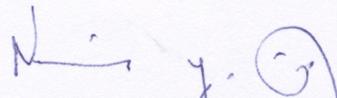


Generalização do ritmo visual e problemas de rastreamento de imagens desportivas

Este exemplar corresponde à redação final da Dissertação devidamente corrigida e defendida por Fábio Augusto Salve Dias e aprovada pela Banca Examinadora.

Campinas, 13 de Fevereiro de 2009.



Neucimar Jerônimo Leite
Instituto de Computação
Universidade Estadual de Campinas
(Orientador)

Dissertação apresentada ao Instituto de Computação, UNICAMP, como requisito parcial para a obtenção do título de Mestre em Ciência da Computação.

**FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DO IMECC DA UNICAMP**

Bibliotecária: Crislene Queiroz Custódio – CRB8 / 7966

Dias, Fábio Augusto Salve

D543g Generalização do ritmo visual e problemas de rastreamento de
imagens desportivas / Fábio Augusto Salve Dias -- Campinas, [S.P. :
s.n.], 2009.

Orientador : Neucimar Jerônimo Leite

Dissertação (Mestrado) - Universidade Estadual de Campinas,
Instituto de Computação.

1. Análise de imagem. 2. Processamento de imagens. 3.
Rastreamento automático. 4. Reconhecimento de padrões. I. Leite,
Neucimar Jerônimo. II. Universidade Estadual de Campinas. Instituto de
Computação. III. Título.

Título em inglês: Generalized visual rhythm and tracking in sports images

Palavras-chave em inglês (Keywords): 1. Image analysis. 2. Image processing. 3. Automatic tracking. 4. Pattern recognition.

Área de concentração: Ciência da Computação

Titulação: Mestre em Ciência da Computação

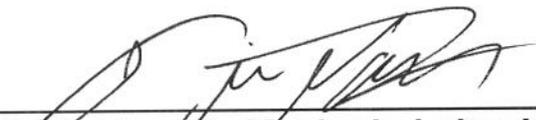
Banca examinadora: Prof. Dr. Ricardo da Silva Torres (IC-Unicamp)
Prof. Dr. Ricardo Machado Leite de Barros (FEF-Unicamp)
Prof. Dr. Hélio Pedrini (IC-Unicamp)
Prof. Dr. Arnaldo de Albuquerque Araújo (UFMG)

Data da defesa: 13/02/2009

Programa de Pós-Graduação: Mestrado em Ciência da Computação

TERMO DE APROVAÇÃO

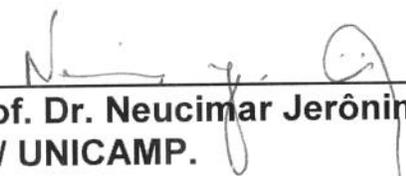
Dissertação Defendida e Aprovada em 13 de fevereiro de 2009, pela
Banca examinadora composta pelos Professores Doutores:



Prof. Dr. Ricardo Machado Leite de Barros
FEF / UNICAMP.



Prof. Dr. Ricardo da Silva Torres
IC / UNICAMP.



Prof. Dr. Neucimar Jerônimo Leite
IC / UNICAMP.

Generalização do ritmo visual e problemas de rastreamento de imagens desportivas

Fábio Augusto Salve Dias¹

Janeiro de 2009

Banca Examinadora:

- Neucimar Jerônimo Leite
Instituto de Computação
Universidade Estadual de Campinas (Orientador)
- Ricardo da Silva Torres
Instituto de Computação
Universidade Estadual de Campinas
- Ricardo Machado Leite de Barros
Faculdade de Educação Física
Universidade Estadual de Campinas
- Hélio Pedrini
Instituto de Computação
Universidade Estadual de Campinas
- Arnaldo de Albuquerque Araújo
Instituto de Ciências Exatas
Universidade Federal de Minas Gerais

¹Suporte financeiro de: Bolsa FAPESP (processo 06/59526-8) 2007–2008.

Resumo

O rastreamento de objetos em imagens de vídeo está se tornando cada vez mais popular, nos mais diversos contextos, possibilitando a obtenção de informações que, de outra maneira, seriam muito trabalhosas ou mesmo impossíveis de serem obtidas.

Entretanto, a grande maioria dos métodos presentes na literatura realiza o rastreamento no domínio da imagem, não no domínio tridimensional natural do problema. No contexto de imagens desportivas, onde movimentos rápidos, similaridades cromáticas entre jogadores do mesmo time e oclusões são significativas, a utilização do domínio tridimensional é particularmente interessante.

Este trabalho tem como premissa básica a realização do rastreamento em um domínio tridimensional, partindo de determinadas hipóteses, plausíveis em um contexto desportivo, visando ao rastreamento automático dos jogadores em atividades desportivas. Generalizamos o conceito de ritmo visual, um método de amostragem vastamente utilizado para detecção de cortes em imagens de vídeos, e o utilizamos para efetuar o rastreamento, resultando em um método simples, intuitivo e automático quando seus pré-requisitos são atendidos.

Para possibilitar a utilização da generalização do ritmo visual, desenvolvemos ainda algoritmos para identificação de objetos em movimento, calibração de câmeras e reconstrução tridimensional, além de um algoritmo baseado em grafos para a segmentação da trajetória de cada objeto em movimento, utilizando a informação obtida através do ritmo visual.

Apresentamos resultados experimentais da utilização do método proposto em vídeos de eventos desportivos e coletados em laboratório, incluindo casos nos quais as condições ideais de utilização não são atendidas, causando falhas no método.

Abstract

The automatic tracking in the video images is becoming increasingly popular, in many different contexts, making available information that, otherwise, would be very laborious or even impossible to be obtained.

However, most of the methods perform the tracking in the image domain not in the natural tridimensional domain. For sports images, with rapid motions, color similarity and oclusions, the tridimensional approach is an interesting alternative.

In this work, we introduce a tracking method in the tridimensional domain, based on some requirements, feasible in the sports context, aiming fully automatic tracking of the players. We also introduce the generalization of the visual rhythm concept and its use to the tracking problem, leading to a simple, intuitive, and automatic method.

To use the generalization of the visual rhythm, we also developed algorithms to detect moving objects, camera calibration and three-dimensional reconstruction, and an algorithm based on graphs for the segmentation of the trajectory of each moving object in the scene, considering information obtained through the visual rhythm.

We also show experimental results of the proposed method, in sports images and controlled environments, including cases where not all the requirements are fulfilled, yielding failures of the proposed approach.

Agradecimentos

Eu gostaria de agradecer a todos que colaboraram, direta ou indiretamente com o desenvolvimento deste trabalho. É provavelmente conhecimento público que tenho péssima memória, incluindo uma séria tendência a trocar o nome das pessoas. Isso, aliado ao fato de eu não querer cometer injustiças e a relativa liberdade que tenho nesta página, me levam a fazer algo incomum. Não citarei nomes. Você, caro leitor, se acha que deve ser agradecido por mim, aqui, provavelmente tem razão, então *Muito Obrigado*, pois este trabalho não seria o mesmo sem sua colaboração.

Espero também que você, estimado leitor e colaborador, se anime a ler o resto do trabalho e espero mais ainda que fique feliz com o que consegui fazer com sua ajuda. A propósito, fique tranquilo, pois o trabalho parece extenso mas é porque existem muitas figuras. Acredito que esta última frase seja também adequada aos caros professores da banca.

Os dois parágrafos acima foram escritos antes da defesa. Por mais politicamente correto que eles sejam, minha consciência ficou pesada ao não citar nomes. Eu sempre fico muito feliz quando me vejo nos agradecimentos dos colegas. Então tentarei ser tradicional e fazer os agradecimentos nominais, mas não serei formal.

Entretanto, se eu não citar seu nome, caro leitor, por favor se inclua através dos dois primeiros parágrafos e não se chateie com isso.

Começarei por ordem cronológica: a sempre presente família: Meu Pai, Vó, Zilda, Roque, Adriana, Gabriela e Isabela. Minha esposa Suellen por toda a ajuda, apoio e broncas.

Os sempre confiáveis companheiros de engenharia, sejam B-zóides ou A-zóides: Toneto, Matemático, Monte Mor, Danilo, Guerreiro, Lufi, Fleury, Dentista, Morpha, Silvio, Aharon, Vivi, Marco, Duck e Tati, assim como vários outros. Também aos ‘bixos’: Murilo, Saffi, Marsola, Choi, Samir, Gui, Victor e Will. Um agradecimento especial ao Enrico, companheiro infalível, seja em idéias doidas, discussões matemáticas ou divertimentos em geral.

Um reconhecimento mais que merecido à Profa. Islene e sua imensa paciência comigo, desde a primeira conversa sobre pesquisa, cujos conselhos foram essenciais para meu per-

curso acadêmico. A insuperável e sempre correta Profa. Claudia, cujo exemplo acadêmico tentarei seguir pelo resto da vida. Os professores Ricardo Torres, Rodolfo e Siome que, sempre muito exigentes e nem sempre compreendidos por mim, ajudaram a moldar o professor exigente que meus alunos de PED alegaram que sou.

Ao Prof. Neucimar, inestimável parceiro e guia neste trabalho, mesmo quando tive a ‘audácia de convidá-lo para participar de uma coleta’. Aos sempre presentes: Daniel Capeleto, Ademilson, Flávio, Furlan, William e Rodrigo Canto.

Aos companheiros de LIS/LIV: Daniel, Nielsen, Luciano, Gilberto, André, Celso, Randall, Claudio, Rafael, Marcelo, Senra, Evandro, Arnaldo, Jefferson, Maurício, Otávio, Javier, Cristiano, Jaudete, Alan, Andréia, Ricardo Freitas, Rodrigo, Papa, Minetto, Danilo Lacerda e muitos outros.

Ao professor Ricardo Barros, por me apresentar à prática e seus problemas imprevisíveis, normalmente divertidos mas algumas vezes frustrantes e toda sua grande equipe de excelentes alunos do LIB: Amanda, Karine, Cadú (que não era oficialmente do LIB), Jerusa, Juliana, Lorena, Luciano, Milton, Pascual, Aline, Angélica, Thiago, Márcio, Mário e vários outros, afinal tem bastante gente trabalhando lá.

Eu tenho absoluta certeza de que esqueci alguém...

Ao Meu Pai.

Sumário

Resumo	vii
Abstract	ix
Agradecimentos	xi
1 Introdução	1
2 Método proposto	7
2.1 Calibração de câmeras	7
2.1.1 Modelo de projeção considerado	8
2.1.2 Estimativa inicial dos parâmetros	9
2.1.3 Otimização dos parâmetros	10
2.1.4 Alguns resultados experimentais	11
2.1.5 Conclusão	12
2.2 Modelagem do fundo e identificação dos objetos em movimento	14
2.2.1 Modelo considerado	14
2.2.2 Aprendizado das informações dos modelos de fundo e objetos	14
2.2.3 Identificação dos objetos em movimento	15
2.2.4 Alguns resultados experimentais	16
2.2.5 Conclusão	20
2.3 Reconstrução Tridimensional	22
2.3.1 Método proposto para a reconstrução tridimensional	22
2.3.2 Alguns resultados experimentais	23
2.3.3 Falsos volumes na reconstrução	30
2.3.4 Conclusão	34
2.4 Amostragem da informação através da generalização do conceito de ritmo visual	36
2.4.1 Ritmo visual	36
2.4.2 Generalizando o ritmo visual	36

2.4.3	Alguns resultados experimentais	38
2.4.4	Conclusão	39
2.5	Filtragem através da generalização do operador direcional multi-escala . . .	44
2.5.1	Definição das direções a serem consideradas	44
2.5.2	Estimativa da orientação dos voxels	44
2.5.3	Filtragem multi-escala	46
2.5.4	Alguns resultados experimentais	47
2.5.5	Conclusão	48
2.6	Identificação das trajetórias	49
2.6.1	Binarização	49
2.6.2	Construção do grafo	49
2.6.3	Conclusão	54
3	Resultados experimentais	59
3.1	Testes em laboratório	59
3.1.1	Calibração de câmeras	59
3.1.2	Identificação dos objetos em movimento	60
3.1.3	Reconstrução Tridimensional	62
3.1.4	Amostragem através do ritmo visual generalizado	63
3.1.5	Identificação das trajetórias	63
3.1.6	Conclusões	64
3.2	Jogo oficial de vôlei feminino - Bulgária x Alemanha	70
3.2.1	Calibração de câmeras	70
3.2.2	Identificação dos objetos em movimento	72
3.2.3	Reconstrução Tridimensional	73
3.2.4	Amostragem através do ritmo visual generalizado	74
3.2.5	Identificação das trajetórias	75
3.2.6	Conclusões	75
3.3	Final do Campeonato Mundial de Futsal para cegos	78
3.3.1	Calibração de câmeras	78
3.3.2	Identificação dos objetos em movimento	78
3.3.3	Reconstrução Tridimensional	79
3.3.4	Conclusões	80
3.4	CVBASE06	87
3.4.1	Identificação dos objetos em movimento	87
3.4.2	Amostragem através do ritmo visual generalizado	88
3.4.3	Identificação das trajetórias	89
3.4.4	Conclusões	89

4	Conclusões	93
5	Trabalhos futuros	95
	Bibliografia	96

Lista de Tabelas

1.1	Resumo dos trabalhos relevantes	4
3.1	Sumário dos erros encontrados na calibração das câmeras.	59
3.2	Sumário dos erros encontrados na calibração das câmeras.	72
3.3	Sumário dos erros encontrados na calibração das câmeras.	78

Lista de Figuras

2.1	Diagrama das etapas envolvidas no método proposto.	8
2.2	Exemplos de imagens para calibração.	11
2.3	Exemplo de imagem de calibração coletada em laboratório.	12
2.4	Coleta controlada.	12
2.5	Ajuste histográfico.	16
2.6	Modelo do fundo e resultado correspondente, considerando 10 quadros e $\epsilon = 5$	17
2.7	Modelo do fundo e resultado correspondente, considerando 10 quadros e $\epsilon = 10$	18
2.8	Modelo do fundo e resultado correspondente, considerando 30 quadros e $\epsilon = 10$	19
2.9	Modelo do fundo e resultado correspondente, utilizando o método apresentado em [13].	21
2.10	Resultado da identificação de fundo - câmera 1.	24
2.11	Resultado da identificação de fundo - câmera 2.	24
2.12	Resultado da identificação de fundo - câmera 3.	25
2.13	Resultado da identificação de fundo - câmera 4.	25
2.14	A reconstrução a partir da interseção de cones de reprojeção.	26
2.15	Resultado da reconstrução.	27
2.16	Resultado da identificação de fundo.	28
2.17	Resultado da reconstrução - Falsos volumes.	29
2.18	Geração de falsos volumes.	30
2.19	Diagrama de posicionamento de câmeras.	31
2.20	Diagrama tridimensional de posicionamento de câmeras.	32
2.21	Diagrama de cobertura de câmeras.	33
2.22	Resultado da reconstrução utilizando o grupo 1.	33
2.23	Resultado da reconstrução utilizando os grupos 1 e 2.	34
2.24	Resultado da reconstrução utilizando as dez câmeras laterais.	34
2.25	Resultado da reconstrução utilizando somente a câmera superior.	35

2.26	Resultado da reconstrução utilizando somente o grupo 1 e a câmera superior.	35
2.27	Exemplo da abordagem tradicional do ritmo visual.	37
2.28	Exemplo da operação de amostragem considerada.	40
2.29	Resultado do ritmo visual para a coleta em laboratório - Quadros 1 a 300.	41
2.30	Resultado do ritmo visual para a coleta real - Quadros 1 a 50.	41
2.31	Reconstrução da sequência sintética - Quadro 1.	42
2.32	Resultado do ritmo visual para o vídeo sintético - Quadros 1 a 200.	43
2.33	Exemplo dos versores perpendiculares considerados - $\delta = 30$ graus.	45
2.34	Exemplo da aplicação do método apresentado.	46
2.35	Exemplo do efeito do operador para fechamento de falhas.	47
2.36	Geração inicial dos vértices para o exemplo considerado.	50
2.37	Vértices iniciais e o grafo correspondente.	51
2.38	Possíveis configurações de trajetórias e suas energias de acordo com a equação 2.6.2.	52
2.39	Após a separação de uma região.	54
2.40	Resultado final para o exemplo considerado.	56
2.41	Áreas correspondentes ao primeiro vértice possível.	57
2.42	Áreas correspondentes ao segundo vértice possível.	57
3.1	Exemplos das imagens consideradas.	60
3.2	Pontos utilizados para calibração.	61
3.3	Resultado da identificação dos objetos em movimento para o quadro 1.	62
3.4	Objetos detectados superpostos com a imagem original.	63
3.5	Resultado da reconstrução tridimensional para o quadro 1.	65
3.6	Resultado da amostragem através do ritmo visual generalizado.	66
3.7	Áreas correspondentes ao primeiro vértice possível - 713.	67
3.8	Áreas correspondentes ao primeiro vértice possível - 714.	67
3.9	Resultado da identificação de trajetórias.	68
3.10	Resultado da identificação de trajetórias.	69
3.11	Exemplos das imagens consideradas.	70
3.12	Pontos utilizados para calibração.	71
3.13	Resultado da identificação dos objetos em movimento para o quadro 30.	73
3.14	Objetos detectados superpostos com a imagem original.	74
3.15	Resultado da reconstrução tridimensional para o quadro 30.	76
3.16	Resultado da amostragem através do ritmo visual generalizado.	77
3.17	Visualização da razão da falha do método apresentado na sequência considerada.	77
3.18	Exemplos das imagens capturadas.	81

3.19	Pontos utilizados para calibração.	82
3.20	Resultado da identificação dos objetos em movimento para o quadro 1. . .	83
3.21	Objetos detectados superpostos com a imagem original.	84
3.22	Resultado da reconstrução tridimensional para o quadro 1.	85
3.23	Cones de reprojeção para o quadro 1.	86
3.24	Exemplos das imagens consideradas.	87
3.25	Resultado da identificação dos objetos em movimento para o quadro 1. . .	88
3.26	Objetos detectados superpostos com a imagem original.	88
3.27	Resultado da amostragem através do ritmo visual generalizado.	90
3.28	Resultado da identificação de trajetórias.	91
3.29	Resultado da identificação de trajetórias.	92

Capítulo 1

Introdução

O rastreamento de objetos em imagens de vídeo tem se tornado uma alternativa cada vez mais atraente para obtenção de informações que, de outra forma, seriam impossíveis ou muito trabalhosas de serem obtidas. A aplicação destes métodos no contexto de análise esportiva tem sido explorada nas mais diversas modalidades, com nível de automatização cada vez melhor.

Em [35], são utilizadas duas câmeras superiores, com grande distorção radial, para filmar toda a quadra durante um jogo de handebol. Três diferentes métodos semi-automáticos para a detecção dos objetos em movimento são considerados: baseado em movimento, baseado em modelos e baseado em informação cromática. No modelo baseado em movimento, a detecção dos objetos em movimento é calculada a partir da diferença entre o quadro considerado e um quadro de referência, seguido pela aplicação de um limiar. Considerando modelos, são utilizados pequenos padrões binários, aplicados em cada canal da imagem. No método baseado em informação cromática, apenas o pixel com maior similaridade ao modelo pré-computado dos jogadores é considerado como posição atual do jogador. Nos testes apresentados, a última abordagem se mostrou mais precisa, necessitando de um menor número de intervenções do usuário.

Pingali et al. [37] apresentam um sistema comercial para obtenção de informações em jogos de tênis, chamado LucentVision. São utilizadas oito câmeras, duas para efetuar o rastreamento dos jogadores através de detecção de movimento e características locais e outras seis para o rastreamento da bola através do movimento, cor e formato da bola. O sistema é utilizado em jogos oficiais desde 1998, com resultados muito precisos.

Pers et al. [36] utilizam uma única câmera superior, em uma quadra de *squash*. O rastreamento é realizado através da subtração do fundo da cena, seguido da aplicação de um limiar. Em seguida, os blobs resultantes são processados, considerando-se a continuidade temporal dos jogadores. Caso os dois jogadores sejam segmentados como um único blob, o maior blob presente, este é dividido com relação à distância das posições anteriores dos

jogadores.

Andrade et al. [1] consideram o conceito de *árvore da imagem*, onde cada imagem é segmentada em regiões similares, hierarquicamente. O rastreamento é efetuado através de um grafo de adjacências, aliado a árvore da imagem, procurando relações entre regiões.

Em [27], o foco da abordagem é a distribuição, de maneira inteligente, dos dados entre diferentes unidades de processamento para efetuar o rastreamento. A detecção dos jogadores é feita através de um modelo computado de fundo, incluindo informações do gradiente da imagem. O rastreamento é feito através da continuidade temporal e, se necessário, informação cromática.

No trabalho apresentado em [22], o rastreamento dos jogadores é realizado através de um filtro de partículas, utilizando informação cromática. A separação de jogadores muito próximos é realizada através de um particionamento de Voronoi. O trabalho considera câmeras superiores, na mesma configuração utilizada em [35]. Um dos principais resultados apresentados neste trabalho é o aumento da automação para o rastreamento quando todas as pessoas em movimento na cena são rastreadas, em comparação ao rastreamento independente e individual de cada um dos jogadores e dos juízes.

Hayet et al [20] consideram imagens de transmissões televisivas dos eventos esportivos, tornando necessário o uso de técnicas de calibração de câmeras que se adaptem ao movimento das câmeras utilizadas. O rastreamento é feito através de características locais, através de pontos de interesse e as relações espaciais entre eles, em um modelo chamado de modelo de distribuição de pontos.

Em [29], os objetos em movimento são detectados através da subtração do fundo, cujo modelo é obtido considerando a mediana com relação ao tempo do vídeo, computada para cada pixel. O resultado é binarizado, filtrado morfológicamente e cada blob é caracterizado através da forma, cor ou tamanho. Estas características são utilizadas para identificar a relação entre blobs em diferentes quadros.

Figuerola et al. [14] modelam o fundo da cena a partir do vídeo do jogo através de operações morfológicas. Os jogadores são detectados através da subtração do fundo obtido com o quadro considerado. O resultado é filtrado morfológicamente, gerando blobs. Os blobs de todos os quadros da sequência são colocados em um grafo cujas arestas tem peso proporcional à distância entre os blobs em um espaço de características. O grafo é percorrido em ambas as direções temporais, eliminando ambiguidades.

No trabalho apresentado em [46], o fundo da cena é modelado como uma mistura de gaussianas. O rastreamento dos jogadores é realizado através de filtros de partículas, utilizando um classificador por vetor de suporte. Este método é utilizado também em [45] que objetiva a extração de informações táticas sobre o jogo. Um conceito relevante deste trabalho é a definição do que foi batizado como trajetória agregada, que é utilizado para a fusão das informações provenientes de várias câmeras ou da mesma câmera, em outra

posição.

Czyz et al. [9] consideram a utilização de filtro de partículas, utilizando somente informações cromáticas para o rastreamento. Os testes apresentados englobam tanto cenas esportivas quanto de monitoramento de ambientes externos.

Bennett et al. [3] consideram somente informações cromáticas. Cada objeto de interesse é modelado através de um histograma. Um modelo bayesiano é utilizado para a localização dos objetos. O rastreamento é efetuado através de um grafo, de uma maneira similar a apresentada em [14].

Leo et al. [24] modelam estatisticamente o fundo e o utiliza para a detecção dos objetos em movimento, de um modo distribuído. Através de homografias, um computador supervisor funde as informações e posiciona os jogadores em um campo virtual. As informações provenientes do rastreamento são utilizadas para detecção de eventos tais como impedimentos.

Miura et al. [26] consideram imagens de transmissões televisivas de jogos de futebol. Os parâmetros da homografia da câmera considerada são atualizados quadro a quadro, além do sistema também possuir técnicas de detecção de cortes através de histogramas. Os jogadores são detectados com base nas cores dos uniformes e a partir destas regiões em cada um dos quadros da sequência, é criado um grafo temporal, cujos vértices contêm uma faixa de número de objetos possíveis. O rastreamento é feito pela aplicação de restrições a este grafo, tais como continuidade temporal e número constante de jogadores em cena.

Um conceito similar foi utilizado em [43], mas tendo como objetivo a contagem de pessoas para fins de monitoramento. Utilizando câmeras estáticas e um modelo do fundo considerado, são obtidos blobs representando as pessoas em movimento. Utilizando estas informações, são definidas regiões que podem conter pessoas. Cada região possui uma faixa associada, indicando o possível número de pessoas contidas naquela região. Estas faixas são processadas em todos os quadros, de maneira similar ao considerado em [26], [3] e [14].

Utilizando técnicas semelhantes, o trabalho apresentado em [16] aborda o rastreamento de pessoas. Sua relevância para nosso trabalho se deve ao fato do rastreamento ser efetuado em um domínio tridimensional, construído a partir dos blobs obtidos pela subtração entre o quadro considerado e o modelo de fundo. A definição das trajetórias é realizada através de programação dinâmica e o método é capaz de tratar corretamente casos patológicos para o rastreamento, como saltos das pessoas envolvidas.

A tabela 1.1 resume as principais características dos trabalhos relevantes encontrados na literatura pesquisada, citando seus pré-requisitos e funcionalidades. Nesta tabela, o campo *Esporte* identifica o principal esporte considerado para o desenvolvimento do trabalho, *3D* identifica se o trabalho considera o espaço tridimensional como domínio para o rastreamento, isto é, efetua o rastreamento não em um domínio bidimensional,

como uma imagem, e depois encontra a posição no espaço 3D através de triangulações a partir de correspondências mas, sim, diretamente no espaço tridimensional. O campo *C.E.* identifica se o trabalho necessita de câmeras estacionárias, *C.S.* indica os trabalhos que necessitam de câmeras superiores, isto é, câmeras situadas acima dos jogadores, *M.F.* identifica quais trabalhos utilizam algum tipo de modelagem do fundo da cena, *Testes* indica a ordem de grandeza da duração dos vídeos utilizados para testes apresentados nos artigos e, finalmente, no campo *Mais informações*, fornecemos alguma informação relativa aos métodos apresentados.

	Esporte	3D	C.E.	C.S.	M.F.	Testes	Mais informações
[35]	Handebol	-	X	X	-	min	Múltiplas características
[37]	Tênis	-	X	-	X	Jogos	Subtração do fundo
[36]	Squash	-	X	X	X	min	Subtração do fundo
[1]	Futebol	-	-	-	-	min	Árvore da imagem
[27]	Futebol	-	X	-	X	Jogos	Sist. Distribuídos
[22]	Handebol	-	X	X	-	min	Filtro de partículas
[20]	Futebol	-	-	-	-	min	Filtro de partículas
[29]	Futebol	-	X	-	X	min	Múltiplas características
[28]	Futebol	-	X	-	X	min	Matriz de proximidade
[14]	Futebol	-	X	-	X	Jogos	Morfologia e grafos
[46]	Futebol	-	-	-	X	min	Filtro de partículas
[45]	Futebol	-	-	-	X	min	Filtro de partículas
[9]	Futebol	-	X	-	-	min	Filtro de partículas
[3]	Basquete	-	X	-	X	min	Grafos e Lei de Bayes
[24]	Futebol	-	X	-	X	Jogos	Sist. Distribuídos
[26]	Futebol	-	-	-	-	min	Grafos
[43]	Pessoas	X	X	-	X		Contagem de pessoas
[16]	Pessoas	X	X	-	X		Mapa probabilístico
Este trabalho	Vários	X	X	-	X	-	

Tabela 1.1: Resumo dos trabalhos relevantes

O principal objetivo deste trabalho é o estudo da utilização de informação tridimensional como entrada para o algoritmo de rastreamento. Adicionalmente, apresentamos generalizações do conceito de ritmo visual e do operador direcional multi-escala, aplicando ambos para a resolução do problema do rastreamento, focando nossos testes em imagens desportivas. O método apresentado possui alguns pré-requisitos, que, uma vez satisfeitos, possibilitam o rastreamento completamente automático, sem mesmo inicialização manual dos objetos em cena.

