os pontos em vermelho são de alta variância, pontos em azul tem variância média e pontos em amarelo variância baixa. Imagens retiradas do artigo de Pingi *et al.* [28].

O descritor de Pingi é uma matriz 13×13 que descreve um ponto em uma superfície de um mapa de profundidade. Em cada posição da matriz está o produto interno entre o vetor normal do ponto descrito com relação ao vizinho daquela posição.

Esse descritor é gerado para um conjunto de pontos de variância média na primeira imagem, e gerado também para todos os pontos da segunda imagem. Cada descritor da primeira imagem é comparado com todos os descritores da segunda e associado com aquele que for mais semelhante. A soma das diferenças entre os descritores é utilizada como critério de semelhança.

De todos esses pares de pontos correspondentes, um conjunto pequeno, normalmente 20 a 30 pares, é selecionado aleatoriamente. Uma combinação 3 a 3 é gerada para esse conjunto e uma transformação é calculada. O erro da transformação é avaliado pela soma das distâncias entre os pares de pontos correspondentes. Caso o erro da transformação

esteja abaixo de um limiar pré-definido o algoritmo encerra com essa transformação como resposta, caso contrário as outras combinações são avaliadas. Se o erro de nenhuma delas for abaixo de certo limiar, outro conjunto de pontos é avaliado até um número fixo de iterações e a melhor transformação obtida até então é adotada como solução.

O método é rápido e eficiente porém o seu descritor só é válido com pares de imagens que possuam pouca rotação entre si, e não é invariante a todo tipo de movimentos rígidos como os demais métodos estudados.

2.6 *Point Signature*

O *Point Signature* [11], semelhante ao *Spin-image*, adota o uso de um descritor de pontos em superfície. Inicialmente proposto para o reconhecimento de imagens, também tem aplicação no alinhamento 3D, o descritor é invariante a movimentos rígidos, o que permite seu uso irrestrito para a busca de correspondências nas mais diversas imagens.

A partir de um ponto p que é centro de uma esfera de raio r , os pontos de interseção da esfera com a superfície formam um contorno C , e os pontos de C são representados em uma novo sistema de coordenadas. A figura 2.8 ilustra a situação. Esse sistema de coordenadas é formado pelos eixos, $n1$ que é o vetor normal de p , $n2$ um vetor "referência" e um terceiro vetor formado pelo produto vetorial de $n1$ e $n2$. Os pontos C são projetados para o plano tangente formando o conjunto C' . O vetor $n2$ é calculado como um vetor unitário de p para um ponto em C' do qual ele possui a maior distância. Cada ponto em C terá duas características que o identifica, a distância positiva ou negativa dele em relação a sua projeção em C' e o ângulo de rotação no sentido horário θ a partir do vetor de referência $n2$.

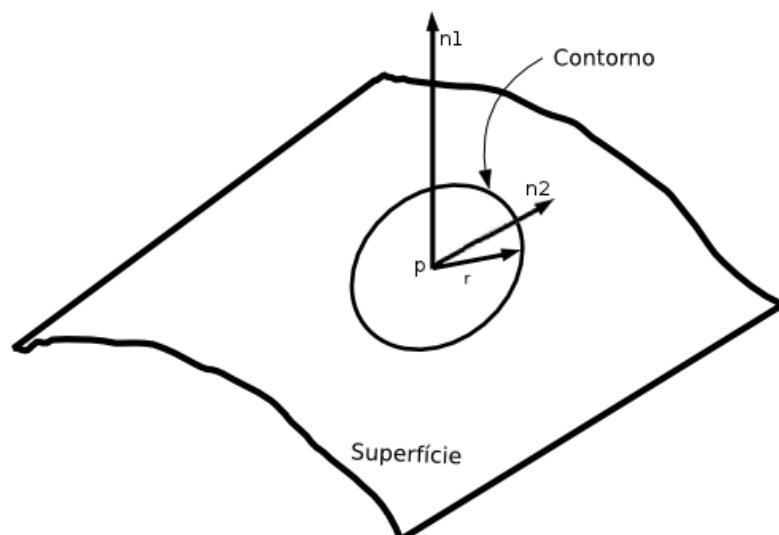


Figura 2.8: Representação do contorno de *Point Signature* em uma superfície.

Os *Point Signatures* das duas vistas são comparados e aqueles que forem semelhantes são considerados como candidatos a correspondentes, essa etapa é bastante eficiente. A representatividade do ponto é reduzida de 3D para 1D bastando calcular a diferença entre 2 vetores para avaliar o quão semelhantes são.

A principal desvantagem deste método é o elevado custo para se calcular os descritores, pois não é trivial encontrar a interseção entre uma esfera e uma superfície formada por pontos e arestas. Neste caso uma interpolação é necessária, o que torna o processo mais trabalhoso e menos preciso, além disso o método é suscetível a ruído no cálculo do vetor referência.

2.7 Comparação

Dentro do nosso escopo precisamos de um método que tenha 3 características: (1) baixo custo computacional; (2) robustez a baixa sobreposição; (3) robustez a rotações. Os métodos apresentados neste capítulo se destacam entre os métodos de pré-alinhamento da literatura porém todos apresentaram limitações dentro dessas 3 características, isso nos motivou a desenvolver um método próprio baseado no conceituado algoritmo de detecção e descrição de pontos característicos *Speeded Up Robust Features* (SURF) [3], o método

proposto será apresentado em detalhes no próximo capítulo. A tabela 2.1 compara os métodos apresentados no que diz respeito as 3 características citadas.

Tabela 2.1: Comparação entre os métodos de pré-alinhamento

| Métodos | Baixo custo computacional | Robustez a pouca sobreposição | Robustez a rotação |
|---------------------|---------------------------|-------------------------------|--------------------|
| 4PCS | X | | X |
| <i>Spin-Image</i> | | X | X |
| DARCES | | X | X |
| PCA | X | | X |
| Pingi <i>et al.</i> | X | X | |
| Point Signature | | X | X |
| Método proposto | X | X | X |

CAPÍTULO 3

IMPLEMENTAÇÕES

Após realizar um estudo sobre os métodos de pré-alinhamento apresentados no último capítulo, foi possível identificar que no geral eles apresentam forte sensibilidade com relação ao tamanho da região de sobreposição. Quanto menor for a região de sobreposição, maior será a tendência para o alinhamento ser impreciso ou ainda lento. E essa limitação foi um fator decisivo para o desenvolvimento da abordagem proposta para o pré-alinhamento.

Objetos de pequeno e grande porte oferecem cenários distintos na tarefa de digitalizá-los. Os objetos de pequeno porte como esculturas de altura inferior a 30 cm em média, tal qual mostrado na figura 3.1, ao serem digitalizados é comum que um grande percentual da superfície total do objeto seja coberta em uma única vista do *scanner*. Isso possibilita, com o uso de poucas vistas, a digitalização de toda a superfície do objeto mantendo alto percentual de sobreposição entre vistas próximas. Outra facilidade na digitalização de acervos de pequeno porte é que normalmente são facilmente manipuláveis simplificando a captura dos ângulos desejados. Para esse cenário os métodos de pré-alinhamento no geral apresentam uma boa taxa de acerto e desempenho.



Figura 3.1: Cerâmica pintada com aproximadamente 20 centímetros de comprimento feita pela comunidade indígena dos Karajas. Objeto do acervo do Museu de Arqueologia e Etnologia da UFPR.

Porém quando se trata de objetos de grande porte, como o da figura 3.2 que mostra uma estátua com aproximadamente 2 metros de altura, garantir que toda a superfície do objeto seja coberta mantendo um alto percentual de sobreposição entre as vistas, se torna muito trabalhoso ou mesmo impossível, dependendo da localização e condições em que o objeto se encontra.

Em média a sobreposição entre os pares utilizados no alinhamento das imagens de profundidade da estátua da figura 3.2 utilizando o pipeline desenvolvido em [35], foi de 35%. Para este cenário os métodos de pré-alinhamento estudados neste trabalho não foram satisfatórios, tanto pela obtenção de resultados imprecisos, quanto pelo elevado custo computacional. No próximo capítulo experimentos que demonstram essa situação foram realizados com alguns dos métodos apresentados.



Figura 3.2: Escultura feita em pedra sabão com aproximadamente 2 metros representando o profeta Jonas, feita por Antônio Francisco Lisboa, "O Aleijadinho". Ela compõe a obra "Os Doze Profetas" e está localizada no Santuário do Bom Jesus de Matosinhos, um dos patrimônios Mundiais da UNESCO, na cidade de Congonhas - MG.

Em função das limitações dos métodos estudados, foi desenvolvido neste trabalho um novo método para o pré-alinhamento de imagens de profundidade, que se mostrou rápido e robusto a baixa taxa de sobreposição. A idéia básica se assemelha ao que foi apresentado por Pingi *et al.* [28] porém utilizando um descritor de pontos mais robusto.

