

UNIVERSIDADE FEDERAL DE SÃO CARLOS

CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

**CAPTURA E RECONSTRUÇÃO DA MARCHA
HUMANA UTILIZANDO MARCADORES
PASSIVOS**

IVAN JOSÉ DOS REIS FILHO

ORIENTADOR: PROF. DR. MARIO AUGUSTO DE SOUZA LIZIÉR

São Carlos – SP

Maio/2016

UNIVERSIDADE FEDERAL DE SÃO CARLOS

CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

**CAPTURA E RECONSTRUÇÃO DA MARCHA
HUMANA UTILIZANDO MARCADORES
PASSIVOS**

IVAN JOSÉ DOS REIS FILHO

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de São Carlos, como parte dos requisitos para a obtenção do título de Mestre em Ciência da Computação, área de concentração: Processamento de Imagens e Sinais

Orientador: Prof. Dr. Mario Augusto de Souza Liziér

São Carlos – SP

Maio/2016

Ficha catalográfica elaborada pelo DePT da Biblioteca Comunitária UFSCar
Processamento Técnico
com os dados fornecidos pelo(a) autor(a)

R375c Reis Filho, Ivan José dos
Captura e reconstrução da marcha humana utilizando
marcadores passivos / Ivan José dos Reis Filho. --
São Carlos : UFSCar, 2016.
70 p.

Dissertação (Mestrado) -- Universidade Federal de
São Carlos, 2016.

1. Captura de movimentos. 2. Marcadores passivos.
3. Detecção de marcadores. 4. Transformada de Hough.
5. Reconstrução 3d. I. Título.



UNIVERSIDADE FEDERAL DE SÃO CARLOS
Centro de Ciências Exatas e de Tecnologia
Programa de Pós-Graduação em Ciência da Computação

Folha de Aprovação

Assinaturas dos membros da comissão examinadora que avaliou e aprovou a defesa de Dissertação de Mestrado do candidato Ivan José dos Reis Filho, realizada em 20/05/2016.

Prof. Dr. Mário Augusto de Souza Liziér
(UFSCar)

Prof. Dr. Ricardo José Ferrari
(UFSCar)

Prof. Dr. João Paulo Gois
(UFABC)

Certifico que a sessão de defesa foi realizada com a participação à distância do membro Prof. Dr. João Paulo Gois, depois das arguições e deliberações realizadas, o participante à distância está de acordo com o conteúdo do parecer da comissão examinadora redigido no relatório de defesa do aluno Ivan José dos Reis Filho.

Prof. Dr. Mário Augusto de Souza Liziér
Presidente da Comissão Examinadora
(UFSCar)

Crie uma meta e estabeleça um limite.

Autor desconhecido.

RESUMO

A análise computacional de movimentos humanos pode auxiliar no tratamento e reabilitação de pessoas que possuem algum tipo de deficiência motora. Estudos na fisioterapia, chamado de avaliação cinemática, tem por objetivo direcionar tratamentos para corrigir as disfunções do corpo, causadas pela má formação genética, doenças degenerativas ou problemas de postura. Um sistema computacional pode ser utilizado para descrever precisamente os movimentos, posição anatômica e termos angulares durante o período da marcha. Para tanto, os movimentos precisam ser capturados e reconstruídos no espaço tridimensional para posteriormente serem avaliados.

Este trabalho consiste no estudo, desenvolvimento e avaliação de um sistema para captura de movimentos humanos baseado em registros de vídeos, utilizando recursos computacionais de baixo custo e adaptado a realidade da análise de marcha. A automatização da captura computacional destes movimentos, além de auxiliar na reabilitação motora neurológica e ou ortopédica de pessoas que possuem alguma deficiência do gênero, ampliará sua aplicação na comunidade. Nesse contexto, foi realizado o levantamento bibliográfico das principais técnicas de captura computacional de movimentos, e no caso dos sistemas ópticos com marcadores passivos, e definido quatro etapas principais: calibração, sincronização das câmeras, detecção e reconstrução tridimensional. Ainda, experimentos computacionais foram realizados para o entendimento prático e teórico do trabalho. Registros de marcha de crianças foram utilizados como estudo de caso para a implementação das etapas de calibração, detecção dos marcadores e reconstrução tridimensional.

Palavras-chave: Captura de Movimentos, Marcadores passivos, Detecção de marcadores, Transformada de *Hough*, Reconstrução 3D.

ABSTRACT

The computational analysis of human mobility can aid in treatment and rehabilitation of people who have some kind of physical disability. Physiotherapeutic studies, called kinematics, aims to correct body's dysfunction, caused by genetic malformations, degenerative diseases or posture problems. A computer system can be used to precisely describe motion, anatomical position and angular terms during gait. Therefore, the movements need to be captured and represented in three-dimensional space to be evaluated.

This work is a study, development and evaluation of a system for capturing human movement based on recorded videos, using low cost computing resources and adapted to the reality of gait analysis. The automation of computing movements capture, as well as aid in neurological, motor or orthopedic rehabilitation will expand its application to community. In this context, a study was conducted about computing methods of human movements capture by passive markers, and defined the four main steps: calibration and synchronization of the cameras, detection and three-dimensional reconstruction. Computer experiments were made to facilitate the theoretical understanding of this work. Kinematic analyzes data were used as a study case for the implementation of the calibration steps, detection of markers and three-dimensional reconstruction.

Keywords: Motion Capture, Passive markers, Markers detection, Hough transform, 3D Reconstruction.

LISTA DE FIGURAS

| | | |
|-----|--|----|
| 2.1 | Dispositivo Mecânico da SouVR | 18 |
| 2.2 | Wii Remote da console da Nintendo | 18 |
| 2.3 | Dispositivo Magnético da Senztech | 19 |
| 2.4 | Rastreamento de movimentos sem marcador (SMINCHISESCU; TRIGGS, 2003). | 20 |
| 2.5 | Console Xbox360 com dispositivos Kinect (ANDREWS; BAKER, 2006) | 21 |
| 2.6 | Pontos infravermelhos distribuídos no ambiente (ANDERSEN et al., 2012) | 21 |
| 2.7 | Marcadores Ativos. Usuário utiliza sensores que emitem raios infravermelhos (KIRK; O'BRIEN; FORSYTH, 2005) | 22 |
| 2.8 | Marcadores passivos. Usuário utiliza uma roupa preta com marcadores reflexivos (CACS, 2007) | 23 |
| 2.9 | Imagem do sistema DVideoW (CARVALHO; TUDELLA; BARROS, 2005) | 24 |
| 3.1 | Etapas para captura de movimentos por dispositivos ópticos. Processos em destaque foram implementados neste trabalho. | 27 |
| 3.2 | Deteção dos marcadores no corpo da pessoa usando a técnica de <i>Thresholding</i> de (GUERRA-FILHO, 2005). | 30 |
| 3.3 | (A) Pontos na circunferência na imagem. (B) Quatro intersecções dos pontos no espaço de <i>Hough</i> | 32 |
| 3.4 | (A) Pontos na circunferência na imagem. (B) Intersecções no espaço de <i>Hough</i> | 33 |
| 3.5 | Ilustração para a correspondência dos marcadores | 34 |
| 3.6 | Estimação de pose pelo método proposto em (LI; MENG; HOLSTEIN, 2003) utilizando 47 pontos que compõe o modelo (esquerda), as poses não identificadas (ao centro) e as mesmas após identificação (direita) (LI; MENG; HOLSTEIN, 2003) | 35 |

| | | |
|------|---|----|
| 3.7 | Abordagem usada no trabalho de (HERDA et al., 2001). | 36 |
| 3.8 | Fluxograma dos processos de calibração e reconstrução. | 37 |
| 3.9 | O ponto 3D \mathbf{x}^w é projetado para \mathbf{x}^l no plano da imagem por meio do centro óptico o^c | 38 |
| 3.10 | Coordenadas da origem do <i>pixel</i> (u^p, v^p) para o canto superior esquerdo | 39 |
| 3.11 | Sistema de referência para calibração do ambiente onde ocorrem as medições . | 43 |
| 4.1 | Ilustração (tela) do sistema DVideoW | 49 |
| 4.2 | Deteção dos marcadores representado por um círculo vermelho do primeiro quadro da avaliação de cada câmera | 50 |
| 4.3 | Binarização da imagem pelos canais RGB. (A) Quadro do Vídeo. (B) Zoom da imagem | 52 |
| 4.4 | Filtro de Sobel. (A) Quadro do vídeo. (B) Zoom da Imagem | 52 |
| 4.5 | Círculos vermelhos representando o resultado da aplicação da transformada de <i>Hough</i> , utilizando o filtro de Sobel como fase de pré processamento | 53 |
| 4.6 | Círculos vermelhos representando a detecção dos marcadores do sistema de referência | 54 |
| 4.7 | Círculos vermelhos representando a detecção dos marcadores no primeiro quadro de avaliação (Criança “A”) | 56 |
| 4.8 | Círculos vermelhos representando a detecção dos marcadores no período da marcha (Criança “A”) | 56 |
| 4.9 | Círculos vermelhos representando a detecção dos marcadores no primeiro quadro de avaliação (Criança “B”) | 57 |
| 4.10 | Círculos vermelhos representando a detecção dos marcadores no período da marcha (Criança “B”) | 58 |
| 4.11 | Espaço de <i>Hough</i> . (A) Quadro do vídeo da figura 4.3. (B) Zoom da Imagem. . | 60 |
| 4.12 | Resultado da Reconstrução do momento exibido na figura 4.7. (A) Vista Lateral; (B) Vista Frontal. | 61 |
| 4.13 | Resultado da Reconstrução do momento exibido na figura 4.8. (A) Vista Lateral; (B) Vista Frontal. | 61 |

| | | |
|------|--|----|
| 4.14 | Resultado da Reconstrução do momento exibido na figura 4.9. (A) Vista Lateral; (B) Vista Frontal. | 62 |
| 4.15 | Resultado da Reconstrução do momento exibido na figura 4.10. (A) Vista Lateral; (B) Vista Frontal. | 62 |

LISTA DE TABELAS

| | | |
|-----|--|----|
| 4.1 | Limiars dos canais RGB da cor dos marcadores | 51 |
| 4.2 | Quantidade de circunferências detectadas pelo sistema. | 55 |
| 4.3 | Comparação dos resultados verdadeiros positivos, falsos negativos e falsos positivos | 55 |
| 4.4 | Quantidade de marcadores registrados pelas câmeras. | 58 |
| 4.5 | Resultados dos marcadores detectados corretamente pelo sistema. | 59 |
| 4.6 | Resultados dos marcadores não detectados pelo sistema | 59 |
| 4.7 | Resultados da detecção que não são marcadores. | 59 |

GLOSSÁRIO

AHT – *Adaptive Hough Transform*

CHT – *Circle Hough Transform*

DEFMH-UFSCar – *Departamento de Educação Física e Motricidade Humana - Universidade Federal de São Carlos*

DLT – *Direct Linear Transformation*

HT – *Hough Transform*

LED – *Light emitter diode*

MoCap – *Captura de Movimentos*

RGB – *Red, Green, Blue*

SVD – *Decomposição de valores singulares*

SUMÁRIO

GLOSSÁRIO

| | |
|--|-----------|
| CAPÍTULO 1 – INTRODUÇÃO | 12 |
| 1.1 Organização do Trabalho | 14 |
| CAPÍTULO 2 – TRABALHOS RELACIONADOS | 16 |
| 2.1 Captura de movimentos | 16 |
| 2.2 Rastreadores por sensores | 17 |
| 2.2.1 Mecânicos | 17 |
| 2.2.2 Magnéticos | 19 |
| 2.3 Rastreadores Ópticos | 19 |
| 2.3.1 Sem Marcadores | 20 |
| 2.3.2 Marcadores Ativos | 21 |
| 2.3.3 Marcadores Passivos | 22 |
| 2.4 DVideoW | 23 |
| 2.5 Considerações Finais | 25 |
| CAPÍTULO 3 – MATERIAIS E MÉTODOS | 26 |
| 3.1 Sincronização | 28 |
| 3.2 Detecção de Marcadores | 29 |
| 3.2.1 Transformada de <i>Hough</i> | 31 |

| | | |
|--------------------------------|---|-----------|
| 3.3 | Correspondência | 33 |
| 3.4 | Rastreamento | 35 |
| 3.5 | Calibração e Reconstrução Tridimensional | 36 |
| 3.5.1 | Conceitos Iniciais | 38 |
| 3.5.2 | Calibração | 42 |
| 3.5.3 | Reconstrução | 45 |
| 3.6 | Considerações Finais | 46 |
| CAPÍTULO 4 – RESULTADOS | | 48 |
| 4.1 | Detecção dos Marcadores | 50 |
| 4.1.1 | Binarização da Imagem | 51 |
| 4.1.2 | Avaliação da Detecção | 53 |
| | Resultados da calibração | 54 |
| | Resultados da criança “A” | 55 |
| | Resultados da criança “B” | 57 |
| | Avaliação dos Resultados das Crianças “A” e “B” | 58 |
| 4.1.3 | Espaço de <i>Hough</i> | 60 |
| 4.2 | Calibração e Reconstrução | 60 |
| 4.3 | Considerações Finais | 62 |
| CAPÍTULO 5 – CONCLUSÃO | | 64 |
| 5.1 | Trabalhos Futuros | 65 |
| REFERÊNCIAS | | 66 |

Capítulo 1

INTRODUÇÃO

Este trabalho consiste no estudo, desenvolvimento e avaliação de um sistema para captura de movimentos humanos, baseado em registros de vídeos.

Nas últimas décadas, têm sido crescente o número de pesquisas envolvendo a análise computacional do movimento humano, sejam motivadas pelo aperfeiçoamento tecnológico dos dispositivos de captura ou por benefícios obtidos com a análise cinemática de movimentos. Com a análise biomecânica, realizada a partir de uma sequência de imagens digitais, pode-se obter, a um custo razoável, informações específicas do desempenho de atletas, ou ainda, avaliar a reabilitação de indivíduos com deficiência. Esses estudos despertam interesse na indústria do entretenimento, esporte, fisioterapia e demais áreas que buscam informações do comportamento da mobilidade e do desempenho do movimento humano.

Na fisioterapia, uma aplicação mais específica, a análise do movimento humano é chamada de avaliação cinemática. Nesse contexto, leva-se em consideração a velocidade, posição e aceleração do movimento para reabilitação (DELISA, 1998). Nessa análise, estruturas anatômicas (anatomia funcional) e análise da causa (cinética) relacionadas ao movimento são estudadas com objetivo de corrigir as disfunções do corpo humano ocasionada pela má formação genética, doenças degenerativas ou problemas de postura causadas na infância.

Na infância, o período de desenvolvimento da marcha é caracterizado pelo controle imaturo da postura, grande variabilidade cinemática, irregularidades em parâmetros espaço-temporais, ampla base de apoio e frequentes quedas (HALLEMANS et al., 2004). O domínio da posição em pé, representa o primeiro marco no desenvolvimento pela estabilidade postural do bebê. Isso indica que o controle da musculatura foi obtida e a força da gravidade não pode mais fixar limitações insuperáveis sobre o movimento, ocorrendo assim a locomoção (HALLEMANS et al., 2004).

Aos três anos de idade, aproximadamente, as instabilidades posturais e a falta de força muscular são superadas e o andar torna-se relativamente mais estável e regular. Contudo, pequenas mudanças continuam ocorrendo até os sete ou oito anos de idade (HALLEMANS et al., 2004) (HAUSDORFF et al., 1999).

A avaliação cinemática pode ser realizada considerando o tempo da marcha (cinemática linear), ou pela descrição do movimento angular (cinemática angular) (DELISA, 1998). Em ambos os casos, a descrição computacional pode ser utilizada para ajudar a descrever o movimento, posição anatômica e termos angulares durante o período da marcha. Para tanto, o primeiro passo para realizar a avaliação cinemática computacional é a captura da marcha humana para reconstrução espacial do movimento.

Nesse contexto, o presente trabalho possui seu foco nos estudos e proposta de um sistema de captura de movimentos para análise da marcha de um indivíduo, com objetivo de auxiliar na reabilitação motora neurológica ou ortopédica de pessoas que possuem alguma deficiência do gênero, utilizando recursos computacionais de baixo custo.

A captura de movimentos (MoCap, do inglês, *Motion Capture*) é o processo de gravar movimentos de objetos ou de pessoas por meio de câmeras ou equipamentos especiais. Esse método é utilizado na indústria do entretenimento de jogos digitais, cinema, esportes, aplicações médicas e visão computacional na robótica. Um exemplo é a captura para simular movimentos de personagens digitais (avatar) em filmes que utilizam recursos de computação gráfica. Por outro lado, os jogos digitais utilizam a captura de movimentos do *joystick* ou de usuários, para controlar os movimentos dos personagens, a partir de movimentos reais realizados pelo jogador.

A representação computacional de movimentos pode ser obtida por meio de dispositivos dedicados ou por sistemas ópticos. Esse primeiro é desenvolvido com a finalidade específica de registrar movimentos utilizando sensores infravermelhos ou equipamentos mecânicos atrelados ao corpo humano. Os sistemas ópticos são *softwares* desenvolvidos que utilizam câmeras de vídeos para capturar o movimento por meio da detecção de marcadores, colocados no corpo do indivíduo. Estes marcadores ficam em destaque nas imagens, tendo em vista que são registrados no ambiente com pouca iluminação.

Os sistemas ópticos que utilizam marcadores possuem a vantagem de operar com maior resolução espacial (FIGUEROA; LEITE; BARROS, 2003), sendo que a identificação dos marcadores pode ser realizada de forma automática por *software*, uma vez que esses estão em destaque nas imagens. Os marcadores são classificados como ativos ou passivos, sendo que esse primeiro emite uma luz própria e os passivos são revestidos por um material reflexivo. Os marcadores ativos podem restringir a execução de alguns movimentos, uma vez que possuem uma pequena

bateria e emissores de luz infravermelho no corpo do indivíduo. Os marcadores passivos permitem maior mobilidade, considerando que esses são fixados no corpo sem usar fios ou sensores. Estes marcadores refletem a luz de fontes emissoras localizadas próximas às câmeras de vídeo.

Uma limitação da captura de movimentos por dispositivos ópticos é o registro da projeção do ambiente tridimensional em um plano 2D, uma vez que a projeção de uma câmera fornece somente um campo de visão 2D do movimento. Por outro lado, se vários registros de vídeos do mesmo ambiente capturarem uma referência conhecida previamente, um processo de calibração das câmeras torna possível calcular a reconstrução tridimensional dessa referência. Para tanto, a calibração consiste em informar ao sistema as coordenadas do plano da imagem 2D e as coordenadas espaciais 3D do objeto conhecido. Após esse processo, se os marcadores forem registrados por no mínimo duas câmeras no mesmo ambiente onde foi realizado a calibração, as coordenadas espaciais dos marcadores podem ser obtidas por meio do método DLT (*Direct Linear Transformation*) (ABDEL-AZIZ; Y.I.KARARA, 1971)(MILLER; SHAPIRO; MCLAUGHLIN, 1980) (WOOD; MARSHALL, 1986) (HATZE, 1988) (CHEN; ARMSTRONG; RAFTOPOULOS, 1994).

Os sistemas de captura de movimentos baseados em câmeras de vídeo possuem como etapa inicial a detecção dos marcadores em cada quadro de cada vídeo. Esse procedimento consiste diretamente na detecção destes marcadores, fornecendo para as etapas seguintes a localização 2D de cada marcador nas respectivas imagens. Essa detecção precisa reconhecer todos os marcadores de interesse e possuir uma precisão aceitável, pois o seu desempenho influencia diretamente as etapas seguintes. Para auxiliar no processo de detecção dos marcadores, em geral, as capturas são realizadas em um ambiente controlado, com fundo escuro e pouca iluminação.

A proposta do presente trabalho consiste no desenvolvimento de um sistema de captura de movimentos, utilizando câmeras de vídeos e marcadores passivos, adaptado a realidade da análise de marcha desenvolvida pelo grupo de pesquisa em Biodinâmica do DEFMH-UFSCar, coordenado pela Professora Doutora Paula Hentschel Lobo da Costa. Essa especificação se deve ao ambiente já construído e a necessidade de melhoria e automatização do processo de detecção dos marcadores, garantindo a detecção de todos os marcadores utilizados e eliminando os falsos marcadores detectados.