O contexto desportivo foi escolhido por apresentar características peculiares, algumas facilitadoras, como uma região de movimento delimitada, um número conhecido de objetos

em movimento e uma relativa estabilidade tanto na iluminação quanto no fundo da cena considerada mas também adversas, como movimentos rápidos, oclusões e contato físico entre os jogadores.

Estas condições facilitadoras foram utilizadas como pré-requisitos do método proposto, possibilitando a completa automação do rastreamento quando totalmente satisfeitas. Além das condições citadas acima, são necessárias uma boa taxa de quadros durante a captura, de modo a suavizar o movimento considerado e uma boa cobertura de câmeras, de modo a cada região ser observada por duas câmeras ou mais, sendo especialmente eficiente a utilização de uma câmera superior.

O restante deste trabalho é organizado da maneira a seguir: no capítulo 2 descrevemos o método proposto, abordando a calibração de câmeras (2.1), identificação dos objetos em movimento (2.2), reconstrução tridimensional (2.3), generalização do conceito de ritmo visual e sua aplicação ao problema do rastreamento (2.4), a generalização do operador direcional multi-escala (2.5) e a definição das trajetórias dos objetos (2.6). No capítulo 3 aplicamos o método apresentado em diversas situações diferentes, apresentando as falhas existentes e suas causas quando o método não consegue efetuar o rastreamento, e também analisamos os resultados apresentados caso o processamento seja completo. No capítulo 4, analisamos o comportamento, vantagens e desvantagens do método apresentado e no capítulo 5 apresentamos alguns possíveis trabalhos futuros que podem aumentar a robustez e o nível de automação do método apresentado.

Capítulo 2

Método proposto

Visando ao estudo do rastreamento em um domínio tridimensional, o método proposto foi dividido em duas partes: etapas que têm como objetivo a reconstrução tridimensional da cena considerada e etapas que realizam o rastreamento.

A figura 2.1 ilustra as etapas e fluxos de informação previstos no método proposto. As etapas de calibração de câmera e modelagem do fundo são executadas somente uma vez para cada vídeo considerado. A etapa de identificação dos objetos em movimento é realizada para todos os quadros do vídeo. A etapa de reconstrução tridimensional utiliza estas informações para recuperar a informação tridimensional da cena considerada.

Como resultado da reconstrução obtemos uma informação 4D, correspondente à evolução temporal da cena tridimensional observada. Esta informação é amostrada a partir do uso do conceito de ritmo visual e pode, opcionalmente, ser filtrada aplicando-se o operador direcional multi-escala. A etapa de definição de trajetórias segmenta a trajetória de cada objeto em movimento, fornecendo o resultado final do método.

2.1 Calibração de câmeras

Um dos passos importantes em nosso trabalho é a calibração das câmeras utilizadas, ou seja, o processo de recuperação dos parâmetros de observação da câmera, que explicitam como a cena observada é transformada para a geração da imagem obtida, possibilitando a recuperação da informação métrica original da cena. Apesar de ser uma área bastante explorada, ainda representa um potencial problema para coletas em campo, onde não se possui controle sobre o posicionamento das câmeras ou acesso irrestrito à região filmada.

Consideramos dois tipos de métodos para calibração, os que utilizam um volume de calibração 3D, como a Transformação Linear Direta (DLT) [38], e métodos que utilizam um elemento planar de calibração, tais como Tsai [40] e Zhang [44].

Utilizando como base a abordagem DLT, adicionamos os parâmetros não lineares

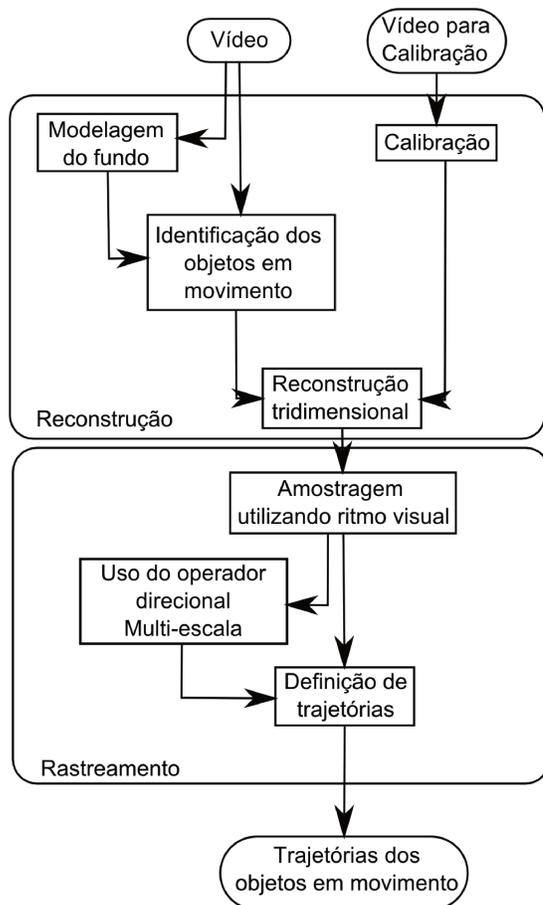


Figura 2.1: Diagrama das etapas envolvidas no método proposto.

a esta, do modo descrito em [42]. Utilizamos o software desenvolvido em [41] como comparação para corretude no cálculo dos parâmetros intrínsecos e também como método auxiliar, caso necessário, pois o trabalho manual para a calibração é razoavelmente maior que nossa abordagem, apesar do método ser incapaz, no nosso contexto, de computar os parâmetros extrínsecos.

Adotamos, opcionalmente, o método descrito em [44], com implementação disponível [6], para efetuar o cálculo inicial dos parâmetros e da distorção radial.

2.1.1 Modelo de projeção considerado

Sejam X a coordenada homogênea de um ponto no mundo real e $x = [u; v; 1]$ sua projeção correspondente. Definimos a seguinte equação de projeção, seguindo o modelo *pinhole* [19]:

$$x = \text{proj}(X) : x = K_{3 \times 3} f_d([R_{3 \times 3} | T_{3 \times 1}]X) \quad (2.1.1)$$

Onde $K_{3 \times 3}$ representa a matriz de parâmetros íntinsecos, f_d representa a função de distorção e $R_{3 \times 3}$ e $T_{3 \times 1}$ representam os parâmetros extrínsecos, rotação e translação, respectivamente. Neste trabalho, consideramos f_d como:

$$f_d(u, v) = [f_u(u, v); f_v(u, v)] \quad (2.1.2)$$

$$f_u = u + k_1 u(u^2 + v^2) + k_2 u(u^2 + v^2)^2 + s_1(u^2 + v^2) + p_1(3u^2 + v^2) + 2p_2 uv \quad (2.1.3)$$

$$f_v = v + k_1 v(u^2 + v^2) + k_2 v(u^2 + v^2)^2 + s_2(u^2 + v^2) + p_2(3v^2 + u^2) + 2p_1 uv \quad (2.1.4)$$

Onde k_1 , k_2 , p_1 , p_2 , s_1 e s_2 são os parâmetros que controlam a distorção. Os dois parâmetros k_1 e k_2 representam a distorção radial, p_1 e p_2 representam distorção *de-centering* e s_1 e s_2 representam a distorção de prisma fino.

2.1.2 Estimativa inicial dos parâmetros

Utilizando o modelo definido na seção 2.1.1, precisamos obter as matrizes K , R e T , além dos coeficientes de distorção para efetuar a calibração.

Inicialmente, consideramos a transformação linear direta [38], visando encontrar a matriz de projeção linear $P_{3 \times 4}$, definida como:

$$P = K[R|T] \quad (2.1.5)$$

Como podemos notar, esta abordagem desconsidera distorções, limitando seu uso para determinadas aplicações. A idéia principal deste algoritmo é que cada ponto adiciona duas equações ao sistema, cujas variáveis são os elementos da matriz P , empilhados e representados por p . As equações utilizadas foram:

$$\begin{bmatrix} X_i^T & 1 & 0 & 0 & 0 & 0 & -x_i(1) * X_i^T \\ 0 & 0 & 0 & 0 & X_i^T & 1 & -x_i(2) * X_i^T \\ & & & \vdots & & & \\ X_n^T & 1 & 0 & 0 & 0 & 0 & -x_n(1) * X_n^T \\ 0 & 0 & 0 & 0 & X_n^T & 1 & -x_n(2) * X_n^T \end{bmatrix} * p = \begin{bmatrix} x_i(1) \\ x_i(2) \\ \vdots \\ x_n(1) \\ x_n(2) \end{bmatrix} \quad (2.1.6)$$

Vale notar que a matriz de projeção é definida com exceção de um fator de escala. Neste passo, assumimos que o quarto elemento da terceira linha da matriz P seja unitário, visando a redução de uma variável no sistema linear apresentado na equação 2.1.6. Esta hipótese também força o último elemento do vetor T a ser unitário, impossibilitando o restante da decomposição desejada.

Todavia, como K deve ser triangular superior e ter o último elemento unitário, os três primeiros elementos da última linha da matriz P correspondem somente a um vetor pertencente à matriz de rotação, que, portanto, deve ser unitário, o que conduz a uma normalização mais apurada.

Após esta normalização, utilizamos a decomposição RQ [17] para obter as matrizes K e R a partir de uma parte da matriz P , e utilizamos estes dados para obter o vetor T . Com isso, obtivemos estimativas para os parâmetros lineares da calibração.

2.1.3 Otimização dos parâmetros

A próxima etapa de nosso método é a obtenção dos parâmetros de distorção e posterior refinamento do resultado obtido. Para tanto, utilizamos uma otimização não linear, visando a minimização do erro de reprojeção.

Sejam X^c o conjunto de pontos de calibração homogêneos, medidos no mundo real a partir de um sistema de coordenadas único, e x^c suas respectivas correspondências medidas manualmente nas imagens, definimos o erro de reprojeção como:

$$e = \sum_{\forall X_i^c \in X^c} (x_i^c - \text{proj}(X_i^c))^2 \quad (2.1.7)$$

Onde proj representa a equação (2.1.1).

Visando melhorar o resultado da otimização, decompomos os valores encontrados e utilizamos os parâmetros reais na otimização, resultando numa diminuição da dimensionalidade dos dados. Os parâmetros utilizados na minimização foram:

- f_x, f_y : distâncias focais em x e y , obtidas a partir de K_{11} e K_{22} .
- c_x, c_y : centros de projeção da imagem em x e y , obtidas a partir de K_{13} e K_{23} .
- $\|r\|, \theta, \varphi$: Utilizando a fórmula de Rodrigues [2], obtemos o vetor de Gibbs, representando a matriz de rotação. Utilizamos θ e φ para representar o versor e $\|r\|$ para representar a quantidade de rotação envolvida.
- T : O vetor de translação T , estimado pela etapa anterior.

- k, p, s : Parâmetros que controlam a distorção. A estimativa inicial pode ser definida pelo usuário ou assumida nula para câmeras com pouca distorção.

Utilizando estas transformações, nosso espaço de busca possuirá 16 dimensões, restritas a variações fisicamente possíveis no modelo considerado.

2.1.4 Alguns resultados experimentais

Nesta seção apresentamos alguns resultados utilizando o método considerado para calibração de câmeras, em duas situações distintas:

- Coleta em laboratório: Apresentamos os resultados de calibração para uma sequência capturada no Laboratório de Instrumentação para Biomecânica - FEF, que possui prumos de calibração, com pontos de calibração a cada 10 centímetros, como ilustrado na figura 2.2(a).
- Coleta em campo utilizando um sistema de pontos como referência: Apresentamos os resultados de calibração para uma sequência filmada no Ginásio Multi-disciplinar da Unicamp, onde efetuamos a calibração filmando um poste de calibração com 6 pontos de calibração, colocado em locais de coordenadas conhecidas na quadra, como ilustrado na figura 2.2(b).

Em todos os testes apresentados neste trabalho, utilizamos câmeras Basler A602fc.



(a) Laboratório.



(b) Coleta controlada.

Figura 2.2: Exemplos de imagens para calibração.

No caso da coleta em laboratório, temos a situação ideal para calibração. Muitos pontos de calibração, medidos com precisão, ocupando um bom volume na cena considerada



Figura 2.3: Exemplo de imagem de calibração coletada em laboratório.



Figura 2.4: Coleta controlada.

e nas imagens obtidas. Um exemplo de imagem para avaliação desta calibração pode ser visto na figura 2.1.4.

No próximo caso, tivemos acesso completo às dependências do ginásio e a quadra. Com isso, colocamos o bastão de calibração em vários pontos medidos previamente da quadra, formando um único sistema de coordenadas, utilizado para a calibração. Um exemplo de imagem para avaliação da calibração neste caso pode ser visto na figura 2.4.

2.1.5 Conclusão

Nesta seção apresentamos a metodologia adotada para a calibração de câmeras considerada neste trabalho, permitindo o uso de câmeras com significativa distorção radial. Como esta técnica, assim como os programas desenvolvidos, são úteis para outros trabalhos de

envolvidos no Laboratório de Instrumentação para a Biomecânica, criamos um projeto de software livre, batizado CaLIB, para garantir a manutenção e reuso destas rotinas. O projeto está atualmente hospedado no servidor de código do Google, disponível em <http://calib.googlecode.com>.

2.2 Modelagem do fundo e identificação dos objetos em movimento

Uma etapa de pré-processamento bastante comum na literatura pesquisada é a identificação do fundo da cena e dos objetos em movimento [16, 14, 29, 28, 27, 36]. Esta etapa também é considerada neste trabalho, e as informações geradas são diretamente utilizadas na reconstrução tridimensional, como valores probabilísticos.

É importante ressaltar que o método aqui considerado foi projetado para possuir poucos falsos negativos, ao custo de muitos falsos positivos, tais como sombras, objetos pertencentes ao fundo, reflexos na iluminação, etc. Este ruído resultante é, em grande parte, eliminado no processo de reconstrução tridimensional.

2.2.1 Modelo considerado

Nas filmagens consideradas, as cameras são estáticas e o fundo da cena é bastante estável. Assumindo esta relativa estabilidade como um pré-requisito, optamos por um modelo probabilístico, modelando cada pixel do fundo como uma gaussiana e utilizando um histograma para representar os objetos não pertencentes ao fundo. Este histograma se faz necessário pelo método que escolhemos para a identificação do objetos em movimento, ou seja, dos objetos não pertencentes ao fundo.

2.2.2 Aprendizado das informações dos modelos de fundo e objetos

Para diminuir o tempo de processamento, procedemos com uma amostragem temporal, considerando somente quadros igualmente separados de acordo com um intervalo pré-definido, gerando um subconjunto Q de quadros do vídeo original. Também utilizamos a moda deste subconjunto como primeira aproximação para o valor esperado das gaussianas que representam o fundo.

Precisamos efetuar uma classificação inicial dos pixels de cada quadro entre as duas categorias consideradas: *fundo* e *não-fundo*. Essa estimativa é dada por:

$$D = \|Q_i - Q_j\| > \epsilon \quad (2.2.1)$$

Onde Q_i e Q_j representam dois quadros distintos pertencentes ao conjunto Q , ϵ é um limiar e a operação de módulo inclui também o cálculo da norma no caso de imagens

com mais de um canal. A matriz binária D gerada por esta operação indica os pixels que devem ser considerados como não fundo.

Utilizando este método, conseguimos gerar um modelo bastante preciso do fundo utilizando somente 0.3% dos quadros da sequência, ou seja, 100 quadros para uma sequência de 34000 quadros, representando aproximadamente 39 minutos de filmagem a 15 quadros por segundo.

O parâmetro ϵ , nestes testes variando entre 5 e 15, representa a expectativa da imutabilidade do fundo, inclusive flexibilizando este conceito, permitindo ligeiras modificações, proporcionais ao limiar. Adicionalmente, o efeito deste limiar pode ser minimizado utilizando-se mais quadros para o cálculo do modelo.

2.2.3 Identificação dos objetos em movimento

Após computado o modelo de fundo para o vídeo considerado, procedemos com a identificação dos objetos em movimento no quadro, em que utilizamos uma abordagem baseada em probabilidades condicionais representados pelo teorema de Bayes. Para tanto, consideramos os seguintes eventos, definidos para cada pixel individualmente:

- C_i : Ocorrência da faixa de cores i no pixel em questão. Esta faixa é definida pelo histograma utilizado.
- F : O pixel em questão pertence ao fundo.

As gaussianas do modelo calculado aproximam a distribuição esperada de cores $P(C_i/F)$ para pixels pertencentes ao fundo. O histograma aproxima a distribuição cromática dos objetos em movimento, ou seja, $P(C_i/\neg F)$. Nosso objetivo, neste momento, é obter a probabilidade do pixel pertencer ao fundo, dada sua informação cromática, ou seja $P(F/C_i)$. Podemos expressar a relação entre estes dois valores através de uma das formas do teorema de Bayes:

$$P(F/C_i) = \frac{P(C_i/F)P(F)}{P(C_i/F)P(F) + P(C_i/\neg F)P(\neg F)} \quad (2.2.2)$$

$$P(F/C_i) = \frac{P(C_i/F)P(F)}{P(C_i/F)P(F) + P(C_i/\neg F)(1 - P(F))} \quad (2.2.3)$$

Nestas equações, $P(F)$ representa a probabilidade de um determinado pixel pertencer ao fundo e é estimada através do mesmo limiar ϵ utilizado na criação do modelo 2.2.3,

sendo o número de pixels que estão abaixo do limiar dividido pelo número total de pixels da cena considerada. A probabilidade condicional $P(C_i/\neg F)$ representa a distribuição cromática dos pixels que não pertencem ao fundo, sendo modelados, juntamente com as gaussianas, através de um histograma calculado considerando os pixels cuja diferença é maior que o limiar.

A probabilidade de um determinado pixel *não pertencer* ao fundo, dado a sua cor atual, é :

$$P(\neg F/C_i) = 1 - P(F/C_i) \quad (2.2.4)$$

2.2.4 Alguns resultados experimentais

Consideremos a figura 2.5(a) que ilustra um quadro do vídeo capturado no campeonato Pan Americano de Handebol, com a distorção radial corrigida através de um procedimento manual, considerando somente o primeiro parâmetro de distorção de grande angular. Podemos observar que as cores presentes na imagem estão escurecidas, um fator que pode comprometer a discernibilidade do fundo na cena considerada. Portanto, aplicamos um ajuste histogrâmico, em cada canal individualmente. O resultado correspondente é ilustrado na figura 2.5(b).



(a) Imagem original.



(b) Após o ajuste.

Figura 2.5: Ajuste histogrâmico.

Utilizando o vídeo com a imagem ajustada, aplicamos o método apresentado nesta seção, com os seguintes parâmetros: $\epsilon = 5$, $q_n = 1$, 128 bins e 10 quadros. A figura 2.6(a)

ilustra a imagem original, a imagem que contém a média das gaussianas está ilustrada na figura 2.6(b) e a correspondente identificação de objetos não pertencentes ao fundo, utilizando este modelo, é ilustrada na figura 2.6(c). A título de comparação, a figura 2.6(d) contém o resultado, não binarizado, do método tradicional de subtração do fundo.



(a) Imagem original.



(b) Média das gaussianas.



(c) Resultado da identificação de objetos.



(d) Resultado da subtração de fundo tradicional.

Figura 2.6: Modelo do fundo e resultado correspondente, considerando 10 quadros e $\epsilon = 5$.

Para ilustrar o efeito do parâmetro de limiarização, consideramos agora $\epsilon = 10$. A figura 2.7(a) ilustra a imagem original, a imagem que contém a média das gaussianas está ilustrada na figura 2.7(b) e a correspondente identificação de objetos não pertencentes ao fundo, utilizando este modelo, é ilustrada na figura 2.7(c). A título de comparação, a

figura 2.7(d) contém o resultado, não binarizado, do método tradicional de subtração do fundo.



(a) Imagem original.



(b) Média das gaussianas.



(c) Resultado da identificação de objetos.



(d) Resultado da subtração de fundo tradicional.

Figura 2.7: Modelo do fundo e resultado correspondente, considerando 10 quadros e $\epsilon = 10$.

Como esperado, ao incrementar ϵ , permitimos uma maior mutabilidade para o modelo de fundo, deixando-o mais abrangente e menos sensível a ruídos de baixa intensidade. Apesar do pequeno número de quadros utilizados, o resultado da identificação dos objetos já é suficientemente melhor que a abordagem tradicional. Mantendo $\epsilon = 10$, aumentamos o número de quadros utilizados como amostras para 30, com os resultados ilustrados na

figura 2.8. A figura 2.8(a) ilustra a imagem original, a imagem que contém a média das gaussianas está ilustrada na figura 2.8(b) e a correspondente identificação de objetos não pertencentes ao fundo, utilizando este modelo, é ilustrada na figura 2.8(c). A título de comparação, a figura 2.8(d) contém o resultado, não binarizado, do método tradicional de subtração do fundo.



(a) Imagem original.



(b) Média das gaussianas.



(c) Resultado da identificação de objetos.



(d) Resultado da subtração de fundo tradicional.

Figura 2.8: Modelo do fundo e resultado correspondente, considerando 30 quadros e $\epsilon = 10$.

Como ilustra a figura 2.8, ao aumentarmos o número de quadros considerados na amostra, o método foi capaz de avaliar melhor os contornos dos objetos e eliminar alguns

elementos do fundo da resposta anterior, ilustrada na figura 2.7.

Observamos que algumas partes do fundo, notoriamente nas partes mais claras como o logotipo no centro da quadra, são identificados como objetos em movimento. Isso se deve à semelhança entre as cores do uniforme do time argentino e as cores destas áreas. As regiões claras nos uniformes causam um alto valor na bin correspondente do histograma que modela os objetos não pertencentes ao fundo. Com isto, o denominador da equação 2.2.3 tende a possuir um valor alto, diminuindo a probabilidade do pixel em questão pertencer ao fundo, levando a um valor alto na equação 2.2.4.

Para efeitos de comparação, aplicamos o método apresentado em [13] no vídeo considerado. A figura 2.9(a) ilustra o quadro original, o fundo obtido é ilustrado na figura 2.9(b), o resultado da subtração tradicional de fundo, utilizando este modelo é ilustrado na figura 2.9(d) e o resultado final do método é ilustrado na figura 2.9(c). Para a geração destes resultados, utilizamos 15 quadros, igualmente espaçados a cada 15 quadros, tolerância de cor igual a 5 e tolerância de comparação igual a 20. O limiar utilizado foi 10, seguido de uma dilatação com elemento estruturante planar, de dimensões 3x3.

2.2.5 Conclusão

Nesta seção apresentamos o método proposto para a identificação dos objetos em movimento. Esta técnica é considerada pois as imagens utilizadas são provenientes de câmeras estáticas e o fundo da cena é relativamente imutável. O método apresentado é mais sensível que o método apresentado em [13], ilustrado para fins de comparação na figura 2.9, levando o resultado a ser menos segmentado e também levando a falsos positivos. Vale ressaltar que o resultado do nosso método possui um significado probabilístico e o valor do limiar envolvido não é crucial para o desempenho do método.



(a) Imagem original.



(b) Modelo de fundo considerado.



(c) Resultado final do método proposto em [13].



(d) Resultado da subtração de fundo tradicional.

Figura 2.9: Modelo do fundo e resultado correspondente, utilizando o método apresentado em [13].

2.3 Reconstrução Tridimensional

Existem na literatura vários métodos para a recuperação da informação tridimensional de uma cena. Podemos classificá-los entre duas categorias bastante abrangentes: os métodos que utilizam correspondências entre as diversas imagens para estimar a profundidade e os que realizam a reconstrução sem a necessidade desta informação. Neste trabalho, consideramos a segunda categoria, pois o cálculo das correspondências pode ser tão complexo quanto o rastreamento em si.

Nesta categoria, se destacam abordagens tais como *Space Carving* [10],[12],[11],[33],[25] e *Shape from silhouette*, também conhecido como *Visual hull* [43],[8]. A abordagem apresentada aqui se assemelha muito com a apresentada em [7], pois ambas consideram as informações através de uma metodologia probabilística. Nossa abordagem, no entanto, é mais simples e menos genérica, pois utiliza as informações resultantes da etapa de identificação de objetos e não as imagens originais diretamente.

É importante observar que estas abordagens utilizam normalmente *short baseline*, ou seja, câmeras próximas entre si, como nos casos mostrados de fachadas de prédios [33] ou muitos pontos de vista diferentes, circundando o objeto em questão, normalmente único [12], casos diferentes do contexto abordado neste trabalho. Algumas aplicações consideram mais objetos na cena, visando por exemplo, contagem de pessoas [43], mas os problemas decorrentes desta abordagem não foram explorados em um contexto de rastreamento.

O principal problema encontrado nesta categoria de métodos é o surgimento de falsos volumes, também presente na abordagem que descrevemos a seguir, ilustrado na figura 2.17(b) e abordado mais detalhadamente na seção 2.3.3. É importante ressaltar também que os vídeos considerados devem ser sincronizados de tal modo que todas as câmeras capturem todos os quadros ao mesmo tempo.

2.3.1 Método proposto para a reconstrução tridimensional

Consideramos a cena atual como um conjunto de voxels de mesmo tamanho. Associamos a cada voxel uma probabilidade de existência de um objeto qualquer, calculada com base nas imagens resultantes da etapa de identificação de objetos em movimento, descrita na seção 2.2, considerando as informações de calibração das câmeras, descrita na seção 2.1, de acordo com a equação seguinte:

$$P(V_i) = \prod_{c=1}^{N_c} \frac{1}{N_p} \sum_{x_i \in \text{proj}(V_i, c)} P(x_i^c) \quad (2.3.1)$$

Onde V_i representa o voxel em questão, N_c o número de câmeras que observam o voxel,

N_p o número de pixels presentes na projeção de V_i na câmera c , também representada por $proj(V_i, c)$ e x_i^c representa um pixel pertencente à projeção, na imagem resultante da identificação dos objetos na câmera c .

A equação 2.3.1 representa uma generalização probabilística da idéia intuitiva de que, para um voxel existir, todas as câmeras devem indicar objetos nas projeções equivalentes. Consideramos também, pela simplicidade computacional, a operação de mínimo entre os valores envolvidos, em substituição ao produtório. Como o mínimo, neste contexto, é sempre menor ou igual ao produtório, esta abordagem é mais restritiva.

Outra restrição que pode melhorar os resultados é considerar um número mínimo de câmeras observando cada voxel. Esta medida é especialmente útil no caso de algumas câmeras estarem próximas entre si, provocando uma grande superposição em seus cones de reprojeção. Adotamos esta restrição, usualmente com um número mínimo de duas câmeras, podendo ser aumentado no caso de uma boa cobertura.

Adicionalmente, podemos definir uma região de interesse para a reconstrução, apoiado na premissa de pouco movimento entre quadros consecutivos, reduzindo o custo computacional do método. Neste trabalho, definimos este volume de interesse baseado no resultado da reconstrução do quadro anterior, após uma dilatação morfológica com elemento estruturante de tamanho definido a partir da velocidade máxima dos objetos no contexto considerado, uma informação de fácil obtenção.