Quando objetos fixos e de grande porte como a estátua da figura 3.2 são digitalizados, no *pipeline* proposto é utilizado um padrão de digitalização de vistas que segue uma orientação vertical.

Inicialmente o *scanner*, que é acoplado a um tripé móvel, é posicionado na altura do objeto à 90° em relação ao chão. A partir desta posição o *scanner* é deslocado para

baixo, capturando vistas do objeto em espaçamentos uniformes de forma que cada nova vista possua sobreposição com sua predecessora formando uma "faixa" vertical. Uma vez obtida as vistas da faixa, o tripé é deslocado ao redor da escultura para a captura de uma nova faixa. Este processo se repete até que todas as superfícies do objeto sejam digitalizadas. A quantidade de deslocamento vai influenciar sobre o percentual médio de sobreposição das vistas vizinhas, no *pipeline* proposto o deslocamento é feito de forma arbitrária se adaptando a cada situação, mas em média mantendo aproximadamente 40% de sobreposição entre vistas consecutivas.

Eventuais regiões que não tenham sido devidamente digitalizadas nesse processo, geralmente regiões de oclusão, são capturadas individualmente na sequência, posicionando mais adequadamente o *scanner* de forma a viabilizar a digitalização de cada uma dessas regiões.

A figura 3.3 mostra em maiores detalhes todo este processo de movimentação do *scanner* para obtenção das vistas do objeto. Essa lógica de digitalização foi adotada por duas razões, primeiro para melhor manter o controle sobre o processo e segundo por ser uma forma mais ágil e prática de se deslocar o *scanner* devido à natureza do tripé em que o movimento vertical de descer o *scanner* é o menos trabalhoso.

Aproveitando essa propriedade de que as vistas são capturadas seguindo uma orientação vertical, horizontal ou circular com pouca rotação, foi utilizado um algoritmo que adota um descritor de ponto 2D chamado *Speeded Up Robust Features* (SURF) [3], que é mais rápido de calcular e comparar do que os descritores de pontos 3D dos algoritmos de pré-alinhamento tradicionais como o *Spin-Images* e *Point Signature*.

O SURF detecta e descreve pontos de interesse, também chamados de pontos chave, em imagens de intensidade luminosa, ou níveis de cinza. A detecção é feita utilizando a matriz de Hessian em uma aproximação da integral da imagem o que dá um grande ganho no custo computacional deste método sem comprometer sua repetibilidade. Já o descritor é feito utilizando um vetor contendo os valores da transformada de Haar em uma região quadrada e orientada ao redor do ponto de de interesse, isso também é feito com o uso da aproximação da integral da imagem para melhorar o desempenho. O *scanner* permite

usar esse tipo de descritor pois junto da imagem em profundidade o *scanner* captura uma imagem colorida da mesma cena e com ela gera-se a imagem de níveis de cinza que é utilizada pelo SURF. A técnica desenvolvida por Pingi *et al.* utiliza um descritor 2D próprio, porém para as vistas geradas pelo *scanner* usado neste trabalho este descritor não apresentou bons resultados nos experimentos realizados.

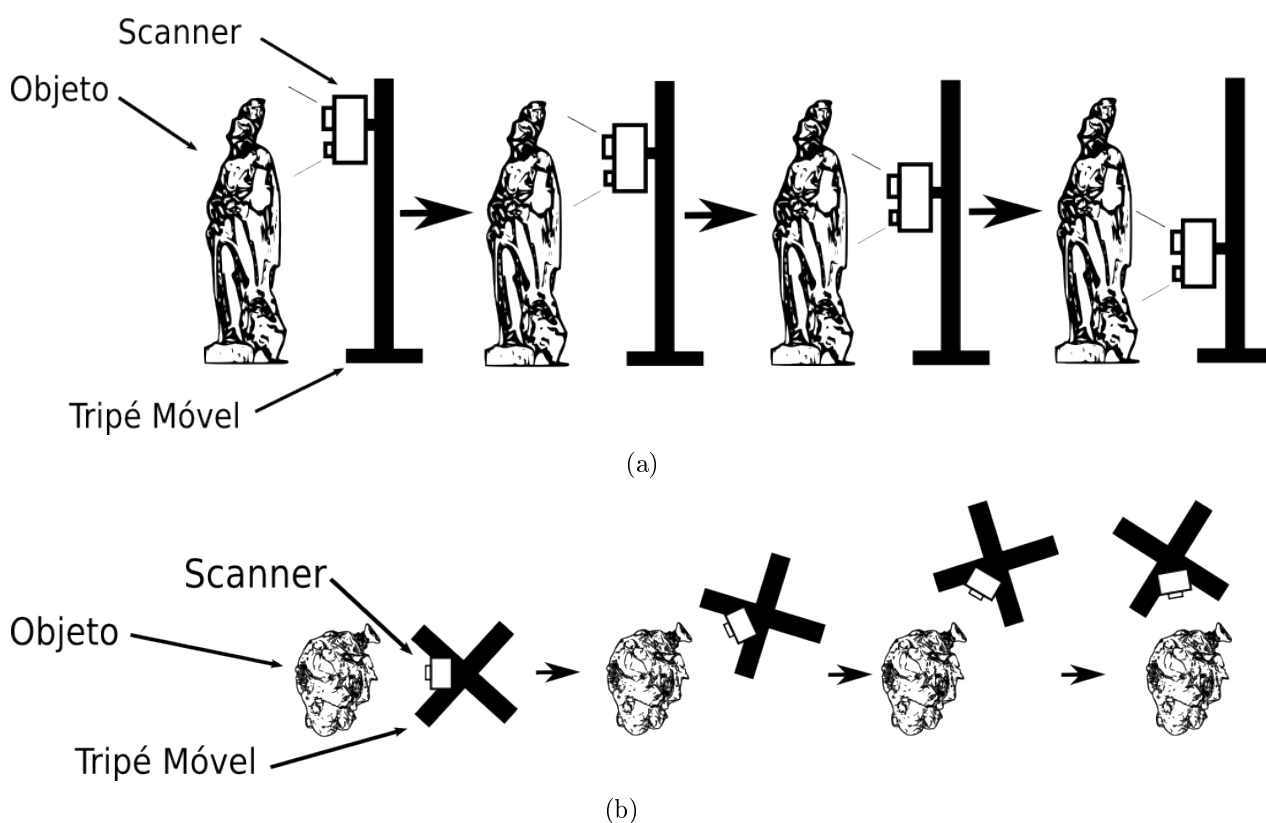


Figura 3.3: Ilustração que demonstra o processo de digitalização de objetos de grande porte no *pipeline* desenvolvido. Digitaliza-se uma "faixa" vertical como mostra a figura (a) e em seguida move-se o tripé em sentido circular como em (b) de forma a obter novas "faixas".

3.1 Algoritmo de pré-alinhamento baseado em descritores SURF

Para alinhar pares de vistas, o método basicamente segue 4 passos:

1. Geração de descritores para pontos chave;
2. Definição de correspondências entre os pontos chave;
3. Pontuação das correspondências;

4. Geração do alinhamento.

Inicialmente aplicamos o método SURF para gerar pontos chave e seus respectivos descritores nas duas imagens e desconsideramos os pontos de intensidade luminosa que não possuam correspondente na imagem de profundidade, ou seja os pontos de fundo. Como é ilustrado na figura 3.4 (a).

Em seguida para cada ponto chave da primeira vista é associado como correspondente o ponto da segunda vista que apresente menor diferença entre os descritores. Como é ilustrado na figura 3.4 (b).

Então para cada par correspondente é dada uma pontuação através de uma técnica que tira proveito da informação 3D da imagem de profundidade. Essa técnica considera que dois pontos chave $p1$ e $p2$ da primeira vista possuem uma distância euclidiana tridimensional D na imagem de profundidade. Portanto é esperado que os pontos correspondentes $pc1$ e $pc2$ referentes a $p1$ e $p2$ respectivamente possuam uma distância D' que seja bem próxima de D . A pontuação de todos os pares é iniciada com zero. Em seguida 2 a 2 todos os pares são agrupados e as distâncias D e D' são calculadas. E para cada dois pares cuja diferença entre D e D' seja inferior a um pequeno limiar, ambos os pares somam 1 em sua pontuação. Como é ilustrado na figura 3.4 (c).

Os pares que possuírem pontuação acima de determinado limiar, definido empiricamente, até um máximo de 15 pares são utilizados para geração do alinhamento. Bastam 3 pares de pontos correspondentes para gerar um alinhamento. Então 3 a 3 as melhores correspondências obtidas são agrupadas e é gerada uma rotação e translação a partir do método de Horn [17]. A transformação gerada é aplicada nos 3 pontos e o erro da transformação é calculado somando as distâncias de cada ponto para seu correspondente na outra imagem. A transformação que apresentar menor erro é considerada como resposta, ou seja o melhor alinhamento. Como é ilustrado na figura 3.4 (d).