1.1 Organização do Trabalho

No capítulo 2 serão abordados os trabalhos disponíveis na literatura e relacionados a este projeto.

No capítulo 3 serão apresentadas as abordagens teóricas para realizar a captura de movimentos que utilizam os sistemas ópticos e marcadores passivos, foco deste trabalho.

No capítulo 4 serão apresentados os experimentos computacionais das etapas efetivamente implementadas, bem como os resultados destas implementações.

No capítulo 5 serão argumentados os resultados obtidos no presente trabalho, incluindo a discussão entre o sistema proposto, implementações realizadas e trabalhos futuros.

Capítulo 2

TRABALHOS RELACIONADOS

Neste capítulo serão apresentados os trabalhos relacionados presentes na literatura e os conceitos teóricos fundamentais para o entendimento técnico desta proposta.

2.1 Captura de movimentos

Captura de movimentos, ou *Motion Capture* (MoCap) consiste no processo de capturar movimentos de objetos ou pessoas. Esses registros são utilizados para análise dos movimentos em diversas áreas de pesquisas, reprodução computacional ou animações digitais.

O rastreamento de movimentos humano por meio de registros de vídeo tem sido estudado extensivamente nos últimos 35 anos, com as primeiras obras datadas no início da década de 1980 (O'ROURKE; BADLER et al., 1980) (HOGG, 1983). A variedade de métodos estatísticos (AGARWAL; TRIGGS, 2004) (SMINCHISESCU et al., 2005), bem como métodos determinísticos (TAYLOR, 2000) (MORI et al., 2004), foram desenvolvidos para rastreamento de pessoas em um único campo de visão (RAMANAN; FORSYTH, 2003) (RONFARD; SCHMID; TRIGGS, 2002) (MORI et al., 2004) ou por vários campos de visão (DEUTSCHER; REID, 2005) (GRAUMAN; SHAKHNAROVICH; DARRELL, 2003).

Para permitir a análise computacional destes movimentos, foram desenvolvidas diversas técnicas de captura, incluindo equipamentos dedicados para este fim (O'BRIEN et al., 1999) (VLASIC et al., 2007). Muitos destes dispositivos foram desenvolvidos para uma aplicação específica, sendo dificilmente encontrados comercialmente. Outros dispositivos são hoje produzidos em grande escala, como por exemplo, o Microsoft Kinect (ANDREWS; BAKER, 2006) e o Nintendo Wii¹, que devido ao foco em jogos digitais possuem um grande mercado consumidor. As abor-

¹Manual da Nintendo - <http://www.nintendo.com/consumer/downloads/WiiChEng.pdf>

dagens que não utilizam dispositivos especiais, em geral, fazem captura por meio óptico (MOESLUND; HILTON; KRÜGER, 2006), utilizando uma ou mais câmeras de vídeo.

Neste trabalho, o sistema utilizado tem por finalidade fornecer informações detalhadas do movimento do andar de uma pessoa, sem utilizar qualquer dispositivo dedicado. Para isso, recursos de visão computacional são utilizados para interpretar corretamente cada quadro de vídeo, considerando as propriedades intrínsecas das câmeras, referentes à sincronização, calibração e a reconstrução tridimensional do movimento.

A seguir, são apresentados os dispositivos e abordagens mais populares encontradas atualmente, contextualizando a escolha deste projeto e suas alternativas de metodologia.

2.2 Rastreadores por sensores

Os sistemas de captura de movimentos por sensores são aqueles que não utilizam câmeras de vídeo e fazem uso de outros dispositivos para registrar informações sobre o posicionamento do corpo. Podemos classificá-los basicamente em dois grupos: mecânicos e magnéticos.

2.2.1 Mecânicos

Os rastreadores mecânicos são equipamentos conectados fisicamente ao corpo e capturam a posição do indivíduo a partir de um ponto, denominado “ponto de referência”. Existe uma conexão entre o alvo e o ponto de referência. Um ator coloca um equipamento esquelético (figura 2.1)² ao longo do seu corpo, de forma que as partes mecânicas se movam simultaneamente aos movimentos realizados pelo corpo. Desta forma, os dispositivos de medição podem fornecer os ângulos das articulações durante o movimento, além de outras propriedades, como velocidade e aceleração. A postura do corpo pode ser determinada combinando as informações dos sensores.

Segundo (VLASIC et al., 2007), os rastreadores mecânicos são mais apropriados quando são exigidos movimentos precisos e em alta velocidade. No entanto, a desvantagem é a mobilidade reduzida que ele oferece ao usuário. Outro problema apontado, é que tecidos moles permitem que a posição do equipamento seja alterada em relação ao corpo, prejudicando a eficiência do dispositivo.

O Wii é um video game da Nintendo em que o movimento do usuário é capturado por meio de sensores que ficam internamente nos controles (*Wii Remote* - figura 2.2)³. Os movi-

²Empresa SouVR - <http://en.souvr.com/product/200712/269.html>

³Nintendo - <http://www.nintendo.com/consumer/downloads/WiiChEng.pdf>



Figura 2.1: Dispositivo Mecânico da SouVR

mentos são capturados por um sensor infravermelho colocado próximo a televisão, responsável por definir a localização do usuário em posse do *Wii Remote*. O dispositivo também possui um sensor de aceleração (acelerômetro), que possibilita registrar a velocidade dos movimentos realizados pelo usuário. Isso significa que a força de um chute, por exemplo, vai ser medida pela intensidade do movimento capturado pelo sensor dedicado.



Figura 2.2: Wii Remote da console da Nintendo

O console possui também outros dispositivos de entrada de dados, entre eles, o *WiiZapper*, *WiiWhell* e o *WiiBalanceBoard*. O *WiiZapper* foi desenvolvido para jogos que envolvem pontaria e tiros, uma espécie de revólver. O *WiiWhell* foi especialmente desenvolvido para o usuário realizar dezenas de exercícios físicos, como alongamento, saltos e flexões, sendo esse sensor vendido como um dos mais revolucionários do console. O *WiiBalanceBoard* é semelhante a uma balança com sensores de pressão que permite o usuário fazer várias atividades como cabecear bolas virtuais, fazer yôga e praticar esportes.

2.2.2 Magnéticos

Os sistemas de captura por rastreadores magnéticos têm sido muito utilizados no rastreamento de movimentos em aplicações de realidade virtual. Com a captura magnética, sensores são colocados na roupa do corpo humano (figura 2.3)⁴ e estes captam baixas frequências emitidas por uma fonte transmissora (O'BRIEN et al., 1999). O sistema utiliza bobinas tanto no emissor quanto no receptor, dispostas ortogonalmente entre si. Estimuladas por corrente elétrica, as bobinas do emissor geram campos magnéticos que serão captados pelas bobinas dos sensores. Desta maneira, as informações sobre a orientação e posicionamento dos sensores, são calculadas e passadas para o controlador.



Figura 2.3: Dispositivo Magnético da Senztech

Devido a estas características, esses sistemas são apropriados para aplicações com necessidade de alta precisão e mobilidade maior que os mecânicos, mas precisam ser utilizados em ambientes sem interferências magnéticas. Segundo (LIVERMAN, 2004), por depender diretamente das características físicas de um campo magnético, os sensores podem perder sua potência à medida que se distanciam da fonte transmissora do campo magnético.

2.3 Rastreadores Ópticos

Os sistemas de reconhecimento de movimentos ópticos utilizam câmeras de vídeo. Uma diferença importante entre as abordagens ópticas é definida pelo uso ou não de marcadores. Para as técnicas que utilizam marcadores, estes podem ser denominadas como ativos ou passivos. A

⁴Empresa Senztech - www.senztech.cc/showpros.aspx?proid=2

principal vantagem dos sistemas ópticos é a maior liberdade de movimentos em comparação ao uso de sensores. Na utilização de câmeras, para a captura dos movimentos, são necessárias a manipulação dos parâmetros intrínsecos e extrínsecos da câmera, bem como a realização de procedimentos de calibração e reconstrução.

2.3.1 Sem Marcadores

O rastreamento de movimentos por sistemas ópticos que não usam marcadores, geralmente são utilizados para estimar a pose do corpo humano ou reconhecer movimentos em um determinado ambiente. A captura pode ser realizada por um sistema monocular ou por várias câmeras (SUNDARESAN; CHELLAPPA, 2005), dependendo da especificação da aplicação.

O método *3D monocular* reconhece movimentos do corpo humano em um ambiente, por meio de imagens de um campo de visão. Tal método é baseado em uma métrica robusta de *matching* que envolve técnicas de detecção de arestas nos limites do corpo que está em movimento no vídeo (SMINCHISESCU; TRIGGS, 2003). Após detectar o indivíduo, estruturas geométricas são modificadas sobrepondo as partes do corpo (figura 2.4).



Figura 2.4: Rastreamento de movimentos sem marcador (SMINCHISESCU; TRIGGS, 2003).

Outras técnicas de rastreamento de movimentos sem marcadores podem ser encontradas em (MOESLUND; GRANUM, 2001) (WANG; SINGH, 2003). No entanto, essas aplicações não dispõem de informações da profundidade do indivíduo, bem como as três dimensões do ambiente e dos movimentos realizados.

O dispositivo Kinect foi desenvolvido para ser utilizado como acessório na plataforma de jogos Xbox360 (ANDREWS; BAKER, 2006), com o intuito de prover recursos de interação que não envolve os controles tradicionais. O Kinect apresenta três componentes em um formato circular (figura 2.5). Da esquerda para direita, estão localizados o emissor de infravermelho (A), a câmera RGB (B) e a câmera de profundidade (C). Além disso, possui quatro microfones (D) localizados na parte inferior do dispositivo. A base (E) é motorizada, permitindo controlar

a inclinação do dispositivo por meio do *software* e um acelerômetro que pode ser utilizado para calcular a intensidade do movimento.

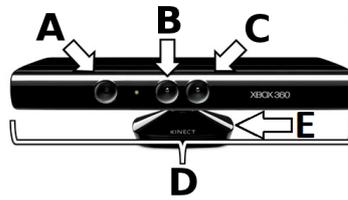


Figura 2.5: Console Xbox360 com dispositivos Kinect (ANDREWS; BAKER, 2006)

A câmera RGB é um dispositivo comum de captura de imagens, como uma câmera convencional, utilizado para fazer o reconhecimento facial e de objetos. A câmera de profundidade detecta a projeção de pontos infravermelhos, realizada pelo projetor infravermelho, distribuídos no ambiente de forma uniforme (figura 2.6). Assim, se algo entrar na frente do projetor, o padrão da distribuição dos pontos projetados será alterado, diferentemente da distribuição da parede de fundo. Dessa forma, o sistema calcula o tamanho dos pontos projetados, bem como o espaçamento entre eles, fornecendo a posição do indivíduo e dos objetos no ambiente. Devido a natureza dos sensores, não é necessário iluminação externa para funcionar. Esse dispositivo não utiliza controle remoto ou marcadores para rastrear os movimentos. No entanto, ele apenas oferece uma projeção do ambiente tridimensional, a partir de um ponto de vista.

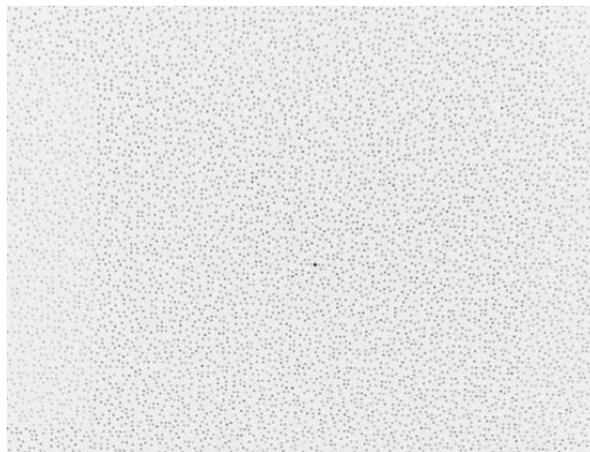


Figura 2.6: Pontos infravermelhos distribuídos no ambiente (ANDERSEN et al., 2012)

2.3.2 Marcadores Ativos

Os marcadores ativos utilizam como referência a iluminação em LED (figura 2.7). O indivíduo transporta a fonte de energia e cabos que conectam os diversos LEDs. As câmeras são equipadas com filtros ópticos sintonizados para capturar o comprimento de onda da fonte de iluminação dos LEDs (KIRK; O'BRIEN; FORSYTH, 2005). Cada marcador é configurado para

emitir seu próprio feixe de luz, modulando os seus pulsos, permitindo a identificação de cada um, por meio de diferentes espectros de luz.



Figura 2.7: Marcadores Ativos. Usuário utiliza sensores que emitem raios infravermelhos (KIRK; O'BRIEN; FORSYTH, 2005)

A utilização do espectro é justificada pela capacidade de isolar as ondas nesta frequência através de um filtro na câmera de captura. Portanto, não é necessário realizar a gravação do vídeo em um ambiente escuro, possibilitando o registro em qualquer lugar. Por outro lado, o uso desse sistema fica restrito a tecnologias de câmeras especializadas e marcadores infravermelhos, além do espaço físico de bateria e fios no corpo da pessoa.

2.3.3 Marcadores Passivos

Esses marcadores são pequenas esferas revestidas por um material reflexivo, colocadas sobre uma roupa escura no corpo do indivíduo (figura 2.8). Cada marcador reflete a luz, gerada nas proximidades das câmeras em um ambiente com pouca iluminação. O vídeo pode ser registrado por uma câmera simples, pois os marcadores serão detectados para fornecer as coordenadas do plano da imagem de cada objeto.

Esse sistema possui a vantagem de operar com maior resolução espacial e temporal nos registros, ainda, permite maior mobilidade dos movimentos do corpo e possui baixo custo operacional. No entanto, a detecção de objetos que não pertencem ao conjunto de marcadores inseridos no indivíduo ou uma identificação errada de um marcador, podem resultar em funcionamento inadequado do sistema.

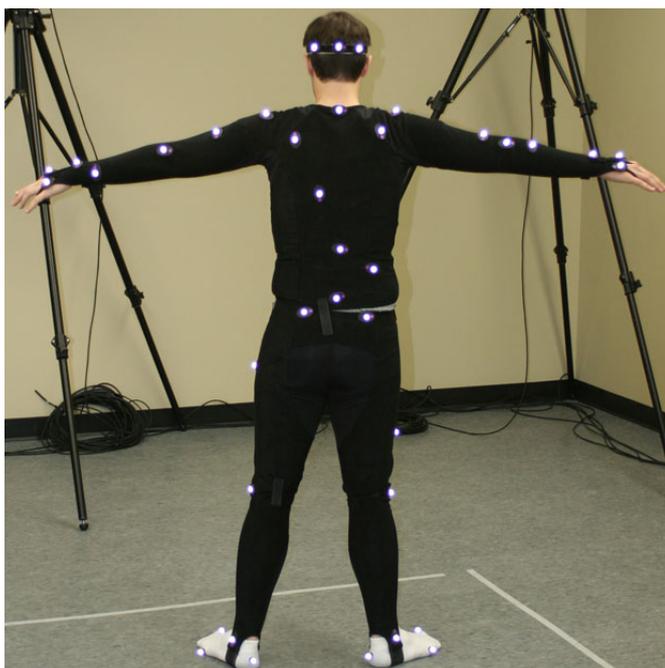


Figura 2.8: Marcadores passivos. Usuário utiliza uma roupa preta com marcadores reflexivos (CACs, 2007)

Devido a flexibilidade fornecida pelos marcadores passivos, como o baixo custo e liberdade dos movimentos, esta abordagem foi escolhida pelo grupo de pesquisa em Biodinâmica do DEFMH-UFSCar (Departamento de Educação Física e Motricidade Humana), coordenado pela Professora Doutora Paula Hentschel Lobo da Costa. Considerando a aplicação específica apresentada por esse grupo, definimos este projeto de mestrado visando tratar o conjunto de vídeos existentes, com o objetivo de realizar a detecção e reconstrução dos movimentos de marcha contidos neles.

2.4 DVideoW

O sistema DVideoW⁵ (*Digital Video for Biomechanics for Windows 32*) foi desenvolvido no laboratório de Instrumentação de Biomecânica (LIB) da Universidade Estadual de Campinas (Unicamp), com cooperação do instituto de computação (IC) e o Instituto de Física “Gleb Wataghin” (IFGW), ambos da Unicamp. Ele foi projetado e amplamente utilizado em diferentes aplicações, entre elas: análise de marcha (ou caminhada) (CASTRO, 2012), análise de movimentos respiratórios (LOULA, 2005), rastreamento automático de jogadores de futebol (MOURA et al., 2007) e reconstrução de superfícies do corpo humano.

Segundo (BRENZIKOFER et al., 2001), o DVideoW é um sistema de análise tridimensional de

⁵<http://revistapesquisa.fapesp.br/2002/12/01/detalhes-do-movimento/>

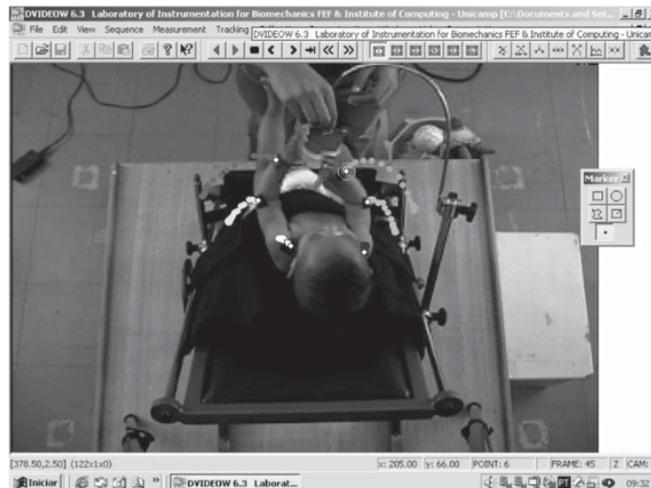


Figura 2.9: Imagem do sistema DVideoW (CARVALHO; TUDELLA; BARROS, 2005)

movimentos humanos, baseados no registro videogramétrico⁶ de imagens por meio de câmeras de vídeo analógicas ou digitais, no qual estão implementados os seguintes procedimentos: captura ou aquisição de imagens de vídeo; medição de pontos nas sequências de imagens; calibração de câmeras e reconstrução tridimensional dos pontos de interesse. A análise dos segmentos (em movimento ou não) representados a partir dos pontos reconstruídos é realizada através de rotinas desenvolvidas na linguagem R e em MATLAB⁷, sendo as demais implementações em C++.

Os trabalhos realizados pelo grupo de pesquisa em Biodinâmica do DEFMH-UFSCar, coordenado pela Professora Doutora Paula Henschel Lobo da Costa, utilizam o *software* DVideoW para o desenvolvimento de pesquisas. Com o objetivo de avaliar a marcha de crianças, o departamento usa quatro câmeras de vídeos e marcadores passivos colocados nas crianças para capturar os movimentos.

Nesse trabalho estudamos a viabilidade de propor uma nova técnica de detecção de marcadores, bem como na automatização dos marcadores no primeiro quadro da avaliação e no período da marcha. A implementação deste trabalho não se trata de uma extensão do *software* DVideoW, sendo apenas uma proposta de melhoria para o sistema utilizado pelo DEFMH-UFSCar.

⁶Tecnologia de se obter informações confiáveis sobre objetos físicos e meio ambiente através de processos de gravação, medição e interpretação de cenas de vídeo.

⁷<http://www.mathworks.com/>

2.5 Considerações Finais

Neste capítulo foram apresentados trabalhos relacionados à captura de movimentos, bem como o *software* utilizado pelo DEFMH-UFSCar. No próximo capítulo serão apresentadas as técnicas necessárias para a implementação das etapas dos sistemas de captura de movimentos por dispositivos ópticos e com marcadores passivos.