2.3.2 Alguns resultados experimentais

Nesta seção, apresentamos alguns resultados da reconstrução tridimensional, primeiramente nas imagens da coleta em laboratório, seguidas da coleta realizada no Campeonato Pan Americano de Handebol.

As figuras 2.10, 2.11, 2.12 e 2.13 ilustram o instante considerado para o primeiro exemplo, bem como o resultado da identificação de objetos correspondente, utilizando os seguintes parâmetros: 100 quadros, $\epsilon = 10$, histograma de 128x128x128 bins e $q_n = 3$. O resultado ainda foi filtrado através de um fechamento seguido de abertura morfológicos, visando a remoção de pequenos ruídos e oscilações.

Uma maneira interessante de entender o funcionamento desta abordagem é apresentada na figura 2.14. Nesta figura, apresentamos a reprojeção da probabilidade de existência dos voxels para cada câmera separadamente, junto com o resultado final da reconstrução. A figura 2.14(a) ilustra os cones de reprojeção para as câmeras 1 e 2, a figura 2.14(b) ilustra os cones de reprojeção para as câmeras 3 e 4, a figura 2.14(c) ilustra todos os cones de reprojeção e a figura 2.14(d) ilustra o resultado da reconstrução, dado pela interseção de todos os cones de reprojeção. Cada voxel possui 10cm x 10cm x 10cm e o número mínimo de câmeras considerado para o resultado final é de 3 câmeras.

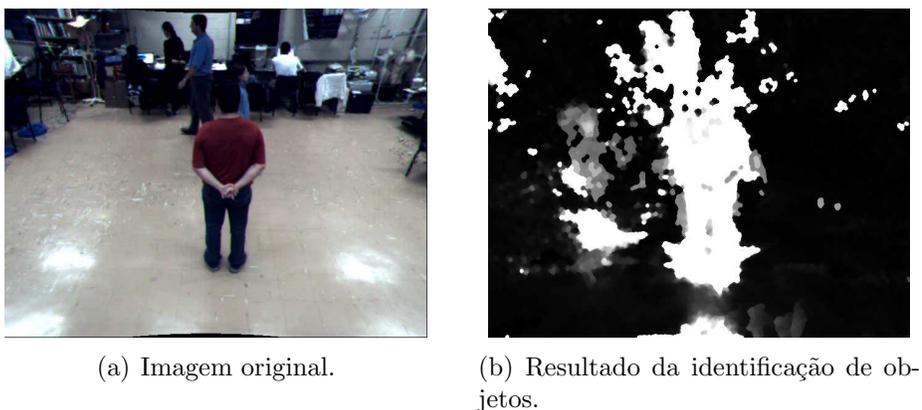


Figura 2.10: Resultado da identificação de fundo - câmera 1.

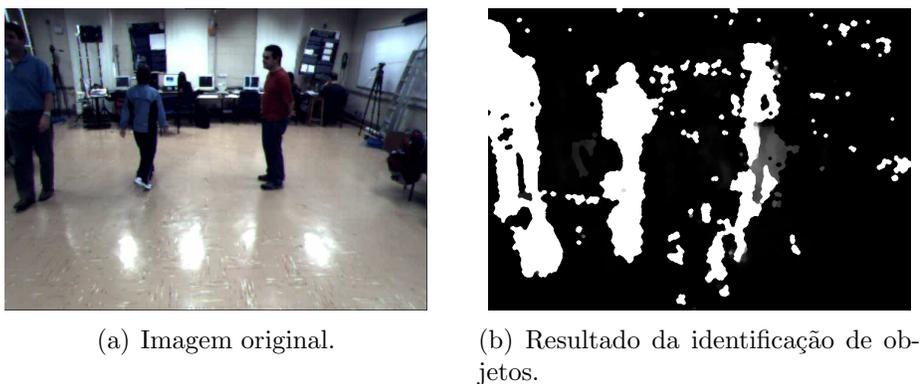


Figura 2.11: Resultado da identificação de fundo - câmera 2.

A figura 2.15 ilustra a reconstrução correspondente, comparando a reconstrução feita após uma seleção manual dos objetos em movimento (2.15(a)), em substituição ao método descrito na seção 2.2, e a reconstrução utilizando a identificação automática, considerando (figura 2.15(c)) ou não (figura 2.15(b)) a aplicação da região de interesse.

A figura 2.16 ilustra os quadros considerados e os respectivos resultados da identificação de objetos. A figura 2.17 ilustra a reconstrução correspondente, comparada ao resultado manual.



(a) Imagem original.



(b) Resultado da identificação de objetos.

Figura 2.12: Resultado da identificação de fundo - câmera 3.

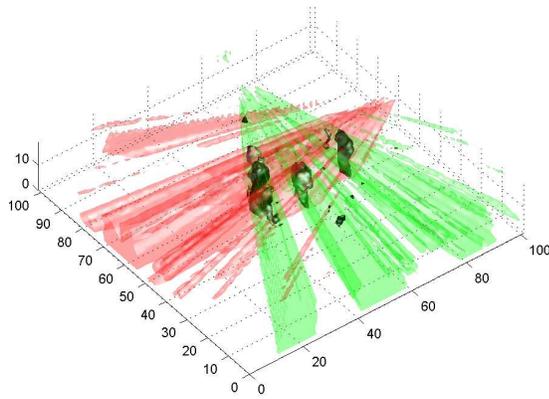


(a) Imagem original.

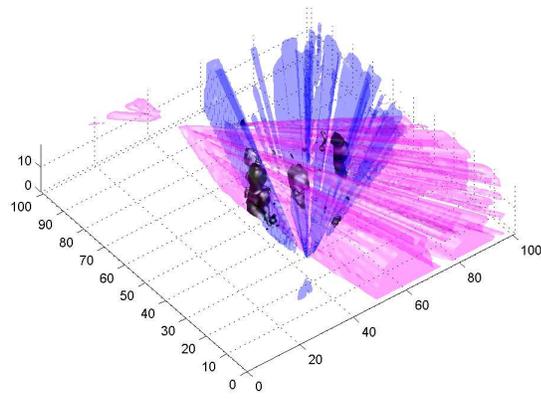


(b) Resultado da identificação de objetos.

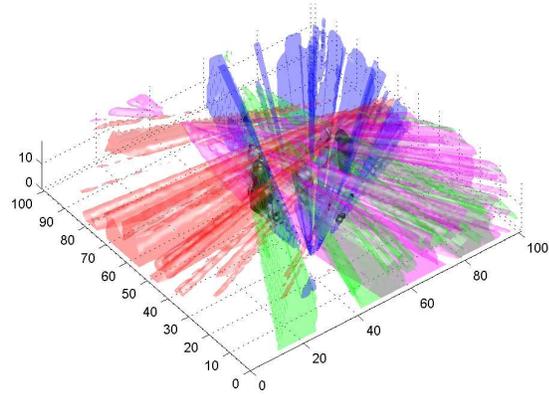
Figura 2.13: Resultado da identificação de fundo - câmera 4.



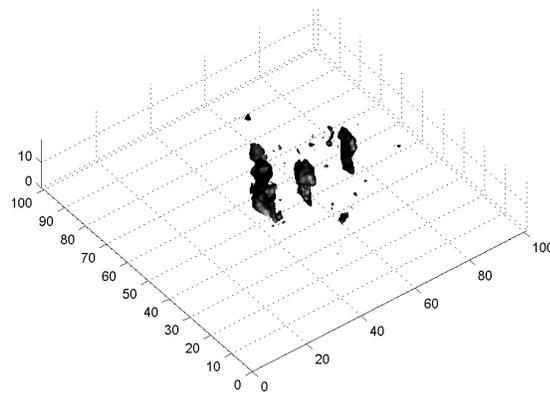
(a) Reprojeção das câmeras 1 e 2.



(b) Reprojeção das cameras 3 e 4.

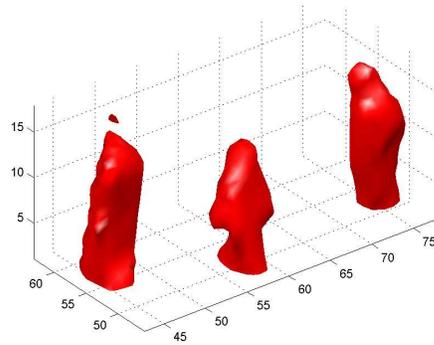


(c) Reprojeção das 4 câmeras.

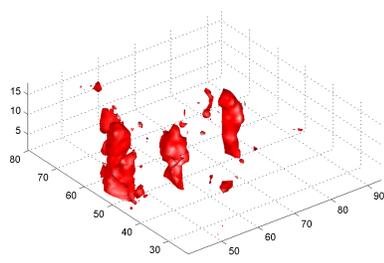


(d) Comparação entre o resultado da reconstrução utilizando a identificação automática ou manual dos objetos em movimento.

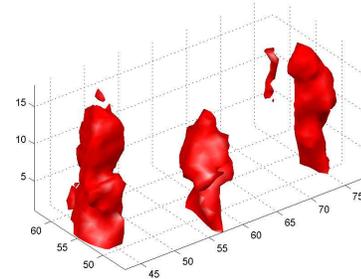
Figura 2.14: A reconstrução a partir da interseção de cones de reprojeção.



(a) Visualização da inicialização manual.



(b) Resultado da reconstrução sem região de interesse.



(c) Resultado da reconstrução com região de interesse.

Figura 2.15: Resultado da reconstrução.



(a) Imagem original. Câmera 1.



(b) Resultado da identificação de objetos. Câmera 1.



(c) Imagem original. Câmera 2.



(d) Resultado da identificação de objetos. Câmera 2.



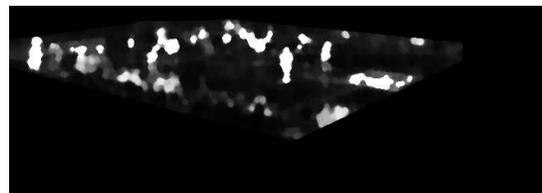
(e) Imagem original. Câmera 3.



(f) Resultado da identificação de objetos. Câmera 3.

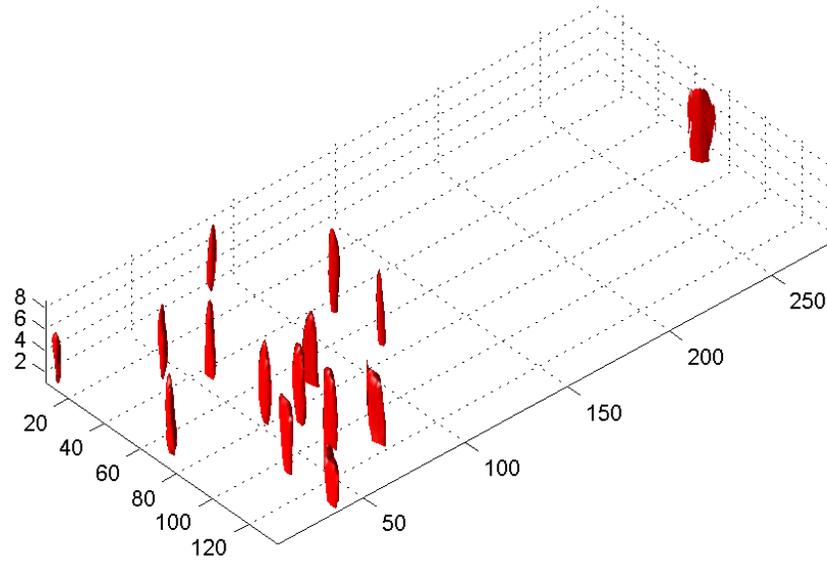


(g) Imagem original. Câmera 4.

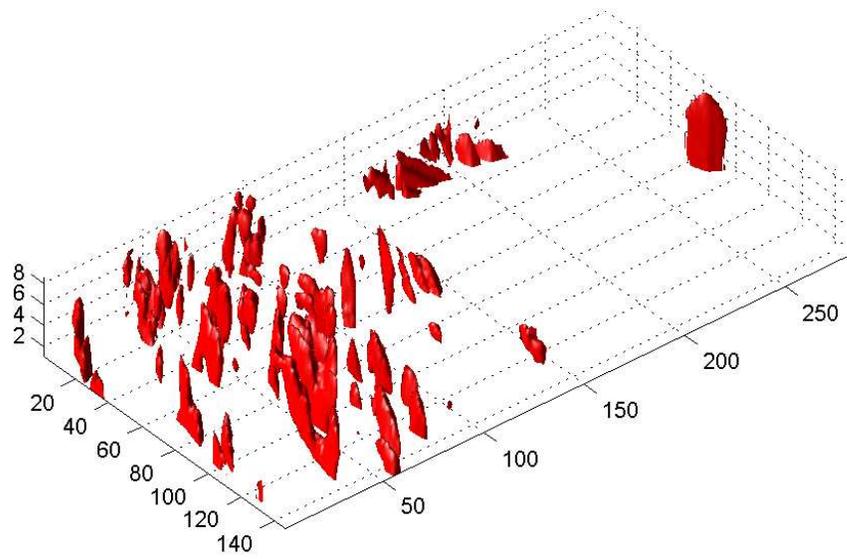


(h) Resultado da identificação de objetos. Câmera 4.

Figura 2.16: Resultado da identificação de fundo.



(a) Visualização da inicialização manual.



(b) Resultado da reconstrução sem região de interesse.

Figura 2.17: Resultado da reconstrução - Falsos volumes.

2.3.3 Falsos volumes na reconstrução

O método adotado para reconstrução pode ser interpretado como a interseção de cones de reprojeção, como ilustrado na figura 2.14. Esta abordagem, como vimos na figura 2.17, gera falsos volumes. Para entender a geração destas regiões espúrias, ilustramos o problema simplificado na figura 2.18.

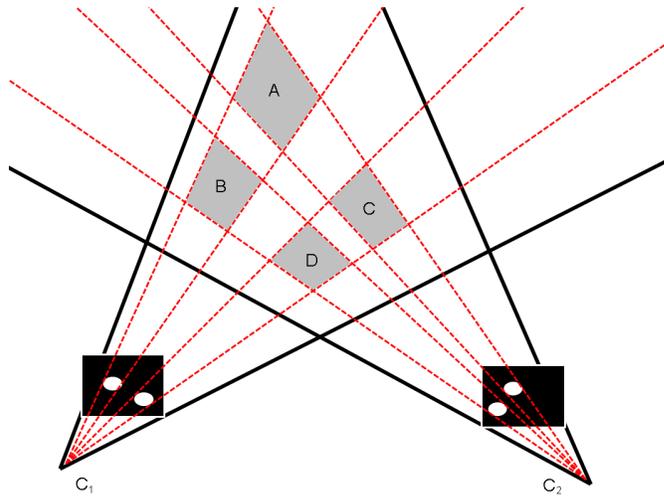


Figura 2.18: Geração de falsos volumes.

A figura 2.18 contém duas câmeras, C_1 e C_2 , com suas respectivas imagens, cada uma com duas regiões identificadas como objetos em movimento. Para simplificação do exemplo, consideramos estas imagens como imagens binárias. Cada região reprojeta um cone, com os limites delimitados pelas linhas tracejadas. Nesta configuração, são reconstruídos quatro objetos, nas regiões A, B, C e D.

Entretanto, isto não implica que existam quatro objetos na cena considerada. Na verdade, além de quatro objetos, um em cada região, todas as combinações possíveis contendo três objetos em cena seriam possíveis, assim como as configurações $((A, D), (B, C))$, contendo somente dois objetos. Este problema também é abordado em [43], sendo a razão da utilização dos limites inferiores e superiores de objetos contidos em cada volume reconstruído, além da continuidade temporal assumida, na qual os polígonos resultantes da reconstrução são considerados somente se a interseção entre o polígono atual e os polígonos presentes no instante anterior é não vazia.

Em uma primeira análise, podemos afirmar que ao aumentarmos o número de câmeras obtendo informações sobre a cena observada, as condições de geração dos falsos volumes seriam diminuídas, melhorando o resultado da reconstrução. Apesar de verdadeira, esta afirmação não aborda completamente o problema, pois não existem garantias que um determinado número de câmeras implique a inexistência de falsos volumes para uma

determinada cena. Adicionalmente, o posicionamento das câmeras também é um fator crucial para o resultado, pois afeta a quantidade de informação relevante que cada câmera fornece para a reconstrução.

Uma abordagem interessante para redução dos falsos volumes, em nosso contexto esportivo, é a elevação das câmeras, reduzindo o volume de cada cone de reprojeção, reduzindo com isto as possíveis interseções. Idealmente, a inclusão de uma câmera superior deve evitar a geração de falsos volumes, pois não existiriam falsas interseções possíveis.

Para ilustrar estes conceitos, criamos uma cena artificial contendo onze câmeras, dez câmeras laterais, divididas em grupos de acordo com seu posicionamento, e uma câmera superior. O posicionamento das câmeras pode ser visto na figura 2.19, com uma visualização tridimensional na figura 2.20 e a respectiva cobertura de cada câmera é ilustrado na figura 2.21, onde ilustramos somente as câmeras de uma das laterais, por simplificação. O posicionamento das câmeras na outra lateral é simétrico. A linha contínua representa os limites do volume de interesse, as linhas tracejadas simples representam a cobertura das câmeras pertencentes ao grupo 1 composto pelas câmeras de número 1, 2, 3 e 4, as linhas pontilhadas representam a cobertura das câmeras do grupo 2, composto pelas câmeras de número 5, 6, 9 e 10 e a linha ponto-traço representa a cobertura das câmeras do grupo 3, composto pelas câmeras de número 7 e 8.

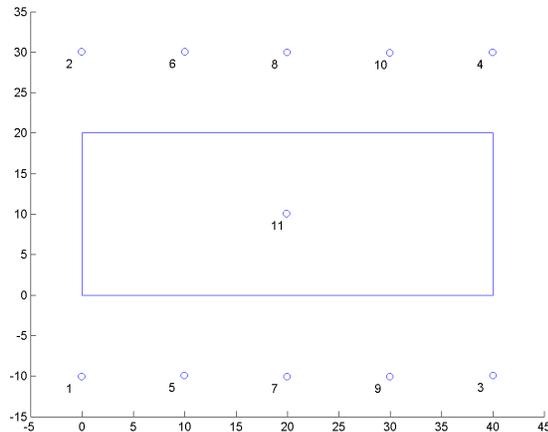


Figura 2.19: Diagrama de posicionamento de câmeras.

Utilizando progressivamente cada grupo de câmeras, tendo como base o grupo 1, por se tratar da configuração de câmeras mais comum na prática, podemos analisar o efeito da adição de mais câmeras na reconstrução da cena. A cena observada possui 40m por 20m e contém 14 cilindros de 1m de diâmetro e 1.80m de altura, representando uma configuração típica de um jogo de handebol.

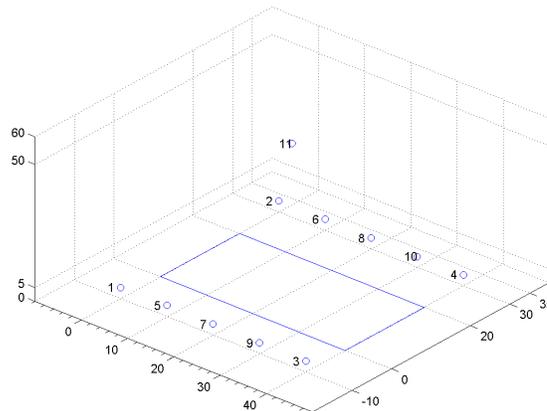


Figura 2.20: Diagrama tridimensional de posicionamento de câmeras.

A figura 2.22 ilustra a reconstrução da cena utilizando somente as câmeras do grupo 1, ou seja, quatro câmeras posicionadas na direção da linha de fundo em cada lado da quadra. Apesar do centro da quadra ser visível a todas as câmeras, existem regiões cobertas por somente duas câmeras. O número de falsos volumes é grande, principalmente considerando-se uma cena sintética, com nenhum ruído na etapa de identificação de objetos. O cilindro que representa o goleiro, localizado aproximadamente na posição (25,100) ilustra o efeito da reconstrução através de duas câmeras paralelas, mesmo em lados opostos da quadra.

Ao adicionarmos mais 4 câmeras, duas em cada lateral, aumentamos muito a informação disponível, diminuindo consideravelmente o número de falsos volumes, como ilustrado na figura 2.23. O resultado é aproximadamente o mesmo quando utilizamos também o último grupo de câmeras laterais, totalizando dez câmeras, cuja reconstrução resultante está ilustrada na figura 2.24.

Para avaliar o efeito da câmera superior, realizamos uma reconstrução utilizando somente as informações provenientes desta. Uma única câmera foi colocada, no centro da quadra, a 55 metros de altura com relação à quadra. Apesar da visualização da figura 2.25, os resultados são cones, cujo cume é o centro da câmera e se estendem infinitamente. Esta informação não é representada na visualização pois limitamos a altura considerada em 2 metros, a partir da quadra. Portanto, o que pode ser visto na figura 2.25 são cortes dos cones de reprojeção, no volume de interesse especificado. Podemos notar que, como previsto, não ocorreram falsos volumes nesta configuração, pela impossibilidade de interseção dos cones de reprojeção. Entretanto esta configuração não é capaz de fornecer informações tridimensionais completas sobre a cena observada, o que limita os resultados

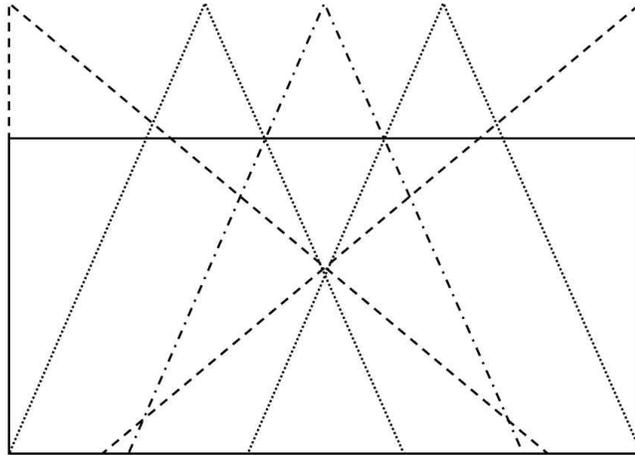


Figura 2.21: Diagrama de cobertura de câmeras.

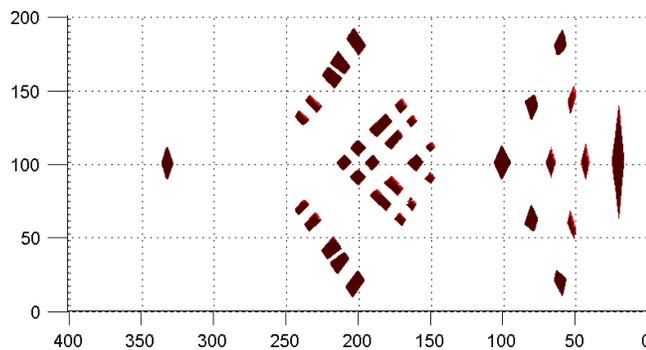


Figura 2.22: Resultado da reconstrução utilizando o grupo 1.

possíveis utilizando somente esta configuração.

Considerando que a utilização de dez câmeras implica uma logística de coleta complexa, potencialmente inviabilizando aplicações práticas, realizamos uma reconstrução utilizando somente o grupo 1, composto de quatro câmeras, e a câmera superior, representando uma configuração relativamente factível na prática. O resultado da reconstrução está ilustrado na figura 2.26.

A reconstrução resultante não possui os falsos volumes presentes na reconstrução utilizando somente o grupo 1, como ilustrado na figura 2.22, eliminados pela câmera superior e ainda é capaz de fornecer informações tridimensionais sobre a cena observada pois a cena é observada por, no mínimo, duas câmeras laterais distintas.

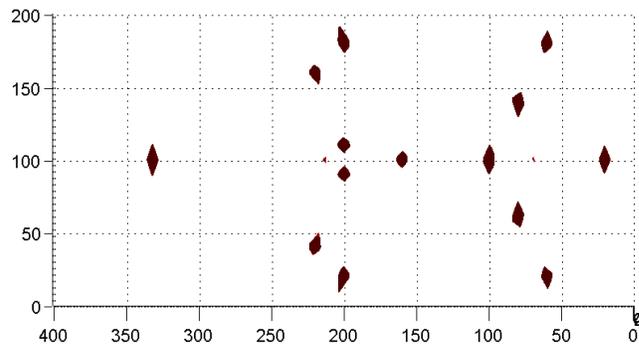


Figura 2.23: Resultado da reconstrução utilizando os grupos 1 e 2.

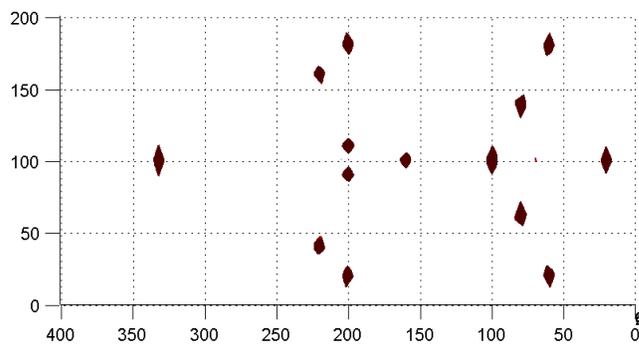


Figura 2.24: Resultado da reconstrução utilizando as dez câmeras laterais.

2.3.4 Conclusão

Nesta seção apresentamos o método considerado para reconstrução tridimensional de uma cena observada por múltiplas câmeras. Apesar de conceitualmente simples, este método é capaz de reconstruir satisfatoriamente uma cena observada mesmo com falsos positivos nas imagens de identificação de objetos, como visto na figura 2.15, desde que algumas condições sejam atendidas, visando diminuir a ocorrência de falsos volumes. Em nosso contexto, a utilização de uma ou mais câmeras superiores adicionais eliminaria a ocorrência de falsos volumes.

Infelizmente, devido a limitações logísticas e de equipamento não pudemos testar esta configuração para uma reconstrução com imagens reais. Entretanto, pelo desempenho na imagens reais processadas, podemos inferir que as cenas sintéticas utilizadas nesta seção são representativas e indicam o bom desempenho do método quando uma boa cobertura de câmeras é utilizada.

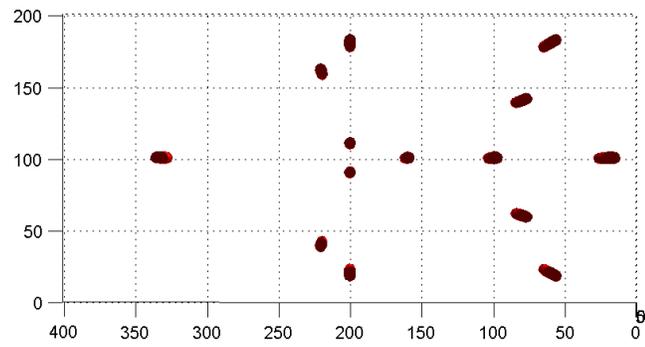


Figura 2.25: Resultado da reconstrução utilizando somente a câmera superior.

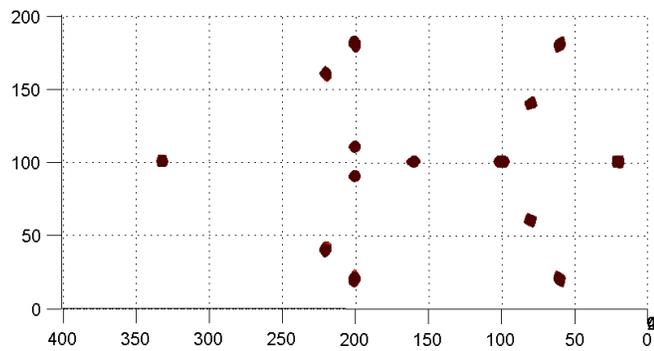


Figura 2.26: Resultado da reconstrução utilizando somente o grupo 1 e a câmera superior.