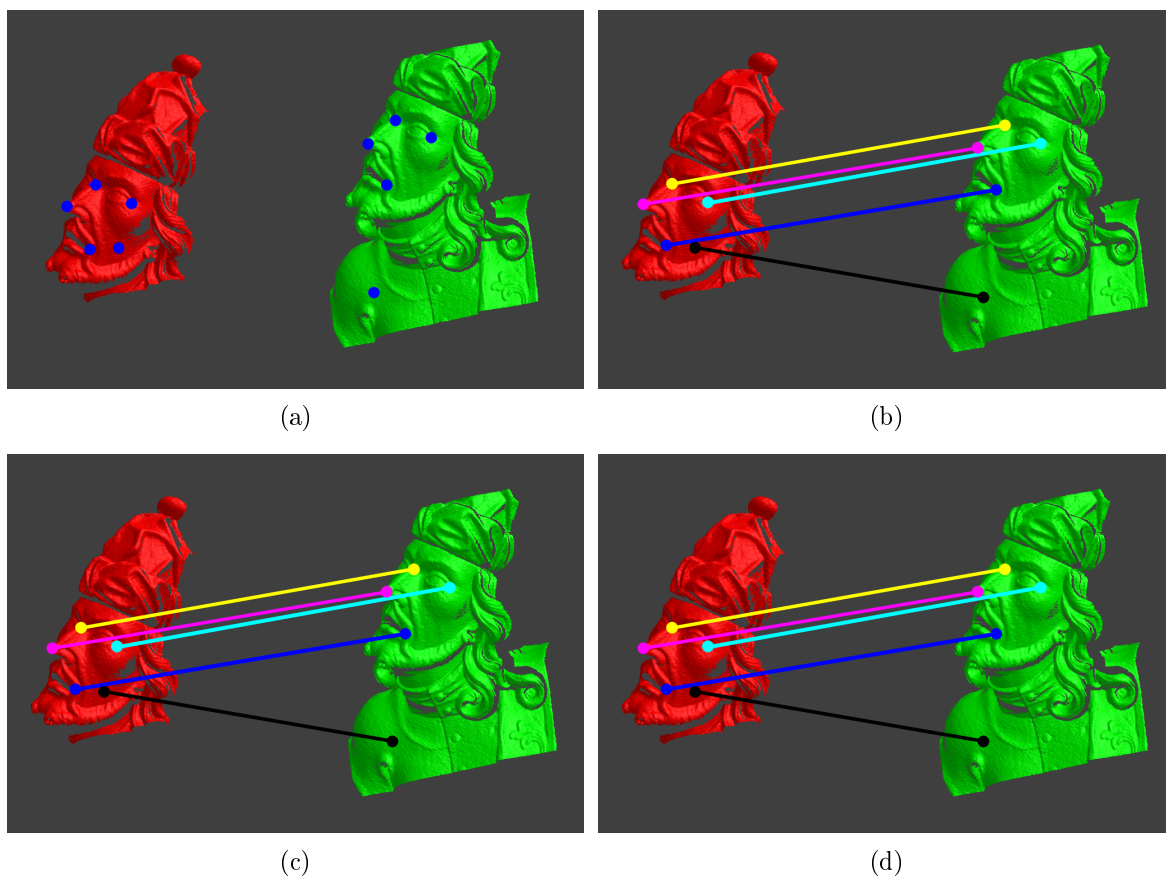


Figura 3.4: Ilustração dos quatro passos principais do método proposto.

Para ilustrar este método o pseudo-algoritmo 3.1 é apresentado a seguir.

Requer: Par de imagens de profundidade A e B

Garante: Matriz T de estimativa de alinhamento.

```

1:  $D_a \leftarrow \text{gera\_descritores\_SURF}(A)$ 
2:  $D_b \leftarrow \text{gera\_descritores\_SURF}(B)$ 
3: para todos  $d_a^i \in D_a$  faça
4:     Encontra o descritor  $d_b^k$  em  $D_b$  que seja mais semelhante ao  $d_a^i$ 
5:      $p_a^i \leftarrow a(d_a^i)$  //  $a(d_a^i) \in A$ , ponto em A correspondente ao descritor  $d_a^i$ 
6:      $p_b^i \leftarrow b(d_b^i)$  //  $b(d_b^i) \in B$ , ponto em B correspondente ao descritor  $d_a^i$ 
7: fim para
8: para todos  $p^i \in P$  faça
9:      $\text{pontos}^i = 0$ 
10:    para todos  $p^j \in P$  faça
11:         $\text{dist} = \text{distancia\_euclidiana}(p_a^i, p_a^j)$ 
12:         $\text{dist}' = \text{distancia\_euclidiana}(p_b^i, p_b^j)$ 
13:        se  $\|\text{dist} - \text{dist}'\| < \text{limiar\_dist}$  então
14:             $\text{pontos}^i ++$ 
15:        fim se
16:    fim para
17: fim para
18: para todos  $p^i \in P$  faça
19:    se  $\text{pontos}^i > \text{limiar\_pontos}$  então
20:         $C \leftarrow p^i$ 
21:    fim se
22:     $C_m \leftarrow \text{busca\_tres}(C)$  // Agrupa pares 3 a 3 e retorna o conjunto que apresentar
    menor erro
23:     $T \leftarrow \text{gera\_alinhamento\_horn}(C_m)$ 
24: fim para

```

3.2 Definição de pares com sobreposição

Definir pares que possuam sobreposição é uma tarefa trabalhosa que também é realizada de forma manual na maioria dos *pipelines* de modelagem 3D. Também é possível adotar os mesmos descritores SURF utilizados na geração do pré-alinhamento para estimar de forma rápida a existência de sobreposição.

Na abordagem proposta, para cada nova vista é gerado um conjunto de descritores SURF que são comparados aos descritores que já foram previamente gerados de todas as demais vistas. E todas as vistas cuja diferença entre descritores somem um valor abaixo de determinado limiar são consideradas vistas com provável sobreposição com a nova vista e o método de pré-alinhamento proposto é aplicado.

Pontos chave semelhantes é um bom indicativo de sobreposição, mas no caso desta técnica falhar o método proposto é aplicado nas demais vistas seguindo a ordem de semelhança dos pontos chave até encontrar uma em que o método consiga encontrar um alinhamento considerado adequado.

3.3 Abordagem proposta

Neste trabalho, foram geradas dentro de uma mesma ferramenta duas contribuições principais para o *pipeline*. Primeiro um método novo e automático de pré-alinhamento, explicado na seção 3.1. A segunda contribuição se trata de um meio de visualizar a superfície adquirida durante a etapa de aquisição das imagens.

No escopo deste trabalho, ver o resultado parcial da digitalização é muito desejável, em especial quando se trabalha com objetos de maior porte. E a forma mais natural para isso consiste em ter uma ferramenta que permita visualizar todas as vistas alinhadas simultaneamente enquanto novas vistas do objeto são adquiridas. Isso permite de forma prática identificar as regiões não-digitalizadas e monitorar eventuais erros para que seja possível tomar providências de forma imediata, que é o momento mais propício para isso.

Para que possa existir esse *feedback* visual, assim que uma nova imagem é digitalizada, é necessário identificar dentro do conjunto das vistas digitalizadas, alguma que possua sobreposição com a nova imagem e obter um alinhamento preciso para este par, para então exibir toda a composição de vistas simultaneamente. Lembrando que esse processo deve ser feito em poucos segundos para não atrasar o trabalho do operador do *scanner*.

O *software* implementado para a abordagem proposta herdou as funcionalidades do *icptool*, que é o nome do aplicativo responsável pelo alinhamento implementado no *pipeline* original e descrito em Vrubel [35]. O *software* possui duas possíveis entradas. No caso de um alinhamento novo ser iniciado é necessário apenas indicar o diretório onde as imagens em profundidade estão ou estarão. Porém no caso de ser dado continuidade a um alinhamento previamente iniciado, é necessário fornecer o arquivo de saída da última execução do alinhamento que se deseja continuar. Trata-se de um arquivo de texto contendo o nome dos arquivos das imagens de profundidade que compõe o objeto, e a relação dos pares de vistas que as alinham com seus respectivos valores de translação e rotação.

O *software* carrega as vistas de um diretório pré-determinado e opcionalmente carrega novas vistas a medida que elas forem capturadas e adicionadas ao diretório. As vistas que forem selecionadas em um *checkbox* são exibidas, e a posição que elas assumem é aquela referente ao seu alinhamento corrente; essa é a funcionalidade que permite visualizar tudo o que já foi digitalizado. Além disso está disponível a principal funcionalidade que é o conjunto de operações de alinhamento. É possível fazer o pré-alinhamento manual ou automático através do método proposto, aplicar o algoritmo do ICP para refinar o alinhamento e também utilizar o alinhamento global, que é feito através do algoritmo de Pulli [29] para distribuir o acúmulo de erro no alinhamento entre todas as vistas.