Capítulo 3

MATERIAIS E MÉTODOS

A utilização de câmeras de vídeo permite maior flexibilidade na seleção dos movimentos estudados, diminui as restrições quanto aos equipamentos de captura e permite maior difusão deste tipo de análise, uma vez que os custos são menores (BARROS et al., 1999). Contudo, a captura de movimentos por uma única câmera não disponibiliza informações tridimensionais do ambiente e dos movimentos realizados. Outra restrição é a oclusão de marcadores no período da marcha, tendo em vista que alguns marcadores são sobrepostos devido aos movimentos do corpo, considerando a limitação do único campo de visão da câmera.

Para capturar todos os marcadores colocados no corpo da criança no período da marcha precisamos utilizar mais de uma câmera, dispostas em locais diferentes, de modo que se o marcador desaparece do campo de visão de alguma câmera, ele aparecerá na visão de outra. Para tanto, processos integrados de calibração, sincronização, detecção, correspondência e *tracking* são necessários para conhecer as informações tridimensionais do ambiente e realizar a reconstrução tridimensional do movimento (figura 3.1). Nesta sequência de etapas, temos basicamente dois fluxos de dados: sequência para calibração das câmeras e sequência do reconhecimento do movimento. Na sequência de calibração o fluxo é definido a partir dos vídeos de calibração, e na sequência do movimento o fluxo é definido a partir dos vídeos dos movimentos capturados, neste caso as marchas.

A etapa de calibração consiste no cálculo das informações tridimensionais do ambiente, bem como na consideração do posicionamento das câmeras. Para realizar esta etapa, um padrão conhecido previamente é capturado pelas câmeras. Assim, a correlação entre as coordenadas 3D definidas para os pontos do padrão conhecido e as coordenadas destes pontos na projeção de cada vídeo é calculada. A descrição deste processo, assim como detalhes da implementação, podem ser encontrados na seção 3.5.

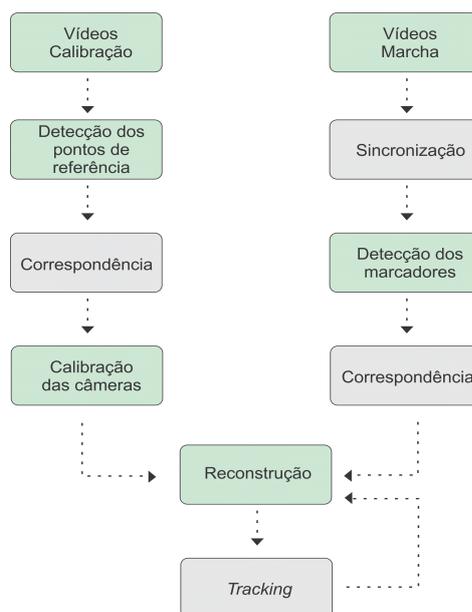


Figura 3.1: Etapas para captura de movimentos por dispositivos ópticos. Processos em destaque foram implementados neste trabalho.

O processo de sincronização garante que todos os vídeos operem na mesma frequência de aquisição, resolução de captura e instante de tempo. A descrição deste processo pode ser encontrados na seção 3.1.

A etapa de detecção dos marcadores é responsável pelo reconhecimento dos marcadores nos quadros (*frames*) dos vídeos, encontrando as coordenadas do plano da imagem (2D) de todos os marcadores em todas as câmeras. A seção 3.2 detalha a metodologia empregada neste trabalho para esta etapa.

O processo de correspondência é necessário na etapa de calibração e no primeiro quadro da avaliação, estabelecendo a relação de um-para-um dos marcadores detectados na etapa anterior, ou seja, a correspondência das coordenadas 2D em todas as câmeras que refere-se a um determinado marcador. Ao longo da trajetória do marcador, a etapa de *tracking* acompanha as correspondências estabelecidas. A descrição destas etapas encontram-se nas seções 3.3 e 3.4, respectivamente.

Na reconstrução são utilizadas as coordenadas 2D dos marcadores para calcular as coordenadas espaciais 3D, juntamente com os dados das câmeras, obtidas na etapa de calibração. Para utilizar este método é preciso ter as coordenadas do marcador em pelo menos duas projeções. A seção 3.5 inclui a descrição da reconstrução implementada.

As seções seguintes abordam detalhadamente os métodos envolvidos para os sistemas de captura de movimentos que utilizam dispositivos ópticos, bem como apresentam as abordagens

disponíveis na literatura e os conceitos proposto no presente trabalho para a etapa de detecção.

3.1 Sincronização

Cada vídeo captura movimentos de uma cena tirando dezenas de fotografias (quadros) por segundo. Diferentes amostragens temporais (quadros por segundo) são utilizadas de acordo com a tecnologia empregada na câmera. Em câmeras comuns, facilmente encontradas no mercado, a taxa de amostragem de trinta quadros por segundo é a mais popular, sendo possível taxas maiores em tecnologias mais avançada. Outra diferença entre os diversos modelos de câmeras de vídeo é a resolução espacial dos quadros.

Para realizar o processo de reconstrução a partir de dois ou mais registros de vídeos, é preciso estabelecer a sincronização temporal das câmeras, isto é, garantir que os quadros sejam simultâneos e registrem o mesmo movimento.

Estas diferenças podem ser resolvidas basicamente por duas abordagens: por *hardware* ou *software*, ou seja, a sincronização pode ser realizada pelo dispositivo de captura ou pelo alinhamento dos vídeos em uma etapa de pré-processamento.

A abordagem caracterizada como “sincronização por *hardware*” consiste no uso de um dispositivo externo que controla todas as câmeras de uma mesma marca e modelo. Dessa forma, o dispositivo comanda apenas o início e término dos registros cinematográficos, tendo em vista que as câmeras operam na mesma frequência (BARROS et al., 1999). Esse conceito é aplicado somente em câmeras especiais, de custo elevado, que permite o controle externo da aquisição.

A abordagem caracterizada como “sincronização por *software*” consiste na utilização de algum método de sincronização dos vídeos que utilizam câmeras comuns, e possivelmente de modelos diferentes. Tendo em vista que as câmeras são ligadas em tempo distintos, os registros possuem uma defasagem temporal entre elas. Dessa forma, para garantir que a frequência e o início da avaliação cinematográfica seja igual para todos os vídeos, procedimentos de interpolação são realizados para garantir a sincronização dos vídeos.

Para sincronizar temporalmente duas câmeras na manipulação por *software*, um método bastante utilizado e simples é a criação de um sinal visual de fácil detecção, como por exemplo o piscar de uma luz no ambiente com pouca iluminação. Esse feixe de luz é capturado por todas as câmeras e posteriormente o algoritmo procura nas transmissões de vídeo, todos os quadros (*frames*) que possuem alternância de luz (KOLAHİ et al., 2007) (CASTRO, 2012). Depois de identificar a etapa anterior, os quadros seguintes das câmeras estão sincronizadas.

Outra abordagem apresenta um método de sincronização pelo uso da técnica DLT (*Direct Linear Transformation*) de (ABDEL-AZIZ; Y.I.KARARA, 1971). Este método calcula a diferença de fases entre duas câmeras, minimizando os erros de reconstrução de um ponto em movimento. Uma vez que o valor da diferença é calculado, uma das gravações é escolhida como referência e a segunda é sincronizada com a primeira, por meio de uma interpolação *spline* cúbica (POURCELOT et al., 2000).

Outro método utilizado para sincronizar duas ou mais câmeras de vídeo é a criação de um sinal de áudio, como por exemplo o apertar de um campainha. Esse áudio sonoro é registrado por todas as câmeras e posteriormente o algoritmo procura nos registros de vídeos, o primeiro quadro (*frame*) que está associado com o áudio da campainha. Depois de identificar o áudio sonoro na etapa anterior, os quadros seguintes das câmeras estão sincronizadas.

O material fornecido pelo DEFMH-UFSCar utilizou o método de ascender a luz no ambiente com pouca iluminação. Dessa forma, este trabalho extraiu os quadros de todos os registros de vídeos das câmeras no momento da avaliação, e manualmente os quadros foram excluídos até o registro do feixe de luz. Os quadros seguintes são considerados sincronizados e a detecção dos marcadores foram realizados a partir desse evento.

Na próxima seção serão abordados as técnicas utilizadas para detecção dos marcadores, bem como os métodos presentes na literatura.

3.2 Detecção de Marcadores

O rastreamento automático de marcadores é um problema bastante desafiador, mesmo quando consideramos sistemas ópticos com marcadores passivos, onde o contraste entre o fundo da imagem e o destaque dos marcadores auxiliam na sua detecção. Este problema se torna mais difícil quando consideramos, por exemplo, as condições de variabilidade do ambiente e os diferentes níveis da complexidade do estudo do movimento. A fim de garantir o rastreamento automático, algumas restrições são impostas para o número, proximidade e o tamanho dos marcadores, bem como na ausência de objetos no ambiente de aquisição, campo de visão da câmera e os segmentos do corpo (CHEN; ARMSTRONG; RAFTOPOULOS, 1994).

Os sistemas que visam capturar os movimentos para análise cinemática, geralmente possuem a especificação de usar uma roupa preta com os marcadores passivos. A localização dos marcadores identificados em cada quadro do registro das câmeras, fornecem a entrada para o processo da estimação do movimento (GUERRA-FILHO, 2005). Essa entrada consiste nas coordenadas 2D do plano da imagem de todos os marcadores, simultaneamente capturados por

todas as câmeras.

Muitas abordagens encontradas na literatura (BARROS et al., 1999) (HERDA et al., 2001) (WANG; SINGH, 2003) (GUERRA-FILHO, 2005)(KOLAHI et al., 2007) (CASTRO, 2012), que usam os sistemas ópticos para avaliação cinemática, detectam os marcadores por uma técnica simples de *thresholding*. Alguns trabalhos calculam no primeiro plano a subtração do fundo, e segmentam as diferenças da roupa do corpo (escuro) e o marcador (brilhante) (GUERRA-FILHO, 2005) (figura 3.2). O resultado do processo é uma imagem binarizada com os componentes em destaque no corpo, representando a localização dos marcadores.

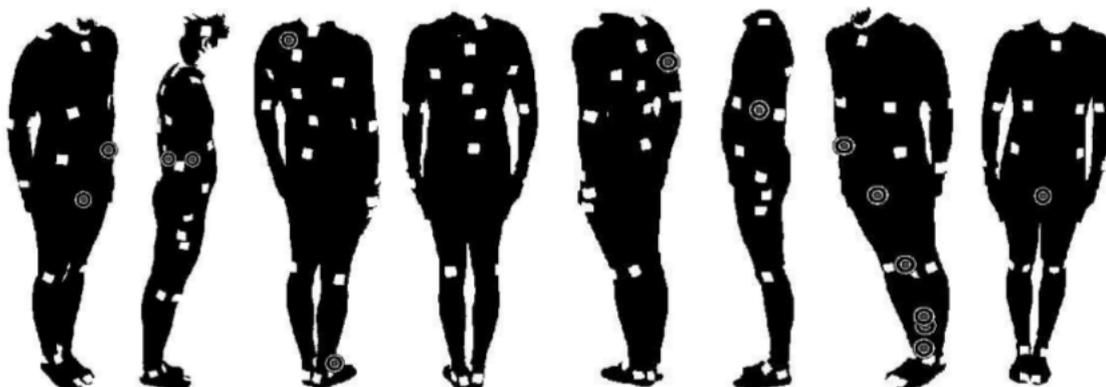


Figura 3.2: Detecção dos marcadores no corpo da pessoa usando a técnica de *Thresholding* de (GUERRA-FILHO, 2005).

O método citado pode ser ineficiente, uma vez que a iluminação vinda das câmeras pode gerar reflexos em objetos ou artefatos no ambiente de registro, ocasionando na detecção de falsos marcadores (*falsos positivos*). Dessa forma, no presente trabalho o uso do *threshold* na detecção dos marcadores foi inviabilizado, uma vez que o material fornecido pelo DEFMH-UFSCar possui objetos no ambiente de registros e que potencialmente são detectados como falsos marcadores. Estes *falsos positivos* resultam no funcionamento inadequado do sistema e comprometem as etapas seguintes de reconstrução, correspondência e *tracking*.

No ambiente de registro, que possui muitos objetos que podem ser detectados como falsos marcadores, uma alternativa seria utilizar métodos que realizam a detecção de estruturas geométricas em imagens digitais. Tendo em vista que os marcadores passivos possuem o formato circular e podem ser diferenciados de outros objetos detectados erroneamente pelo sistema, no presente trabalho estudou-se a viabilidade de fazer a detecção dos marcadores pela transformada de *Hough*, uma vez que esse método localiza as circunferências em imagens digitais. Os detalhes sobre a transformada de *Hough* foram apresentadas na próxima seção.

3.2.1 Transformada de *Hough*

A detecção de um objeto localizado no ambiente que possui várias formas diferentes, pode se tornar um problema complexo em visão computacional. Para casos onde o objeto detectado se assemelha a uma estrutura geométrica simples, como retas, circunferências e elipses, a transformada de *Hough* (HT) pode ser apropriada. Seu desenvolvimento iniciou-se na localização e detecção de retas (HOUGH; PAUL, 1962), posteriormente foi estendida para possibilitar a localização de outras formas geométricas que podem ser parametrizadas, tais como circunferências (*Circles Hough transform* - CHT) e elipses (DUDA; HART, 1972) (TSUJI; MATSUMOTO, 1978) (BALLARD, 1981).

Neste trabalho foi utilizado a CHT para detectar os marcadores passivos colocados no corpo da criança. Tendo em vista que os marcadores possuem o formato circular a CHT pode ser aplicado para detectar os objetos e fornecer as coordenadas 2D do plano da imagem para a etapa de reconstrução. Para tanto, uma circunferência na imagem digital pode ser descrita por:

$$(x - a)^2 + (y - b)^2 = r^2 \quad (3.1)$$

onde (a, b) são as coordenadas do centro da circunferência e r o raio da circunferência. A CHT usa a equação para cada ponto (*pixel*) (x, y) mapeado na imagem, isto é, para os pontos realçados por algum detector de borda (BALLARD, 1981). Para cada ponto (*pixel* da borda) (x, y) na circunferência original (figura 3.3a)¹, pode ser definida uma circunferência centrada em (x, y) com raio r , segundo a equação 3.1. As intersecções dos pontos de todas as circunferências no espaço de *Hough* (figura 3.3b), corresponde ao ponto central da circunferência na imagem original (figura 3.3a).

A transformada é mostrada no espaço de *Hough* (figura 3.3b). Com o raio r fixado, as quatro intersecções dos pontos definem o centro da circunferência no espaço original (figura 3.3a). Na equação 3.1, pode ser observado que possui três parâmetros a ser calculados (x, y) , (a, b) e (r) . Quando o raio (r) não é conhecido, uma matriz 3D é necessária para acumular todos os cálculos da CHT, o que seria altamente ineficaz quanto ao tempo de processamento da transformada (ILLINGWORTH; KITTLER, 1988). Se o raio é fixado, então os parâmetros espaciais podem ser reduzidos para 2D (a posição do centro da circunferência).

Múltiplas circunferências com os raios conhecidos também podem ser encontrados usando a mesma técnica. Para tanto, uma matriz é usada para acumular as intersecções e os pontos (x, y) de todas as circunferências no espaço de *Hough*.

¹<http://www.mathworks.com/help/images/ref/imfindcircles.html>

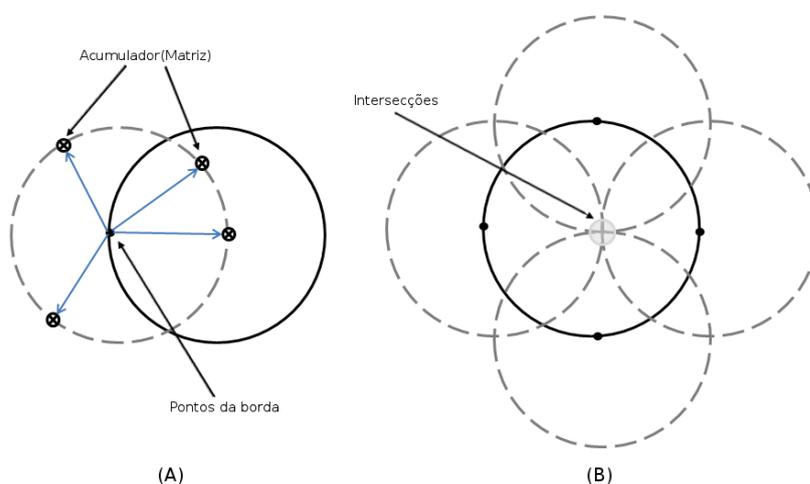


Figura 3.3: (A) Pontos na circunferência na imagem. (B) Quatro intersecções dos pontos no espaço de Hough

O método original da CHT (DUDA; HART, 1972), requiere um grande número de pontos para ser acumulado no espaço de *Hough*, uma vez que cada ponto na imagem é considerada como possível borda da circunferência. Outra abordagem para detectar circunferências em imagens digitais também utiliza os pré-requisitos proposto por (DUDA; HART, 1972), isto é, passa por um detector de borda e posteriormente o método elimina os *pixels* de bordas que possuem ligações quebradas com outros *pixels*. Após esse processo, as equações trigonométricas 3.2 e 3.3 são calculadas para cada *pixel* de borda (KIMME; BALLARD; SKLANSKY, 1975) (DAVIES, 2012).

$$x_c = x - R\left(\frac{g_x}{g}\right) \quad (3.2)$$

$$y_c = y - R\left(\frac{g_y}{g}\right) \quad (3.3)$$

onde x_c e y_c representam a localização das coordenadas obtidas pelas equações, x e y são as coordenadas 2D na imagem para cada *pixel* de borda, R o raio estimado e $\left(\frac{g_x}{g}, \frac{g_y}{g}\right)$ são os valores de $\cos\theta$ e $\sin\theta$, respectivamente (DAVIES, 2012). Esse método ainda estima a quantidade de raios necessários para localizar uma circunferência com o determinado R . Por exemplo, se o R for 10 (dez) a função estima que são necessários 200 (duzentos) raios para se formar uma circunferência. Para cada *pixel* de borda (figura 3.4a) o método encontra um ponto central e localiza todos os agrupamentos que possui intersecções nas mesmas coordenadas (figura 3.4b) (KIMME; BALLARD; SKLANSKY, 1975) (DAVIES, 2012).

Os sistemas de captura de movimentos usam marcadores passivos, os quais possuem o formato circular, que podem ser diferenciados de outros objetos pela forma, evitando que outros

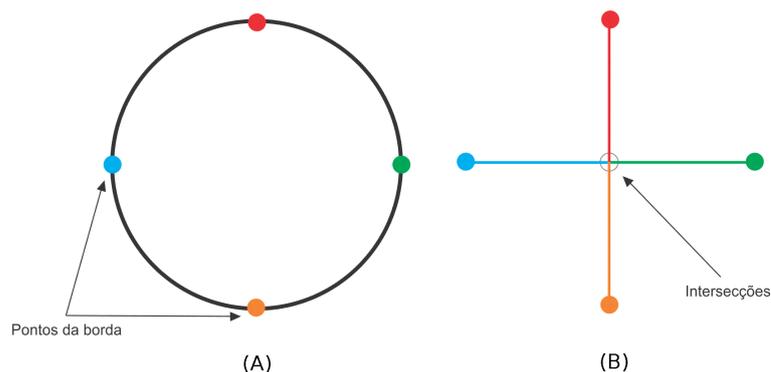


Figura 3.4: (A) Pontos na circunferência na imagem. (B) Interseções no espaço de *Hough*.

objetos sejam detectados erroneamente pelo sistema. Portanto, este projeto implementou e incorporou a transformada de *Hough* no processo de detecção dos marcadores e foi avaliado o desempenho dessa abordagem no ambiente de teste.