2.4 Amostragem da informação através da generalização do conceito de ritmo visual

Nas seções anteriores, apresentamos os passos considerados para realizar a reconstrução tridimensional da cena observada, que será utilizada como informação inicial para o algoritmo de rastreamento. Neste sentido, a abordagem que consideramos para o rastreamento é similar a alguns trabalhos presentes na literatura [16], [43]. Entretanto, optamos pela simplicidade, evitando a utilização de programação dinâmica ou aproximações, como apresentado em [16] ou mesmo *voting* [23]. A técnica de amostragem utilizada neste trabalho se assemelha com a técnica apresentada em [45], utilizada para a extração de informações dos resultados do rastreamento realizado previamente [46].

Nesta seção apresentamos uma das principais contribuições deste trabalho, a generalização do conceito de ritmo visual e sua aplicação para o rastreamento em um contexto tridimensional. Esta abordagem, além de simplificar a operação, reduz o volume de informação considerado e permite a visualização da informação proveniente do rastreamento de uma maneira mais rápida e significativa.

2.4.1 Ritmo visual

O conceito de ritmo visual é muito utilizado para identificação de cortes em imagens de vídeo [34], [30], [21], [18], [4], [5] e [39], com resultados bastante satisfatórios. De um modo genérico, esta abordagem consiste em tomar amostras da informação a ser analisada e posteriormente agrupá-las de uma maneira ordenada. No contexto de identificação de cortes em vídeos, que pode ser considerado como uma informação tridimensional, tomamos amostras unidimensionais, sejam colunas, linhas ou outro modelo, de cada quadro considerado e agrupamos estas amostras formando uma imagem bidimensional. A figura 2.27 ilustra esta abordagem. É importante ressaltar que a escolha do método de amostragem é o principal fator para o comportamento do método.

2.4.2 Generalizando o ritmo visual

A generalização do conceito de ritmo visual para o nosso contexto é bastante direta. Neste trabalho, nossa informação a ser processada é quadridimensional, composta pela evolução da cena reconstruída no tempo. A cada instante considerado, tomamos uma amostra bidimensional. Agrupamos estas amostras, obtendo uma informação tridimensional.

Genericamente, os dados manipulados pelo método podem ser definidos como:

$$D_n = \left\{ \bigcup f \mid f : \mathbb{N}^n \rightarrow \mathbb{R} \right\} \quad (2.4.1)$$

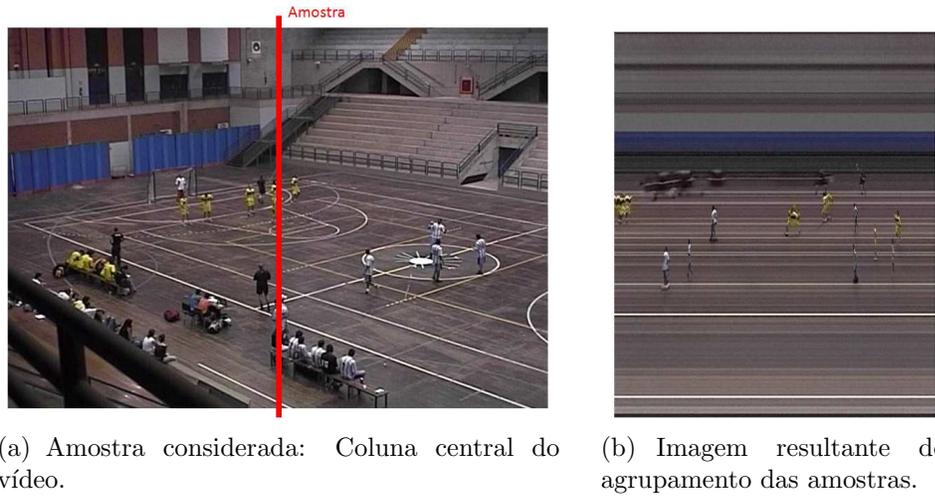


Figura 2.27: Exemplo da abordagem tradicional do ritmo visual.

Onde n indica o número de dimensões da informação considerada. Definimos duas operações básicas, amostragem e agrupamento:

$$S : D_i \rightarrow P(D_k) \tag{2.4.2}$$

$$A : P(D_k) \rightarrow D_j \tag{2.4.3}$$

Onde S indica a operação de amostragem, $P(D_k)$ indica o conjunto potência do conjunto D_k , A indica a operação de agrupamento e $i > j > k > 0$.

Neste contexto, para definir a operação de ritmo visual, necessitamos apenas definir duas operações que atendam às condições impostas nas equações 2.4.2 e 2.4.3. A operação de amostragem deve gerar um conjunto de amostras com dimensão menor em relação à informação original, a operação de agrupamento deve agrupar estas amostras gerando um conjunto de dimensão maior que o das amostras mas ainda menor do que a informação original.

Neste trabalho, escolhemos a função de amostragem ilustrada na equação 2.4.4, que representa um achatamento da dimensão Z , onde S_q representa a amostra correspondente ao instante q , T_q representa a cena tridimensional no instante q e X , Y e Z correspondem aos eixos. Essencialmente, a equação 2.4.4 cria uma imagem bidimensional contendo o valor máximo entre os possíveis voxels com mesmas coordenadas X e Y .

$$S_q(X, Y) = \max_{Z_i} (T_q(X, Y, Z)) \tag{2.4.4}$$

Esta função de amostragem tem como motivação o fato que, no handebol, os jogadores raramente se sobrepõem verticalmente. Em contextos nos quais esta hipótese não seja válida, outra função de amostragem deve ser considerada.

Como função de agrupamento, empilhamos as amostras obtidas, adicionando uma dimensão aos dados, explicitando a continuidade temporal intrínseca ao contexto.

2.4.3 Alguns resultados experimentais

Nesta seção apresentamos alguns resultados experimentais da aplicação da generalização do ritmo visual ao nosso contexto. A figura 2.28 ilustra o processo de amostragem para o quadro 100 para a coleta realizada em laboratório. A figura 2.29 ilustra o resultado do agrupamento das amostras, no intervalo entre os quadros 1 e 200. A figura 2.30 ilustra o resultado para a coleta real no campeonato Pan Americano, entre os quadros 1 e 50.

A figura 2.30 ilustra a geração de falsos volumes durante a reconstrução, numa situação desfavorável, tanto pelo posicionamento dos jogadores, próximos entre si, quanto pela grande área a ser coberta por poucas câmeras, com posicionamento muito distante do ideal. A coleta em laboratório, mais controlada, apresenta menor ocorrência destes falsos volumes, como pode ser visto na figura 2.29. Estas imagens ilustram o quanto o número de câmeras e os posicionamentos considerados são importantes para o bom desempenho do método.

Para tornar possível a análise num ambiente maior, impossível com a quantidade disponível de câmeras, criamos um vídeo sintético, simulando o contexto de um jogo de handebol. A título de simplificação, não consideramos aqui a etapa de identificação dos objetos em movimento. Nosso objetivo com este experimento sintético é analisar o comportamento do restante da metodologia proposta, assumindo uma boa cobertura das câmeras e um bom resultado na identificação de objetos. Os demais passos, tais como calibração, foram realizados exatamente como apresentado anteriormente.

Neste exemplo, consideramos 10 câmeras, posicionadas ao longo das laterais da quadra, em ambos os lados. Existem 14 objetos em cena, cilindros de mesmo tamanho, com as dimensões aproximadas de uma pessoa. Os voxels possuem 20cm x 20cm x 50cm. A figura 2.31 ilustra um quadro da câmera 1 e o resultado da reconstrução, incluindo um falso volume, próximo das coordenadas (40,40). O resultado do ritmo visual é ilustrado na figura 2.32.

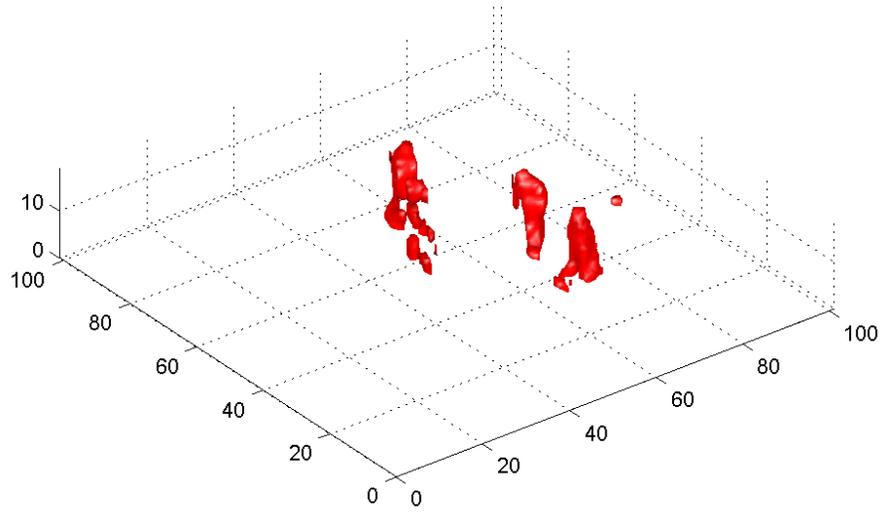
A informação resultante do ritmo visual, ilustrada na figura 2.32, engloba 200 quadros de informação tridimensional, apresentandos em uma única imagem. É importante ressaltar que a informação mostrada na figura corresponde a uma versão binarizada, para fins de visualização, dos dados considerados.

2.4.4 Conclusão

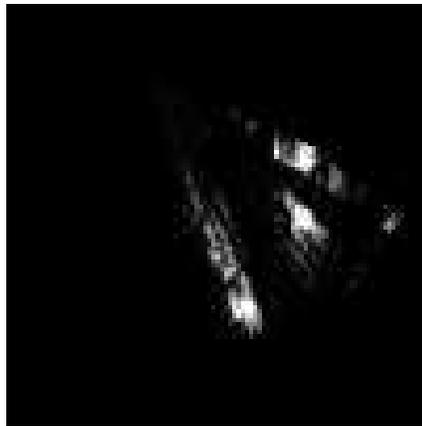
Nesta seção apresentamos uma generalização do conceito de ritmo visual, um recurso bastante explorado para detecção de cortes em vídeos. Através desta generalização, transformamos o problema de rastreamento temporal no domínio tridimensional em um problema de segmentação tridimensional. Indo além, as estruturas a serem segmentadas possuem formato e características específicas, que podem facilitar a segmentação.

Esta técnica ainda é interessante para visualização das informações resultantes do rastreamento. Como a informação de movimentação dos objetos em um intervalo de tempo é ilustrada como uma imagem tridimensional, a análise destas informações é facilitada.

Aplicando as idéias abordadas nesta seção, pode-se facilmente generalizar o conceito de ritmo visual para um número arbitrário de dimensões.



(a) Resultado da reconstrução - Quadro 100.



(b) Amostra correspondente.

Figura 2.28: Exemplo da operação de amostragem considerada.

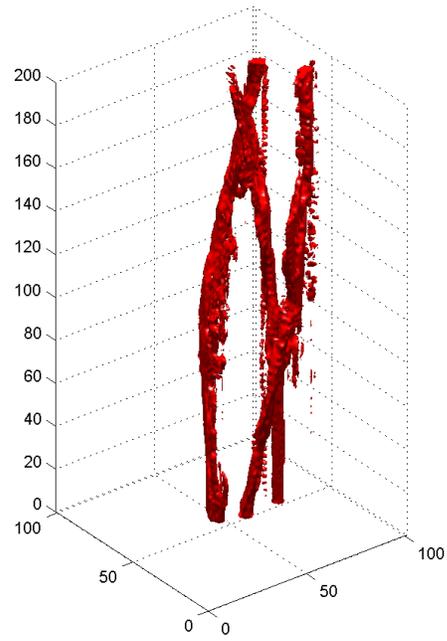


Figura 2.29: Resultado do ritmo visual para a coleta em laboratório - Quadros 1 a 300.

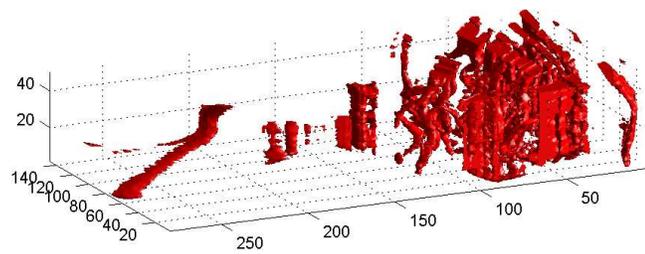
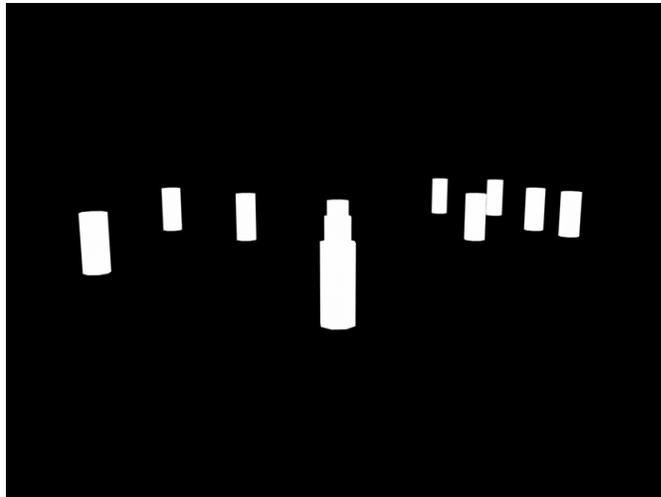
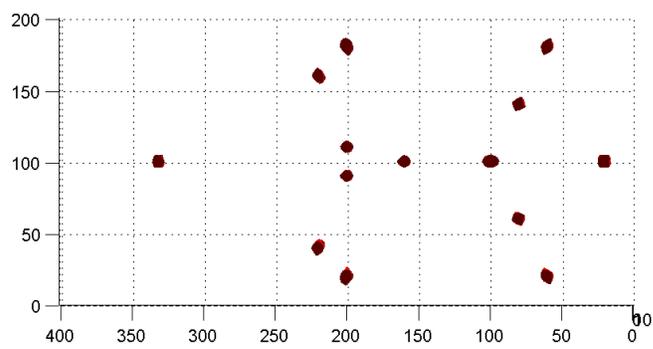


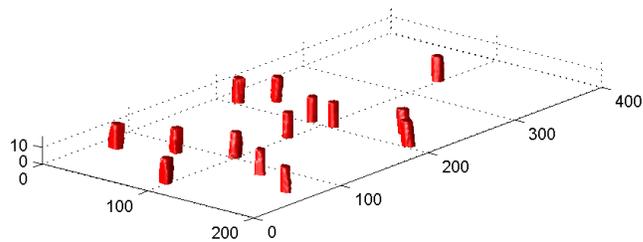
Figura 2.30: Resultado do ritmo visual para a coleta real - Quadros 1 a 50.



(a) Imagem original.



(b) Reconstrução correspondente.



(c) Reconstrução correspondente.

Figura 2.31: Reconstrução da sequência sintética - Quadro 1.



Figura 2.32: Resultado do ritmo visual para o vídeo sintético - Quadros 1 a 200.

2.5 Filtragem através da generalização do operador direcional multi-escala

Em [31] foi apresentado um método de filtragem específico para impressões digitais, baseado no operador direcional multi-escala. Este método infere a orientação de cada pixel da imagem considerada e a utiliza para completar pequenas falhas nas impressões digitais.

Em nosso trabalho, a informação resultante do ritmo visual é abstratamente semelhante a uma impressão digital. Procedemos então com a generalização do método apresentado para um ambiente tridimensional, deixando esta opção de filtragem disponível caso seja necessário.

2.5.1 Definição das direções a serem consideradas

No trabalho original [31], as direções consideradas são discretas e definidas por um ângulo em coordenadas polares. No domínio tridimensional, consideramos dois ângulos, θ e ϕ , em coordenadas esféricas, definindo um versor, com $0 \leq \theta < 180^\circ$ e $0 \leq \phi < 180^\circ$. Definimos então um passo δ , em graus, que será considerado em ambos os parâmetros, para definir os versores que serão utilizados no método. O parâmetro δ deve ser uma fração inteira do ângulo reto, de modo que outros versores considerados pertençam ao plano perpendicular de um dado versor.

Para cada versor considerado, associamos dois versores perpendiculares, tanto ao versor quanto entre si, que serão utilizados no cálculo das orientações dos voxels. Como existem infinitos versores que atendem a estas condições, optamos por versores que pertençam ao subconjunto considerado. Nesse subconjunto, ainda existem $180/\delta$ versores que atendem às condições impostas. Escolhemos então o primeiro versor com menor ϕ . Com isso, o terceiro versor fica restrito a uma única possibilidade. Um exemplo trivial destes versores pode ser visto na figura 2.33¹. O versor considerado em vermelho, os escolhidos em verde, as demais possibilidades em azul.

2.5.2 Estimativa da orientação dos voxels

Utilizando a inicialização manual do método proposto de rastreamento, obtemos facilmente o tamanho esperado dos objetos em movimento, estimando então um raio médio, r_m . O próximo passo para estimar a orientação dos voxels é computar o que foi chamado no trabalho original de *informação*, que consiste em aplicar uma determinada função aos voxels posicionados em uma dada direção (θ_i, ϕ_i) , distantes no máximo r_m voxels do voxel

¹Apresentamos vetores ao invés de versores para fins de visualização.

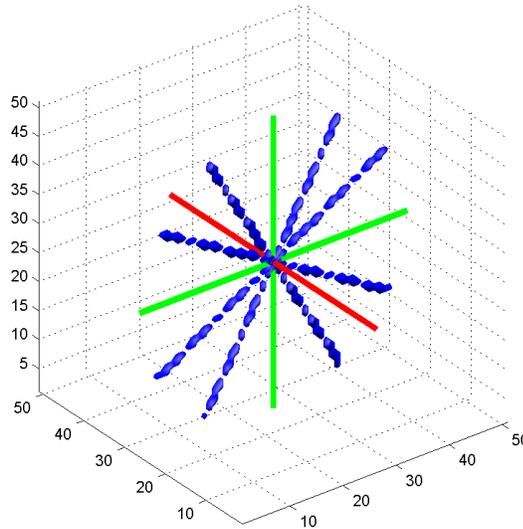


Figura 2.33: Exemplo dos versores perpendiculares considerados - $\delta = 30$ graus.

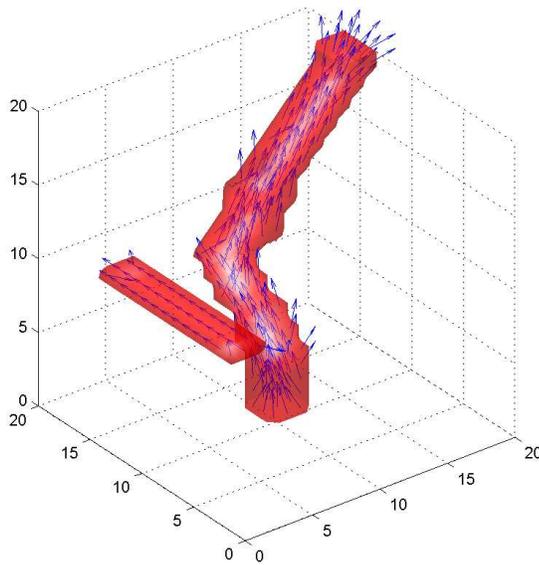
considerado e também nas direções perpendiculares escolhidas. Neste trabalho, consideramos esta função como a função soma. A equação 2.5.1 ilustra a equação considerada neste trabalho.

$$I_{\theta_i, \phi_i} = 2(O \otimes h_{\theta_i, \phi_i}) - (O \otimes h_{\theta_i, \phi_i}^{p1}) - (O \otimes h_{\theta_i, \phi_i}^{p2}) \quad (2.5.1)$$

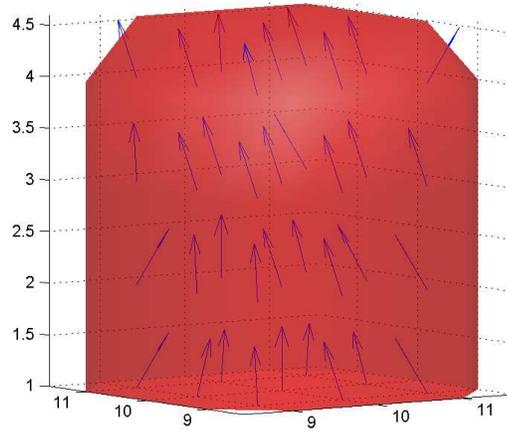
Onde I_{θ_i, ϕ_i} representa a informação na direção (θ_i, ϕ_i) , O representa a informação original, h_{θ_i, ϕ_i} representa uma máscara de convolução de tamanho $2r_m$, possuindo números 1 na direção (θ_i, ϕ_i) , $h_{\theta_i, \phi_i}^{p1}$ e $h_{\theta_i, \phi_i}^{p2}$ também representam máscaras idênticas, mas possuindo 1 apenas nas duas direções perpendiculares a (θ_i, ϕ_i) escolhidas previamente e \otimes representa a operação de convolução. Visualmente, estas máscaras podem ser representadas como ilustra a figura 2.33, com o versor vermelho representando h_{θ_i, ϕ_i} e os verdes representando $h_{\theta_i, \phi_i}^{p1}$ e $h_{\theta_i, \phi_i}^{p2}$.

Com a aplicação da equação 2.5.1, para todas as direções consideradas, possuiremos, para cada voxel, um valor de informação associado para cada possível direção. A orientação do voxel é definida como a direção com maior informação associada. Isto, pois, para obter um valor maior, a direção considerada deve possuir mais voxels com maior valor do que suas direções perpendiculares. A figura 2.34 ilustra um exemplo do resultado da abordagem considerada para um exemplo sintético, assim como uma visualização da instabilidade do método em torno dos limites da matriz considerada.

A instabilidade ilustrada na figura 2.34(b) é causada pela ausência de valores abaixo do voxel considerado, nos primeiros níveis do volume considerado. Uma possível solução, bastante simples, para esta questão é a replicação do primeiro e último níveis r_m vezes,



(a) Exemplo de resultado do cálculo das orientações - $\delta = 30$ graus.



(b) Visualização da região instável.

Figura 2.34: Exemplo da aplicação do método apresentado.

para que todos os voxels considerados possam ter sua orientação calculada sem a utilização de valores não presentes na matriz.

2.5.3 Filtragem multi-escala

Por se tratar de uma abordagem aplicada independentemente a cada voxel, este operador direcional é bastante sensível a ruídos. Para diminuir esta influência e conseguir estimar a orientação dos voxels mesmo em caso de pequenas falhas no volume considerado, aplicamos operações de moda nas orientações resultantes.

Assim como no trabalho original [31], utilizamos duas janelas de tamanhos diferentes, W_s e W_b ao utilizarmos esta operação. Estes parâmetros são definidos manualmente neste trabalho, mas podem ser inferidos a partir do volume considerado, com $W_b > W_s$. Os resultados destas operações são interpretados como as orientações através de diferentes escalas.

A orientação resultante da escala mais grosseira, W_b , é menos sensível a ruídos, mas também pode ocultar pequenas regiões de orientação diferente, presentes em resoluções mais finas, como W_s . Aplicamos então a mesma operação sugerida no trabalho original, ajustando o resultado da resolução mais fina com o auxílio da resolução mais grosseira.

Este ajuste consiste em substituir, na resolução mais fina, as orientações que se mostram perpendiculares às encontradas na escala mais grosseira. Esta operação se baseia no

princípio de que pequenas falhas, menores que W_b , não terão influência no resultado da escala maior mas deverão ser identificadas em uma escala menor, pois nesta escala a orientação identificada se posicionará na fronteira do volume, resultando em uma orientação perpendicular.

Procedemos, então, com uma operação morfológica de dilatação, utilizando para cada voxel um elemento estruturante baseado na orientação encontrada.

2.5.4 Alguns resultados experimentais

A figura 2.35 ilustra, em um exemplo sintético, a operação de filtragem descrita na seção 2.5.3. O número considerado de dilatações deve ser proporcional ao tamanho da falha a ser reconstruída.

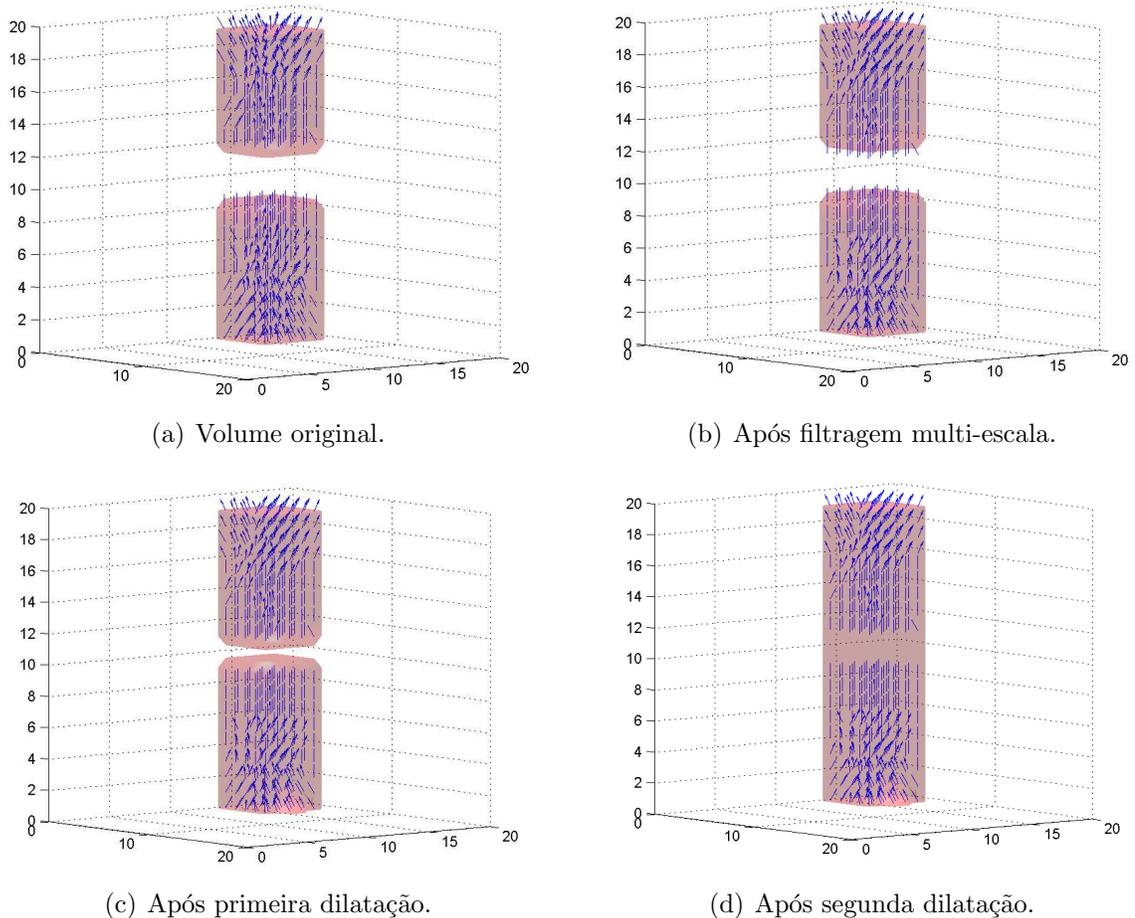


Figura 2.35: Exemplo do efeito do operador para fechamento de falhas.