O *software* trabalha com as imagens reduzidas, que são originalmente imagens de profundidade de 640x480, porém foi escolhido utilizar uma redução através da média para 160x120. Essa redução é muito rápida para se computar e pouco interfere no alinhamento, pois como a densidade de pontos adquirida pelo *scanner* é muito grande aplicar esta redução não altera significativamente a estrutura topológica da superfície das vis-

tas. Esta redução representa uma grande melhora em desempenho tanto na questão de processamento quanto em memória, afinal se trata de quase 20 vezes menos dados para ser armazenado em memória, processado e renderizado. Isso é muito significativo para objetos de maior porte e para garantir o melhor resultado possível no alinhamento, posteriormente é possível refinar os pares de alinhamento obtidos utilizando as vistas em sua forma não reduzida. Na figura 3.5 pode-se ver um par de vistas alinhadas no seu tamanho original e reduzido.

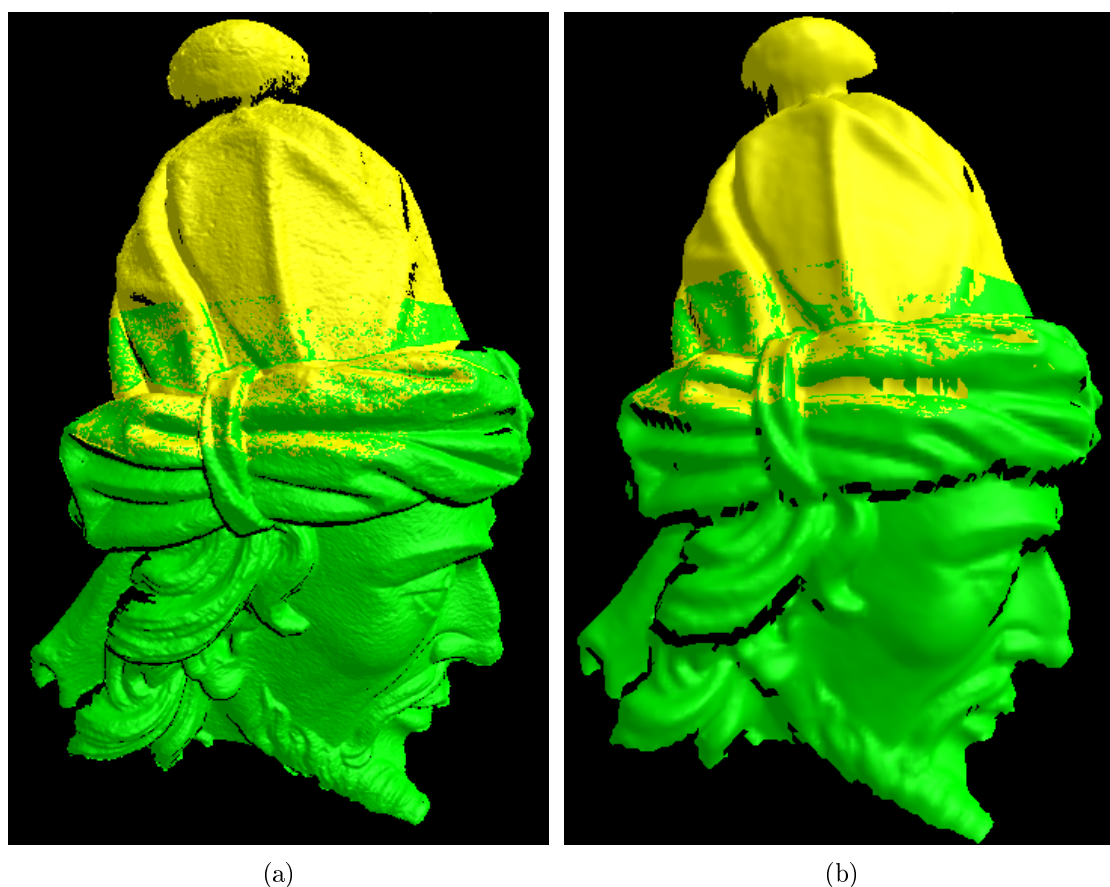


Figura 3.5: Exemplo de vistas alinhadas em diferentes resoluções: (a) par de vistas alinhado com resolução de 640x480; (b) O mesmo par de vistas em resolução reduzida 160x120. Imagens da escultura da representação do profeta Joel, parte da obra: "Os Doze Profetas" do escultor Antônio Francisco Lisboa, "O Aleijadinho".

Quando uma vista nova é carregada é possível utilizar a funcionalidade de automaticamente gerar um alinhamento com as vistas que encontrar maior similaridade através dos descritores SURF. E caso o método de pré-alinhamento falhe em conseguir uma resposta adequada é possível fazer o alinhamento manual de forma análoga ao que era feito no

icptool com uma técnica que consiste em movimentar a última vista selecionada enquanto as demais vistas permanecem fixas. Essa movimentação é feita utilizando comandos do mouse, clicar e arrastar o botão esquerdo para rotações, e clicar e arrastar o botão direito para translações.

O alinhamento fino e global pode ser feito exatamente como é feito no *icptool*. Com a vantagem de que a nova interface permite adicionar novos pares de vistas e os diversos parâmetros do ICP que podem ser configurados na ferramenta e não somente através de valores definidos em um arquivo texto de configuração antes de iniciar o aplicativo. A figura 3.6 mostra a interface do *software*.

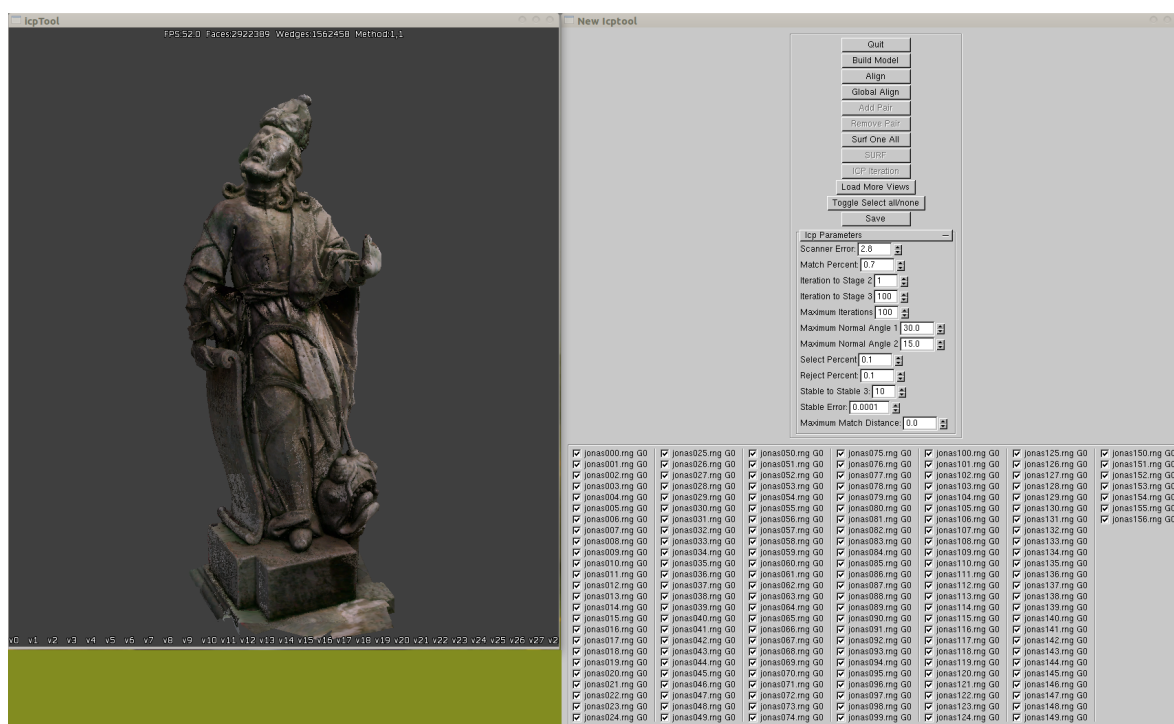


Figura 3.6: Interface do *software* de alinhamento implementado neste trabalho. As vistas exibidas são referentes a escultura da representação do profeta Jonas, parte da obra: "Os Doze Profetas" do escultor Antônio Francisco Lisboa, "O Aleijadinho".

CAPÍTULO 4

EXPERIMENTOS

Neste capítulo são apresentados resultados experimentais comparativos sobre quatro abordagens de pré-alinhamento: o método de alinhamento baseado em PCA [12], 4PCS [1], o método de Pingi *et al* [28] além da nova abordagem baseada em SURF apresentada neste trabalho. Estes algoritmos foram selecionados por exibirem alto desempenho computacional. Já os demais métodos apresentados no capítulo 2 possuem custo computacional inviável para o objetivo deste trabalho e não foram considerados nos experimentos.