3.3 Correspondência

O problema da correspondência entre as diferentes projeções capturadas é encontrado em diversas áreas de pesquisas (LI; MENG; HOLSTEIN, 2003), incluindo visão computacional (MOESLUND; HILTON; KRÜGER, 2006) (DAVID et al., 2003), reconhecimento de padrões, geometria computacional (DAESIK, 2006) e registro de imagens (ZITOVA; FLUSSER, 2003). O *matching* entre os vídeos busca a correspondência de imagens em diferentes ângulos, tiradas do mesmo ambiente, resultando em imagens panorâmicas (BROWN; LOWE, 2003). As características de *matching* podem ser encontradas por meio de pontos, linhas, curvas ou superfícies que geralmente são formuladas por um problema de reconhecimento de padrão geométrico (LEWIS, 1995) (LI; MENG; HOLSTEIN, 2003).

Para os sistemas de captura de movimentos onde o marcador é detectado por várias câmeras, as correspondências das coordenadas dos marcadores precisam ser estabelecidas. No exemplo ilustrado pela figura 3.5, no primeiro quadro da avaliação as câmeras registram cinco marcadores. O resultado da detecção de cada câmera é uma lista que possui as coordenadas 2D da tela. Em outras palavras, quando um marcador é detectado por uma câmera, sua localização no quadro é determinada no plano da imagem 2D. Quando esse mesmo marcador for detectado por outra câmera é necessário estabelecer a correspondência entre esses dois pontos 2D (de uma câmera para a outra), para ser calculado na etapa de reconstrução.

Em muitas abordagens as correspondências dos marcadores para os sistemas de captura de movimentos são realizados manualmente pelo usuário que controla o sistema (HERDA et al.,

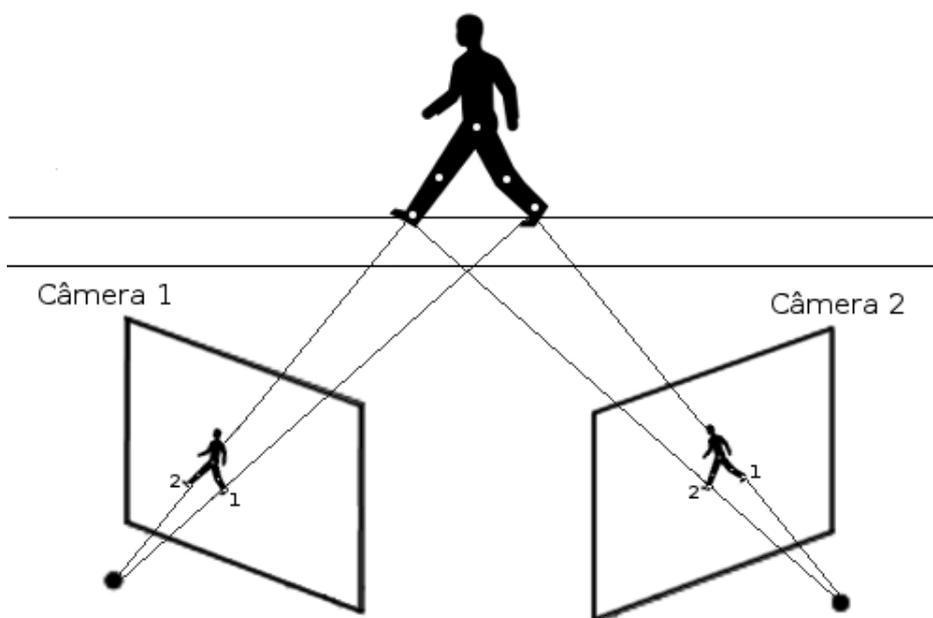


Figura 3.5: Ilustração para a correspondência dos marcadores

2001) (GUERRA-FILHO, 2005) (CASTRO, 2012), isto é, a correspondência está relacionada na sequência que o usuário seleciona os marcadores. Esse processo de correspondência “manual” acontece sucessivamente para todos os marcadores em todas as câmeras no primeiro quadro da avaliação. Após essa etapa, o usuário deve verificar se os marcadores reconstruídos no primeiro quadro é o número de marcadores conhecidos, colocados no indivíduo. Como todo processamento restante é automático, é absolutamente essencial que todos os marcadores sejam detectados no primeiro quadro da análise. Qualquer marcador não detectado por no mínimo duas câmeras e não reconstruído no primeiro quadro de análise é perdido na sequência de *tracking* da avaliação.

Existem várias abordagens que procuram melhorar a correspondência dos marcadores (LI; MENG; HOLSTEIN, 2003) (HERDA et al., 2001). No entanto, a maioria delas usam a correspondência manual e precisam de alguma intervenção do usuário. O grande desafio de realizar a correspondência dos marcadores para os sistemas de captura de movimentos é estabelecer a correspondência entre eles sem intervenção manual e sem informação prévia do usuário. Alguns trabalhos propõem algoritmos para a estimação da pose (LI; MENG; HOLSTEIN, 2003) (WADA et al., 2000), de forma arbitrária, que utilizam a correspondência do tipo usual “de um para um” entre pontos de um modelo e um conjunto de pontos observados, sem informação prévia e sem a inicialização do sistema (figura 3.6). No entanto, essas abordagens apenas estabelecem a pose e não fornecem a relação de correspondência dos marcadores no primeiro quadro de análise.

Neste trabalho a correspondência dos marcadores no primeiro quadro da avaliação foram

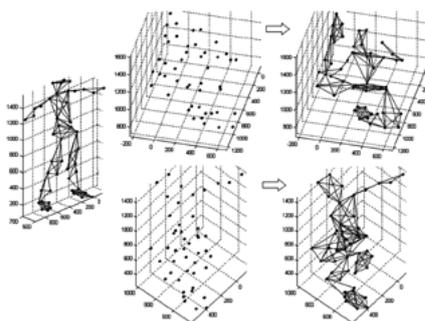


Figura 3.6: Estimação de pose pelo método proposto em (LI; MENG; HOLSTEIN, 2003) utilizando 47 pontos que compõe o modelo (esquerda), as poses não identificadas (ao centro) e as mesmas após identificação (direita) (LI; MENG; HOLSTEIN, 2003)

realizados manualmente, ou seja, depois que os marcadores foram detectados automaticamente, o usuário estabelece as correspondências das coordenadas em todas as câmeras.

A correspondência é realizada somente no primeiro quadro de avaliação, sendo esta tarefa obrigatória para a etapa seguinte informar um índice das coordenadas 2D e seus respectivos pontos 3D reconstruídos, referente aos marcadores detectados no primeiro quadro de análise. Nas próximas seções serão abordadas como a trajetória (*tracking*) e a reconstrução tridimensional dos marcadores são realizadas ao longo dos quadros.

3.4 Rastreamento

No primeiro quadro de avaliação, os marcadores são detectados automaticamente e o processo de *matching* estabelece a correspondência entre eles. A partir do segundo quadro a tarefa de *tracking* garante a trajetória do marcador de um quadro para o próximo, ao longo de toda a sequência do vídeo. A reconstrução tridimensional dos marcadores (3D) é realizada logo em seguida ao do rastreamento dos pontos no plano da imagem (2D), por causa das sequências 2D serem obrigadas a fornecer a continuidade das sequências 3D reconstruídas (HERDA et al., 2001).

É realizado o rastreamento dos marcadores a partir do segundo quadro, uma ligação quebrada na trajetória de *tracking* de um marcador pode implicar na perda de sua identidade, e intervenção subsequente do usuário que controla o sistema. Uma abordagem para determinar a trajetória de um marcador em quadro subsequentes é observar o deslocamento do marcador por meio de uma janela deslizante entre quadros (MAAS; GRUEN; PAPANTONIOU, 1993). A associação básica é que o deslocamento é mínimo de um quadro para o próximo, assim a ideia é prever e confirmar a posição de um marcador de um quadro para o próximo (HERDA et al., 2001) (MAAS; GRUEN; PAPANTONIOU, 1993).

Existem muitos casos onde os marcadores estão oclusos ou quando o algoritmo confunde a trajetória de um marcador com outro. Isto requer muito trabalho por parte dos usuários que controlam os sistemas de captura, antes do processo de animação e reconstrução dos movimentos. Para aumentar a robustez dos sistemas, existem abordagens que usam um modelo anatômico humano para evitar que os marcadores oclusos, por partes do corpo, não sejam confundidos com outros marcadores que estão próximos (figura 3.7) (HERDA et al., 2001) (SILAGHI et al., 1998) (GALL et al., 2009).

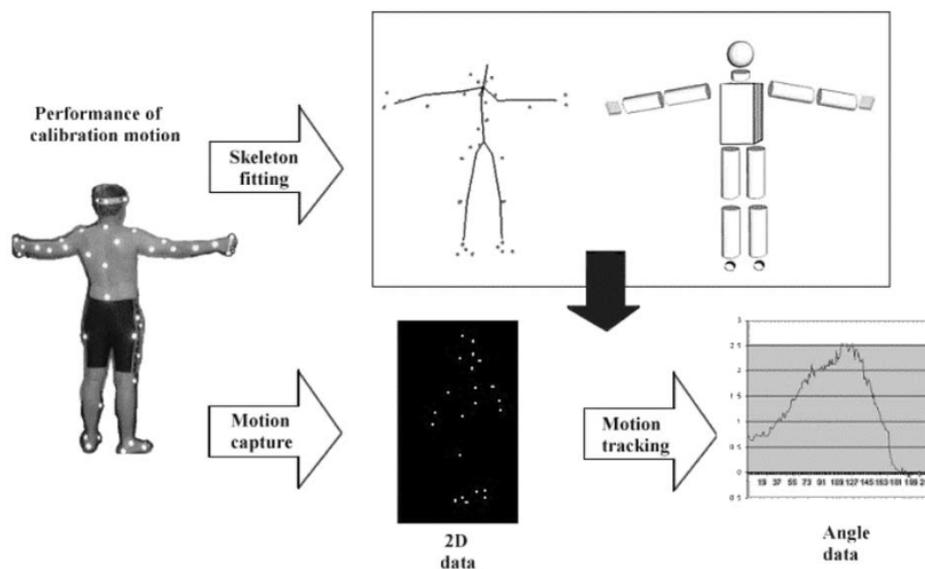


Figura 3.7: Abordagem usada no trabalho de (HERDA et al., 2001).

Essa abordagem inclui uma descrição razoavelmente precisa da mobilidade do esqueleto e as distâncias aproximadas dos pontos articulados do corpo. Esse modelo permite prever com precisão a localização 3D e a visibilidade de marcadores, aumentando significativamente a robustez do rastreamento do marcador e atribuição, reduzindo ou mesmo eliminando a necessidade de intervenção humana, durante o processo de reconstrução 3D.

Neste trabalho o rastreamento de marcadores na sequência dos quadros não foi realizado de forma automática.

Nas próximas seções serão apresentadas as técnicas necessárias para se realizar a calibração das câmeras e o método utilizado para calcular a reconstrução tridimensional dos marcadores.

3.5 Calibração e Reconstrução Tridimensional

A câmera de vídeo registra apenas uma projeção do ambiente tridimensional, em um plano bidimensional. Essa projeção é insuficiente para a descrição do movimento do espaço, sendo

impossível a reconstrução e trajetória de uma marcha humana em três dimensões. A calibração é um passo necessário em visão computacional, a fim de conhecer o posicionamento e orientação da câmera. Em um sistema de várias câmeras podemos estabelecer uma relação de profundidade, mas precisamos de um referência para determinar as coordenadas tridimensionais.

A trajetória ocorrida em três dimensões pode ser obtida por meio de duas etapas integradas. A primeira é a calibração das câmeras, e a segunda é a reconstrução tridimensional. Este método é conhecido na literatura por método DLT (*Direct Linear Transformation*) (ABDEL-AZIZ; Y.I.KARARA, 1971)(MILLER; SHAPIRO; MCLAUGHLIN, 1980) (WOOD; MARSHALL, 1986) (HATZE, 1988) (CHEN; ARMSTRONG; RAFTOPOULOS, 1994). A primeira etapa permite obter o posicionamento e a orientação de duas ou mais câmeras de vídeo em relação a um ou vários pontos de referências. Obtidas tais informações, a segunda etapa consiste em calcular os valores das coordenadas dos marcadores no plano de projeção de todas as câmeras, resultando nas coordenadas espaciais dos pontos de interesse (figura 3.8).

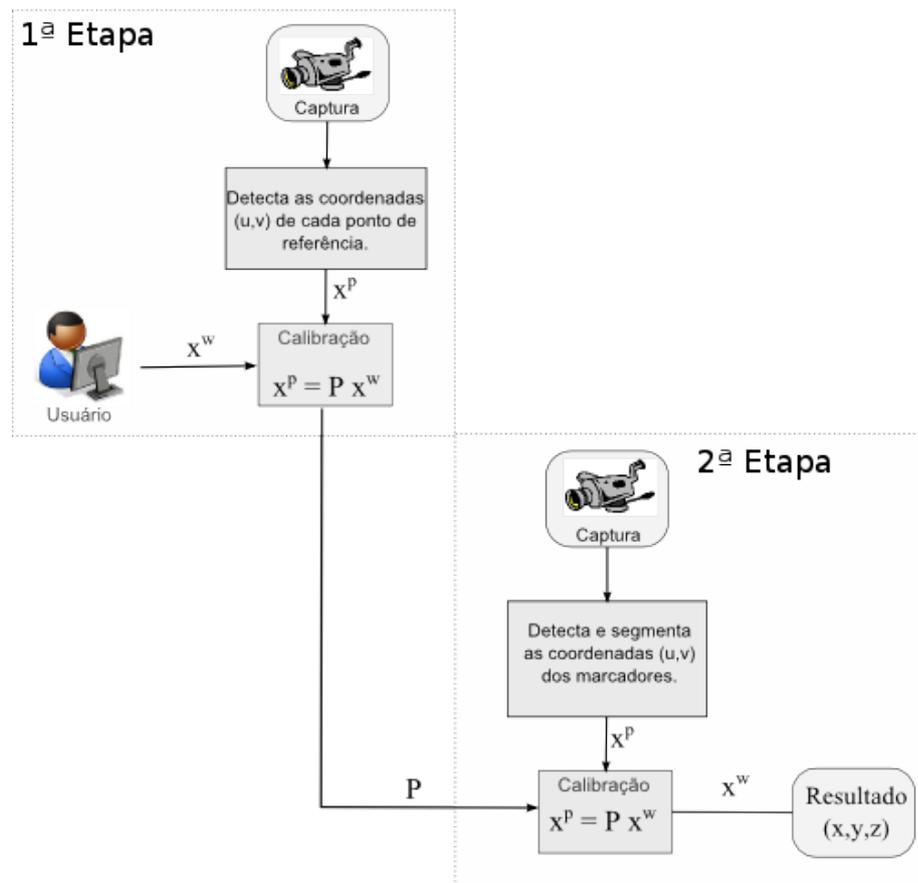


Figura 3.8: Fluxograma dos processos de calibração e reconstrução.

3.5.1 Conceitos Iniciais

Para obter as coordenadas dos marcadores, é necessário conhecer as informações do ambiente, bem como os métodos que determinam a relação entre os pontos no plano da imagem e os pontos espaciais. A figura 3.9 demonstra os atributos da imagem no plano π , tendo como referência o centro de projeção o^c e a linha perpendicular ao plano da imagem π que se encontra a uma distância f entre eles (DAESIK, 2006). O ponto detectado no espaço \mathbf{x}^w possui a sua localização no plano de imagem em \mathbf{x}^l .

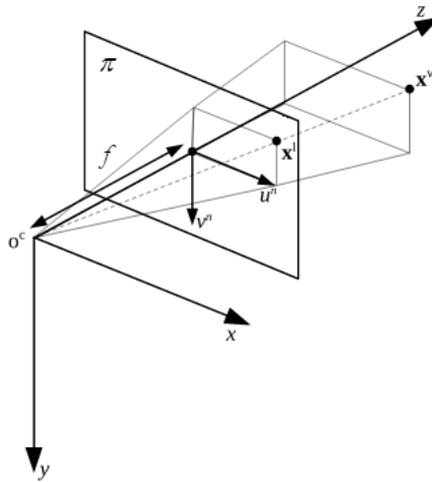


Figura 3.9: O ponto 3D \mathbf{x}^w é projetado para \mathbf{x}^l no plano da imagem por meio do centro óptico o^c .

O ponto $\mathbf{x}^l = (u^n, v^n)^T$ está posicionado na linha reta $\mathbf{x}^w = (x, y, z)^T$ com a origem do centro de projeção o^c , intersectando com o plano da imagem π . A distância de \mathbf{x}^l para \mathbf{x}^w é proporcional a de f para z , assim como o valor de v^n para y é proporcional a de f para z (DAESIK, 2006). Assim, a posição normalizada do ponto (u^n, v^n) no plano da imagem é dada pelas seguintes equações 3.4 e 3.5.

$$u^n = f \frac{x}{z} \quad (3.4)$$

$$v^n = f \frac{y}{z} \quad (3.5)$$

A projeção demonstrada acima, é realizada pela lente da câmera. Se pelo menos duas imagens registram um marcador, com visões das câmeras em diferentes ângulos e posições, sua

posição 3D pode ser calculada pelo método da triangulação (DAESIK, 2006). A relação das equações 3.4 e 3.5 pode ser representada pela equação 3.6, em termos de matrizes de coordenadas homogêneas.

$$\begin{bmatrix} u^n \\ v^n \\ 1 \end{bmatrix} \cong \begin{bmatrix} fx \\ fy \\ z \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \quad (3.6)$$

em que a notação \cong significa que ele é igual até que o fator de escala do ponto seja único. Se este fator f é desconhecido, não podemos inferir diretamente as posições e comprimentos dos objetos em cena a partir dos pontos correspondentes no plano da imagem, uma vez que a dimensão das coordenadas da imagem é menor do que a dimensão das coordenadas da câmera. Em outras palavras, as coordenadas do plano da imagem podem representar mais de um ponto na mesma linha de projeção, sendo desconhecida a aproximação entre eles.

A origem das coordenadas da geometria da imagem é o ponto principal da figura 3.10 (u^n, v^n). No entanto, no caso da captura de imagem por uma câmera digital, o ponto de origem da imagem localiza-se no canto esquerdo superior e as coordenadas são especificadas em termos de *pixel* (u^p, v^p). Portanto, a fim de representar o relacionamento entre a geometria da imagem e as coordenadas do quadro do vídeo em *pixel*, dois passos são necessários. O primeiro é mudar as coordenadas do quadro especificado em unidade métricas para as coordenadas do quadro especificado em *pixel*. O segundo passo é transladar a origem da coordenada do ponto principal para a origem do *pixel* no canto esquerdo superior.

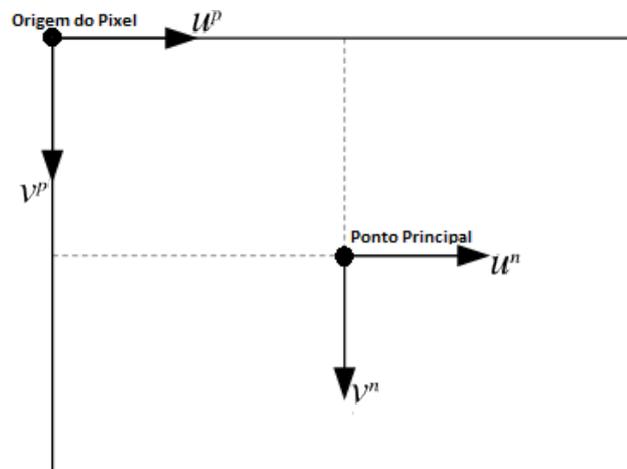


Figura 3.10: Coordenadas da origem do *pixel* (u^p, v^p) para o canto superior esquerdo

Assim as coordenadas dos quadros dos vídeos podem ser escritas da seguinte forma:

$$u^p = \frac{u^n}{s_u} + u_0 \quad (3.7)$$

e

$$v^p = \frac{v^n}{s_v} + v_0$$

onde o u^n e v^n são as coordenadas projetadas; s_u e s_v são os tamanhos dos *pixels* nas direções u e v ; u_0 e v_0 são as coordenadas do ponto principal no sistema de coordenadas em *pixel*. A equação 3.7 pode ser reescrita em termos de coordenadas homogêneas 3.8:

$$\begin{bmatrix} u^p \\ v^p \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{s_u} & \gamma & u_0 \\ 0 & \frac{1}{s_v} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u^n \\ v^n \\ 1 \end{bmatrix} \quad (3.8)$$

Em geral, o *pixel* representado no plano da imagem pode não ser retangular (DAESIK, 2006). Dessa forma, na matriz da equação 3.8 foi considerado o termo γ , que se refere a um fator de inclinação proporcional a $\cot(\theta)$, onde θ é o ângulo entre os eixos da imagem u^p e v^p . Tipicamente, o ângulo θ é muito perto de 90° , sendo então o valor de γ próximo a 0 (zero).