2.5.5 Conclusão

Nesta seção apresentamos a generalização do operador direcional multi-escala, aplicando este método para completar pequenas falhas no resultado da amostragem através do ritmo visual generalizado. Consideramos também sua utilização para obter uma estimativa robusta da orientação de movimento dos objetos. Este método pode ser simplesmente estendido a qualquer dimensão e aplicado sempre que a informação considerada possuir uma informação direcional significativa.

2.6 Identificação das trajetórias

Após a reconstrução, amostragem e filtragem da cena considerada, devemos efetuar o rastreamento propriamente dito, ou seja, separar as trajetórias de cada um dos objetos, no intervalo de tempo considerado.

Em nosso contexto, isto significa segmentar cada cilindro distorcido que representa a trajetória de um objeto. Entretanto, nossa informação considerada é probabilística, o que dificulta a definição de fronteiras e, conseqüentemente, a identificação dos objetos. Assumindo uma reconstrução tridimensional razoável, com poucos falsos volumes, o que depende fortemente do desempenho das etapas de identificação dos objetos em movimento e de calibração, adotamos uma abordagem simples para a identificação da trajetória de cada objeto em movimento.

Inicialmente binarizamos a informação considerada e adotamos procedimentos simples, utilizando um grafo, para identificar cada trajetória. No caso de ocorrência de falsos volumes que provoquem ambigüidades não solucionáveis com as informações aqui consideradas, a intervenção do usuário é requerida.

A utilização de grafos para o rastreamento é uma idéia bastante adotada na literatura pesquisada, seja para auxiliar a organização semântica da cena observada [1], ou efetuar o rastreamento em si [15], [26], [43]. A abordagem considerada neste trabalho é similar à presente em [3], porém simplificada.

2.6.1 Binarização

No caso de uma coleta com boas condições, uma simples limiarização deve ser suficiente para a identificação das regiões correspondentes aos objetos em movimento. Entretanto, desenvolvemos também um método alternativo baseado na informação do gradiente, aplicável quando a reconstrução oscilar entre quadros consecutivos. Neste caso, estamos interessados em regiões com uma diferença significativa com relação a sua vizinhança.

Inicialmente, computamos o módulo do gradiente do volume considerado, utilizando operadores de Sobel em três dimensões. Aplicamos uma limiarização neste resultado, com um limiar pequeno, proporcional à diferença desejada no gradiente, ou seja, à discrepância desejada da região sobre seus vizinhos. Finalmente, aplicamos um fechamento morfológico, utilizando um elemento estruturante de tamanho proporcional a r_m , o raio médio dos objetos em cena.

2.6.2 Construção do grafo

Consideramos um grafo orientado simples, com os vértices representando regiões no espaço da informação considerada e adicionamos arestas representando possíveis transições entre

estas regiões. Idealmente, após o processamento, existirão somente um vértice para cada objeto em cena. Adicionalmente, podemos processar separadamente cada componente conexo da informação considerada, para acelerar o processo.

Inicialmente, criamos um vértice para cada região bidimensional conexa, para todas as amostras consideradas, como ilustrado na figura 2.36. Criamos arestas entre os vértices vizinhos verticalmente, desconsiderando automaticamente regiões que não possuem arestas de entrada, pois representam o aparecimento repentino de falsos volumes, como ilustrado na figura 2.36(b), com relação ao volume original ilustrado na figura 2.36(a).

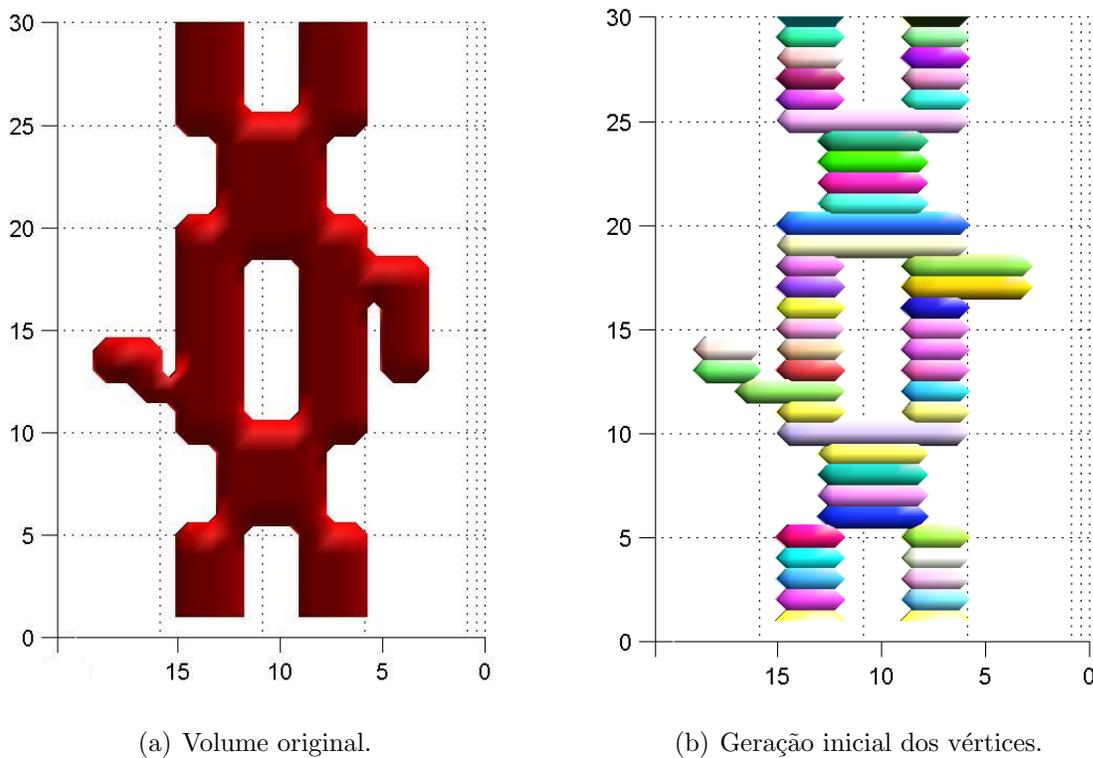


Figura 2.36: Geração inicial dos vértices para o exemplo considerado.

Em seguida agrupamos todos os vértices equivalentes de acordo com a seguinte regra:

$$V_i \equiv V_j \leftrightarrow (Adj^+(V_i) = (V_j)) \wedge (Adj^-(V_j) = (V_i)) \quad (2.6.1)$$

Onde V_i e V_j representam dois vértices distintos, $Adj^+(V_i)$ representa o conjunto de vértices apontados pelo vértice V_i e $Adj^-(V_j)$ representa o conjunto de vértices que apontam para o vértice V_j . Em outras palavras, dois vértices V_i e V_j são equivalentes se V_i

aponta somente para V_j e V_j só é apontado por V_i . Também removemos, repetindo a operação até a idempotência, vértices que não apontam para nenhum outro vértice, mas mantendo os vértices correspondentes ao final da trajetória.

Adicionamos a cada vértice uma sequência de propriedades, representando a probabilidade de um determinado objeto ser encontrado neste vértice, ou seja, a probabilidade da trajetória de cada determinado objeto conter a região correspondente ao vértice considerado. Considerando o exemplo da figura 2.36, a figura 2.37 ilustra o resultado da fusão de vértices equivalentes 2.37(a) e o grafo correspondente pode ser visto na figura 2.37(b).

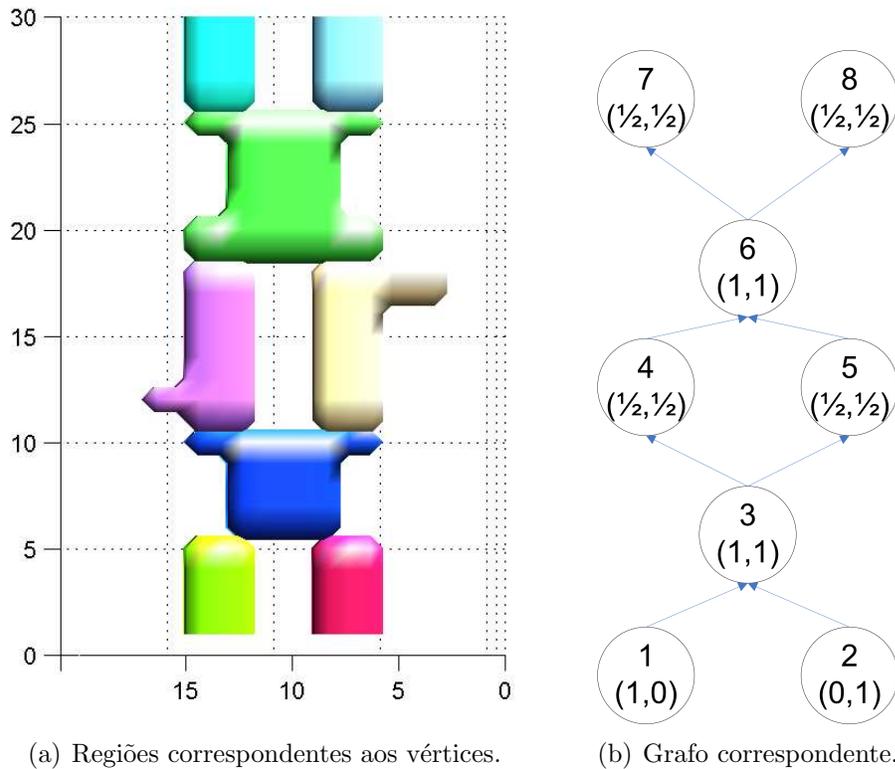


Figura 2.37: Vértices iniciais e o grafo correspondente.

No grafo da figura 2.37(b), a região correspondente ao vértice número 3 pertence, obrigatoriamente, a trajetória dos dois objetos presentes na cena. Precisamos, portanto, dividir esta região, de modo a separar a trajetória de cada um dos objetos. Sem uma análise mais fina da cena, não possuímos informação suficiente para tal operação. Entretanto, possuímos informações sobre a trajetória anterior e possíveis trajetórias posteriores dos objetos. Procedemos, então, com uma heurística, considerando todas as trajetórias possíveis e escolhendo a configuração de menor energia.

Para simplificar a operação, consideramos pontos ao invés de regiões. Cada aresta de entrada do vértice a ser separado gera um possível ponto de entrada para as trajetórias e cada aresta de saída gera um possível ponto de saída. A localização destes pontos é calculada como o centro da interseção entre o dilatado morfológico da região a ser dividida e a região do vértice de interesse.

Geramos então, todas as trajetórias possíveis, lembrando que podem existir mais opções de saída do que de entrada no vértice considerado, causado pela ocorrência de falsos volumes. Consideramos a trajetória entre um ponto de entrada I_i e um ponto de saída F_j como a spline cúbica gerada considerando estes pontos como nós e com derivada proporcional aos vetores de orientação, obtidos através do operador apresentado na seção 2.5. A figura 2.38 ilustra as duas possíveis configurações de trajetórias para dois pontos de entrada e dois pontos de saída. Na figura 2.38(a), mostramos a configuração 1, com as trajetórias representadas pelas linhas atravessando o sólido. A outra configuração possível é ilustrada na figura 2.38(b).

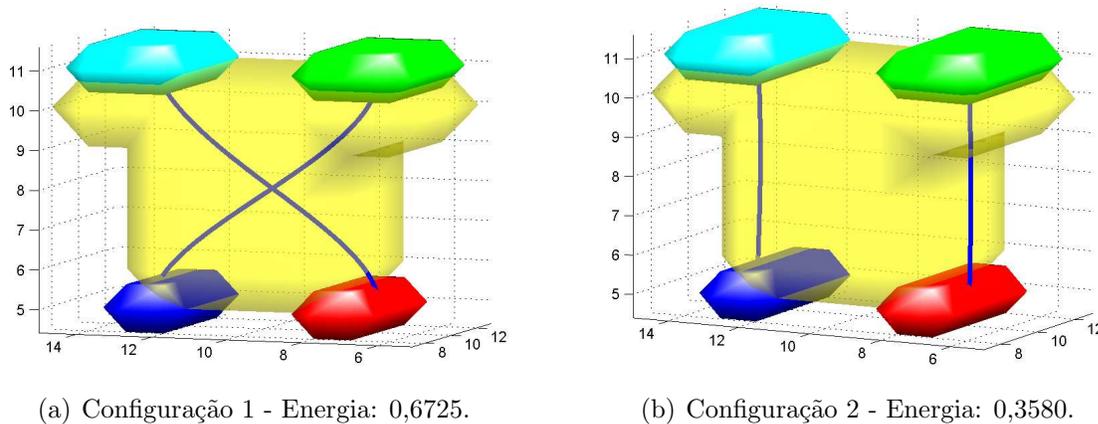


Figura 2.38: Possíveis configurações de trajetórias e suas energias de acordo com a equação 2.6.2.

Com todas as possíveis trajetórias devidamente geradas, devemos escolher a configuração que representa a menor energia. Em nosso caso, consideramos como menor energia a configuração que necessita do menor trabalho externo para ser realizada. Sendo o trabalho realizado proporcional à força realizada, que por sua vez é proporcional à aceleração, aproximamos este conceito como:

$$E(Conf_i) = \sum_{\forall P_i \in Conf_i} (\|a_i\|) \quad (2.6.2)$$

$$\vec{a}_i = \vec{v}_i - v_{i-1} \quad (2.6.3)$$

$$\vec{v}_i = \vec{P}_i - P_{i-1} \quad (2.6.4)$$

Onde $Conf_i$ representa uma determinada configuração de pontos de entrada e saída, P_i representa um ponto pertencente a trajetória de algum objeto, \vec{a} representa uma aproximação da aceleração, definida na equação 2.6.3 como a variação do vetor velocidade \vec{v}_i , cujo valor é dado pela equação 2.6.4 como o vetor entre dois pontos consecutivos da trajetória considerada.

A configuração de menor energia é escolhida, as trajetórias correspondentes são colocadas no volume, substituindo a região anterior, e o grafo é reconstruído. Seguindo o exemplo da figura 2.37, a figura 2.39 ilustra as regiões (2.39(a)) e o grafo correspondente (2.39(b)).

Repetindo estas operações até a idempotência, como explicitado pelo algoritmo 1, obtemos o resultado final. Neste caso, como não possuímos falsos volumes na cena, o método foi completamente automático. A figura 2.40 ilustra a informação original(2.40(a)) e o respectivo resultado final(2.40(b)).

Nos casos em que a existência de falsos volumes induza ambiguidades, como um vértice apontando para outros dois vértices que apontam para um único vértice, a intervenção do usuário é requerida. Mas, ao contrário da maioria dos métodos existentes, esta intervenção não solicita ao usuário que informe, usualmente clicando, a correta posição do objeto para os quadros não resolvidos.

Para a resolução destas ambiguidades, escolhemos um instante aproximadamente no meio do intervalo de quadros no qual está a ambiguidade, reprojetaamos, separadamente, uma aproximação do volume correspondente a cada caminho possível em todas as câmeras, e o usuário informa se as regiões indicadas correspondem ou não ao mesmo objeto. Com isso, o caminho correto é identificado e corretamente rastreado, independente do número de quadros envolvidos. Portanto, o número de intervenções do usuário é proporcional ao número de ambiguidades e não ao número de quadros presentes na ambiguidade.

Adicionalmente, pode-se adotar outros tipos de classificadores automáticos, considerando que a decisão a ser tomada é se as regiões fornecidas correspondem ou não ao mesmo objeto e, obrigatoriamente, uma das configurações é correta. Não implementamos esta abordagem neste trabalho, por considerarmos fora do escopo atual, caracterizando um desejável trabalho futuro. As figuras 2.41 e 2.42 ilustram as imagens mostradas ao usuário para auxiliar o processo de decisão, assumindo cada uma das possibilidades como

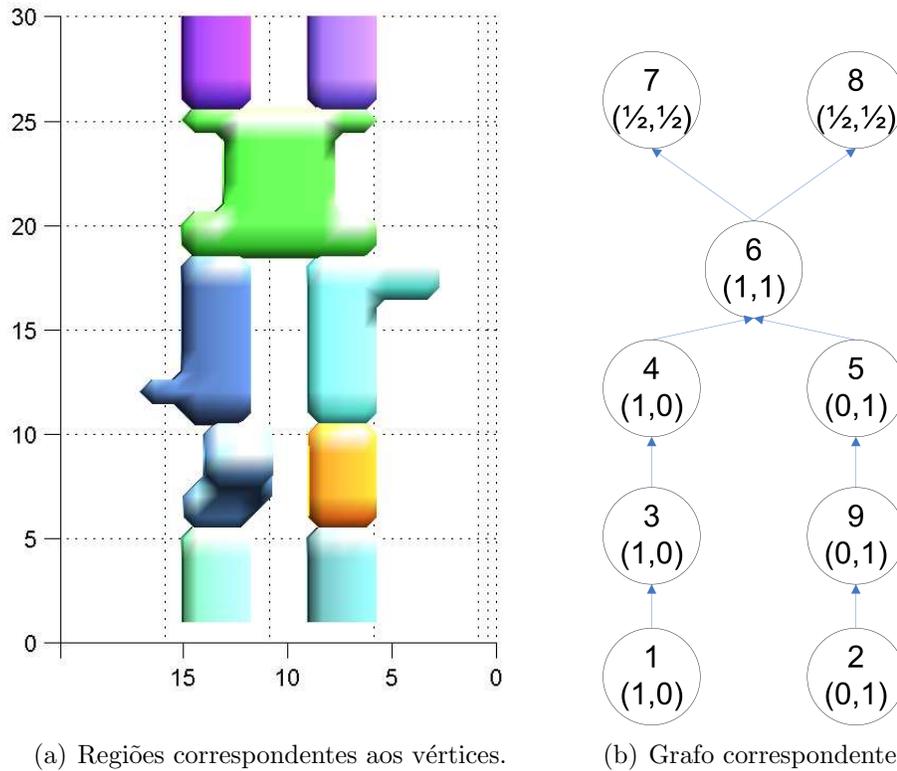


Figura 2.39: Após a separação de uma região.

correta. As linhas delimitadoras foram manualmente engrossadas para facilitar a visualização neste documento.

Através destas imagens, podemos claramente afirmar que as regiões delimitadas na figura 2.41 não correspondem ao mesmo objeto, ao contrário da imagem 2.42. Neste caso específico, um classificador automático baseado em informação cromática provavelmente seria suficiente para a resolução da ambiguidade.

2.6.3 Conclusão

Nesta seção apresentamos o método considerado para a identificação das trajetórias de cada objeto a partir da informação resultante do ritmo visual generalizado. Apesar de bastante simples e similar a diversos trabalhos presentes na literatura, pois grafos são uma ferramenta particularmente útil na resolução deste tipo de problema, obtemos resultados interessantes, não somente em relação ao rastreamento automático, cujo desempenho pode ser melhorado, mas também com relação ao tipo de intervenção requerida do usuário em

Entrada: Grafo construído a partir da amostragem através do ritmo visual.

Saída: Grafo contendo as trajetórias de cada objeto separadamente.

Repita

Realize a fusão de vértices equivalentes;

Remova os ramos parasitas;

Esvazie a fila;

Coloque todas as fontes na fila;

Repita

Remova o primeiro elemento para ser processado;

Coloque todos os vértices apontados por ele na fila;

se *Duas trajetórias utilizam o vértice* **então**

se *Os vértices anteriores são utilizados por somente uma trajetória cada*

então

| Divida o vértice;

fim

senão

| Resolva a ambigüidade entre os vértices anteriores e o atual;

fim

fim

até *efetuar uma mudança* ;

até *idempotência* ;

Algoritmo 1: Algoritmo para definição das trajetórias.

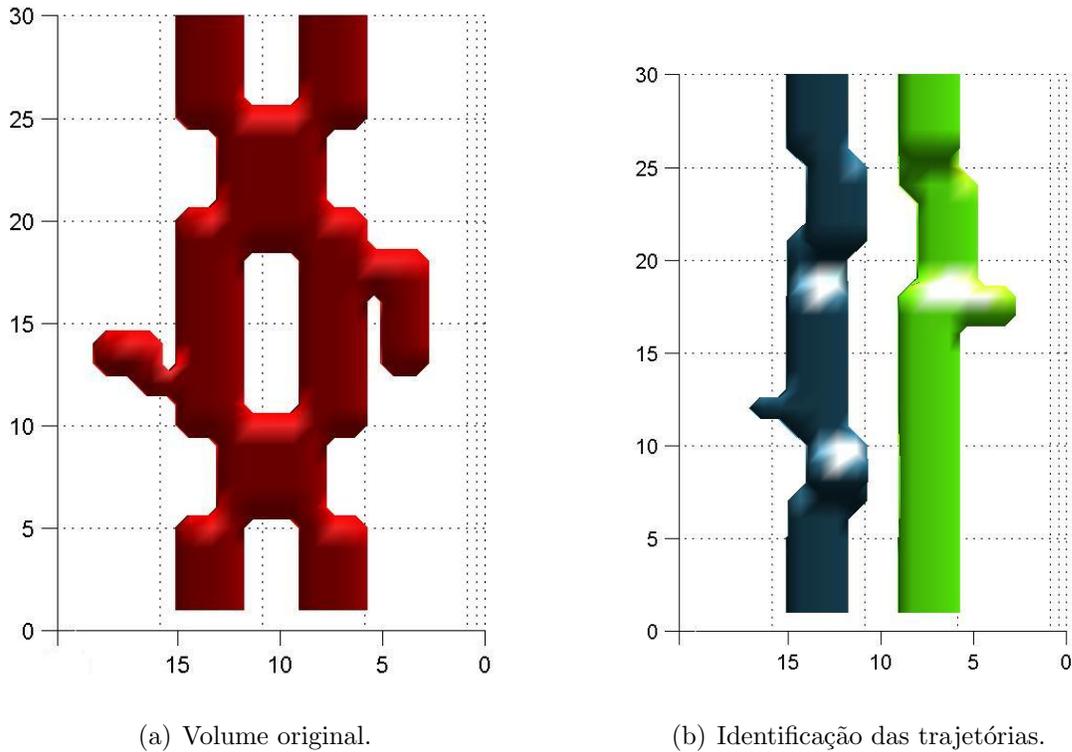


Figura 2.40: Resultado final para o exemplo considerado.

caso de falha do método automático.

Tradicionalmente, em aplicações práticas de rastreamento, um usuário supervisiona o progresso do método e intervém quando necessário, indicando o posicionamento correto do objeto considerado. Em nosso caso, como não existem oclusões no domínio tridimensional, o objeto rastreado sempre será detectado pelo método, existindo somente a necessidade de intervenção quando um falso volume provocar ambiguidades não solucionáveis. O próprio método detecta estas situações e apresenta as informações necessárias para a tomada de decisão ao usuário, como ilustrado nas figuras 2.41 e 2.42, eliminando a necessidade de monitoramento constante. Indo além, o número de intervenções do usuário é igual ao número de ambiguidades existentes na sequência considerada e não ao número de quadros rastreados incorretamente, o que reduz o trabalho manual envolvido.

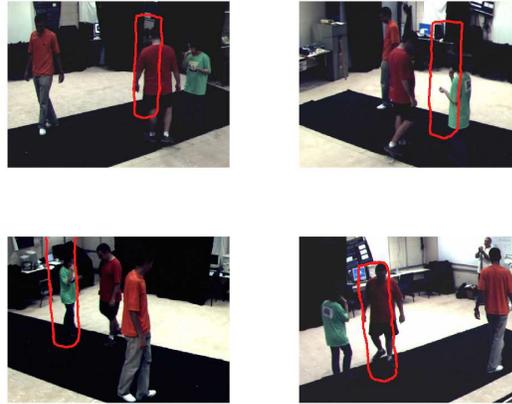


Figura 2.41: Áreas correspondentes ao primeiro vértice possível.



Figura 2.42: Áreas correspondentes ao segundo vértice possível.

Capítulo 3

Resultados experimentais

Neste capítulo detalharemos todos os procedimentos considerados para inúmeros vídeos experimentais, discutindo cada opção feita e os resultados obtidos.

3.1 Testes em laboratório

Nesta seção consideramos uma sequência de testes, feita no Laboratório de Instrumentação para Biomecânica - FEF - Unicamp, que consiste em três pessoas caminhando no laboratório. Utilizamos quatro câmeras Basler A602f, sincronizadas via hardware, capturando a 35 quadros por segundo, totalizando 2100 quadros. A figura 3.1 ilustra as imagens capturadas por cada câmera.

3.1.1 Calibração de câmeras

Para a calibração de câmeras, utilizamos seis prumos de calibração, com esferas reflexivas a cada 10 centímetros. A figura 3.2 ilustra os pontos consideradas em cada câmera e a tabela 3.1 apresenta os erros de calibração encontrados. Neste caso, devido à abundância de pontos de calibração, optamos por realizar somente a calibração incluindo os parâmetros de distorção radial.

Câmera	Erro quadrático	Pontos	Erro Médio
1	478,652	114	4,1987
2	259,362	115	2,2553
3	652,885	104	6,2777
4	491,137	116	4,2339

Tabela 3.1: Sumário dos erros encontrados na calibração das câmeras.



(a) Câmera 1.



(b) Câmera 2.



(c) Câmera 3.



(d) Câmera 4.

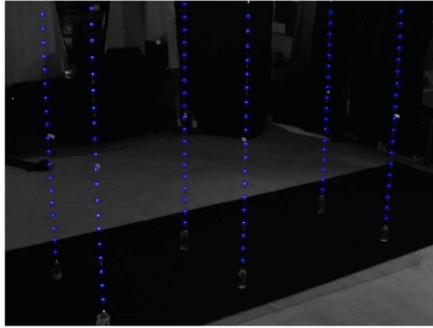
Figura 3.1: Exemplos das imagens consideradas.

Considerando os resultados apresentados na tabela 3.1, podemos afirmar que a calibração obteve resultados satisfatórios, como esperado numa situação com tantos pontos de calibração medidos com razoável precisão.

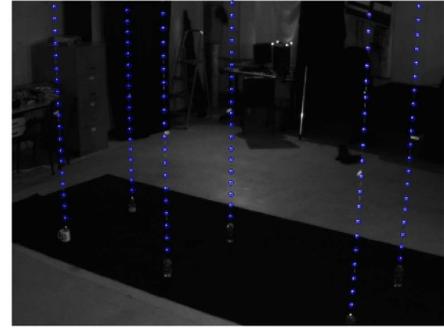
3.1.2 Identificação dos objetos em movimento

Procedemos com a construção do modelo do fundo, conforme descrito na seção 2.2, utilizando como parâmetros:

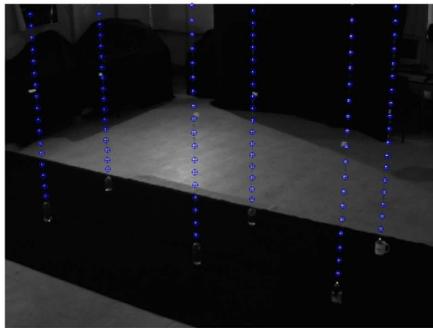
- **Número de quadros para a inicialização via moda:** 600 quadros, uniformemente distribuídos entre os quadros 100 e 2100.