Os experimentos foram conduzidos sobre imagens adquiridas de 8 objetos diferentes: 3 esculturas indígenas de pequeno porte mostradas nas figuras 4.2, 4.3 e 4.4; e 5 esculturas de pedra sabão com aproximadamente 2 metros de altura representando profetas bíblicos nas figuras 4.5, 4.6, 4.7, 4.8 e 4.9. As esculturas fazem parte da obra "Os Doze Profetas" de Antônio Francisco Lisboa, "O Aleijadinho" e estão localizadas no Santuário do Bom Jesus de Matosinhos, um dos patrimônios mundiais pela UNESCO.

Os resultados experimentais foram aqui agrupados em duas classes, as imagens dos acervos de grande e pequeno porte. Isso foi feito para que se possa visualizar a diferença no desempenho dos métodos nesses dois cenários. Os 3 objetos de pequeno porte somam 105 vistas agrupadas em 173 pares, já os 5 objetos de grande porte somam 717 vistas agrupadas em 1027 pares, todos definidos manualmente. O percentual de sobreposição entre as vistas é bem variado mas em média é de 70% para os objetos de pequeno porte e 40% para os de grande porte.

A forma como é realizada a captura afeta os métodos de pré-alinhamento, alguns métodos são fortemente dependentes do tamanho da região de sobreposição entre as vistas enquanto outros são afetados pela diferença no ângulo de captura, pois a rotação modifica

a projeção 2D de pontos característicos na imagem. Para demonstrar isso os experimentos foram divididos em dois grupos de imagens. Os objetos de pequeno porte possuem pares de imagem com bastante sobreposição porém ângulos de captura diversos. Já os de grande porte na maior parte possuem menos sobreposição, porém há um padrão de aquisição que pouco afeta a projeção dos ponto característicos entre as imagens.

As imagens foram adquiridas utilizando o Scanner Vivid 910 desenvolvido pela Konica Minolta, o mesmo utilizado no desenvolvimento do *pipeline*, a figura 4.1 mostra uma foto deste equipamento junto a uma imagem do software que o controla. Este sensor captura imagens com resolução 640x480, porém visando tornar o pré-alinhamento mais ágil, as imagens foram reduzidas para 160x120, o que nos resulta em imagens contendo até 19200 pontos. O computador utilizado para avaliação dos algoritmos foi um *Intel i7 Quad-Core* 1.6GHz, 4GB de memória.

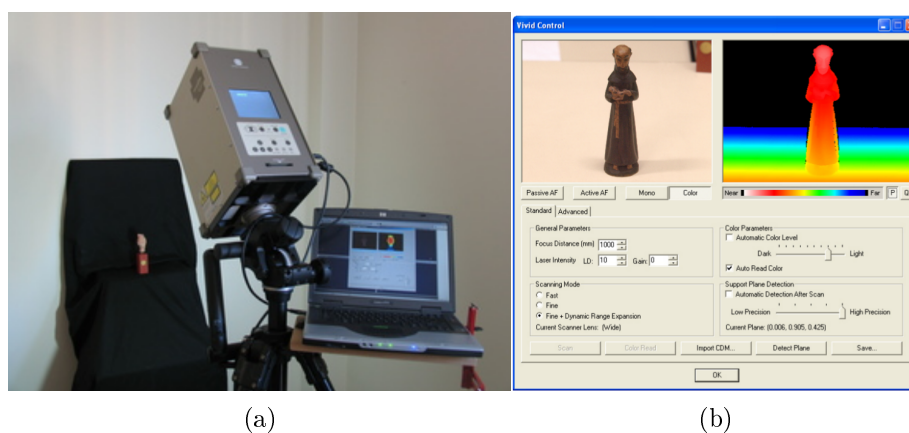


Figura 4.1: (a) Scanner Vivid 910 desenvolvido pela Konica Minolta. (b) Imagem do software desenvolvido pelo grupo IMAGO para controlar o equipamento.



Figura 4.2: Escultura feita na comunidade indígena Wauja, emprestado pelo Museu de Arqueologia e Etnologia da UFPR. (a) mostra a foto deste objeto e (b) mostra o modelo 3D do objeto renderizado.



Figura 4.3: Escultura feita na comunidade indígena Karajá, emprestado pelo Museu de Arqueologia e Etnologia da UFPR: (a) mostra a foto deste objeto e (b) mostra o modelo 3D do objeto renderizado.



Figura 4.4: Zoólito em forma de pássaro, escultura em osso de baleia feita em uma comunidade indígena brasileira, emprestado pelo Museu de Arqueologia e Etnologia da UFPR: (a) mostra a foto deste objeto e (b) mostra o modelo 3D do objeto renderizado.



Figura 4.5: Escultura representando o profeta Joel: (a) mostra a foto deste objeto e (b) mostra o modelo 3D do objeto renderizado.



Figura 4.6: Escultura representando o profeta Daniel: (a) mostra a foto deste objeto e (b) mostra o modelo 3D do objeto renderizado.



Figura 4.7: Escultura representando o profeta Jonas: (a) mostra a foto deste objeto e (b) mostra o modelo 3D do objeto renderizado.

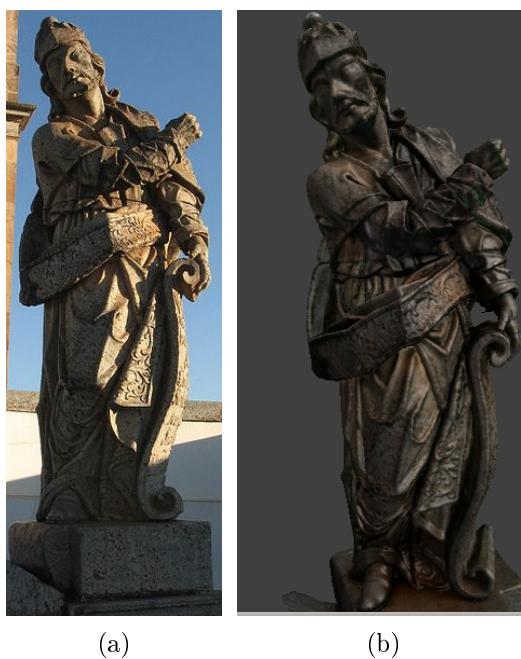


Figura 4.8: Escultura representando o profeta Ezequiel: (a) mostra a foto deste objeto e (b) mostra o modelo 3D do objeto renderizado.



Figura 4.9: Escultura representando o profeta Oséias: (a) mostra a foto deste objeto e (b) mostra o modelo 3D do objeto renderizado.

Neste experimento foram avaliados os seguintes itens.

- Acerto: É a porcentagem de acerto do método, neste experimento um pré-

alinhamento foi considerado correto quando a estimativa de alinhamento obtida permitiu ao ICP convergir para uma transformação em que pelo menos 99% dos pares de correspondência ICP ficaram com erro de alinhamento abaixo do 1 milímetro em menos de 50 iterações.

- **Iterações:** É a média do número de iterações que foram necessárias para o algoritmo ICP convergir para um alinhamento correto. Este é um bom indicador da precisão do pré-alinhamento obtido, visto que obter uma estimativa de alinhamento para o ICP é objetivo do pré-alinhamento neste *pipeline*.
- **RMS:** *Root Mean Square*, é o valor quadrático médio da distância entre um conjunto de pontos correspondentes nas imagens, esta é uma medida de erro bastante popular e é avaliado aqui para indicar a precisão média dos alinhamentos, (dado em milímetros).
- **Tempo:** a média do tempo das execuções em segundos.

4.1 Avaliação do método de alinhamento de Pingi *et al.*

É um método que depende fortemente da forma de captura para trabalhar com seus descritores adequadamente. O método pode ser compreendido nos seguintes passos:

- Um descritor de pontos em superfície, semelhante a uma máscara de tamanho pré-definido contendo a diferença entre a normal do ponto em cada posição da máscara em relação a normal do ponto central, é gerado para um conjunto de pontos na primeira imagem e para todos os pontos na segunda imagem.
- Para cada descritor na primeira imagem é gerado um par correspondente com o descritor mais semelhante na segunda imagem.
- Um conjunto pequeno de pares correspondentes é selecionado aleatoriamente.

- 3 a 3 os pares são combinados e têm uma transformação calculada. Caso o erro de alguma dessas transformações esteja abaixo de um limiar pré-definido essa transformação é considerada resposta. Caso contrário, outro grupo pequeno de pares é selecionado aleatoriamente e esta etapa é repetida.

A implementação foi realizada conforme indicado pelos autores do método [28], assim como os parâmetros adotados foram os mesmos propostos, o descritor possui tamanho 13 por 13, número máximo de iterações sendo 20, e o limiar de erro sendo 1,0 mm.

A tabela 4.1 traz o resultado do experimento realizado.