Na projeção realizada pelas lentes da câmera, na transformação entre as coordenadas do espaço $(x, y, z, 1)^T$ e as coordenadas projetadas $(x^c, y^c, z^c, 1)^T$, é realizada a transformação indicada na equação 3.9. Os valores $[r_{11} \dots r_{33}]$ representa a matriz de rotação (3x3) e de translação $[t_x, t_y, t_z]$ de (3x1);

$$\begin{bmatrix} x^c \\ y^c \\ z^c \\ 1 \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \quad (3.9)$$

No modelo compacto, \mathbf{x}^c representa as coordenadas dos pontos no plano da imagem, \mathbf{R} a matriz de rotação (3x3), \mathbf{t} refere-se a matriz de translação (3x1) e \mathbf{x}^w as coordenadas espaciais dos pontos, expressadas na forma compacta na equação 3.10:

$$\mathbf{x}^c = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ 0 & 1 \end{bmatrix} \mathbf{x}^w \quad (3.10)$$

O modelo completo da equação que descreve a relação entre o ponto 3D $\mathbf{x}^w = (x, y, z, 1)^T$ e

a sua projeção $\mathbf{x}^p = (u^p, v^p, 1)^T$ pode ser calculado da seguinte forma no modelo compacto.

$$\mathbf{x}^p \cong \mathbf{T}_n^p \mathbf{T}_c^n \mathbf{T}_w^c \mathbf{x}^w \quad (3.11)$$

$$\cong \mathbf{K} [\mathbf{R} \mid \mathbf{t}] \mathbf{x}^w \quad (3.12)$$

$$= \mathbf{P} \mathbf{x}^w \quad (3.13)$$

O ponto \mathbf{x}^p corresponde às coordenadas de $(u^p, v^p, 1)^T$ do *pixel* no canto esquerdo superior, \mathbf{T}_n^p representa a matriz que descreve a imagem normalizada p referente as coordenadas do *pixel* do *frame* n . A matriz \mathbf{T}_c^n descreve a transformação do *frame* n e do ponto c . A matriz \mathbf{T}_w^c refere-se aos parâmetros w de rotação e translação espacial do c ponto espacial. O ponto \mathbf{x}^w representa as coordenadas do espaço 3D do objeto. A equação 3.11 é reescrita pela matriz da equação 3.14.

$$\begin{bmatrix} u^p \\ v^p \\ 1 \end{bmatrix} \cong \begin{bmatrix} \frac{1}{s_u} & \gamma & u_o \\ 0 & \frac{1}{s_v} & v_o \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \quad (3.14)$$

No modelo compacto da equação 3.12, o \mathbf{K} representa a matriz dos parâmetros intrínsecos da câmera, ou seja, as características e configurações da lente focal da câmera, escalas, fatores de localização e informações da geometria da imagem. A $[\mathbf{R} \mid \mathbf{t}]$ é a matriz dos parâmetros extrínsecos da imagem, que corresponde às transformações entre as coordenadas do quadro e as coordenadas do mundo, onde \mathbf{R} descreve a matriz de rotação e \mathbf{t} é o vetor de translação. As equações 3.12 e 3.13 estão reescritas na forma de matrizes de coordenadas homogêneas pelas equações 3.15 e 3.16, respectivamente.

$$\begin{bmatrix} u^p \\ v^p \\ 1 \end{bmatrix} \cong \begin{bmatrix} \frac{f}{s_u} & \gamma & u_o \\ 0 & \frac{f}{s_v} & v_o \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \quad (3.15)$$

$$\begin{bmatrix} u^p \\ v^p \\ 1 \end{bmatrix} \cong \begin{bmatrix} n_{11} & n_{12} & n_{13} & n_{14} \\ n_{21} & n_{22} & n_{23} & n_{24} \\ n_{31} & n_{32} & n_{33} & n_{34} \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \quad (3.16)$$

Assim, podemos concluir que a perspectiva da matriz de projeção \mathbf{P} da equação 3.13 é composta pelos parâmetros extrínsecos e intrínsecos da câmera. O \mathbf{x}^w representa as coordenadas espaciais 3D do objeto.

3.5.2 Calibração

Conforme descrito na seção anterior, a transformação entre as coordenadas 3D do ponto $\mathbf{x}^w = (x, y, z)^T$ do espaço e as coordenadas 2D de sua projeção $\mathbf{x}^p = (u, v, 1)^T$ no plano de imagem, pode ser obtida por:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} n_{11} & n_{12} & n_{13} & n_{14} \\ n_{21} & n_{22} & n_{23} & n_{24} \\ n_{31} & n_{32} & n_{33} & n_{34} \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

Na forma compacta:

$$\mathbf{x}^p = \mathbf{P}\mathbf{x}^w \quad (3.17)$$

O processo de calibração consiste no registro de duas ou mais câmeras no ambiente onde ocorrerá a análise dos movimentos. Nesse local, as câmeras podem ficar posicionadas em qualquer lugar, de modo que o sistema consiga detectar os marcadores. Um dos modelos de calibração mais utilizados (HATZE, 1988) (WOOD; MARSHALL, 1986) é formado por pontos espaçados regularmente (semelhante a figura 3.11) no ambiente, que são os marcadores passivos detectados e associadas às coordenadas \mathbf{x}^p de cada plano de projeção, em duas dimensões. Após esse processo, as câmeras não podem ser movimentadas.

Para realizar esse processo é necessário utilizar doze parâmetros n_i de transformação. Em cada câmera, obtemos as coordenadas \mathbf{x}^w e \mathbf{x}^p para cada ponto de referência. Segundo (WOOD; MARSHALL, 1986), é necessário no mínimo seis pontos de referências para realizar o processo de calibração. No entanto, (CHEN; ARMSTRONG; RAFTOPOULOS, 1994) mostram que a precisão

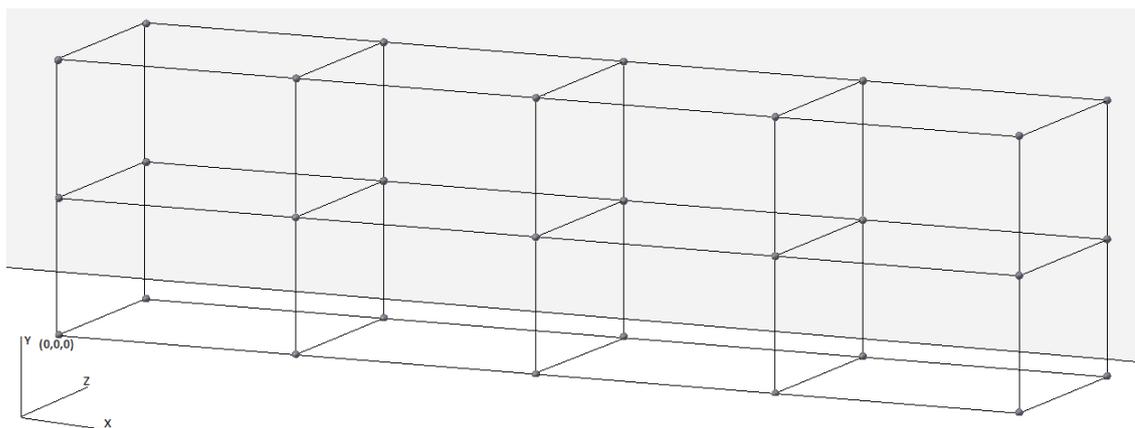


Figura 3.11: Sistema de referência para calibração do ambiente onde ocorrem as medições

da calibração está relacionada ao número e posicionamento dos pontos no sistema de referência, sendo recomendados no mínimo 16 pontos.

É necessário definir as coordenadas \mathbf{x}^w dos pontos do sistema de referência (figura 3.11). Por exemplo, o primeiro ponto possui as coordenadas $(0,0,0)$, o segundo ponto $(0,0,1)$. Em seguida, após a detecção dos pontos das coordenadas \mathbf{x}^p da câmera, são atribuídos todos os pontos de referência. Dessa forma, o sistema de calibração fornece o resultado de \mathbf{P} , que refere-se ao posicionamento tridimensional naquele ambiente (figura 3.8).

A multiplicação das matrizes $\mathbf{P}\mathbf{w}^x$ da equação 3.17, pode ser representada pela seguinte forma:

$$u = \frac{n_{11}x + n_{12}y + n_{13}z + n_{14}}{n_{31}x + n_{32}y + n_{33}z + n_{34}} \quad (3.18)$$

$$v = \frac{n_{21}x + n_{22}y + n_{23}z + n_{24}}{n_{31}x + n_{32}y + n_{33}z + n_{34}} \quad (3.19)$$

A partir das equações 3.18 e 3.19, obtemos as equações 3.20 e 3.21, respectivamente:

$$u(n_{31}x + n_{32}y + n_{33}z + n_{34}) = n_{11}x + n_{12}y + n_{13}z + n_{14} \quad (3.20)$$

$$v(n_{31}x + n_{32}y + n_{33}z + n_{34}) = n_{21}x + n_{22}y + n_{23}z + n_{24} \quad (3.21)$$

Realocando as variáveis das equações 3.20 e 3.21, obtemos as equações 3.22 e 3.23, respectivamente:

$$n_{11}x + n_{12}y + n_{13}z + n_{14} - uxn_{31} - uyn_{32} - uz n_{33} - un_{34} = 0 \quad (3.22)$$

$$n_{21}x + n_{22}y + n_{23}z + n_{24} - vxn_{31} - vyn_{32} - vz n_{33} - vn_{34} = 0 \quad (3.23)$$

Essas equações podem ser representadas na forma de matriz:

$$\begin{bmatrix} x & y & z & 1 & 0 & 0 & 0 & 0 & -ux & -uy & -uz & -u \\ 0 & 0 & 0 & 0 & x & y & z & 1 & -vx & -vy & -vz & -v \end{bmatrix} \begin{bmatrix} n_{11} \\ n_{12} \\ \cdot \\ \cdot \\ \cdot \\ n_{34} \end{bmatrix} = 0 \quad (3.24)$$

Após incluir todos os pontos de referência no processo de calibração, a equação 3.24 se transforma na equação 3.25. O conjunto de duas linhas, da primeira matriz da equação 3.25, representa um marcador detectado.

$$\begin{bmatrix} x_1 & y_1 & z_1 & 1 & 0 & 0 & 0 & 0 & -u_1x_1 & -u_1y_1 & -u_1z_1 & -u_1 \\ 0 & 0 & 0 & 0 & x_1 & y_1 & z_1 & 1 & -v_1x_1 & -v_1y_1 & -v_1z_1 & -v_1 \\ x_2 & y_2 & z_2 & 1 & 0 & 0 & 0 & 0 & -u_2x_2 & -u_2y_2 & -u_2z_2 & -u_2 \\ 0 & 0 & 0 & 0 & x_2 & y_2 & z_2 & 1 & -v_2x_2 & -v_2y_2 & -v_2z_2 & -v_2 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ x_n & y_n & z_n & 1 & 0 & 0 & 0 & 0 & -u_nx_n & -u_ny_n & -u_nz_n & -u_n \\ 0 & 0 & 0 & 0 & x_n & y_n & z_n & 1 & -v_nx_n & -v_ny_n & -v_nz_n & -v_n \end{bmatrix} \begin{bmatrix} n_{11} \\ n_{12} \\ n_{13} \\ n_{14} \\ n_{21} \\ n_{22} \\ n_{23} \\ n_{24} \\ n_{31} \\ n_{32} \\ n_{33} \\ n_{34} \end{bmatrix} = 0 \quad (3.25)$$

Na forma compacta.

$$\mathbf{Ap} = 0$$

Desta forma, é possível estimar as coordenadas dos marcadores por meio da decomposição

SVD (Decomposição de valores singulares). Essa técnica consiste na fatoração dos valores de A em três matrizes (U, D, V^T) , onde U é a matriz unitária quadrada $m \times m$, D é uma matriz $m \times n$ retangular e possui valores somente na sua diagonal, e V^* (conjugada transposta de V) é uma matriz unitária $n \times n$. A multiplicação das três matrizes (U, D, V^T) resulta no valor de A . Dessa forma, o vetor \mathbf{p} pode ser recuperado pela última coluna de V , correspondendo ao menor valor singular de $A = UDV^T$ que são os elementos da matriz \mathbf{p} .

3.5.3 Reconstrução

Após o processo de calibração, a posição das câmeras não pode mais ser alterada e no lugar dos marcadores de referência, o movimento pode ser realizado. O ambiente de registro está calibrado, e pronto para reconstruir os movimentos no espaço tridimensional.

A etapa de reconstrução utiliza os valores da matriz de projeção \mathbf{P} , calculados no processo de calibração. As coordenadas \mathbf{x}^p de todos os marcadores detectados no ambiente de projeção são adicionados na mesma equação $\mathbf{x}^p = \mathbf{P}\mathbf{x}^w$. O resultado é o ponto \mathbf{x}^w que corresponde as coordenadas espaciais dos marcadores.

Para tornar o procedimento de reconstrução possível, cada marcador deve ser detectado por no mínimo duas câmeras, tendo em vista que é impossível estimar as coordenadas espaciais por uma única câmera. Dessa forma, o sistema precisa dispor de no mínimo quatro equações, citadas aqui como 3.26, 3.27, 3.28 e 3.29 (DAESIK, 2006), duas para cada câmera, referente ao marcador detectado em (u_α, v_α) e (u_β, v_β) , respectivamente câmeras α e β .

$$u_\alpha = \frac{n_{11}^\alpha x + n_{12}^\alpha y + n_{13}^\alpha z + n_{14}^\alpha}{n_{31}^\alpha x + n_{32}^\alpha y + n_{33}^\alpha z + n_{34}^\alpha} \quad (3.26)$$

$$v_\alpha = \frac{n_{21}^\alpha x + n_{22}^\alpha y + n_{23}^\alpha z + n_{24}^\alpha}{n_{31}^\alpha x + n_{32}^\alpha y + n_{33}^\alpha z + n_{34}^\alpha} \quad (3.27)$$

$$u_\beta = \frac{n_{11}^\beta x + n_{12}^\beta y + n_{13}^\beta z + n_{14}^\beta}{n_{31}^\beta x + n_{32}^\beta y + n_{33}^\beta z + n_{34}^\beta} \quad (3.28)$$

$$v_\beta = \frac{n_{21}^\beta x + n_{22}^\beta y + n_{23}^\beta z + n_{24}^\beta}{n_{31}^\beta x + n_{32}^\beta y + n_{33}^\beta z + n_{34}^\beta} \quad (3.29)$$

Realizando o cálculo das equações 3.26, 3.27, 3.28 e 3.29, realocamos as variáveis da seguinte forma:

$$(n_{11}^{\alpha} - u_{\alpha}n_{31}^{\alpha})x + (n_{12}^{\alpha} - u_{\alpha}n_{32}^{\alpha})y + (n_{13}^{\alpha} - u_{\alpha}n_{33}^{\alpha})z + (n_{14}^{\alpha} - u_{\alpha}n_{34}^{\alpha}) = 0 \quad (3.30)$$

$$(n_{21}^{\alpha} - v_{\alpha}n_{31}^{\alpha})x + (n_{22}^{\alpha} - v_{\alpha}n_{32}^{\alpha})y + (n_{23}^{\alpha} - v_{\alpha}n_{33}^{\alpha})z + (n_{24}^{\alpha} - v_{\alpha}n_{34}^{\alpha}) = 0 \quad (3.31)$$

$$(n_{11}^{\beta} - u_{\beta}n_{31}^{\beta})x + (n_{12}^{\beta} - u_{\beta}n_{32}^{\beta})y + (n_{13}^{\beta} - u_{\beta}n_{33}^{\beta})z + (n_{14}^{\beta} - u_{\beta}n_{34}^{\beta}) = 0 \quad (3.32)$$

$$(n_{21}^{\beta} - v_{\beta}n_{31}^{\beta})x + (n_{22}^{\beta} - v_{\beta}n_{32}^{\beta})y + (n_{23}^{\beta} - v_{\beta}n_{33}^{\beta})z + (n_{24}^{\beta} - v_{\beta}n_{34}^{\beta}) = 0 \quad (3.33)$$

As equações 3.30, 3.31, 3.32 e 3.33 podem ser expressas na forma matricial, em termos de coordenadas homogêneas:

$$\begin{bmatrix} n_{11}^{\alpha} - u_{\alpha}n_{31}^{\alpha} & n_{12}^{\alpha} - u_{\alpha}n_{32}^{\alpha} & n_{13}^{\alpha} - u_{\alpha}n_{33}^{\alpha} & n_{14}^{\alpha} - u_{\alpha}n_{34}^{\alpha} \\ n_{21}^{\alpha} - v_{\alpha}n_{31}^{\alpha} & n_{22}^{\alpha} - v_{\alpha}n_{32}^{\alpha} & n_{23}^{\alpha} - v_{\alpha}n_{33}^{\alpha} & n_{24}^{\alpha} - v_{\alpha}n_{34}^{\alpha} \\ n_{11}^{\beta} - u_{\beta}n_{31}^{\beta} & n_{12}^{\beta} - u_{\beta}n_{32}^{\beta} & n_{13}^{\beta} - u_{\beta}n_{33}^{\beta} & n_{14}^{\beta} - u_{\beta}n_{34}^{\beta} \\ n_{21}^{\beta} - v_{\beta}n_{31}^{\beta} & n_{22}^{\beta} - v_{\beta}n_{32}^{\beta} & n_{23}^{\beta} - v_{\beta}n_{33}^{\beta} & n_{24}^{\beta} - v_{\beta}n_{34}^{\beta} \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = 0$$

Na forma compacta

$$\mathbf{M}\mathbf{x}^w = 0$$

Assim, o ponto pode ser estimado por meio da decomposição SVD. Os quatro últimos elementos da coluna de \mathbf{V} , da fatoração da matriz de $\mathbf{M} = \mathbf{U}\mathbf{D}\mathbf{V}^T$ são considerados as coordenadas homogêneas de \mathbf{x}^w .

3.6 Considerações Finais

Neste capítulo foram abordadas as etapas para os sistemas de captura de movimentos por dispositivos ópticos. Na etapa de detecção dos marcadores foram abordadas as técnicas de *thresholding* e a transformada de *Hough*. Na etapa de *matching* os métodos utilizados na literatura e os grandes desafios foram discutidos para estabelecer a correspondência dos marcadores. No processo de *tracking* foi apresentado como a trajetória dos marcadores ao longo da sequência dos quadros são realizados. Por último, as etapas integradas de calibração e reconstrução foram

abordados, com objetivos de obter o posicionamento das câmeras e a reconstrução tridimensional dos marcadores.

Nesse contexto, o próximo capítulo demonstra os experimentos computacionais que envolvem as etapas de captura de movimentos por dispositivos ópticos, bem como apresentam os resultados obtidos no presente trabalho.

Capítulo 4

RESULTADOS

Neste capítulo serão apresentados os experimentos computacionais realizados. Foi utilizado o material fornecido pelo grupo de pesquisa em Biodinâmica do DEFMH-UFSCar (Departamento de Educação Física e Motricidade Humana).

Atualmente, o sistema DVideoW (CARVALHO; TUDELLA; BARROS, 2005) é utilizado pelo grupo de Biomecânica do DEFMH-UFSCar para captura de movimentos, permitindo assim sua análise (CASTRO, 2012). Para tanto, o departamento utiliza quatro câmeras para registrar os movimentos em um ambiente com pouca iluminação e marcadores passivos colocados no corpo da criança. O sistema consiste em detectar os marcadores e calcular a reconstrução tridimensional dos movimentos realizados no período da marcha. Posteriormente, a análise cinemática é realizada pelos especialistas.