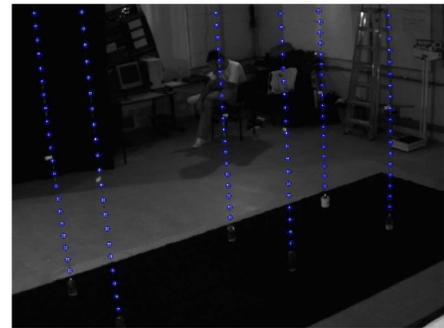
(a) Câmera 1.



(b) Câmera 2.



(c) Câmera 3.



(d) Câmera 4.

Figura 3.2: Pontos utilizados para calibração.

- **Número de quadros para a estimativa do modelo:** 21 quadros, uniformemente distribuídos entre os quadros 100 e 2100.
- **Número de bins no histograma que modela os objetos em movimento:** 128 bins por dimensão.
- **Limiar ϵ :** 20 para a câmera 4, 15 para todas as outras câmeras consideradas.

A figura 3.3 ilustra o resultado da identificação de objetos para o primeiro quadro da sequência. Para facilitar a visualização, a figura 3.4 contém a imagem original, com a região de interesse considerada e todas as regiões de probabilidade maior que zero demarcadas.

Considerando os dados da figura 3.4, podemos afirmar que o resultado não possui falsos negativos, atendendo um dos pré-requisitos para o funcionamento do método.

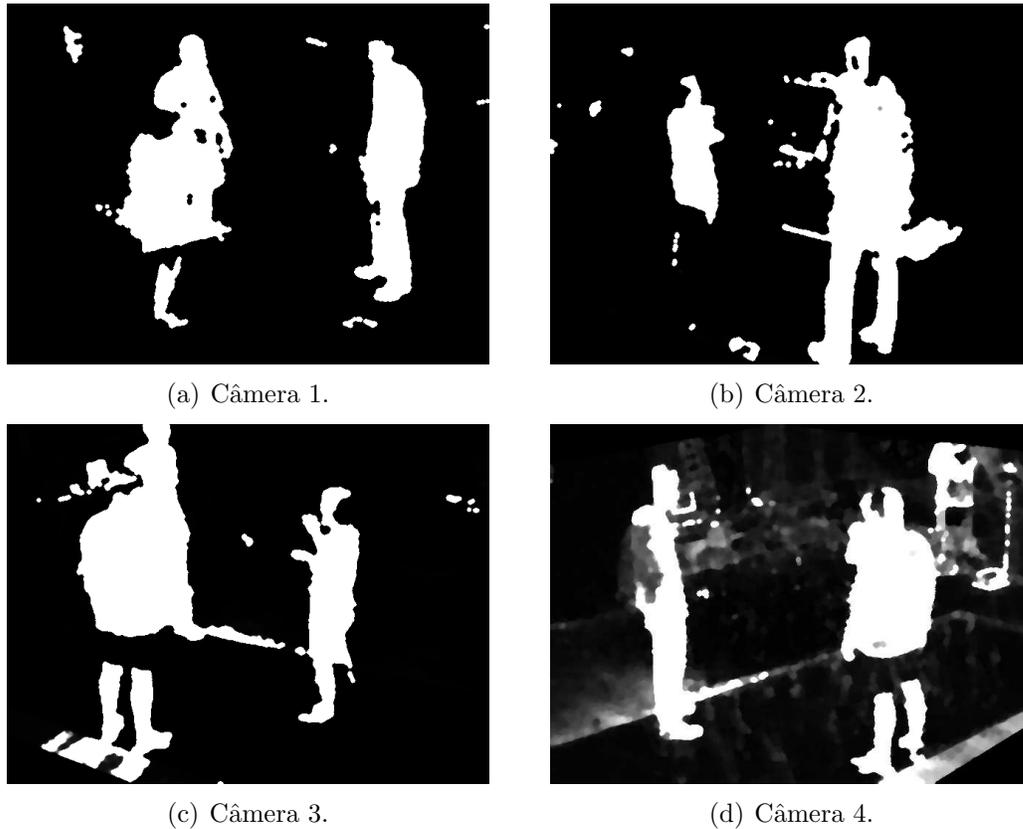


Figura 3.3: Resultado da identificação dos objetos em movimento para o quadro 1.

3.1.3 Reconstrução Tridimensional

Utilizando o método descrito na seção 2.3, considerando as imagens da figura 3.3 e as informações de calibração obtidas, obtivemos a reconstrução tridimensional da cena considerada, ilustrada na figura 3.5. O processo de reconstrução considerou a quadra, com altura entre 0,2 metro e 1,8 metro, cada voxel possui 10 centímetros de comprimento, 10 centímetros de largura e 10 centímetros de altura.

Considerando a figura 3.5(a), em comparação com o resultado da reconstrução utilizando a identificação manual dos objetos em movimento, ilustrado na figura 3.5(b), podemos afirmar que o resultado da reconstrução é bastante satisfatório, não existindo, neste quadro, nenhum falso volume ou falha na reconstrução. Entretanto, falsos volumes ocorrem neste vídeo, devido à proximidade entre os objetos.

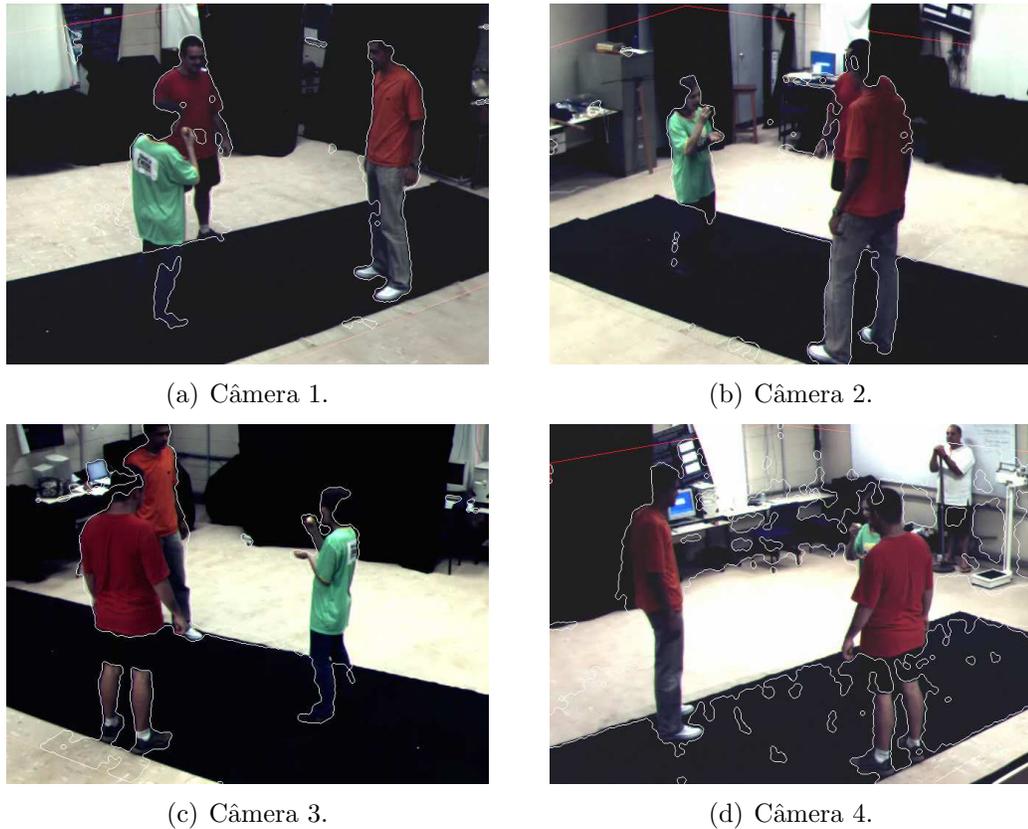


Figura 3.4: Objetos detectados superpostos com a imagem original.

3.1.4 Amostragem através do ritmo visual generalizado

A figura 3.6 ilustra a aplicação do ritmo visual generalizado, utilizando a função de amostragem definida na equação 2.4.2, para 500 quadros da sequência considerada, divididos em dois gráficos: entre os quadros 1 e 250(3.6(a)) e entre 251 e 500(3.6(b)). Após o junção das amostras, foram aplicados filtros morfológicos de fechamento e abertura, nesta ordem, utilizando como elemento estruturante um cubo de dimensão 3x3x3. O limiar considerado, tanto para visualização quanto para definição de trajetórias, foi 0,4.

3.1.5 Identificação das trajetórias

Utilizando a informação mostrada na figura 3.6, utilizamos o método apresentado na seção 2.6 para identificação de trajetórias. Durante o processamento, o método encontrou uma única ambiguidade, ilustrada nas figuras 3.7 e 3.8.

Ao informarmos o número 713 como vértice correto para a correção da ambiguidade,

o método finalizou o processamento, apresentando como resultado três regiões distintas, apresentadas na figura 3.9, dividido em duas imagens para facilitar a visualização: entre 1 e 250(3.9(a)) e entre 251 e 500(3.9(b)).

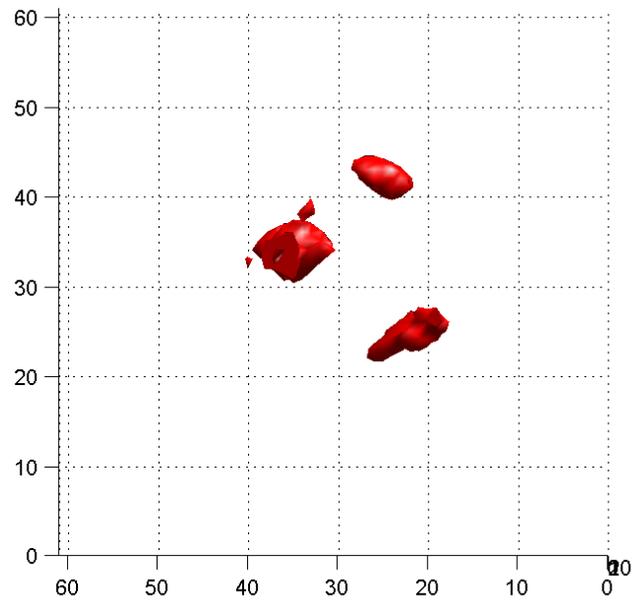
Na figura 3.9, podemos notar alguns trechos no qual a espessura da trajetória diminui drasticamente. Estes trechos são resultados da divisão de vértices, quando a região correspondente a um vértice é substituída pelas trajetórias inferidas pela heurística considerada, como apresentado na figura 2.6.

Para confirmar se as trajetórias detectadas correspondem aos objetos certos, reprojeteamos os volumes correspondentes nas imagens dos vídeos, verificamos que objeto teve sua trajetória confundida com outro. A figura 3.10 ilustra alguns quadros contendo os limites dos objetos rastreados.

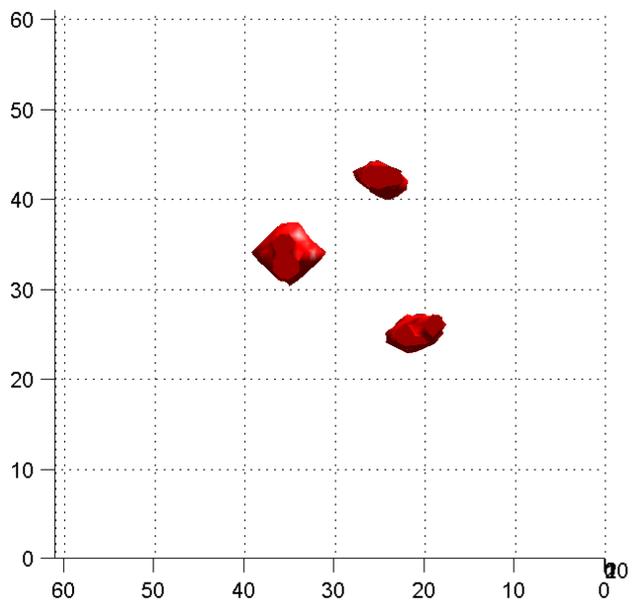
3.1.6 Conclusões

Considerando os resultados apresentados nas figuras 3.9 e 3.10, podemos concluir que o método obteve um desempenho adequado, apesar de considerarmos somente 550 quadros, dos 2100 presentes na sequência. Isto se deve a limitações de memória presentes na implementação atual que, de modo geral, não comprometem a idéia apresentada neste trabalho.

Esta sequência atende a todos os requisitos do nosso método, ao menos em parte, pois ainda estão presentes alguns falsos volumes, e o resultado foi satisfatório, sendo capaz de rastrear os três objetos em cena, mesmo com oclusões em todas as câmeras e similaridades cromáticas entre objetos.



(a) Utilizando identificação automática dos objetos em movimento.



(b) Utilizando identificação manual dos objetos em movimento.

Figura 3.5: Resultado da reconstrução tridimensional para o quadro 1.

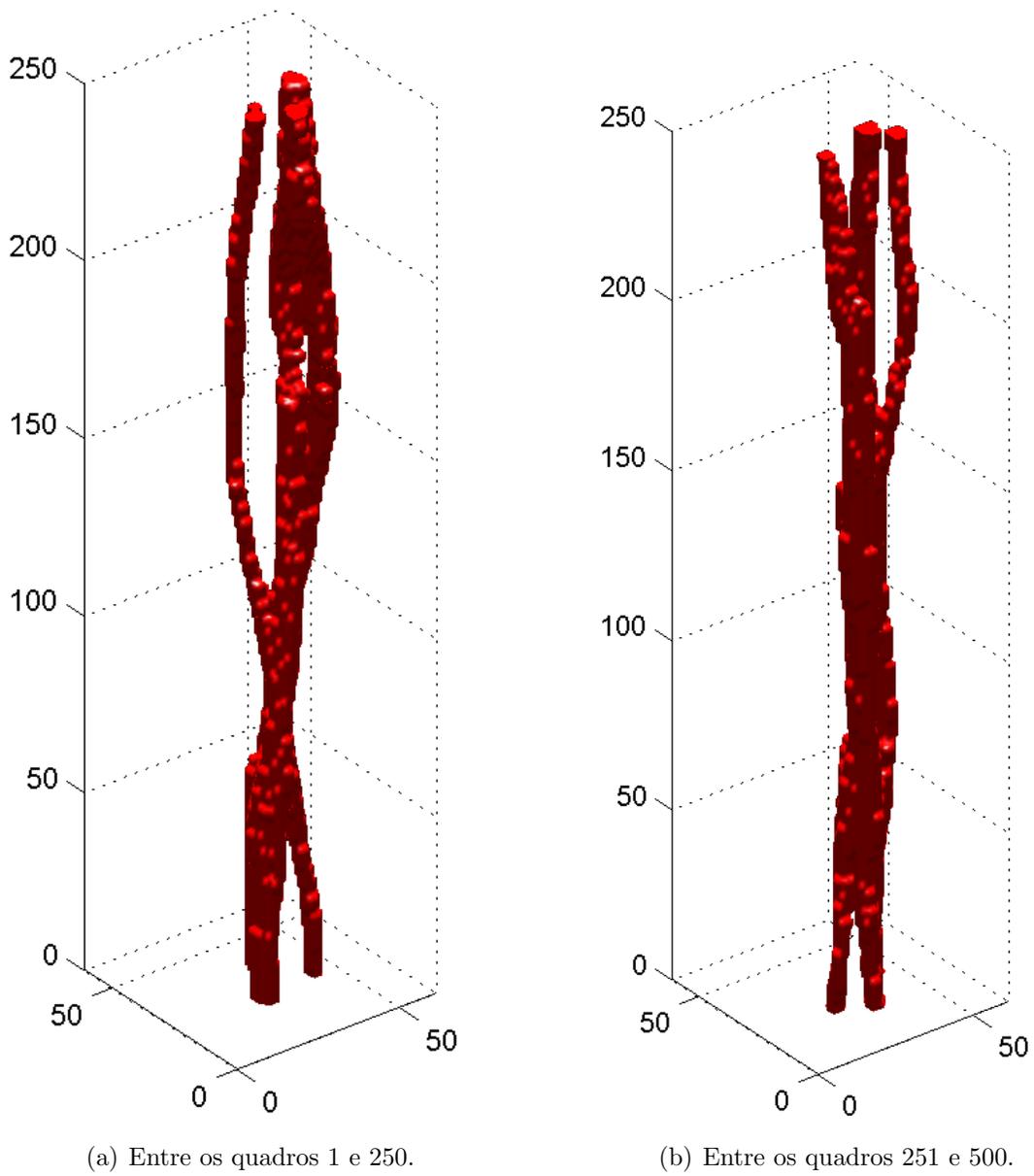


Figura 3.6: Resultado da amostragem através do ritmo visual generalizado.



Figura 3.7: Áreas correspondentes ao primeiro vértice possível - 713.

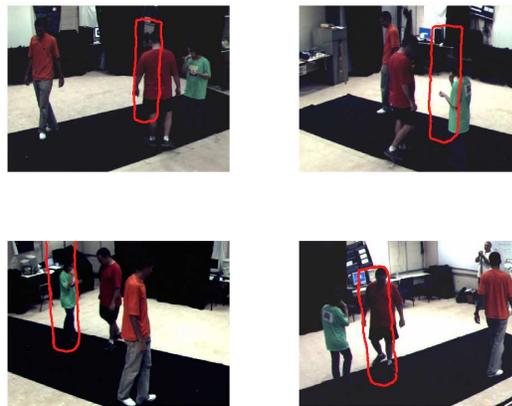
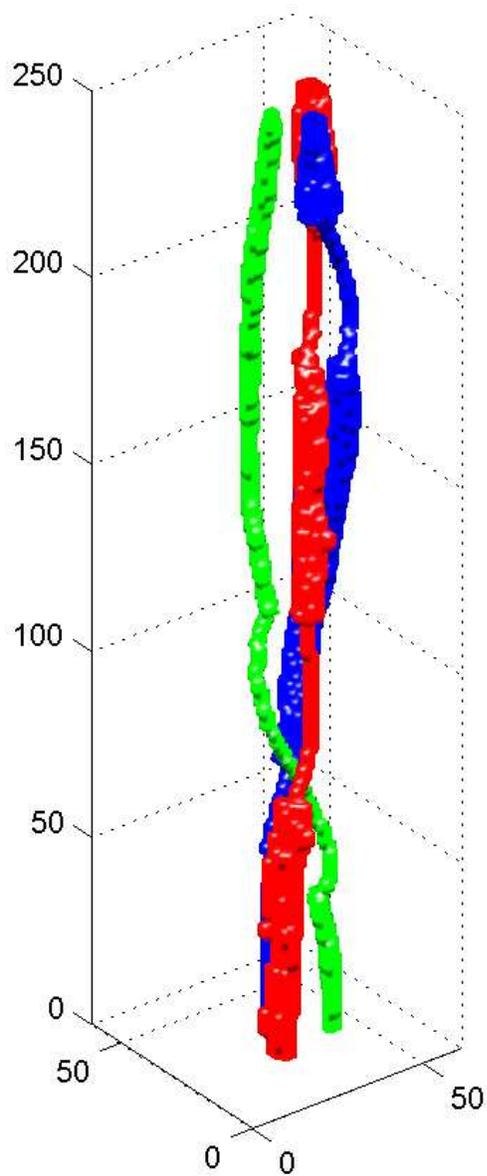
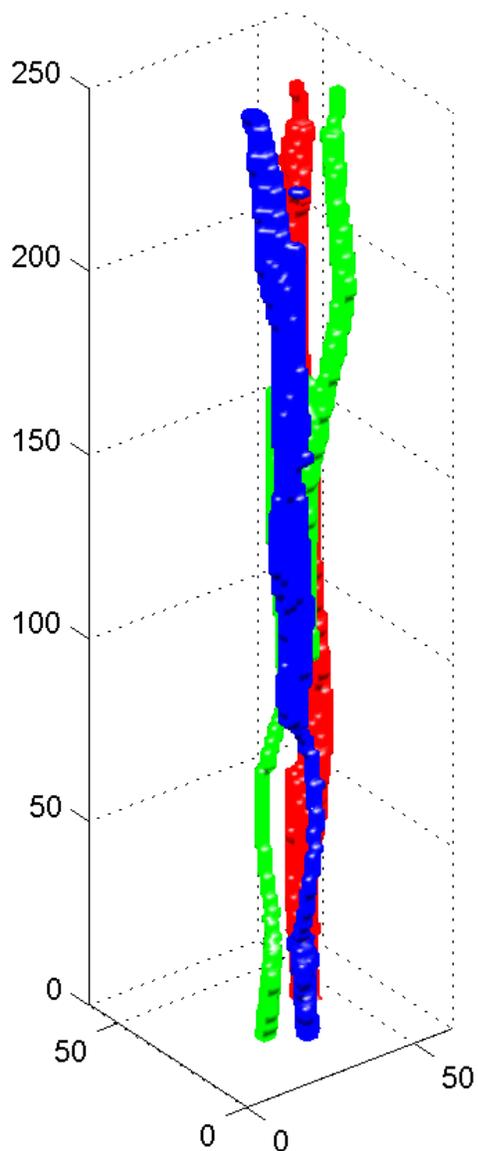


Figura 3.8: Áreas correspondentes ao primeiro vértice possível - 714.

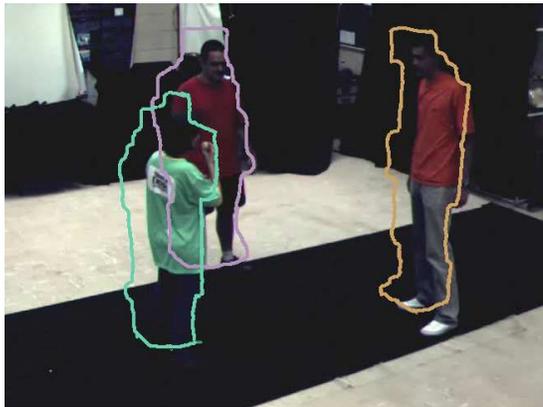


(a) Entre os quadros 1 e 250.

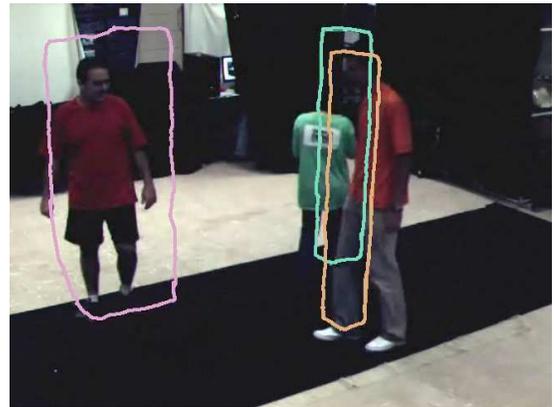


(b) Entre os quadros 251 e 500.

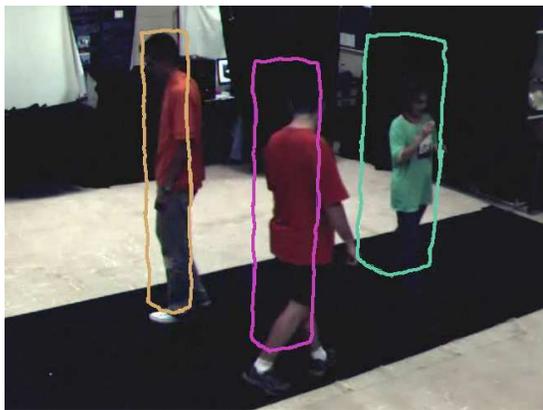
Figura 3.9: Resultado da identificação de trajetórias.



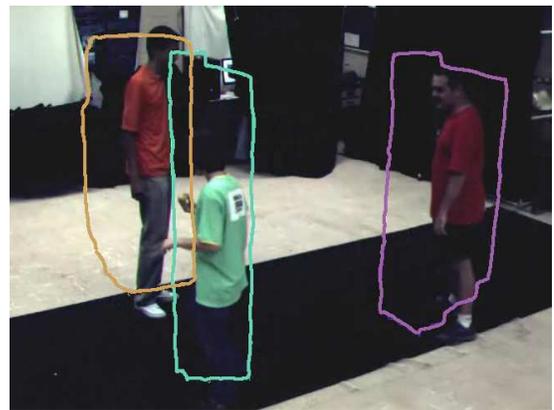
(a) Quadro 1.



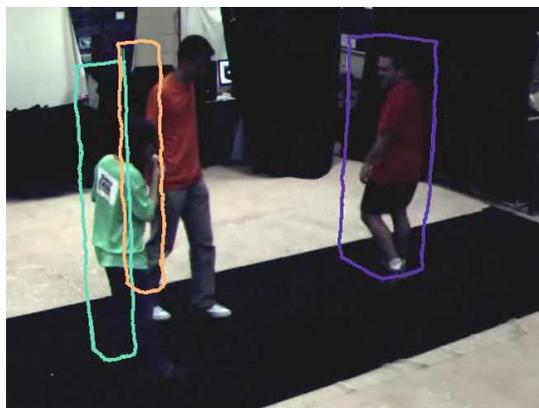
(b) Quadro 111.



(c) Quadro 221.



(d) Quadro 441.



(e) Quadro 500.

Figura 3.10: Resultado da identificação de trajetórias.

3.2 Jogo oficial de vôlei feminino - Bulgária x Alemanha

Esta coleta possui aproximadamente 600 quadros, correspondentes a um lance do jogo. A figura 3.11 ilustra as imagens capturadas por cada câmera.



(a) Câmera 1.



(b) Câmera 2.



(c) Câmera 3.



(d) Câmera 4.

Figura 3.11: Exemplos das imagens consideradas.

3.2.1 Calibração de câmeras

Para a calibração de câmeras, utilizamos somente características presentes na quadra, de acordo com a padronização internacional. A figura 3.12 ilustra os pontos considerados em

cada câmera e a tabela 3.2 apresenta os erros de calibração encontrados, onde *Sem Dist.* representa a abordagem apresentada na seção 2.1, mas desconsiderando os parâmetros de distorção de câmeras, *Com Dist.* representa a utilização dos parâmetros de distorção de câmera. Os erros apresentados são calculados de acordo com a equação 2.1.7 e correspondem a erros quadráticos.



(a) Câmera 1.



(b) Câmera 2.



(c) Câmera 3.



(d) Câmera 4.

Figura 3.12: Pontos utilizados para calibração.

Considerando os resultados apresentados na tabela 3.2, podemos concluir que o posicionamento dos pontos de calibração, na câmera 4, são mal-condicionados para o cálculo utilizado desconsiderando distorção, o que não acontece quando consideramos a distorção, reduzindo significativamente o erro de reprojeção.

Câmera	Sem Dist.	Com Dist.	Pontos	Erro Médio S.D.	Erro Médio C.D.
1	1230,1	81,556	37	33,25	2,2042
2	315,4	306,222	30	10,51	10,2074
3	120,5	115,874	25	4,85	4,6350
4	186576,0	211,833	22	8480,73	9,6288

Tabela 3.2: Sumário dos erros encontrados na calibração das câmeras.

3.2.2 Identificação dos objetos em movimento

Procedemos com a construção do modelo do fundo, conforme descrito na seção 2.2, utilizando como parâmetros:

- **Número de quadros para a inicialização via moda:** 500 quadros, uniformemente distribuídos entre os quadros 1 e 600.
- **Número de quadros para a estimativa do modelo:** 21 quadros, uniformemente distribuídos entre os quadros 1 e 600.
- **Número de bins no histograma que modela os objetos em movimento:** 128 bins por dimensão.
- **Limiar ϵ :** 45, para a câmera 1, 50 para todas as outras câmeras consideradas.