Tabela 4.1: Avaliação do método de alinhamento desenvolvido por Pingi *et al.*

| Objetos | Número de iterações | Porcentagem de acerto | RMS | Tempo (s) |
|---------------------|---------------------|-----------------------|-------|-----------|
| Anta | 4.72 | 37% | 12.57 | 0.061 |
| Jacare | 2.58 | 79% | 10.78 | 0.012 |
| Zoólito | 4.06 | 78% | 9.89 | 0.012 |
| Total pequeno porte | 3.95 | 63% | 11.04 | 0.029 |
| Jonas | 16.45 | 16% | 46.48 | 0.773 |
| Joel | 8.45 | 27% | 44.63 | 0.754 |
| Daniel | 15.16 | 28% | 45.66 | 0.731 |
| Ezequiel | 16.51 | 27% | 44.27 | 0.913 |
| Oséias | 14.33 | 18% | 47.49 | 0.878 |
| Total grande porte | 13.64 | 23% | 45.76 | 0.795 |
| Total geral | 10.65 | 29% | 40.75 | 0.684 |

O método, apesar de parecer promissor por ser aplicado com sucesso em um sistema de modelagem 3D semelhante ao que é utilizado neste trabalho, possui uma elevada taxa de erro por conta da rotação entre o ponto de vista dos pares de imagens, gerando muita diferença entre os descritores de pontos correspondentes. A média de iterações do ICP para corrigir o alinhamento foi alto, indicando que mesmo quando o método gerou uma transformação suficiente para o ICP convergir adequadamente ela não foi tão precisa.

4.2 Avaliação do método de alinhamento baseado em PCA

O método de pré-alinhamento baseado em PCA é um dos mais rápidos da literatura. Ele pode ser resumido nos passos:

- O centro de massa é calculado para as duas imagens.
- O eixo principal é calculado para as duas imagens.
- A rotação que alinha os eixos principais das duas imagens e a translação que alinha os dois centros de massa formam a matriz de transformação que alinha as duas imagens.

A implementação foi realizada conforme indicado pelos autores do método [12].

A tabela 4.2 traz o resultado da avaliação realizada sobre este método.

Tabela 4.2: Avaliação do método de alinhamento baseado em PCA

| Objetos | Número de iterações | Porcentagem de acerto | RMS | Tempo (s) |
|---------------------|---------------------|-----------------------|-------|-----------|
| Anta | 4.30 | 30% | 11.47 | 0.007 |
| Jacare | 2.68 | 79% | 15.32 | 0.007 |
| Zoólito | 5.47 | 79% | 9.18 | 0.007 |
| Total pequeno porte | 4.74 | 60% | 10.90 | 0.007 |
| Jonas | 12.81 | 26% | 25.02 | 0.007 |
| Joel | 7.31 | 27% | 26.10 | 0.007 |
| Daniel | 14.37 | 32% | 23.83 | 0.007 |
| Ezequiel | 16.44 | 25% | 25.33 | 0.007 |
| Oséias | 17.11 | 33% | 28.34 | 0.007 |
| Total grande porte | 13.25 | 29% | 25.62 | 0.007 |
| Total geral | 11.05 | 33% | 23.5 | 0.007 |

O sucesso deste método é totalmente dependente do tamanho da região de sobreposição. Sua grande vantagem fica por conta do baixo custo computacional, como os casos de uso deste trabalho envolvem muitos pares de vista com pouca sobreposição, aplicar este método sozinho não é adequado. A porcentagem de acerto foi maior do que Pingi *et al.*, mas com menos precisão.

4.3 Avaliação do método 4PCS

O 4PCS é um método recente que se propõe a ser rápido e robusto a ruídos e *outliers*. Seus passos resumidamente consistem em:

- Definir um conjunto base de 4 pontos coplanares na primeira imagem e fazer o cálculo de 2 pontos intermediários.
- Duas proporções são calculadas entre os 4 pontos.
- Na segunda imagem os pontos são combinados em pares e proporções são calculadas para eles.
- Cada dois pares de ponto cujas proporções combinem com as proporções dos 4 pontos da primeira imagem são considerados candidatos a correspondentes e tem uma transformação calculada.

O código para este método foi disponibilizado pelo próprio autor Niloy J. Mitra [1] com algumas pequenas adaptações e correções. O principal parâmetro para este método é a estimativa de sobreposição, quanto maior for a estimativa maior será a distância entre os pontos que irá definir a base de 4 pontos. E quanto maior é a distância entre os pontos da base, menor é o custo computacional do método, pois isso diminui a quantidade de candidatos correspondentes na outra imagem, porém diminui também as chances dos 4 pontos correspondentes estarem na região de sobreposição na outra vista. O padrão 70% de sobreposição foi adotado, valores menores acresciam muito o custo computacional e inviabilizava este método para os casos de uso deste trabalho.

A tabela 4.3 traz o resultado da avaliação realizada sobre o método 4PCS.

Tabela 4.3: Avaliação do método 4PCS

| Objetos | Número de iterações | Porcentagem de acerto | RMS | Tempo (s) |
|---------------------|---------------------|-----------------------|-------|-----------|
| Anta | 0.66 | 66% | 5.61 | 1.955 |
| Jacare | 0.00 | 91% | 4.87 | 0.417 |
| Zoólito | 2.19 | 86% | 3.76 | 1.349 |
| Total pequeno porte | 1.36 | 79% | 4.62 | 1.349 |
| Jonas | 8.77 | 33% | 20.97 | 1.084 |
| Joel | 4.52 | 32% | 22.33 | 1.246 |
| Daniel | 7.60 | 38% | 19.58 | 1.244 |
| Ezequiel | 9.17 | 36% | 22.07 | 1.111 |
| Oséias | 10.81 | 34% | 22.81 | 1.122 |
| Total grande porte | 7.96 | 34% | 21.42 | 1.169 |
| Total geral | 6.12 | 41% | 19 | 1.21 |

O desempenho deste método foi afetado pelo tamanho da região de sobreposição entre as vistas por duas razões. Primeiro pela distância dos 4 pontos que formam a base ser muito grande para os pares com pouca sobreposição. E a segunda razão foi a métrica de erro adotada pelo algoritmo, o LCP (*Largest Common Point set*), maior quantidade de pontos em comum, não é apropriada nos casos em que a sobreposição é pequena. Por vezes transformações erradas que colocam uma vista sobre a outra cobrem mais pontos e possuem um valor de LCP maior que as transformações adequadas que sobrepõe menos pontos, já que a região de sobreposição é pequena.

A taxa de acerto e precisão deste método foram significativamente superiores nos experimentos realizados ao método baseado em PCA e o método de Pingi *et al.*. O tempo de execução foi superior mas ainda assim se manteve baixo, na média sempre inferior a 2 segundos.

4.4 Método desenvolvido

A abordagem apresentada neste trabalho possui uma estratégia semelhante à abordagem de Pingi *et al.*, com o principal diferencial de que é utilizado um descritor mais robusto e uma técnica que tira vantagem da informação 3D para pontuar pares de pontos correspondentes. A principal limitação do método proposto são as vistas que possuem regiões

homogêneas, como planos sem textura onde não é possível encontrar pontos característicos adequadamente.

O método consiste dos passos:

- Descritores SURF são gerados nas duas imagens
- Para cada descritor na primeira imagem é feita uma associação com o descritor mais semelhante da segunda imagem
- É atribuída uma pontuação para cada par correspondente, a pontuação é aumentada em 1 quando dois pares de pontos tem distâncias semelhantes tanto na primeira quanto na segunda imagem.
- Os 15 pares correspondentes de maior pontuação são agrupados 3 a 3. Para cada agrupamento uma transformação é calculada e tem seu erro avaliado, a que tiver menor erro é adotada como resposta.

A tabela 4.4 trás o resultado da avaliação realizada sobre o método baseado em SURF.

Tabela 4.4: Avaliação do método de desenvolvido

| Objetos | Número de iterações | Porcentagem de acerto | RMS | Tempo (s) |
|---------------------|---------------------|-----------------------|-------|-----------|
| Anta | 0.19 | 71% | 6.33 | 0.909 |
| Jacare | 0.00 | 95% | 0.82 | 0.833 |
| Zoólito | 1.13 | 81% | 4.70 | 0.747 |
| Total pequeno porte | 0.62 | 79% | 4.78 | 0.821 |
| Jonas | 0.39 | 78% | 6.73 | 1.647 |
| Joel | 0.91 | 58% | 10.45 | 1.608 |
| Daniel | 1.07 | 82% | 7.02 | 1.860 |
| Ezequiel | 0.38 | 83% | 6.33 | 1.444 |
| Oséias | 1.00 | 71% | 8.74 | 1.333 |
| Total grande porte | 0.77 | 74% | 7.97 | 1.608 |
| Total geral | 0.74 | 74% | 7.51 | 1.495 |

Como o desempenho do método proposto não depende diretamente do tamanho da região de sobreposição, ele foi muito bem sucedido em comparação aos demais métodos no caso das imagens dos objetos de grande porte, que em média possuem pouca sobreposição.

O método também foi muito mais preciso que os demais, os alinhamentos gerados por ele precisaram de pouca correção do ICP. Porém o 4PCS foi igualmente bem sucedido para os objetos de pequeno porte. O tempo de execução em média foi muito semelhante ao 4PCS.