Para capturar os movimentos no sistema DVideoW é necessário especificar qual tipo de objeto vai ser considerado como um marcador. Dessa forma, no primeiro quadro da avaliação o usuário escolhe manualmente os marcadores presentes na imagem, selecionando os pontos que são considerados como marcadores. Caso o usuário tenha dificuldades na visualização dos objetos, o operador do sistema pode aumentar a imagem (*zoom*) para auxiliar a seleção da região da imagem que está localizado o marcador (figura 4.1). A etapa anterior informa ao sistema a identificação de cada marcador utilizado e as coordenadas 2D no plano da imagem dos pontos selecionados.

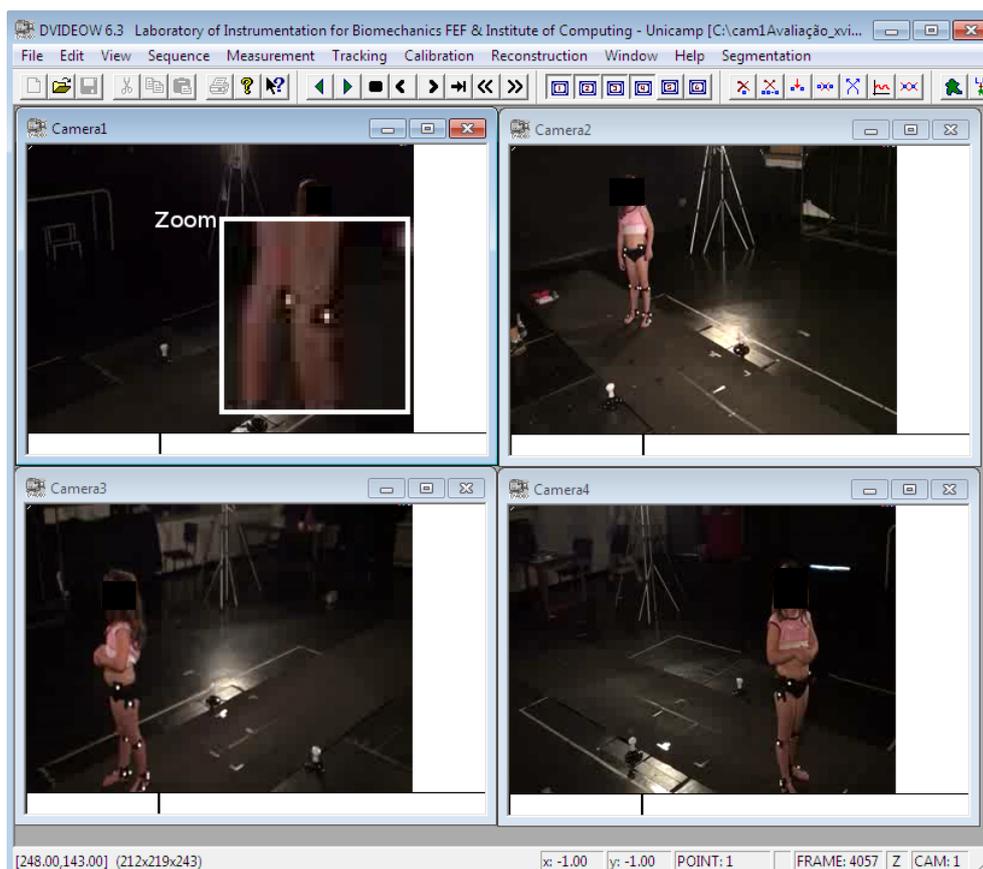


Figura 4.1: Ilustração (tela) do sistema DVideoW

A correspondência está definida na ordem em que o usuário seleciona os marcadores, por exemplo, se na câmera 1 (um) o marcador colocado no joelho esquerdo da criança foi o primeiro selecionado pelo o usuário, na câmera 2 (dois) o mesmo marcador também tem que ser o primeiro a ser selecionado. Esse processo de correspondência “manual” acontece sucessivamente para todos os marcadores em todas as câmeras no primeiro quadro da avaliação. A partir do segundo quadro da avaliação a detecção é realizada de forma automática pelo sistema, uma vez que o sistema dispõe das características de cada marcador.

Considerando a abordagem utilizada pelo DVideoW no contexto desta aplicação, o presente trabalho realiza a detecção automática dos marcadores no primeiro quadro da avaliação cinemática, bem como realiza a detecção dos marcadores na sequência dos quadros.

Este trabalho foi implementado utilizando o MATLAB¹, tanto na etapa de detecção dos marcadores, quanto nas etapas de calibração e reconstrução. As demais etapas, sincronização e correspondência, foram executadas manualmente.

Nas próximas seções serão apresentados os resultados obtidos no presente trabalho.

¹<http://www.mathworks.com/>

4.1 Detecção dos Marcadores

Com o objetivo de rastrear os marcadores passivos em destaque no vídeo, alguns experimentos foram realizados para detectar os marcadores da avaliação cinemática do trabalho de (CASTRO, 2012). Após a etapa de sincronização, o primeiro quadro de registro da avaliação de todas as câmeras foram extraídos, resultando em quatro imagens (figura 4.2).

Para calcular a transformada de *Hough* foram consideradas as dimensões dos marcadores passivos nas imagens digitais, isto é, os raios em *pixels* dos marcadores registrados nos quadros de vídeos. Com a avaliação experimental, o intervalo dos raios procurados pela transformada foi fixado em 5 a 20 *pixels*. Dessa forma, a transformada de *Hough* foi aplicada nos quadros de vídeos utilizando três parâmetros: imagem, raio mínimo e máximo. O resultado são as coordenadas 2D dos objetos detectados e o raio estimado, correspondente a cada circunferência detectada (figura 4.2).

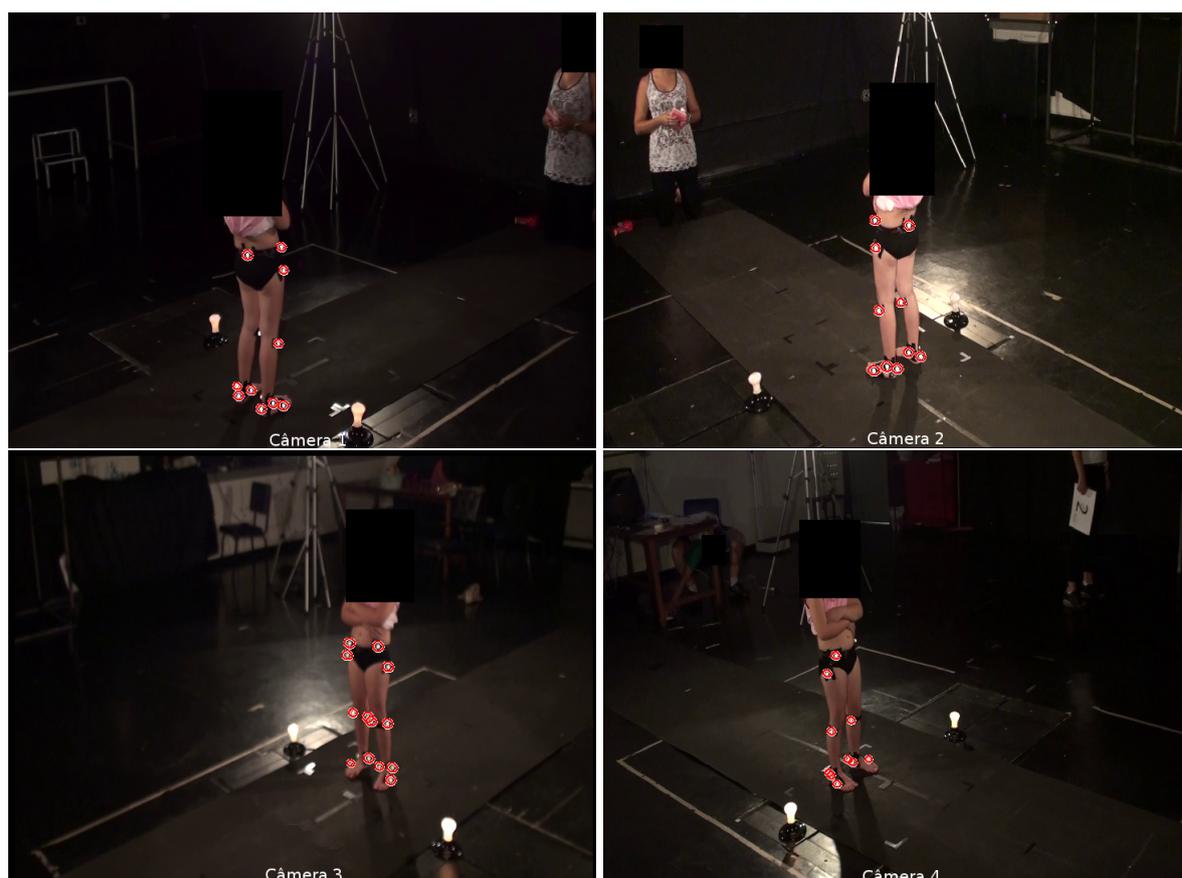


Figura 4.2: Detecção dos marcadores representado por um círculo vermelho do primeiro quadro da avaliação de cada câmera

A transformada de *Hough*, aplica as equações (KIMME; BALLARD; SKLANSKY, 1975) descrita na subseção 3.2.1 para cada *pixel* (x, y) mapeados na imagem, isto é, os pontos destacados

por algum detector de borda. Portanto, se a imagem for colorida ou em níveis de cinza, uma fase de pré processamento é necessário para realizar a detecção das bordas.

Considerando que este trabalho possui imagens coloridas, com muitos objetos presentes no ambiente e que podem ser detectados como falsos marcadores (*falsos positivos*), antes de aplicar a transformada de *Hough*, foi realizada a binarização da imagem de maneira que favoreçam na detecção de bordas dos marcadores, por meio das cores dos canais RGB que caracterizam os marcadores. O método utilizado para binarização da imagem está descrito na próxima seção.

4.1.1 Binarização da Imagem

Para destacar os marcadores nesse trabalho, pesquisamos a viabilidade de binarizar a imagem usando as cores RGB. Para tanto, foram conhecidos os limiares dos canais RGB que representam a cor dos marcadores (tabela 4.1), na escala de 0 a 255. Os pontos (*pixels*) da imagem que estão entre os limiares dos canais RGB, são fixados para a cor branca e os que estão fora para a cor preta (figura 4.3). O resultado é uma imagem binária que vai ser utilizada para calcular a transformada de *Hough* no processo de detecção dos marcadores. Esse tipo de binarização da imagem prioriza as bordas dos marcadores e diminui drasticamente a detecção de bordas que não pertence aos objetos de interesse.

| | R | G | B |
|-----|-----|-----|-----|
| Min | 120 | 115 | 110 |
| Max | 254 | 252 | 252 |

Tabela 4.1: Limiares dos canais RGB da cor dos marcadores



Figura 4.3: Binarização da imagem pelos canais RGB. (A) Quadro do Vídeo. (B) Zoom da imagem

No contexto deste trabalho, se aplicado a conversão da imagem colorida para níveis de cinza e em seguida aplicá-las por algum detector de borda (figura 4.4), o processo realça as bordas de todos objetos presentes na imagem, isto é, bordas como por exemplo, o dedo do pé, lâmpadas e reflexos em objetos vindo da iluminação das câmeras. Se todos esses pontos detectados forem processados pela transformada de *Hough*, ocorre na detecção de falsos marcadores (figura 4.5).

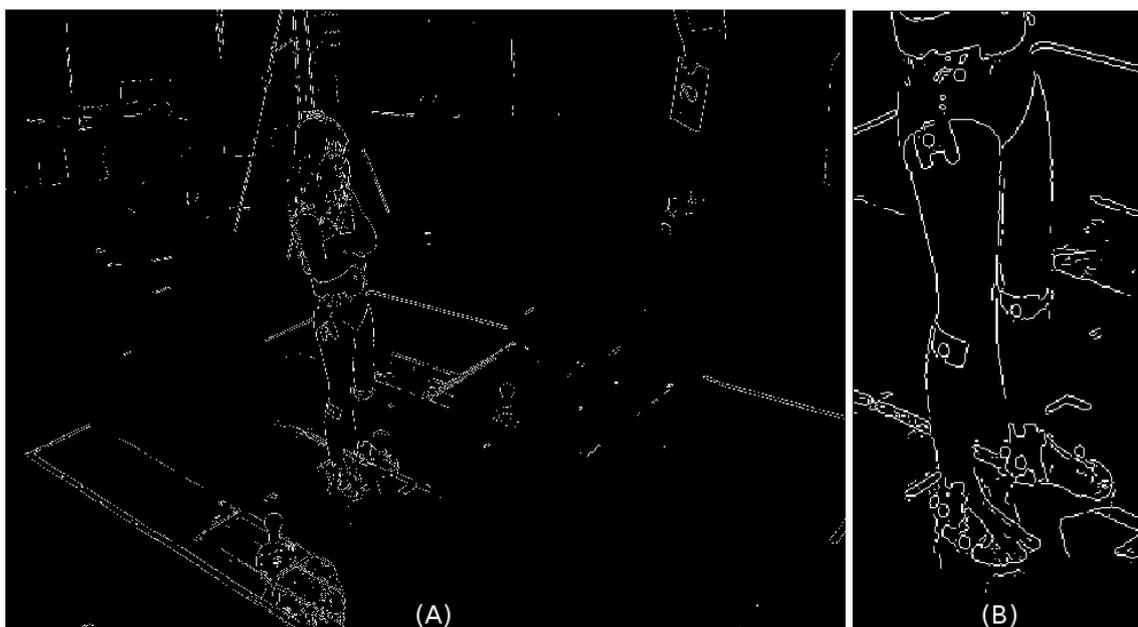


Figura 4.4: Filtro de Sobel. (A) Quadro do vídeo. (B) Zoom da Imagem

Os resultados demonstrados pelas figuras 4.4 e 4.3, apresentam as diferentes abordagens

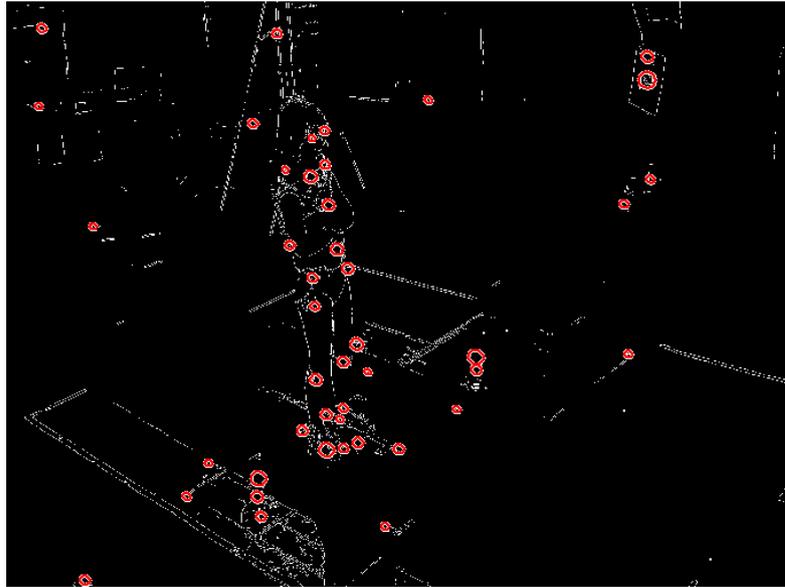


Figura 4.5: Círculos vermelhos representando o resultado da aplicação da transformada de *Hough*, utilizando o filtro de Sobel como fase de pré processamento

que podem ser utilizadas pelo cálculo da transformada de *Hough*. Observa-se que a binarização da imagem por meio das cores dos canais RGB prioriza a detecção das bordas características dos marcadores passivos, alcançando assim melhores resultados na detecção dos marcadores. Na próxima seção será avaliado quantitativamente o desempenho da detecção .

4.1.2 Avaliação da Detecção

Os resultados obtidos pela detecção dos marcadores foram avaliados obtendo-se os verdadeiros positivos (*VP*), falsos negativos (*FN*) e falsos positivos (*FP*). Como medida de verdadeiro positivo *VP*, foi considerada a proporção entre a quantidade *a* de todos os marcadores presentes na imagem pela quantidade *b* de marcadores detectados corretamente (equação 4.1). Os falsos negativos *FN* referem-se a proporção entre os marcadores que não foram detectados *c* pela quantidade de marcadores presentes na imagem *a* (equação 4.2). Como falsos positivos *FP* foram considerados a proporção entre os marcadores erroneamente detectados *d* pela quantidade total de pontos detectados *e* pelo sistema no quadro da imagem (equação 4.3).

$$VP = \frac{b * 100}{a} \quad (4.1)$$

$$FN = \frac{c * 100}{a} \quad (4.2)$$

$$FP = \frac{d * 100}{e} \quad (4.3)$$

Para avaliar o desempenho da detecção dos marcadores, quadros de momentos distintos de duas análises cinemáticas e do processo de calibração das câmeras foram retirados dos vídeos. As próximas seções consideram três avaliações caracterizadas como resultados da: calibração, criança “A” e criança “B”.

Resultados da calibração

Um quadro de cada vídeo que registra o sistema de referência foi extraídos com o objetivo de detectar os marcadores para a etapa de calibração das câmeras (figura 4.6). O sistema de referência possui sessenta e três marcadores, distribuídos uniformemente no ambiente onde será capturado os movimentos das crianças. A quantidade de pontos detectados e a avaliação dos resultados da detecção dos marcadores são expressos nas tabelas 4.2 e 4.3, respectivamente.

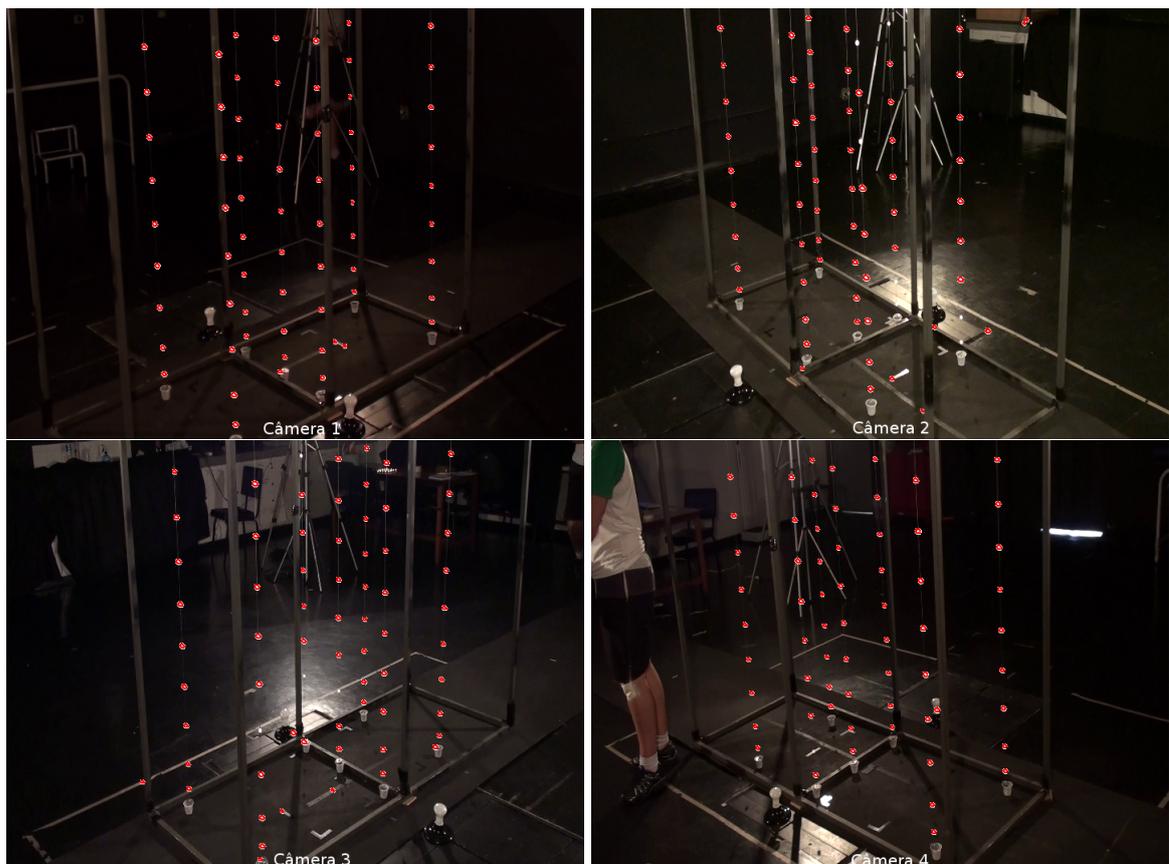


Figura 4.6: Círculos vermelhos representando a detecção dos marcadores do sistema de referência

| Câmera 1 | Câmera 2 | Câmera 3 | Câmera4 |
|----------|----------|----------|---------|
| 66 | 63 | 64 | 63 |

Tabela 4.2: Quantidade de circunferências detectadas pelo sistema.

| % | Câmera 1 | Câmera 2 | Câmera 3 | Câmera4 |
|----|----------|----------|----------|---------|
| VP | 100% | 88,89% | 90,47% | 98,41% |
| FN | 0% | 11,11% | 9,53% | 1,59% |
| FP | 4,54% | 11,12% | 10,76% | 1,56% |

Tabela 4.3: Comparação dos resultados verdadeiros positivos, falsos negativos e falsos positivos

Os resultados apresentados pela tabela 4.3 referem-se a proporção dos verdadeiros positivos, falsos negativos e falsos positivos. As câmeras dois e três tiveram resultados inferiores aos das câmeras um e quatro. Esse desempenho ruim se deve ao reflexo da luz no chão, nas imagens das câmeras dois e três. Pode ser observado que os marcadores não detectados nas câmeras dois e três pertencem a região da imagem que possuem o reflexo da luz, vindo da iluminação das câmeras.