A figura 3.13 ilustra o resultado da identificação de objetos para o primeiro quadro da sequência. Para facilitar a visualização, a figura 3.14 contém a imagem original, com a região de interesse considerada e todas as regiões de probabilidade maior que zero demarcadas.

Considerando os resultados ilustrados na figura 3.14, podemos afirmar que o resultado não possui falsos negativos, atendendo a um dos pré-requisitos para o funcionamento do método. Determinados efeitos de iluminação foram detectados, assim como algumas regiões nas placas de publicidade. Com um bom número de câmeras na cena, estes falsos positivos não constituem problema pois, mesmo se consistentemente presentes em todas as câmeras, possuem uma posição tridimensional bem definida, assim como as sombras dos jogadores, e são reconstruídos em locais facilmente identificáveis, como o chão ou o exterior da quadra.

Entretanto, as sequências consideradas são muito curtas, possuindo apenas 600 quadros, um valor longe do ideal para uma estimativa estatística do fundo da cena, causando um grande número de falsos positivos, mesmo com um valor alto de limiar ϵ .

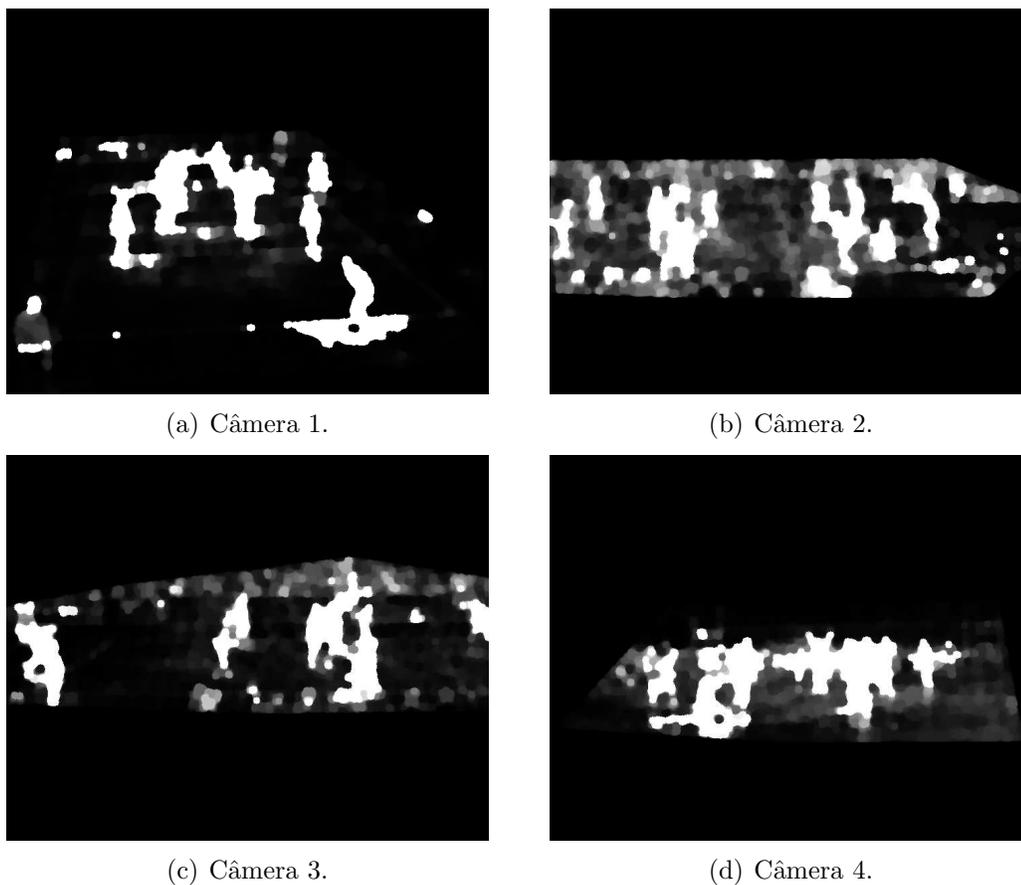


Figura 3.13: Resultado da identificação dos objetos em movimento para o quadro 30.

3.2.3 Reconstrução Tridimensional

Utilizando o método descrito na seção 2.3, considerando as imagens da figura 3.13 e as informações de calibração obtidas, obtivemos a reconstrução tridimensional da cena considerada, ilustrada na figura 3.15. O processo de reconstrução considerou a quadra, com altura entre 0,2 metro e 2,2 metro, cada voxel possui 20 centímetros de comprimento, 20 centímetros de largura e 10 centímetros de altura.

Considerando a figura 3.15(a), em comparação com o resultado manual ilustrado na figura 3.15(b), constatamos a presença de muitos falsos volumes na reconstrução. A partir da abordagem apresentada na seção 2.3.3, podemos afirmar que o posicionamento das câmeras é a causa dos falsos volumes. Mesmo desconsiderando uma câmera superior, os resultados poderiam ser melhores caso as câmeras estivessem numa posição mais elevada, como a câmera 1. Com isto, a câmera 4, além de não fornecer muita informação sobre



(a) Câmera 1.



(b) Câmera 2.



(c) Câmera 3.



(d) Câmera 4.

Figura 3.14: Objetos detectados superpostos com a imagem original.

o outro lado da quadra, tem cones de reprojeção muito longos, o que, considerando as câmeras laterais, provoca muitas interseções e, conseqüentemente, muitos falsos volumes.

3.2.4 Amostragem através do ritmo visual generalizado

Apesar da grande quantidade de falsos volumes presentes na reconstrução, os objetos presentes na cena também foram reconstruídos, o que permite, com ressalvas, que continuemos com a aplicação do método proposto. Entretanto, não devemos esperar um resultado positivo, pois a grande quantidade de falsos volumes presentes na cena dificulta enormemente a etapa de identificação das trajetórias.

A figura 3.16 ilustra a aplicação do ritmo visual generalizado, utilizando a função de amostragem definida na equação 2.4.2, para 50 quadros da sequência considerada. Após o agrupamento das amostras, foram aplicados filtros morfológicos de fechamento

e abertura, nesta ordem, utilizando como elemento estruturante um cubo de dimensão 3x3x3. O limiar considerado, tanto para visualização quanto para definição de trajetórias, foi 0,5.

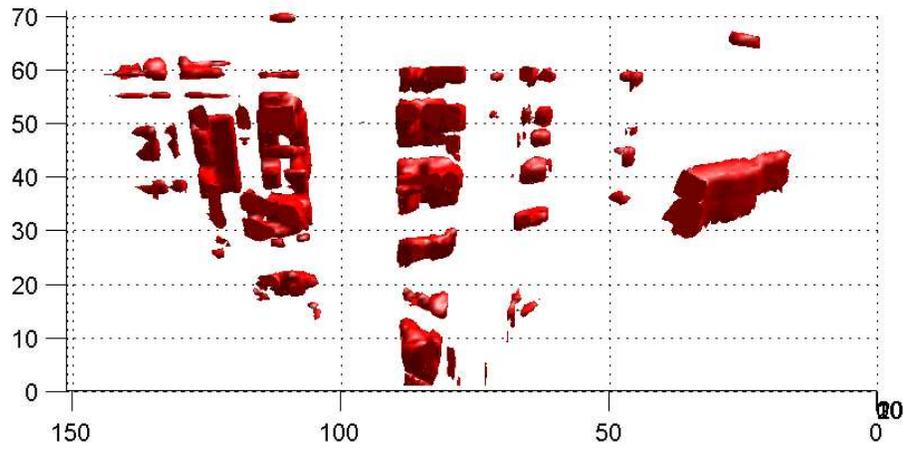
3.2.5 Identificação das trajetórias

Utilizando a informação mostrada na figura 3.16, utilizamos o método apresentado na seção 2.6, para identificação de trajetórias. O excesso de falsos volumes gerou casos não previstos no atual método, tal como o ilustrado na figura 3.17, com um trecho do grafo correspondente ilustrado na figura 3.17(b). O método para separação de regiões foi projetado para separar 2 ou mais objetos, baseado em possíveis pontos de entrada e saída. Neste caso, como existem 2 pontos de entrada e somente um de saída, ambos os objetos, obrigatoriamente, deveriam sair pelo mesmo ponto de saída, o que contraria diretamente o conceito físico que impede dois corpos de ocuparem o mesmo lugar no espaço ao mesmo tempo. O tratamento deste tipo de dificuldades, causadas por um posicionamento não ideal das câmeras, aumentaria significativamente a robustez e flexibilidade do método, mas consideramos fora do escopo deste trabalho, sendo, no entanto, um interessante trabalho futuro.

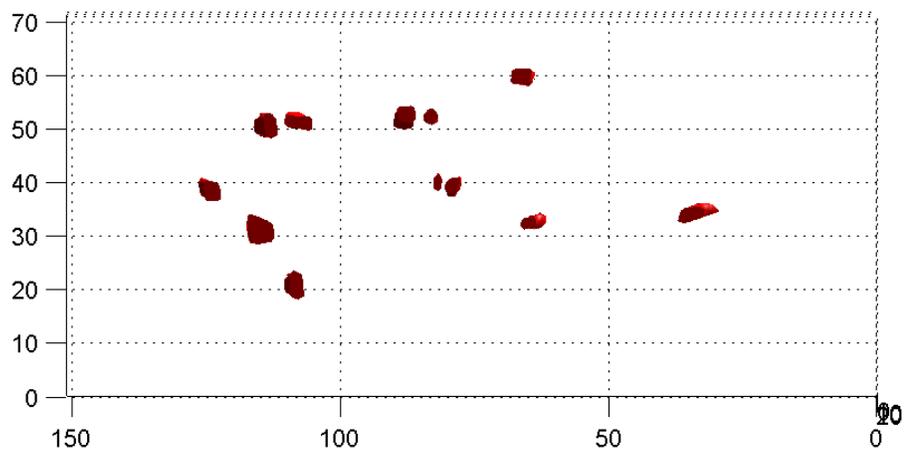
3.2.6 Conclusões

Esta sequência não pode ser processada pelo método apresentado, principalmente devido ao posicionamento das câmeras, muito distante do ideal, resultando em mais falsos volumes do que o método apresentado é capaz de tratar. Além disso, a curta duração do vídeo comprometeu o desempenho do método de identificação dos objetos em movimento, resultando em muitos falsos positivos, aumentando o problema de falsos volumes.

Para o funcionamento do método proposto, um melhor posicionamento das câmeras seria crucial. Numa posição mais elevada com relação à quadra, os cones de reprojeção de cada câmera seriam mais curtos, diminuindo a ocorrência de falsos volumes e possibilitando o rastreamento. A utilização de uma ou mais câmeras superiores também seria altamente recomendada, pois eliminaria a ocorrência dos falsos volumes.



(a) Utilizando identificação automática dos objetos em movimento.



(b) Utilizando identificação manual dos objetos em movimento.

Figura 3.15: Resultado da reconstrução tridimensional para o quadro 30.

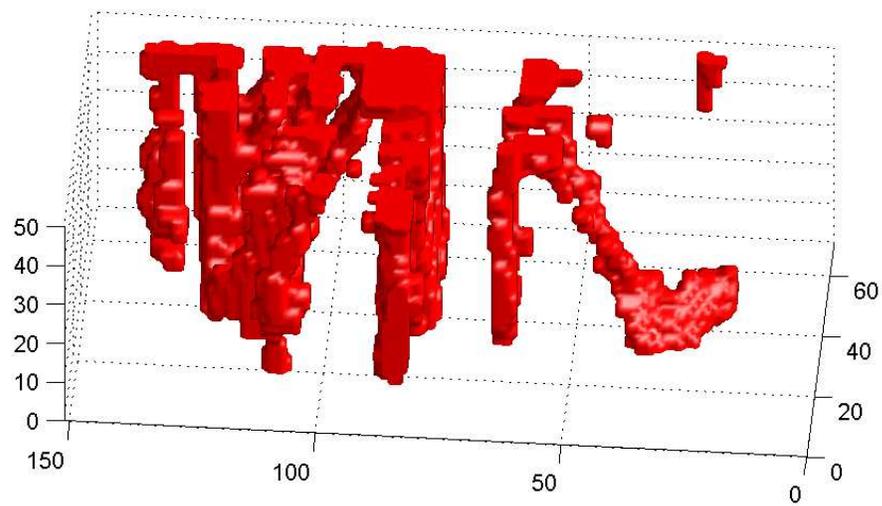
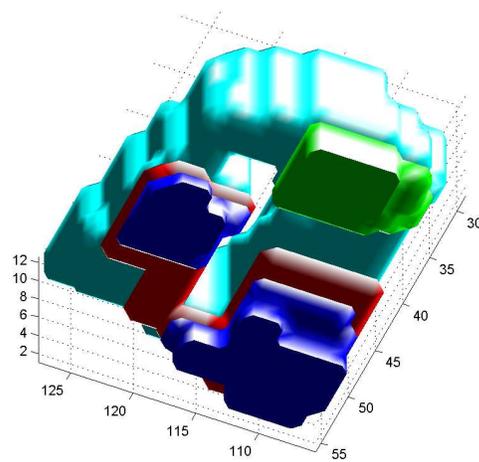
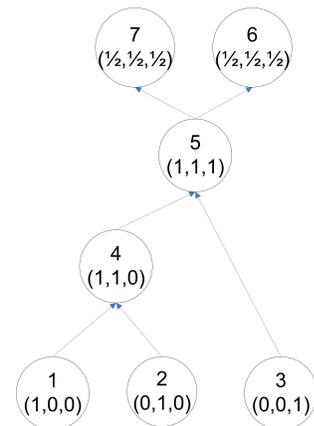


Figura 3.16: Resultado da amostragem através do ritmo visual generalizado.



(a) Visualização tridimensional das regiões envolvidas.



(b) Grafo correspondente.

Figura 3.17: Visualização da razão da falha do método apresentado na sequência considerada.

3.3 Final do Campeonato Mundial de Futsal para cegos

Nesta coleta, utilizamos cinco câmeras Basler A602f, não sincronizadas por hardware. A figura 3.18 ilustra as imagens capturadas por cada câmera. A cobertura das câmeras não é ideal e a rede de segurança, que se movimenta, dificulta o processamento do vídeo. Utilizamos aproximadamente 18.000 quadros, correspondentes aos 10 minutos da primeira prorrogação da final do evento.

3.3.1 Calibração de câmeras

Para a calibração de câmeras, utilizamos marcações presentes na cena, sem o uso de um objeto de calibração. A maior parte dos pontos utilizados pertence à quadra, mas também consideramos as traves e alguns pontos medidos nas placas de publicidade. A figura 3.19 ilustra o pontos considerados em cada câmera e a tabela 3.3 apresenta os erros de calibração encontrados, onde *Sem Dist.* representa a abordagem apresentada na seção 2.1, mas desconsiderando os parâmetros de distorção de câmeras, *Com Dist.* representa a utilização dos parâmetros de distorção de câmera. Os erros apresentados são calculados de acordo com a equação 2.1.7 e correspondem a erros quadráticos.

Câmera	Sem Dist.	Com Dist.	Pontos	Erro Médio S.D.	Erro Médio C.D.
1	627,025	521,749	24	26,1260	21,7395
2	4288,060	1643,392	16	268,0038	102,7450
3	1041,990	874,762	24	43,4162	36,4484
4	355,535	271,740	23	15,4580	11,8148
5	620,681	461,409	23	26,9861	20,0611

Tabela 3.3: Sumário dos erros encontrados na calibração das câmeras.

Considerando os resultados apresentados na tabela 3.3, optamos por desconsiderar a câmera de número 2, que apresentou um erro de calibração muito mais alto que as demais câmeras. Mesmo a câmera 4, que apresentou o menor erro de reprojeção quadrático, em todas as situações, ainda possui um erro muito alto, que pode comprometer o resultado da reconstrução, principalmente considerando um posicionamento de câmeras no qual os jogadores ocupam pouco espaço na imagem, como é o caso desta coleta.

3.3.2 Identificação dos objetos em movimento

Procedemos com a construção do modelo do fundo, conforme descrito na seção 2.2, utilizando como parâmetros:

- **Número de quadros para a inicialização via moda:** 400 quadros, uniformemente distribuídos entre os quadros 100 e 18000.
- **Número de quadros para a estimativa do modelo:** 11 quadros, uniformemente distribuídos entre os quadros 100 e 18000.
- **Número de bins no histograma que modela os objetos em movimento:** 128 bins por dimensão.
- **Limiar ϵ :** 15, em todas as câmeras consideradas.

A figura 3.20 ilustra o resultado da identificação de objetos para o primeiro quadro da sequência. Para facilitar a visualização, a figura 3.21 contém a imagem original, com a região de interesse considerada e todas as regiões de probabilidade maior que zero demarcadas.

Considerando os resultados ilustrados na figura 3.21, podemos afirmar que o resultado não possui falsos negativos, atendendo um dos pré-requisitos para o funcionamento do método. Determinados efeitos de iluminação foram detectados, assim como algumas regiões nas placas de publicidade. Com um bom número de câmeras na cena, estes falsos positivos não são problema pois, mesmo se consistentemente presentes em todas as câmeras, possuem uma posição tridimensional bem definida, assim como as sombras dos jogadores, e são reconstruídos em locais facilmente identificáveis, como o chão ou o exterior da quadra.

Entretanto, na situação considerada nesta sequência, estes falsos positivos aumentam o número de falsos volumes presentes na cena, muitas vezes conectando vários volumes distintos, impossibilitando o rastreamento considerado neste trabalho.

3.3.3 Reconstrução Tridimensional

Utilizando o método descrito na seção 2.3, considerando as imagens da figura 3.20 e as informações de calibração obtidas, obtivemos a reconstrução tridimensional da cena considerada, ilustrada na figura 3.22. O processo de reconstrução considerou a quadra, com altura entre 0,2 metro e 1,8 metro, cada voxel possui 20 centímetros de comprimento, 20 centímetros de largura e 40 centímetros de altura.

Comparando a figura 3.22(c), em comparação com as reconstruções automáticas, ilustradas nas figuras 3.22(a) e 3.22(b), podemos concluir que a reconstrução possui muitos falsos volumes e que, graças à uma calibração de câmeras que está longe da ideal, os volumes atualmente presentes na cena não foram reconstruídos. Apesar de utilizar a mesma informação de calibração, a identificação manual dos objetos em movimento teve um

resultado minimamente satisfatório, pois as regiões escolhidas manualmente foram significativamente maiores do que as identificadas pelo método descrito na seção 2.2, reduzindo o impacto dos erros de calibração.

3.3.4 Conclusões

Esta sequência não pode ser processada pelo método apresentado, principalmente por não atender dois pré-requisitos do método: uma boa cobertura de câmeras e uma boa calibração. Adicionalmente, as imagens não são sincronizadas por hardware, o que poderia causar falhas na reconstrução de movimentos rápidos.

Para a utilização do método apresentado, seriam necessárias mais câmeras, com mais superposição das imagens, para reduzir falsos volumes. A utilização da câmera desconsiderada pela calibração falha poderia melhorar os resultados mas, como é necessário mais de uma câmera para recuperar a informação tridimensional, as outras câmeras não poderiam fornecer boas informações sobre a região central da quadra, por estarem muito afastadas, o que resultaria numa reconstrução muito sensível a pequenos erros de calibração.

A calibração considerada também gerou problemas, principalmente pelo fato da maioria dos pontos de calibração estarem no mesmo plano, o que, aliado à presença de uma leve distorção radial, causam instabilidade no processo de calibração. A distorção em si não configuraria um problema, caso mais pontos de calibração estivessem disponíveis, cobrindo todo o volume considerado. Em tais situações, o método recomendado para a calibração seria através de um passeio com um bastão de calibração em pontos conhecidos da quadra, como proposto na seção 2.1, fornecendo, no mínimo 100 pontos de calibração.



(a) Câmera 1.



(b) Câmera 2.



(c) Câmera 3.



(d) Câmera 4.



(e) Câmera 5.

Figura 3.18: Exemplos das imagens capturadas.



(a) Câmera 1.



(b) Câmera 2.



(c) Câmera 3.



(d) Câmera 4.



(e) Câmera 5.

Figura 3.19: Pontos utilizados para calibração.

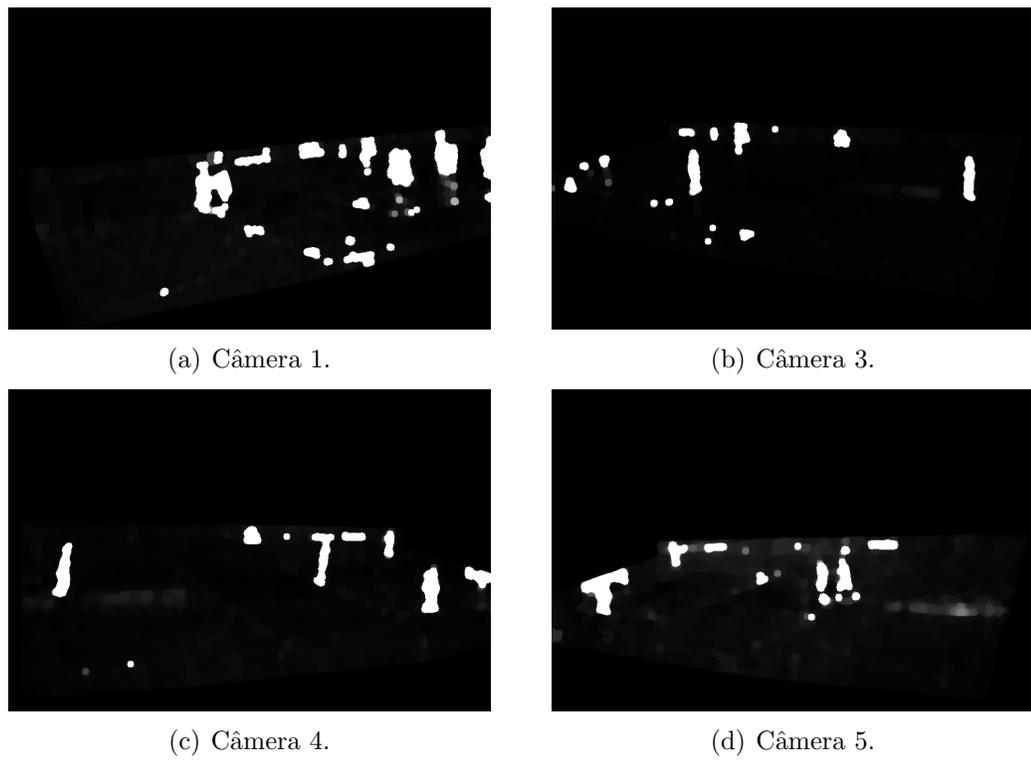


Figura 3.20: Resultado da identificação dos objetos em movimento para o quadro 1.



(a) Câmera 1.



(b) Câmera 3.

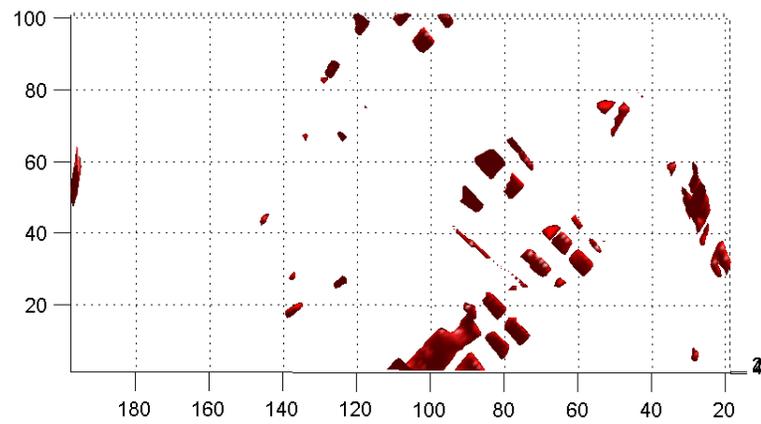


(c) Câmera 4.

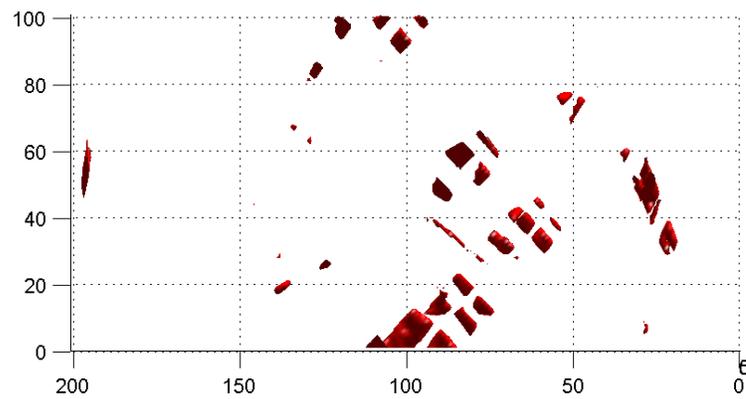


(d) Câmera 5.

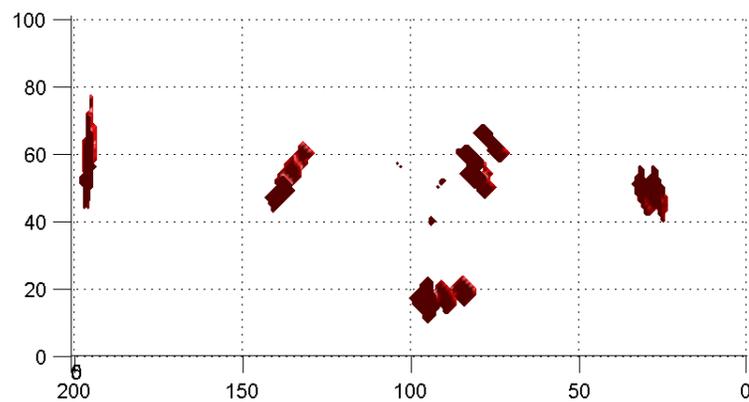
Figura 3.21: Objetos detectados superpostos com a imagem original.



(a) Desconsiderando os parâmetros de distorção.



(b) Considerando os parâmetros de distorção.



(c) Considerando os parâmetros de distorção - Inicialização Manual.

Figura 3.22: Resultado da reconstrução tridimensional para o quadro 1.

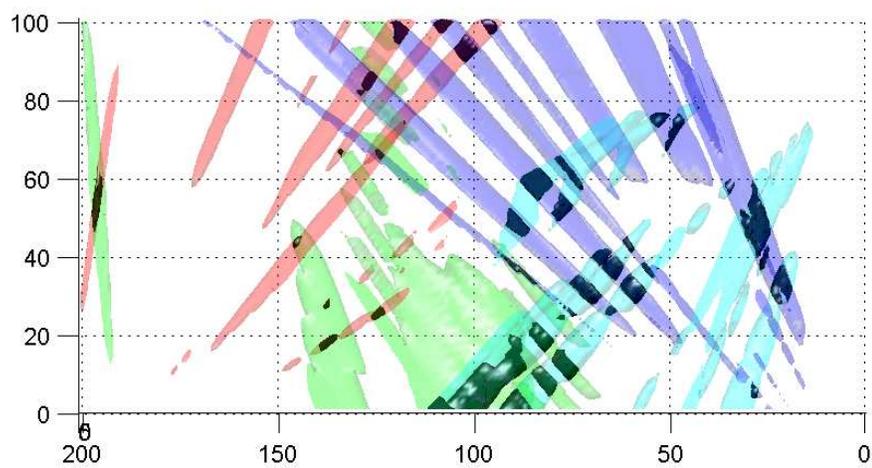


Figura 3.23: Cones de reprojeção para o quadro 1.