4.5 Comparação geral

Para melhor visualizar os resultados obtidos por todos os métodos dentro de cada critério as tabelas 4.5, 4.6 e 4.7 mostram os resultados experimentais para os objetos de pequeno porte, grande porte e o total respectivamente.

Tabela 4.5: Avaliação do total das vistas de pequeno porte de todos os métodos.

| Método | Número de iterações | Porcentagem de acerto | RMS | Tempo (s) |
|---------------------|---------------------|-----------------------|-------|-----------|
| Pingi <i>et al.</i> | 3.95 | 63% | 11.04 | 0.029 |
| PCA | 4.74 | 60% | 10.90 | 0.007 |
| 4PCS | 1.36 | 79% | 4.62 | 1.349 |
| Método proposto | 0.62 | 79% | 4.78 | 0.821 |

Tabela 4.6: Avaliação do total das vistas de grande porte de todos os métodos.

| Método | Número de iterações | Porcentagem de acerto | RMS | Tempo (s) |
|---------------------|---------------------|-----------------------|-------|-----------|
| Pingi <i>et al.</i> | 13.64 | 23% | 45.76 | 0.795 |
| PCA | 13.25 | 29% | 25.62 | 0.007 |
| 4PCS | 7.96 | 34% | 21.42 | 1.169 |
| Método proposto | 0.77 | 74% | 7.97 | 1.608 |

Tabela 4.7: Avaliação do total geral de todos os métodos.

| Método | Número de iterações | Porcentagem de acerto | RMS | Tempo (s) |
|---------------------|---------------------|-----------------------|-------|-----------|
| Pingi <i>et al.</i> | 10.65 | 29% | 40.75 | 0.684 |
| PCA | 11.05 | 33% | 23.5 | 0.007 |
| 4PCS | 6.12 | 41% | 19 | 1.21 |
| Método proposto | 0.74 | 74% | 7.51 | 1.495 |

O método 4PCS e o método proposto foram significativamente superiores quanto a taxa de acerto e precisão em relação aos outros métodos nos experimentos realizados.

Para os objetos de grande porte o método proposto se destacou apresentando uma taxa de acerto 40 pontos percentuais superior ao 4PCS e apresentou tempo médio de execução bem próximo.

Ilustrando a precisão dos métodos a figura 4.10 mostra o mesmo par de vistas pré-alinhados através dos 4 métodos. Em 4.10 (a) o método de Pingi *et al.* falhou no alinhamento por ter associado pares de pontos de correspondência equivocadamente. Na 4.10 (b) o método baseado em PCA acertou na rotação mas errou na translação por assumir os centros de massa como sendo semelhantes. Em 4.10 (c) o método 4PCS conseguiu uma boa aproximação da transformação que alinha as duas vistas mas sem muita precisão. Em 4.10 (d) o método proposto conseguiu com boa precisão gerar o alinhamento para as duas vistas.

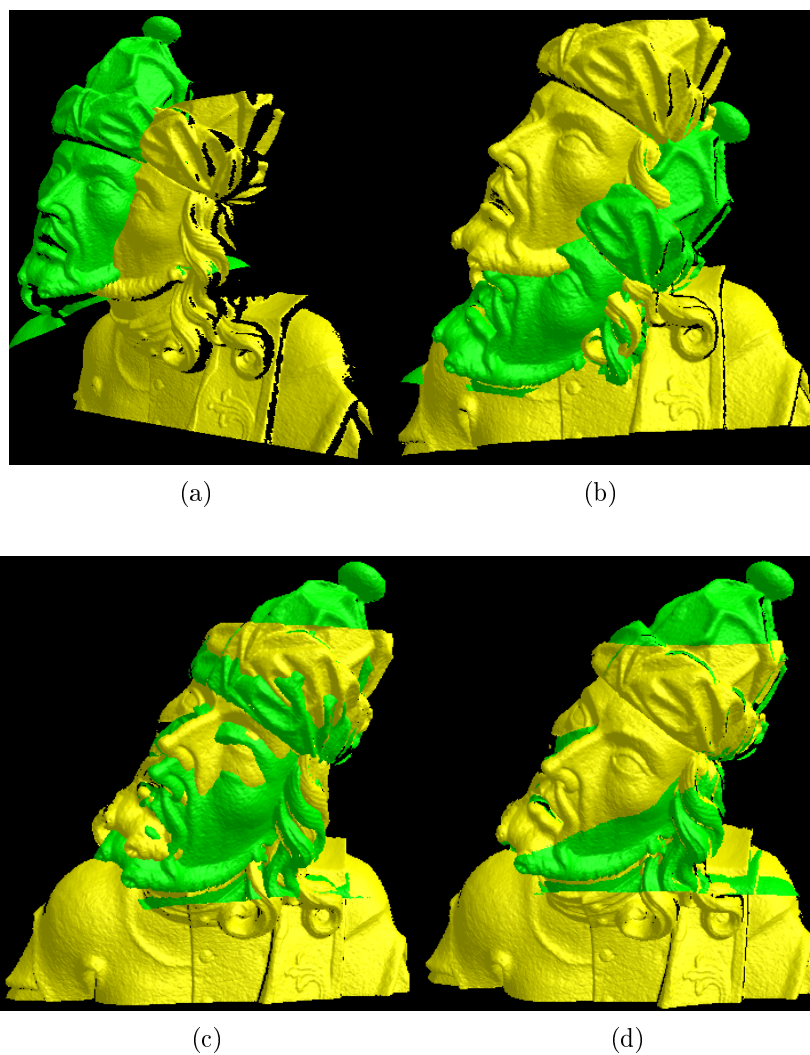


Figura 4.10: Par de vistas renderizadas em cores diferentes com aproximadamente 50% de sobreposição. As vistas são retiradas da escultura do profeta Jonas, feita por Aleijadinho. (a) mostra o par de vistas pré-alinhada com o método de Pingi *et al.*, (b) foi pré-alinhado com o método baseado em PCA, (c) foi pré-alinhado com o método 4PCS e (d) foi pré-alinhado com o método proposto.

CAPÍTULO 5

CONCLUSÃO

Este trabalho teve como objetivo conhecer os métodos automáticos de pré-alinhamento gerando soluções dentro de um *pipeline* de modelagem 3D já existente [35]. Apesar de um dos principais objetivos da preservação de acervos culturais e naturais é gerar modelos com a mais alta fidelidade, é importante considerar que além da fidelidade a praticidade é fundamental para viabilizar que acervos maiores sejam digitalizados, diminuindo a necessidade de interferência humana. Além da importância cultural agregada neste trabalho a principal contribuição científica está na apresentação de um método de pré-alinhamento que gera soluções precisas de forma rápida mesmo nos casos em que existe baixa sobreposição.

Recapitulando os objetivos deste trabalho:

- Realizar um estudo sobre os principais métodos de pré-alinhamento.
- Implementar um método de pré-alinhamento automático adaptado ao escopo do *pipeline*.
- Implementar uma ferramenta que facilite a identificação de regiões não digitalizadas.

Os três objetivos aos quais este trabalho se propôs foram concluídos, diversas abordagens de pré-alinhamento estado-da-arte foram estudadas e algumas aplicadas em testes experimentais. Os experimentos mostraram vantagens e desvantagens dos métodos dentro do escopo deste trabalho. O 4PCS apresentou um desempenho equivalente ao método proposto para os objetos de pequeno porte. Porém o método desenvolvido foi superior para os pares de imagens dos objetos de grande porte aproveitando-se do padrão executado na captura.

Os objetos de grande porte são o principal caso de uso para o *pipeline* desenvolvido no grupo IMAGO [35]. Por esta razão foi desenvolvido um método de pré-alinhamento rápido e que tira proveito da estratégia adotada para a captura sem a dependência de possuir uma grande região de sobreposição.

O aplicativo *icptool* que no *pipeline* original recebia os alinhamentos para refinar com o ICP, ganhou três grandes melhorias. A primeira foi uma interface visual para controlar todos os parâmetros do programa. A segunda foram os métodos automáticos de pré-alinhamento que foram agregados ao *software*. E a terceira foi a funcionalidade que permite exibir diversas vistas simultaneamente em tempo de aquisição, o que entre outras vantagens facilita a localização de regiões não digitalizadas. Essa funcionalidade contribui para aumentar a fidelidade e a agilidade da aquisição e modelagem 3D em si, pois facilita que o objeto tenha toda sua superfície capturada, diminuindo o uso de informações estimadas para preencher eventuais buracos na superfície do objeto digitalizado, e também auxilia o operador do *scanner* não adquirir vistas redundantes.

5.1 Trabalhos Futuros

Uma limitação existente no método proposto é a impossibilidade de gerar correspondências quando não existem pontos característicos na região de sobreposição das duas imagens. Realizar um estudo sobre formas de contornar esse problema, talvez gerando um segundo descritor especializado em regiões com essa característica, ou alguma outra abordagem robusta a regiões homogêneas, seria uma contribuição interessante para o método proposto. Outra contribuição seria fazer um estudo mais profundo sobre diversos descritores de *features* e suas vantagens para o método em conjunto, ou substituição ao SURF.