Resultados da criança “A”

Um quadro de cada câmera em dois momentos distintos da análise cinemática foram retirados dos vídeos e aplicados ao processo de detecção dos marcadores (figuras 4.7 e 4.8). A análise possui dezessete marcadores colocados no corpo da criança “A”. No entanto, cada câmera registra uma quantidade diferente de marcadores (tabela 4.4), tendo em vista que alguns são ocluídos por membros do corpo.

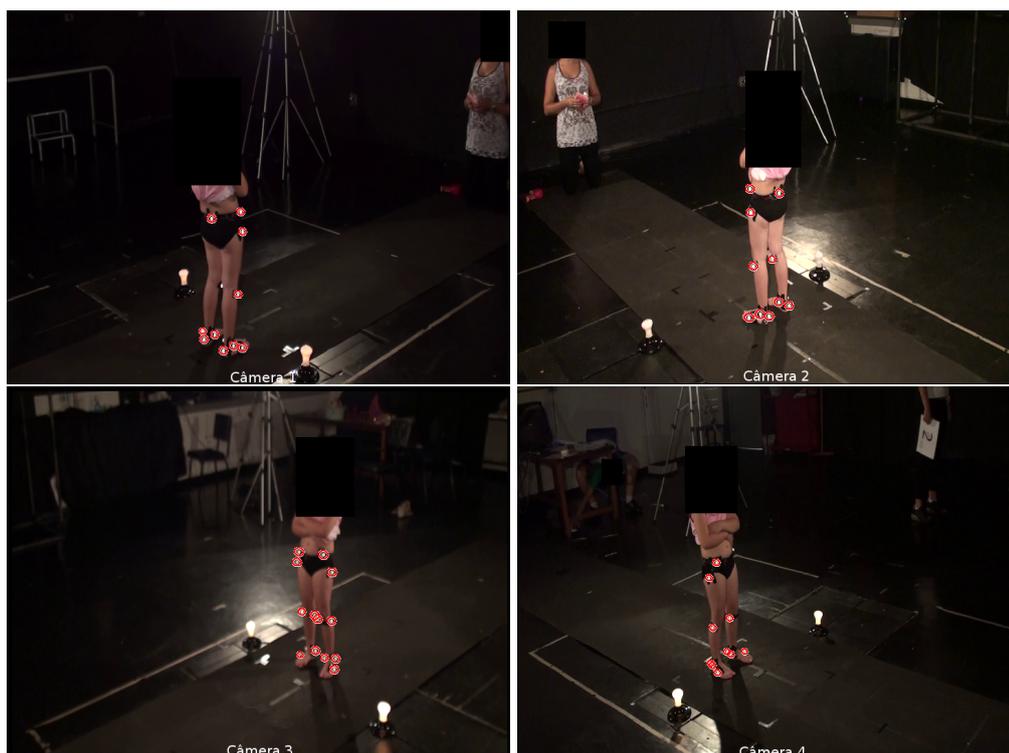


Figura 4.7: Círculos vermelhos representando a detecção dos marcadores no primeiro quadro de avaliação (Criança “A”)

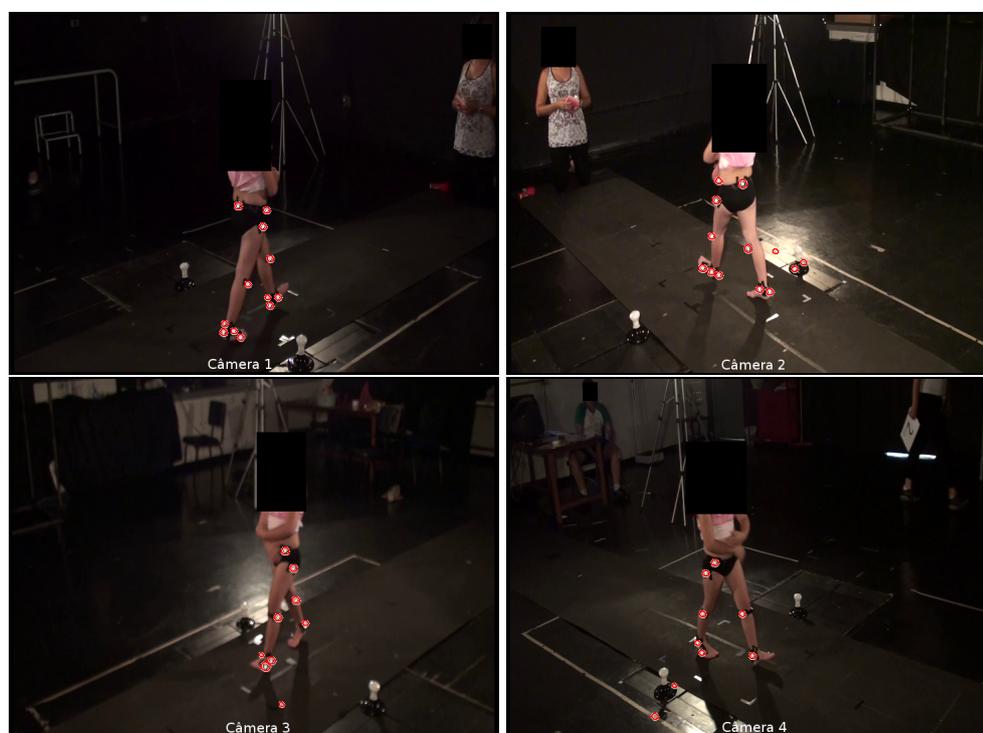


Figura 4.8: Círculos vermelhos representando a detecção dos marcadores no período da marcha (Criança “A”)

Resultados da criança “B”

Quadros de dois momentos distintos de outra análise cinemática também foram retirados dos vídeos e aplicados ao processo de detecção dos marcadores (figuras 4.9 e 4.10). A análise possui dezessete marcadores colocados no corpo da criança “B”. No entanto, cada câmera registra uma quantidade diferente de marcadores (tabela 4.4), uma vez que alguns são ocluídos por membros do corpo.

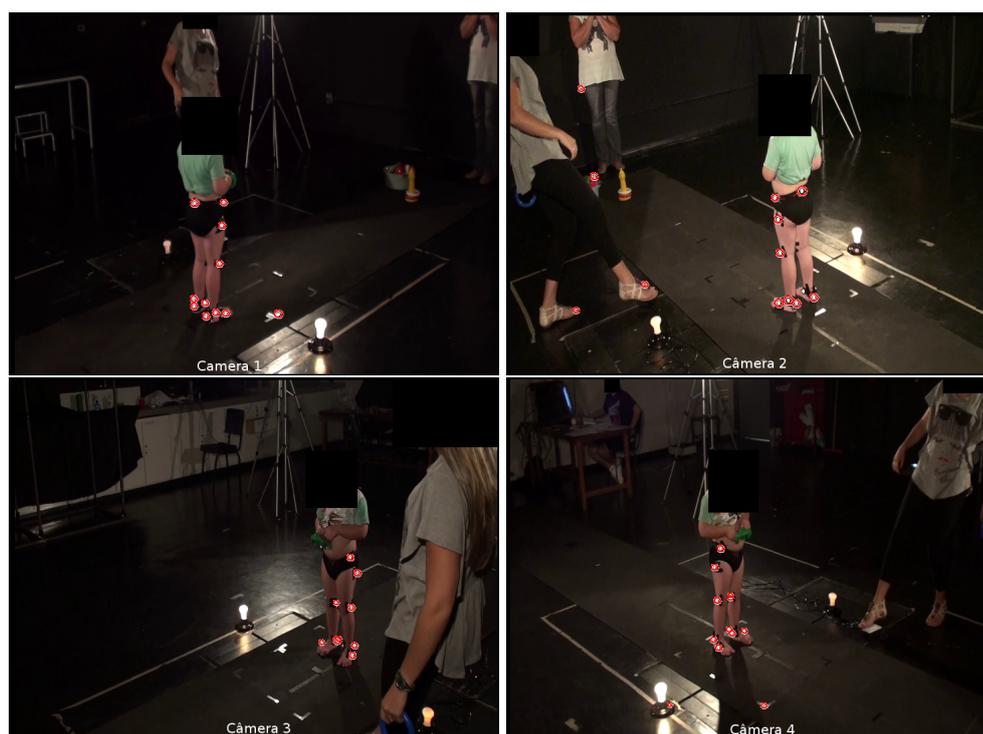


Figura 4.9: Círculos vermelhos representando a detecção dos marcadores no primeiro quadro de avaliação (Criança “B”)

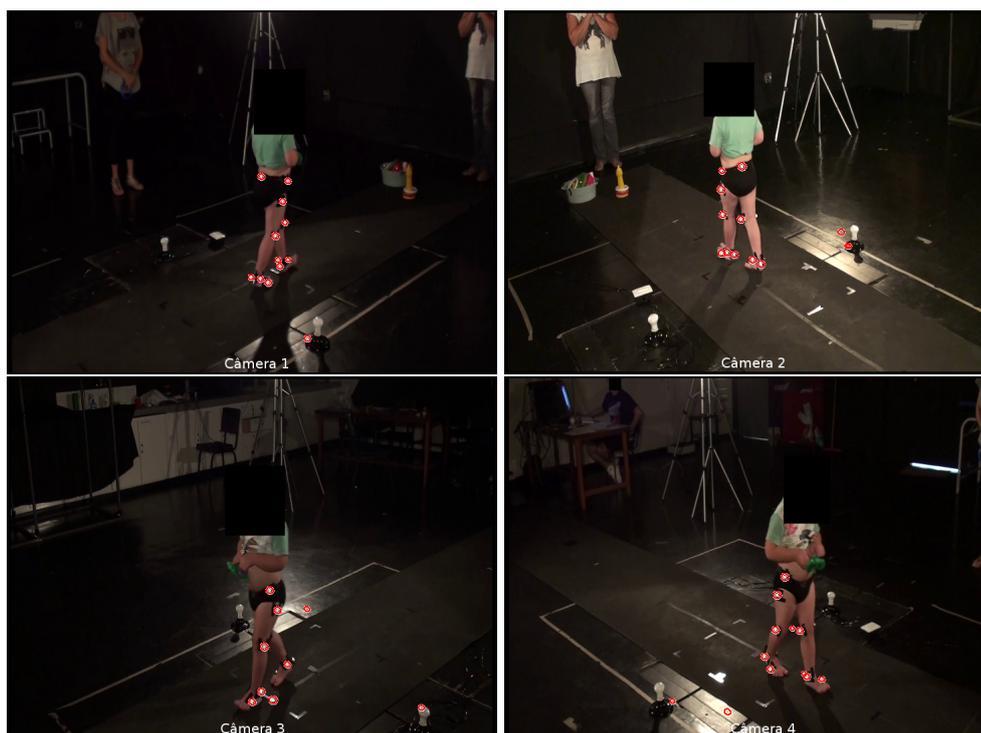


Figura 4.10: Círculos vermelhos representando a detecção dos marcadores no período da marcha (Criança “B”)

Avaliação dos Resultados das Crianças “A” e “B”

Os resultados demonstrados na tabela 4.4 referem-se a quantidade de marcadores registrados pelos quadros durante o período da avaliação cinemática. Pode ser observado que os valores são diferentes, uma vez que a posição das câmeras e os movimentos realizados podem ocorrer na oclusão dos marcadores.

| Quantidade | Câmera 1 | Câmera 2 | Câmera 3 | Câmera 4 |
|-------------|----------|----------|----------|----------|
| Figura 4.7 | 12 | 10 | 14 | 11 |
| Figura 4.8 | 12 | 10 | 10 | 9 |
| Figura 4.9 | 10 | 10 | 12 | 10 |
| Figura 4.10 | 11 | 10 | 6 | 10 |

Tabela 4.4: Quantidade de marcadores registrados pelas câmeras.

| Verdadeiros Positivos | Câmera 1 | Câmera 2 | Câmera 3 | Câmera4 |
|-----------------------|----------|----------|----------|---------|
| Figura 4.7 | 83,34% | 100% | 92,85% | 90,90% |
| Figura 4.8 | 100% | 100% | 80% | 77,78% |
| Figura 4.9 | 100% | 88,88% | 80% | 75% |
| Figura 4.10 | 100% | 90,90% | 75% | 90,90% |

Tabela 4.5: Resultados dos marcadores detectados corretamente pelo sistema.

Os resultados apresentados na tabela 4.5 demonstram as porcentagens dos verdadeiros positivos, isto é, marcadores registrados pelas câmeras e que foram detectados corretamente pelo sistema. Nota-se que a grande maioria dos marcadores registrados pelas câmeras foram detectados corretamente.

| Falsos Negativos | Câmera 1 | Câmera 2 | Câmera 3 | Câmera4 |
|------------------|----------|----------|----------|---------|
| Figura 4.7 | 16,66% | 0% | 7,15% | 9,10% |
| Figura 4.8 | 0% | 0% | 20% | 22,22% |
| Figura 4.9 | 0% | 11,12% | 20% | 25% |
| Figura 4.10 | 0% | 9,10% | 25% | 9,10% |

Tabela 4.6: Resultados dos marcadores não detectados pelo sistema

Os resultados apresentados na tabela 4.6 demonstram as porcentagens dos falsos negativos, ou seja, os marcadores que foram registrados pelas câmeras e não foram detectados pelo sistema. A maioria destes marcadores estavam parcialmente ocluídos por membros do corpo.

| Falsos Positivos | Câmera 1 | Câmera 2 | Câmera 3 | Câmera4 |
|------------------|----------|----------|----------|---------|
| Figura 4.7 | 0% | 0% | 0% | 0% |
| Figura 4.8 | 0% | 23,08% | 11,12% | 22,23% |
| Figura 4.9 | 9,09% | 33,34% | 10% | 16,67% |
| Figura 4.10 | 8,34% | 16,67% | 25% | 16,67% |

Tabela 4.7: Resultados da detecção que não são marcadores.

Os resultados apresentados na tabela 4.7 demonstram as porcentagens dos falsos positivos, isto é, os pontos que não são marcadores e que foram detectados pelo sistema. Esse desempenho ruim se deve à detecção de objetos que refletem a luz vinda das câmeras.

4.1.3 Espaço de *Hough*

Para detectar os marcadores neste trabalho foi utilizado as equações trigonométricas 3.2 e 3.3 descritas na subseção 3.2.1. Para cada *pixel* de borda na imagem binarizada (figura 4.3) o método projeta um raio com o tamanho R especificado. Dessa forma, a transformada representa no “espaço dos parâmetros” (figura 4.11) as projeções dos raios.

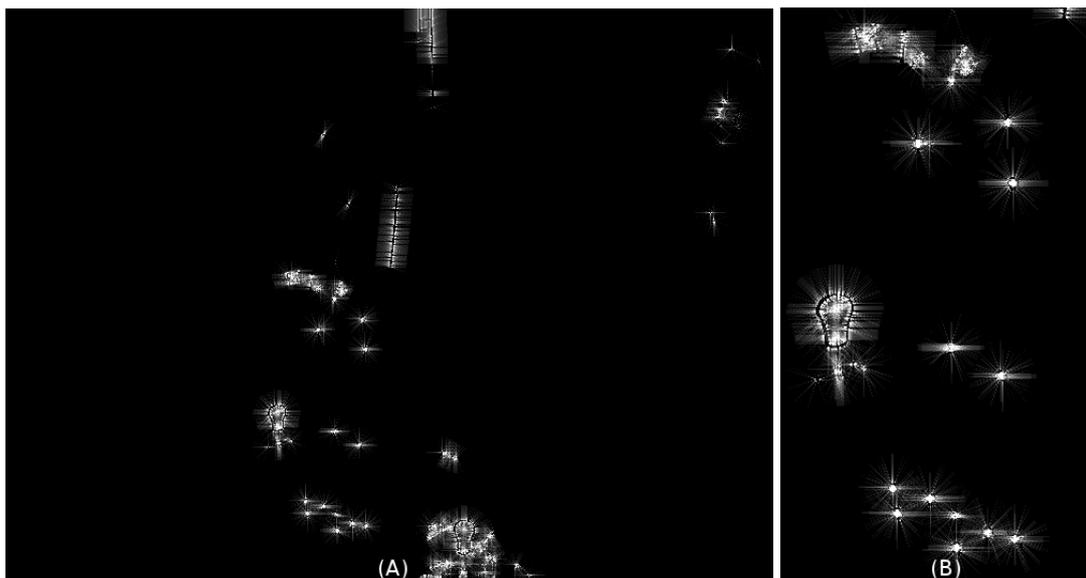


Figura 4.11: Espaço de *Hough*. (A) Quadro do vídeo da figura 4.3. (B) Zoom da Imagem.

O intervalo dos raios procurados pela transformada foram de 5 a 20 *pixels*. Pode ser observado que o centro dos marcadores possuem muitas intersecções dos raios, sendo considerado o ponto central de uma circunferência para um determinado raio.

O método descrito acima funciona como deve se esperar, um problema decorrente são para os marcadores com pequenos raios (menor que 10 *pixels*) para uma imagem de alta resolução. Esse problema refere-se a falta de discriminação nas formas de objetos pequenos, ou seja, para um pequeno raio poucas intersecções podem considerar um circunferência pequena, resultando na detecção de um falso positivo.

Na próxima seção serão apresentados os resultados dos processos de calibração e reconstrução tridimensional dos marcadores detectados.

4.2 Calibração e Reconstrução

Para obter as informações sobre o posicionamento das câmeras e dimensões do ambiente onde ocorrerá a gravação, é necessário realizar o processo de calibração, utilizando o sistema

de referência (figura 4.6). Realizado o processo de calibração, onde a matriz \mathbf{P} da equação $\mathbf{x}^p = \mathbf{P}\mathbf{x}^w$ foi obtida, a próxima etapa é calcular as coordenadas (x, y, z) dos marcadores colocados no corpo da criança.

Os resultados das reconstruções tridimensionais dos quadros dos vídeos exibidos nas figuras 4.7, 4.8, 4.9 e 4.10, utilizados na etapa de detecção dos marcadores (seção 4.1), são demonstrados pelos gráficos nas figuras 4.12, 4.13, 4.14 e 4.15, respectivamente.