3.4 CVBASE06

A seguir, consideramos um dos vídeos disponibilizados para o CVBASE06 [32], contendo dois jogadores de squash, vistos de cima por uma única câmera, sem informações sobre a calibração. Para utilizar esta sequência, modificamos ligeiramente o método, suprimindo a etapa de calibração e a etapa de reconstrução, que seriam impossíveis no atual contexto. As amostras que seriam geradas a partir da reconstrução da cena são obtidas através da identificação dos objetos em movimento. Nesta configuração, nenhuma informação métrica a respeito da cena pode ser obtida, mas o método de rastreamento utilizado é o mesmo, explorando a utilização da generalização do ritmo visual.

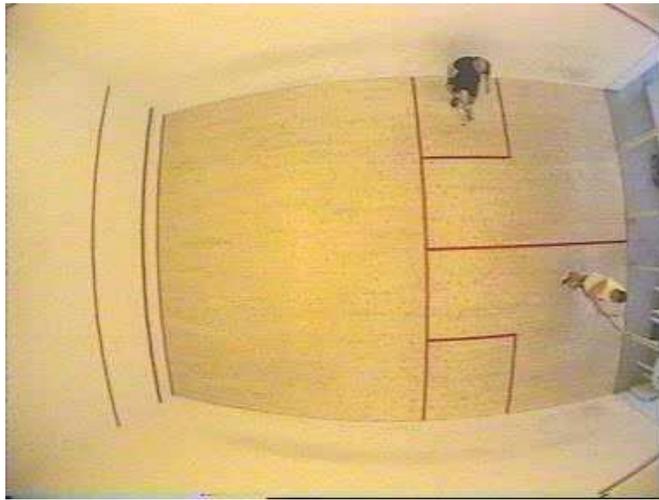


Figura 3.24: Exemplos das imagens consideradas.

3.4.1 Identificação dos objetos em movimento

Procedemos com a construção do modelo do fundo, conforme descrito na seção 2.2, utilizando como parâmetros:

- **Número de quadros para a inicialização via moda:** 800 quadros, uniformemente distribuídos entre os quadros 100 e 2100.
- **Número de quadros para a estimativa do modelo:** 41 quadros, uniformemente distribuídos entre os quadros 100 e 2100.
- **Número de bins no histograma que modela os objetos em movimento:** 128 bins por dimensão.
- **Limiar ϵ :** 25 .

A figura 3.25 ilustra o resultado da identificação de objetos para o primeiro quadro da sequência. Para facilitar a visualização, a figura 3.26 contém a imagem original, com a região de interesse considerada e todas as regiões de probabilidade maior que zero demarcadas.

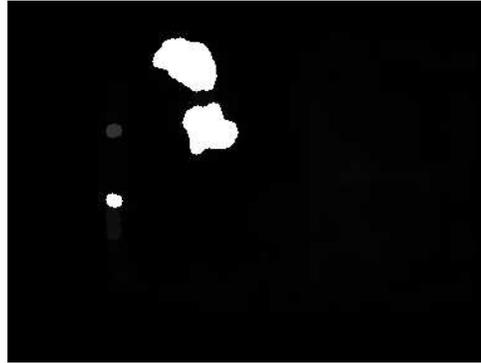


Figura 3.25: Resultado da identificação dos objetos em movimento para o quadro 1.

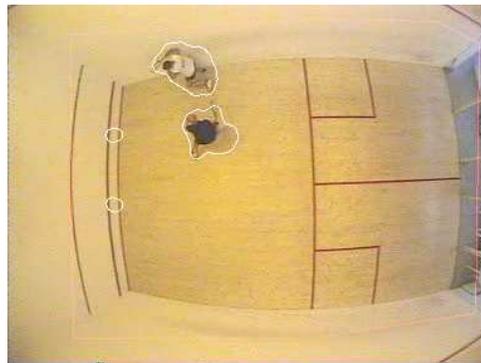


Figura 3.26: Objetos detectados superpostos com a imagem original.

Considerando os resultados ilustrados na figura 3.26, podemos afirmar que o resultado não possui falsos negativos, atendendo um dos pré-requisitos para o funcionamento do método.

3.4.2 Amostragem através do ritmo visual generalizado

A figura 3.27 ilustra a aplicação do ritmo visual generalizado, utilizando a função de amostragem 2.4.2, para 600 quadros da sequência considerada, divididos em dois gráficos: entre os quadros 1 e 300(3.27(a)) e entre 301 e 600(3.27(b)). Após o junção das amostras, foram aplicados filtros morfológicos de fechamento e abertura, nesta ordem, utilizando

como elemento estruturante um cubo de dimensões $3 \times 3 \times 3$. O limiar considerado, tanto para visualização quanto para definição de trajetórias, foi 0,4.

3.4.3 Identificação das trajetórias

Considerando a informação mostrada na figura 3.27, utilizamos o método apresentado na seção 2.6, para identificação de trajetórias.

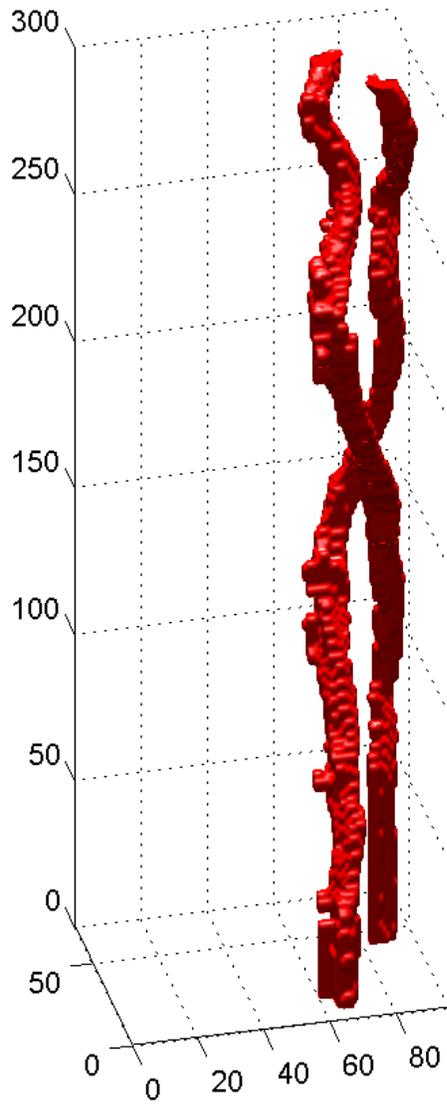
Na figura 3.28, podemos notar alguns trechos no qual a espessura da trajetória diminui significativamente. Estes trechos são resultados da divisão de vértices, quando a região correspondente a um vértice é substituída pelas trajetórias inferidas pela heurística considerada, como apresentado na figura 2.6.

Para confirmar se as trajetórias detectadas correspondem aos objetos certos, reprojeteamos os volumes correspondentes nas imagens dos vídeos, verificamos que objeto teve sua trajetória confundida com outro. A figura 3.29 ilustra alguns quadros contendo os limites dos objetos rastreados.

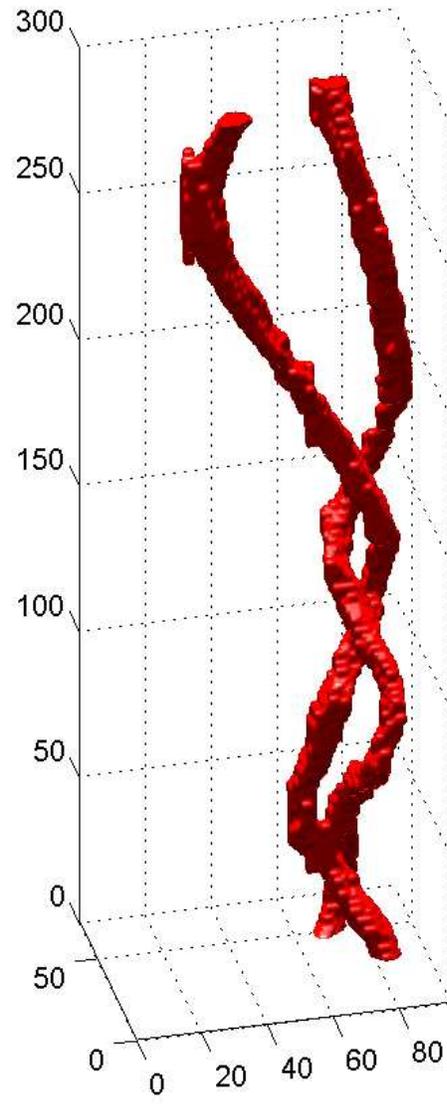
Como pode ser visto na figuras 3.29(f) e 3.29(e), as regiões identificadas como correspondentes aos jogadores não estão corretamente posicionadas. Esta discrepância é resultado da influência das sombras detectadas pela identificação de fundo, que não são desconsideradas por utilizarmos somente uma câmera. As regiões demarcadas correspondem, portanto, a uma trajetória suavizada do objeto em questão, que inclui as sombras correspondentes.

3.4.4 Conclusões

Nesta sequência, utilizamos uma variação do método proposto, não considerando a utilização de múltiplas câmeras, calibração de câmeras e reconstrução. Deste modo, o método não é capaz de obter informações tridimensionais, nem informações métricas sobre a cena considerada. Efetuamos o rastreamento no domínio da imagem, o que foge ligeiramente do proposto sobre rastreamento no domínio tridimensional, mas demonstra a flexibilidade do método, uma vez que seus requisitos são atendidos.



(a) Entre os quadros 1 e 300.



(b) Entre os quadros 301 e 600.

Figura 3.27: Resultado da amostragem através do ritmo visual generalizado.

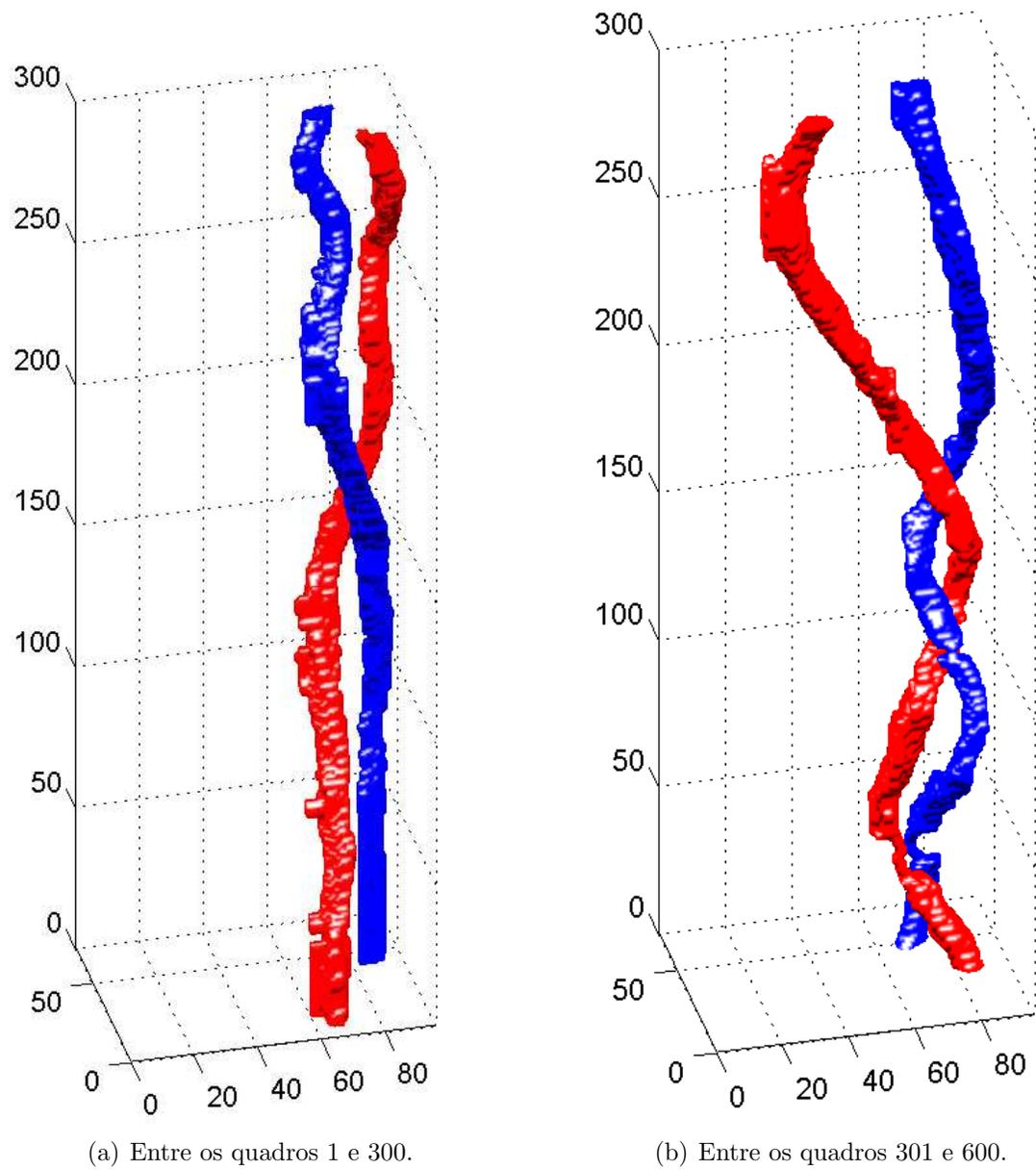


Figura 3.28: Resultado da identificação de trajetórias.



(a) Quadro 1.



(b) Quadro 101.



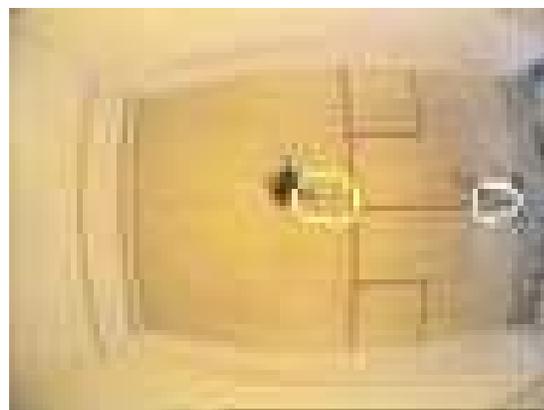
(c) Quadro 201.



(d) Quadro 301.



(e) Quadro 401.



(f) Quadro 501.

Figura 3.29: Resultado da identificação de trajetórias.

Capítulo 4

Conclusões

Neste trabalho apresentamos um método para o rastreamento de objetos em movimento, considerando a entrada do método no domínio tridimensional, um conceito não totalmente explorado na literatura pesquisada, que pode aumentar a robustez e flexibilidade dos algoritmos de rastreamento, uma vez que neste domínio não existem oclusões. O método apresentado, entretanto, se demonstrou de difícil aplicação prática, pois as condições a serem satisfeitas para o perfeito funcionamento deste são, muitas vezes, impraticáveis.

Para efetuar automaticamente o rastreamento, tendo como região de interesse uma quadra de aproximadamente 20 metros por 40 metros, são necessárias, no mínimo 4 câmeras laterais e duas superiores para obter informações tridimensionais completas da cena considerada. Considerando somente as câmeras superiores, podemos efetuar o rastreamento automático e recuperar a posição dos jogadores em quadra com um alto nível de automação.

A calibração ideal das câmeras envolveria o passeio de um bastão de calibração, através de pontos medidos na cena considerada, de modo a fornecer uma boa quantidade de informação, principalmente se considerarmos a utilização de lentes com grande angular.

A identificação de objetos em movimento necessita de um bom número de quadros, maior ou igual a 2000 quadros, dependendo da cena considerada, para obter uma boa estimativa da distribuição cromática do fundo e dos objetos em movimento. Os resultados são melhores quando os objetos em movimento e o fundo não possuem cores semelhantes, apesar de esta ser uma condição não necessária, pois o método como um todo foi planejado para conseguir lidar com falsos positivos.

Considerando um bom posicionamento das câmeras, uma calibração precisa e uma boa identificação dos objetos em movimento, a reconstrução da cena é feita de maneira satisfatória, sem falsos volumes, possibilitando o rastreamento completamente automático após a amostragem através do ritmo visual. O algoritmo de rastreamento é baseado em grafos, uma abordagem bastante simples, com técnicas similares presentes na literatura.

Quando consideramos uma cena que não atende os requisitos estabelecidos, o rastreamento pode ser feito de maneira semi-automática, com a intervenção do usuário para o solucionamento de ambiguidades causadas por falsos volumes presentes na cena.

Capítulo 5

Trabalhos futuros

Neste capítulo apresentaremos algumas extensões possíveis ou desejáveis para o método apresentado. Ainda existe muito a ser feito para aumentar a robustez e flexibilidade do mesmo.

A calibração das câmeras pode ser melhorada através da flexibilização do método. Muitas vezes não é possível ou praticável um passeio com um calibrador através da região a ser filmada. O emprego de outros métodos de calibração poderia flexibilizar a utilização do método proposto.

A identificação dos objetos em movimento, apesar de atender o requisito de não possuir falsos negativos, possui muitos falsos positivos, o que gera falsos volumes na reconstrução. A própria reconstrução pode ser melhorada ao nos aproximarmos mais de técnicas como *space carving*, que utiliza mais informações do que somente quais objetos estão em movimento, potencialmente diminuindo a ocorrência de falsos volumes.

Entretanto, as contribuições mais desejáveis referem-se ao algoritmo de definição de trajetórias, a partir da informação resultante do ritmo visual generalizado. A remoção das binarizações pode aumentar a robustez do método, removendo um parâmetro crítico a ser ajustado. Adicionalmente, existem vantagens semânticas ao considerarmos esta informação em seu domínio natural, ou seja, informações probabilísticas sobre a existência de um objeto em um determinado local.

Referências Bibliográficas

- [1] E.L. Andrade, E. Khan, J.C. Woods, and M. Ghanbari. Player identification in interactive sport scenes using region space analysis prior information and number recognition. In *Visual Information Engineering, 2003. VIE 2003. International Conference on*, pages 57 – 60, July 2003.
- [2] S. Belongie. Rodrigues' rotation formula. MathWorld—A Wolfram Web Resource, created by Eric W. Weisstein., 2008.
- [3] B. Bennett, D. R. Magee, A. G. Cohn, and D. C. Hogg. Enhanced tracking and recognition of moving objects by reasoning about spatio-temporal continuity. *Image and Vision Computing*, 26(1):67–81, January 2008.
- [4] F. N. Bezerra and N. J. Leite. Using string matching to detect video transitions. *Pattern Analysis & Applications*, 10(1):45–54, October 2006.
- [5] F. N. Bezerra and E. Lima. Low cost soccer video summaries based on visual rhythm. In *MIR '06: Proceedings of the 8th ACM international workshop on Multimedia information retrieval*, pages 71–78, New York, NY, USA, 2006. ACM.
- [6] J. Bouguet. Camera calibration toolbox for matlab. http://www.vision.caltech.edu/bouguetj/calib_doc, October 2008.
- [7] A. Broadhurst, T.W. Drummond, and R. Cipolla. A probabilistic framework for space carving. *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, 1:388–393 vol.1, 2001.
- [8] G.K.M. Cheung, S. Baker, and T. Kanade. Visual hull alignment and refinement across time: a 3d reconstruction algorithm combining shape-from-silhouette with stereo. *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, 2:II–375–82 vol.2, June 2003.
- [9] J. Czyz, B. Ristic, and B. Macq. A particle filter for joint detection and tracking of color objects. *Image and Vision Computing*, 25(8):1271–1281, August 2007.

- [10] G. Dainese, M. Marcon, A. Sarti, and S. Tubaro. 3d object modeling with a voxelset carving approach. In *IEEE International Conference on Image Processing*, pages 1053–1056, 2005.
- [11] M. Daum and G. Dudek. On 3-d surface reconstruction using shape from shadows. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 461–468, 1998.
- [12] C. H. Esteban and F. Schmitt. Multi-stereo 3d object reconstruction. In *Proc. of 3D Data Processing Visualization and Transmission*, pages 159–167, Padova, Italy, June 2002.
- [13] P. J. Figueroa, N. J. Leite, and R. M. L. Barros. Background recovering in outdoor image sequences: An example of soccer players segmentation. *Image Vision Comput*, 24(4):363–374, 2006.
- [14] P. J. Figueroa, N. J. Leite, and R. M. L. Barros. Tracking soccer players aiming their kinematical motion analysis. *Comput. Vis. Image Underst.*, 101(2):122–135, 2006.
- [15] P. J. Figueroa, N. J. Leite, R. M. L. Barros, I. Cohen, and G. G. Medioni. Tracking soccer players using the graph representation. In *International Conference on Pattern Recognition*, pages 787–790, 2004.
- [16] F. Fleuret, J. Berclaz, R. Lengagne, and P. Fua. Multicamera people tracking with a probabilistic occupancy map. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(2):267–282, Feb. 2008.
- [17] Charles F Golub, Gene H.; Van Loan. *Matrix Computations*. Johns Hopkins, 1996.
- [18] S. J. F. Guimarães, M. Couprie, A. A. Araújo, and N. J. Leite. Video segmentation based on 2d image analysis. *Pattern Recognition Letters*, 24(7), April 2003.
- [19] R. Hartley and A. Zisserman. *Multiple View Geometry in computer vision*. Cambridge University Press, 2003.
- [20] J.B. Hayet, Mathes, J. T. Czyz, J. Piater, J. and Verly, and B. Macq. A modular multi-camera framework for team sports tracking. In *Advanced Video and Signal Based Surveillance*, pages 493– 498. AVSS, September 2005.
- [21] H. Kim, J. Lee, J. Yang, S. Sull, W. M. Kim, and S. M. Song. Visual rhythm and shot verification. *Multimedia Tools and Applications*, 15(3):227–245, December 2001.

- [22] M. Kristan, J. Pers, M. Perse, S. Kovacic, and M. Bon. Multiple interacting targets tracking with application to team sports. In *Image and Signal Processing and Analysis, 2005. ISPA 2005. Proceedings of the 4th International Symposium on*, pages 322 – 327, September 2005.
- [23] S. J. Krotosky and M. M. Trivedi. Mutual information based registration of multi-modal stereo videos for person tracking. *Computer Vision and Image Understanding*, 106(2-3):270–287, May-June 2007.
- [24] M. Leo, N. Mosca, P. Spagnolo, P. L. Mazzeo, T. D’Orazio, and A. Distanti. Real-time multiview analysis of soccer matches for understanding interactions between ball and players. In *Proceedings of the 2008 international conference on Content-based image and video retrieval*, pages 525–534, New York, NY, USA, 2008.
- [25] S. Liu, K. Kang, J.-P. Tarel, and D.B. Cooper. Free-form object reconstruction from silhouettes, occluding edges and texture edges: A unified and robust operator based on duality. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(1):131–146, Jan. 2008.
- [26] J. Miura and H. Kubo. Tracking players in highly complex scenes in broadcast soccer video using a constraint satisfaction approach. In *Proceedings of the 2008 international conference on Content-based image and video retrieval*, pages 505–514, New York, NY, USA, 2008.
- [27] B. Müller and R. O. Anido. Distributed real-time soccer tracking. In *Proceedings of the ACM 2nd international workshop on Video surveillance & sensor networks*, pages 97–103, New York, NY, USA, 2004.
- [28] A. P. P. Mya, T. L. L. Thein, and M. M. Sein. Extracting the motion pattern of the players from a video stream of the football game. In *SICE-ICASE, 2006. International Joint Conference*, pages 5624 – 5627, October 2006.
- [29] W. C. Naidoo and J. R. Tapamo. Soccer video analysis by ball, player and referee tracking. In *SAICSIT ’06: Proceedings of the 2006 annual research conference of the South African institute of computer scientists and information technologists on IT research in developing countries*, pages 51–60, , Republic of South Africa, 2006. South African Institute for Computer Scientists and Information Technologists.
- [30] C. Ngo. A robust dissolve detector by support vector machine. In *MULTIMEDIA ’03: Proceedings of the eleventh ACM international conference on Multimedia*, pages 283–286, New York, NY, USA, 2003.

- [31] M.D. Oliveira and N.J. Leite. Reconnection of fingerprint ridges based on morphological operators and multiscale directional information. *Computer Graphics and Image Processing, 2004. Proceedings. 17th Brazilian Symposium on*, pages 122–129, Oct. 2004.
- [32] Workshop on Computer Vision Based Analysis in Sport Environments. Cvbase'06. <http://vision.fe.uni-lj.si/cvbase06/downloads.html>, December 2008.
- [33] S. Paris and F. Sillion. Robust acquisition of 3d informations from short image sequences. *Computer Graphics and Applications, 2002. Proceedings. 10th Pacific Conference on*, pages 127–136, 2002.
- [34] S. J. Park, K. Seo, J. Kim, and S. M. Song. *Advances in Multimedia Information Processing - PCM 2005*, volume 3767/2005 of *Lecture Notes in Computer Science*, chapter Automatic Dissolve Detection Scheme Based on Visual Rhythm Spectrum, pages 787–798. Springer Berlin / Heidelberg, 2005.
- [35] J. Pers and S. Kovacic. Computer vision system for tracking players in sports games. In *Proceedings of the First Int'l Workshop on Image and Signal Processing and Analysis*, pages 81–86, 2000.
- [36] J. Pers, G. Vuckovic, S. Kovacic, and B. Dezman. A low-cost real-time tracker of live sport events. In *In 2nd international symposium on image and signal processing and analysis*, pages 362–365, June 2001.
- [37] G. Pingali, A. Opalach, Y. Jean, and I. Carlbom. Visualization of sports using motion trajectories: providing insights into performance, style, and strategy. In *Visualization, 2001. VIS '01. Proceedings*, pages 75 – 83, October 2001.
- [38] Robert Shapiro. Direct linear transformation method for three-dimensional cinematography. *Research Quarterly*, 49(2):197–205, May 1978.
- [39] N. C. Simões. Detecção de algumas transições abruptas em sequências de imagens. Master's thesis, Institute of Computing, UNICAMP, 2003.
- [40] R. Tsai. A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses. *Robotics and Automation, IEEE Journal of [legacy, pre - 1988]*, 3(4):323–344, Aug 1987.
- [41] V. Vezhnevets. Gml matlab camera calibration toolbox. <http://research.graphicon.ru/calibration/gml-matlab-camera-calibration-toolbox.html>, October 2008.

- [42] J. Wang, F. Shia, J. Zhanga, and Y. Liu. A new calibration model of camera lens distortion. *Pattern Recognition*, 41(2):607–615, February 2008.
- [43] D.B. Yang, H.H. Gonzalez-Banos, and L.J. Guibas. Counting people in crowds with a real-time network of simple image sensors. *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pages 122–129 vol.1, Oct. 2003.
- [44] Z. Zhang. A flexible new technique for camera calibration. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(11):1330–1334, Nov 2000.
- [45] G. Zhu, Q. Huang, C. Xu, Y. Rui, S. Jiang, W. Gao, and H. Yao. Trajectory based event tactics analysis in broadcast sports video. In *MULTIMEDIA '07: Proceedings of the 15th international conference on Multimedia*, pages 58–67, New York, NY, USA, 2007.
- [46] G. Zhu, C. Xu, Q. Huang, and W. Gao. Automatic multi-player detection and tracking in broadcast sports video using support vector machine and particle filter. *Multimedia and Expo, 2006 IEEE International Conference on*, pages 1629–1632, July 2006.