Outro trabalho futuro interessante seria adaptar os algoritmos de alinhamento existentes para as novas plataformas que trabalham com paralelismo. Algoritmos que possuem um custo computacional proibitivo quando programados de forma serial poderiam ser re-visitados utilizando esse novo paradigma. A maior parte dos algoritmos de alinhamento

tem um forte potencial para o paralelismo e os benefícios merecem ser avaliados.

BIBLIOGRAFIA

- [1] D. Aiger, N. J. Mitra, e D. Cohen-Or. 4-points congruent sets for robust surface registration. *ACM Transactions on Graphics*, vol. 27 num. 3 pp. 1–10, 2008.
- [2] B. T. Andrade, O. R. P. Bellon, L. Silva, e A. Vrubel. Enhancing color texture quality of 3d models for digital preservation of indigenous ceramic artworks. *Proceedings of the IEEE 12th International Conference on Computer Vision Workshops*, pp. 980–987, 2009.
- [3] H. Bay, T. Tuytelaars, e L. Van Gool. SURF: Speeded Up Robust Features. *Proceedings of the 9th European Conference on Computer Vision*, Graz Austria, 2006.
- [4] O. R. P. Bellon. *Imagens de Profundidade: Segmentação e Representação por superfícies Planares*. Tese de Doutorado, Universidade Estadual de Campinas, Campinas, São Paulo, Brasil, 1997.
- [5] F. Bernardini e H. Rushmeier. The 3D model acquisition pipeline. *Computer Graphics Forum*, vol. 21 num. 2 pp. 149–172, 2002.
- [6] P. J. Besl e N. D. McKay. A method for registration of 3D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14 num. 2 pp. 239–256, 1992.
- [7] G. Blais e M. D. Levine. Registering Multiview Range Data to Create 3D Computer Objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17 pp. 820–824, 1993.
- [8] C. Chen, Y. Hung, e J. Cheng. Ransac-based darces: A new approach to fast automatic registration of partially overlapping range images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21 pp. 1229–1234, 1999.
- [9] Y. Chen e G. Medioni. Object modelling by registration of multiple range images. *Image and Vision Computing*, vol. 10 num. 3 pp. 145–155, 1992.

- [10] A. Chiuso, P. Favaro, H. Jin, e S. Soatto. 3D Motion and Structure from 2D Motion Causally Integrated over Time: Implementation. *Proceedings of the IEEE Transactions on Robotics and Automation*, pp. 734–750, 2000.
- [11] C. S. Chua e R. Jarvis. Point Signatures: A New Representation for 3D Object Recognition. *International Journal of Computer Vision*, pp. 63–85, 1997.
- [12] D. H. Chung, I. D. Yun, e S. U. Lee. Registration of multiple-range views using the reverse-calibration technique. *Pattern Recognition*, vol. 31 num. 4 pp. 457–464, 1998.
- [13] B. Curless e M. Levoy. A volumetric method for building complex models from range images. *Proceedings of the 23rd annual Conference on Computer graphics and interactive techniques*, pp. 303–312, New York, NY, USA, 1996. ACM.
- [14] J. Davis, S. R. Marschner, M. Garr, e M. Levoy. Filling holes in complex surfaces using volumetric diffusion. *Proceedings of the 1st International Symposium on 3D Data Processing, Visualization and Transmission. Special Interest Group on Graphics and Interactive Techniques*, pp. 428–861, 2002.
- [15] M. Farouk, I. El-Rifai, S. El-Tayar, H. El-Shishiny, M. Hosny, M. El-Rayes, J. Gomes, F. Giordano, H. Rushmeier, F. Bernardini, e K. Magerlein. Scanning and processing 3d objects for web display. *Proceedings of the 4th International Conference on 3-D Digital Imaging and Modeling*, pp. 310–317, 2003.
- [16] M. A. Fischler e R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Readings in computer vision: issues, problems, principles, and paradigms*, pp. 726–740, 1987.
- [17] B. K. P. Horn. Closed-form solution of absolute orientation using unit quaternions. *Journal of the Optical Society of America A*, vol. 4 num. 4 pp. 629–642, 1987.
- [18] K. Ikeuchi, T. Oishi, J. Takamatsu, R. Sagawa, A. Nakazawa, R. Kurazume, K. Nishino, M. Kamakura, e Y. Okamoto. The great buddha project: Digitally

- archiving, restoring, and analyzing cultural heritage objects. *International Journal of Computer Vision*, vol. 75 num. 1 pp. 189–208, 2007.
- [19] A. Johnson. *Spin-Images: A Representation for 3-D Surface Matching*. Tese de Doutorado, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, 1997.
- [20] M. Levoy, K. Pulli, B. Curless, S. Rusinkiewicz, D. Koller, L. Pereira, M. Ginzton, S. Anderson, J. Davis, J. Ginsberg, J. Shade, e D. Fulk. The Digital Michelangelo Project: 3D Scanning of Large Statues. *Proceedings of the International Conference on Computer Graphics and Interactive Techniques*, pp. 131–144, 2000.
- [21] H. Li, R. Straub, e H. Prautzsch. Structured light based reconstruction under local spatial coherence assumption. *International Symposium on 3D Data Processing Visualization and Transmission*, pp. 575–582, 2006.
- [22] S. B. Liu, K. B. Kang, J. P. Tarel, e D. B. Cooper. Free-form object reconstruction from silhouettes, occluding edges and texture edges: A unified and robust operator based on duality. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30 num. 1 pp. 131–146, 2008.
- [23] W. E. Lorensen e H. E. Cline. Marching cubes: A high resolution 3D surface construction algorithm. *Proceedings of the Annual Conference on Computer Graphics and Interactive Techniques*, pp. 163–169, 1987.
- [24] T. Masuda. Generation of geometric model by registration and integration of multiple range images. *Proceedings of the 3rd International Conference on 3D Digital Imaging and Modeling, 2001.*, pp. 254 –261, 2001.
- [25] T. Masuda. Object shape modelling from multiple range images by matching signed distance fields. *3D Data Processing Visualization and Transmission, International Symposium on*, pp. 439, 2002.

- [26] C. M. Mendes, D. R. Drees, O. R. P. Bellon, e L. Silva. Sistema para visualização de museu virtual 3D. *Proceedings of the 5th Workshop of Undergraduated Work. Special Interest Group on Graphics and Interactive Techniques*, 2007.
- [27] G. Pavlidis, A. Koutsoudis, F. Arnaoutoglou, V. Tsioukas, e C. Chamzas. Methods for 3D digitalization of Cultural Heritage. *Journal of Cultural Heritage*, vol. 8 num. 1 pp. 93–98, 2007.
- [28] P. Pingi, A. Fasano, P. Cignoni, C. Montani, e R. Scopigno. Exploiting the scanning sequence for automatic registration of large sets of range maps. *Computer Graphics Forum*, vol. 24 num. 3 pp. 517–526, 2005.
- [29] K. Pulli. Multiview registration for large data sets. *Proceedings of the International Conference on 3D Digital Imaging and Modeling*, pp. 160–168. IEEE Computer Society, 1999.
- [30] C. C. Queirolo, L. Silva, O. R. P. Bellon, e M. Pamplona Segundo. 3D Face Recognition Using Simulated Annealing and the Surface Interpenetration Measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32 num. 2 pp. 206–219, 2010.
- [31] S. Rusinkiewicz e M. Levoy. Efficient variants of the ICP algorithm. *Proceedings of the International Conference on 3D Digital Imaging and Modeling*, pp. 145–152, 2001.
- [32] J. Salvi, C. Matabosch, D. Fofi, e J. Forest. A review of recent range image registration methods with accuracy evaluation. *Image Vision Computing*, vol. 25 num. 5 pp. 578–596, 2007.
- [33] L. Silva, O.R.P. Bellon, e K.L. Boyer. Precision range image registration using a robust surface interpenetration measure and enhanced genetic algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27 num. 5 pp. 762–776, 2005.

- [34] A. J. M. Traina, A. H. M. A. Prado, e J. M. Bueno. 3D reconstruction of magnetic resonance imaging using largely spaced slices. *Proceedings of the 10th IEEE Symposium on Computer-Based Medical Systems*, pp. 114, Washington, DC, USA, 1997. IEEE Computer Society.
- [35] A. Vrubel, O.R.P. Bellon, e L. Silva. A 3d reconstruction pipeline for digital preservation. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 2687–2694, 2009.
- [36] M. Wheeler, Y. Sato, e K. Ikeuchi. Consensus surfaces for modeling 3D objects from multiple range images. *Proceedings of the International Conference on Computer Vision*, pp. 917–924, 1998.
- [37] L. Zhang, B. Curless, A. Hertzmann, e S. M. Seitz. Shape and motion under varying illumination: Unifying structure from motion, photometric stereo, and multi-view stereo. *Proceedings of the 9th IEEE International Conference on Computer Vision*, vol. 1 pp. 618, 2003.