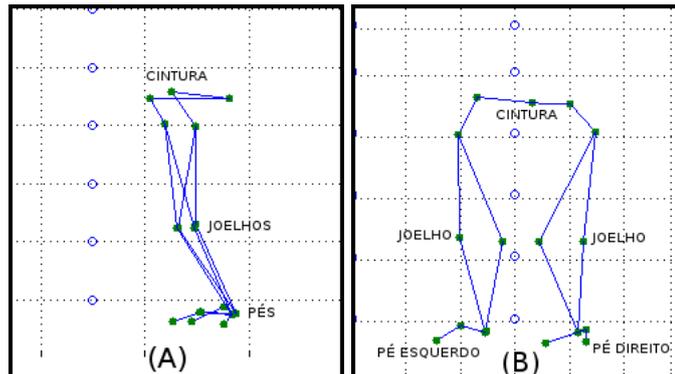


Figura 4.12: Resultado da Reconstrução do momento exibido na figura 4.7. (A) Vista Lateral; (B) Vista Frontal.

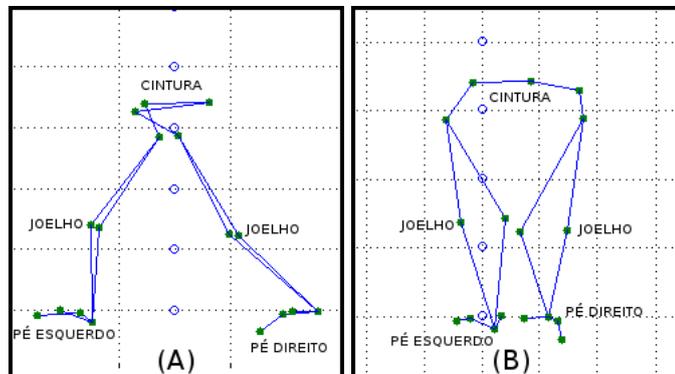


Figura 4.13: Resultado da Reconstrução do momento exibido na figura 4.8. (A) Vista Lateral; (B) Vista Frontal.

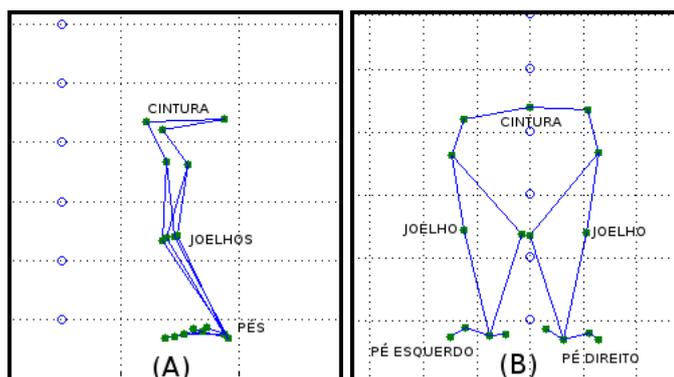


Figura 4.14: Resultado da Reconstrução do momento exibido na figura 4.9. (A) Vista Lateral; (B) Vista Frontal.

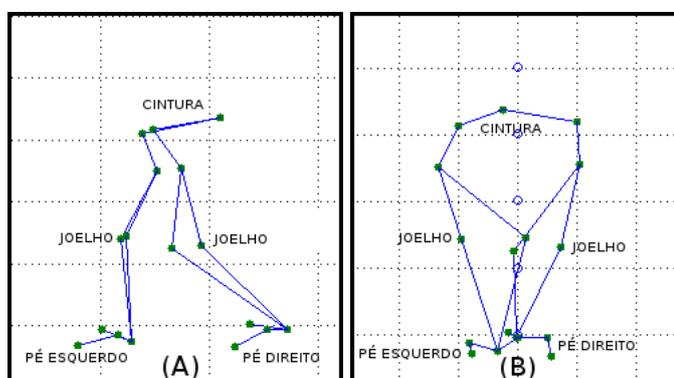


Figura 4.15: Resultado da Reconstrução do momento exibido na figura 4.10. (A) Vista Lateral; (B) Vista Frontal.

Não foi possível realizar a avaliação destes resultados, pois não podemos comparar as coordenadas calculadas com o experimento real. Visualmente, podemos notar que a reconstrução foi obtida conforme o esperado.

4.3 Considerações Finais

Neste capítulo foram apresentados os métodos utilizados nesse trabalho na detecção dos marcadores passivos para os sistemas de captura de movimentos por dispositivos ópticos. A transformada de *Hough* foi aplicada nos quadros de vídeos usando como parâmetros de busca as dimensões dos raios em *pixels*, com objetivo de localizar as coordenadas dos marcadores no plano da imagem 2D. Foi constatado que se a imagem for transformada em níveis de cinza e depois aplicada um detector de borda, a função de *Hough* detecta falsos marcadores. Dessa forma, foi necessário realizar a binarização da imagem usando os limiares dos canais RGB. Essa binarização favorece destacar as bordas dos marcadores passivos, diminuindo drasticamente a

detecção de bordas de outros objetos registrados pelas câmeras. Os resultados foram avaliados por métricas estatísticas, em seguida, os marcadores detectados foram reconstruídos no espaço tridimensional.

No próximo capítulo, será concluído o presente trabalho e apresentadas as indicações de trabalho futuro.

Capítulo 5

CONCLUSÃO

Este trabalho apresentou uma proposta de um sistema de captura de movimentos baseados em câmeras de vídeos e marcadores passivos que são detectados e utilizados na etapa de reconstrução de movimentos. Neste sistema foram implementadas as etapas de detecção dos marcadores, calibração das câmeras e reconstrução tridimensional dos marcadores passivos. As implementações consideraram a realidade do ambiente já construído para captura de marcha do grupo de Biomecânica do DEFMH-UFSCar. Este grupo utiliza atualmente o sistema DVideow (CARVALHO; TUDELLA; BARROS, 2005) para todas as etapas da análise computacional. Considerando os resultados apresentados neste trabalho, a etapa de detecção dos marcadores, realizada manualmente, pode ser substituída pela detecção implementada, que utiliza a transformada de *Hough*.

Para alcançar os objetivos, um levantamento bibliográfico foi realizado referente as técnicas que envolvem os sistemas de captura de movimentos por dispositivos ópticos e marcadores passivos. Essas etapas consistem na sincronização dos vídeos, detecção dos marcadores, correspondência, *tracking*, calibração e reconstrução tridimensional dos movimentos.

Na etapa de sincronização dos vídeos, todos os quadros foram extraídos e salvos em arquivos no momento da avaliação cinemática. O evento de ascender a luz no ambiente com pouca iluminação foi realizado no início de cada marcha. Dessa forma, foram localizados o primeiro quadro de todas as câmeras que registrou a alternância de luz no ambiente. Depois de identificar a etapa anterior, os quadros seguintes das câmeras são considerados sincronizados e o processo de detecção dos marcadores é realizado. Este processo, atualmente manual, pode ser substituído por uma etapa totalmente automática em um trabalho futuro.

Na etapa de detecção dos marcadores verificou-se que a transformada de *Hough* poderia ser utilizada para extrair as coordenadas 2D dos marcadores registrados pelas câmeras, uma

vez que estes possuem um formato circular. Embora os resultados apresentem alguns erros, a automatização gerada por esta etapa acelera a execução da análise computacional, pois evita que o usuário marque manualmente cada marcador em cada vídeo.

Para estabelecer a correspondência dos marcadores no primeiro quadro da avaliação, técnicas utilizadas na literatura foram pesquisadas para realizar esse processo automaticamente. Uma abordagem de correspondência chegou a ser implementada, mas sua investigação ainda não foi concluída, tornando inviável a inclusão desta etapa de forma automática. Em um momento futuro esta etapa será totalmente automatizada.

O processo de *tracking* não foi implementado e a reconstrução foi realizada em alguns quadros, não completando o movimento integral da marcha. Esta etapa precisa ser melhor investigada.

O método proposto nesse trabalho para realizar a detecção dos marcadores passivos por meio da transformada de *Hough* demonstraram resultados muito positivos. Os resultados também podem ser melhorados se algumas características do ambiente forem alteradas, por exemplo, na eliminação dos reflexos no chão e nos objetos no período de registro da avaliação. As imposições oriundas das capturas já realizadas previamente limitaram algumas decisões que poderiam otimizar o processo de análise computacional.

5.1 **Trabalhos Futuros**

Como trabalhos futuros, é necessário investigar técnicas para otimizar as etapas de sincronização, correspondência e rastreamento dos marcadores. Na etapa de sincronização dos registros um algoritmo pode ser desenvolvido para capturar as alternâncias de luz nas lâmpadas de sincronização, tornando o processo de sincronização dos vídeos automático.

Outra abordagem interessante é a detecção dos marcadores somente na região da imagem que possui o movimento realizado pela pessoa. Esse método pode diminuir a detecção de falsos marcadores, minimizar o tempo de processamento da transformada de *Hough* e otimizar qualitativamente a detecção dos marcadores.

REFERÊNCIAS

- ABDEL-AZIZ; Y.I.KARARA, H. Direct linear transformation from comparator coordinates into object-space coordinates in close-range photogrammetry. *Proceeding of the ASP/UI Symposium on Close-Range Photogrammetry. American Society of Photogrammetry, Falls Church, VA*, p. 1–18, 1971.
- AGARWAL, A.; TRIGGS, B. 3D human pose from silhouettes by relevance vector regression. In: IEEE. *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*. [S.l.], 2004. v. 2, p. II–882.
- ANDERSEN, M. R. et al. Kinect depth sensor evaluation for computer vision applications. *Technical report ECE-TR-6*, p. 37, February, 2012. Disponível em: <"<http://www.eng.au.dk>">.
- ANDREWS, J.; BAKER, N. Xbox 360 system architecture. *IEEE ASSP Magazine*, v. 26-N.2, p. 25–37, 2006.
- BALLARD, D. H. Generalizing the hough transform to detect arbitrary shapes. *Pattern Recognition*, Elsevier, v. 13, n. 2, p. 111–122, 1981.
- BARROS, R. et al. Desenvolvimento e avaliação de um sistema para análise cinemática tridimensional de movimentos humanos. *SBEB - Sociedade Brasileira de Engenharia Biomédica*, v. 15 n. 1-2, p. 79–86, 1999.
- BRENZIKOFER, R. et al. Desenvolvimento de uma metodologia para reconstrução tridimensional e análise de superfícies do corpo humano. 2001. Disponível em: <<http://www.ime.unicamp.br/calculo/mostra/outros/paulo/resumo.htm>>.
- BROWN, M.; LOWE, D. G. Recognising panoramas. In: *ICCV*. [S.l.: s.n.], 2003. v. 3, p. 1218.
- CACS. The center for advanced computer studies. university of louisiana at lafayette. *University of Louisiana at Lafayette*, 2007. Disponível em: <<http://www.cacs.louisiana.edu/labs/ecrg/vicon/>>.
- CARVALHO, R. P.; TUDELLA, E.; BARROS, R. Utilização do sistema dvideow na análise cinemática do alcance manual de lactentes. *Revista Brasileira de Fisioterapia*, v. 9, n. 1, p. 41–7, 2005.
- CASTRO, K. C. de. *Características Biomecânicas dos pés durante a marcha de crianças típicas e com síndrome de down*. Dissertação (Mestrado) — Universidade Federal de São Carlos. C355cb CDD: 615.82(20), 2012.

- CHEN, L.; ARMSTRONG, C. W.; RAFTOPOULOS, D. D. An investigation on the accuracy of three-dimensional space reconstruction using the direct lineal transformation technique. *Biomechanics*, v. 27, p. 493–500, 1994.
- DAESIK, K. Imaging geometry and camera calibration and 3d reconstruction. *Intelligent Systems Research Center - Sunghyunkwan University - Suwon, Korea*, 2006. 2006-SN-003-EN.
- DAVID, P. et al. Simultaneous pose and correspondence determination using line features. In: IEEE. *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*. [S.l.], 2003. v. 2, p. II–424.
- DAVIES, E. Computer and machine vision: Theory. *Algorithms, Practicalities*, v. 4, p. 912, 2012.
- DELISA. *Gait Analysis in the Science of Rehabilitation*. Department of Veterans Affairs, Veterans Health Administration, Rehabilitation Research and Development Service, Scientific and Technical Publications Section, 1998. (Collection of Jamie and Michael Kassler). ISBN 9780756700218. Disponível em: <http://books.google.com.br/books?id=e_WKGCDF0xgC>.
- DEUTSCHER, J.; REID, I. Articulated body motion capture by stochastic search. *International Journal of Computer Vision*, Springer, v. 61, n. 2, p. 185–205, 2005.
- DUDA, R. O.; HART, P. E. Use of the hough transformation to detect lines and curves in pictures. *Communications of the ACM*, ACM, v. 15, n. 1, p. 11–15, 1972.
- FIGUEROA, P. J.; LEITE, N. J.; BARROS, R. M. L. A flexible software for tracking of marker used in human motion analysis. *Computer Methods and Programs in Biomedicine* 72, v. 72, p. 155–165, 1 August 2002 2003.
- GALL, J. et al. Motion capture using joint skeleton tracking and surface estimation. In: IEEE. *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. [S.l.], 2009. p. 1746–1753.
- GRAUMAN, K.; SHAKHNAROVICH, G.; DARRELL, T. Inferring 3d structure with a statistical image-based shape model. In: IEEE. *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*. [S.l.], 2003. p. 641–647.
- GUERRA-FILHO, G. Optical motion capture: Theory and implementation. *RITA(Revista de Informática Teórica e Aplicada)*, Citeseer, v. 12, n. 2, p. 61–90, 2005.
- HALLEMANS, A. et al. 3d joint dynamics of walking in toddlers a cross-sectional study spanning the first rapid development phase of walking. *Gait & Posture*, v. 22, p. 107–118, 2004.
- HATZE, H. High-precision three-dimensional photogrammetric calibration and object space reconstruction using a modified dlt-approach. *Journal Biomechanics*, v. 21, n. 7, p. 533–538, May 1987 1988.
- HAUSDORFF, J. M. et al. Maturation of gait dynamics: stride-to-stride variability and its temporal organization in children. *Journal of Applied Physiology*, v. 86, p. 1040–1047, 1999.

- HERDA, L. et al. Using skeleton-based tracking to increase the reliability of optical motion capture. *Human movement science*, Elsevier, v. 20, n. 3, p. 313–341, 2001.
- HOGG, D. Model-based vision: a program to see a walking person. *Image and Vision computing*, Elsevier, v. 1, n. 1, p. 5–20, 1983.
- HOUGH, V.; PAUL, C. *Method and means for recognizing complex patterns*. 1962. US Patent 3,069,654.
- ILLINGWORTH, J.; KITTLER, J. A survey of the hough transform. *Computer vision, graphics, and image processing*, Elsevier, v. 44, n. 1, p. 87–116, 1988.
- KIMME, C.; BALLARD, D.; SKLANSKY, J. Finding circles by an array of accumulators. *Communications of the ACM*, ACM, v. 18, n. 2, p. 120–122, 1975.
- KIRK, A. G.; O'BRIEN, J. F.; FORSYTH, D. A. Skeletal parameter estimation from optical motion capture data. v. 2, n. ISSN : 1063-6919, p. 782–788, 2005.
- KOLAH, A. et al. Design of a marker-based human motion tracking system. *Biomedical Signal Processing and Control*, v. 2, p. 59–67, 2007.
- LEWIS, J. P. Fast template matching. In: *Vision interface*. [S.l.: s.n.], 1995. v. 95, n. 120123, p. 15–19.
- LI, B.; MENG, Q.; HOLSTEIN, H. Point pattern matching and applications-a review. In: IEEE. *Systems, Man and Cybernetics, 2003. IEEE International Conference on*. [S.l.], 2003. v. 1, p. 729–736.
- LIVERMAN, M. *The Animator's Motion Capture Guide: Organizing, Managing, and Editing*. Editora Charles River Media, 2004. (Charles River Media Game Development). ISBN 9781584502913. Disponível em: <http://books.google.com.br/books?id=Sa9NtqT_YxQC>.
- LOULA, C. M. A. Análise de volumes parciais do tronco durante a respiração por videogrametria. Biblioteca Digital da Unicamp, 2005.
- MAAS, H.; GRUEN, A.; PAPANTONIOU, D. Particle tracking velocimetry in three-dimensional flows. *Experiments in Fluids*, Springer, v. 15, n. 2, p. 133–146, 1993.
- MILLER, N. R.; SHAPIRO, R.; MCLAUGHLIN, T. M. A technique for obtaining spatial kinematic parameter of segments of biomechanical system from cinematographic data. *Journal of Biomechanics*, v. 13, p. 535–547, 1980.
- MOESLUND, T. B.; GRANUM, E. A survey of computer vision-based human motion capture. *Computer Vision and Image Understanding*, v. 81, p. 231–268, 2001.
- MOESLUND, T. B.; HILTON, A.; KRÜGER, V. A survey of advances in vision-based human motion capture and analysis. *Computer vision and image understanding*, Elsevier, v. 104, n. 2, p. 90–126, 2006.
- MORI, G. et al. Recovering human body configurations: Combining segmentation and recognition. In: IEEE. *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*. [S.l.], 2004. v. 2, p. II–326.

- MOURA, F. A. et al. Analysis of the shots to goal strategies of first division brazilian professional soccer teams. In: *ISBS-Conference Proceedings Archive*. [S.l.: s.n.], 2007. v. 1, n. 1.
- O'BRIEN, J. F. et al. Automatic joint parameter estimation from magnetic motion capture data. Georgia Institute of Technology, 1999.
- O'ROURKE, J.; BADLER, N. et al. Model-based image analysis of human motion using constraint propagation. *IEEE Transactions on, Pattern Analysis and Machine Intelligence*, n. 6, p. 522–536, 1980.
- POURCELOT, P. et al. A method to synchronise cameras using the direct linear transformation technique. *Journal of Biomechanics*, v. 33, p. 1751–1754, 2000.
- RAMANAN, D.; FORSYTH, D. A. Finding and tracking people from the bottom up. In: IEEE. *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*. [S.l.], 2003. v. 2, p. II–467.
- RONFARD, R.; SCHMID, C.; TRIGGS, B. Learning to parse pictures of people. In: *Computer Vision - ECCV 2002*. [S.l.]: Springer, 2002. p. 700–714.
- SILAGHI, M.-C. et al. Local and global skeleton fitting techniques for optical motion capture. In: *Modelling and Motion Capture Techniques for Virtual Environments*. [S.l.]: Springer, 1998. p. 26–40.
- SMINCHISESCU, C. et al. Discriminative density propagation for 3d human motion estimation. In: IEEE. *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. [S.l.], 2005. v. 1, p. 390–397.
- SMINCHISESCU, C.; TRIGGS, B. Estimating articulated human motion with covariance scaled sampling. *The International Journal of Robotics Research*, v. 22, n. 6, p. 371–391, June 2003.
- SUNDARESAN, A.; CHELLAPPA, R. Markerless motion capture using multiple cameras. In: IEEE. *Computer Vision for Interactive and Intelligent Environment, 2005*. [S.l.], 2005. p. 15–26.
- TAYLOR, C. J. Reconstruction of articulated objects from point correspondences in a single uncalibrated image. In: IEEE. *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on*. [S.l.], 2000. v. 1, p. 677–684.
- TSUJI, S.; MATSUMOTO, F. Detection of ellipses by a modified hough transformation. *IEEE transactions on computers*, C-27, n. 8, p. 777–781, 1978.
- VLASIC, D. et al. Practical motion capture in everyday surroundings. *ACM Transactions on Graphics*, v. 26, n. 3, p. 35–1 a 35–10, 2007.
- WADA, T. et al. Homography based parallel volume intersection: Toward real-time volume reconstruction using active cameras. In: IEEE. *Computer Architectures for Machine Perception, 2000. Proceedings. Fifth IEEE International Workshop on*. [S.l.], 2000. p. 331–339.
- WANG, J. J.; SINGH, S. Video analysis of human dynamics - a survey. *Real-time Imaging*, v. 9, p. 321–346, 2003.

WOOD, G. A.; MARSHALL, R. N. The accuracy of dlt extrapolation in three-dimensional film analysis. *J. Biomechanics*, v. 19, n. 9, p. 781–785, April 1986 1986.

ZITOVA, B.; FLUSSER, J. Image registration methods: a survey. *Image and vision computing*, Elsevier, v. 21, n. 11, p. 977–1000, 2003.