

Perseguição de Marcadores para Análise de
Movimentos Humanos

Pascual Jovino Figueroa Rivero

Dissertação de Mestrado

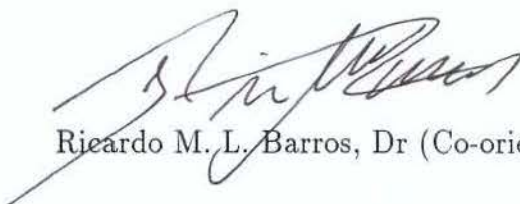
Perseguição de Marcadores para Análise de Movimentos Humanos

Este exemplar corresponde à redação final da Dissertação devidamente corrigida e defendida por Pascual Jovino Figueroa Rivero e aprovada pela Banca Examinadora.

Campinas, 22 de julho de 1998.

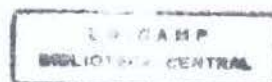


Neucimar Jerônimo Leite, Dr (Orientador)



Ricardo M. L. Barros, Dr (Co-orientador)

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



Perseguição de Marcadores para Análise de Movimentos Humanos

Pascual Jovino Figueroa Rivero¹

Julho de 1998

Banca Examinadora:

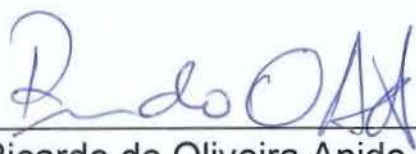
- Neucimar Jerônimo Leite, Dr (Orientador)
- Arnaldo de Albuquerque Araújo, Dr
- Ricardo Anido, PhD
- Alexandre Xavier Falcão, Dr (Suplente)

¹Supported in part by CAPES and FAPESP

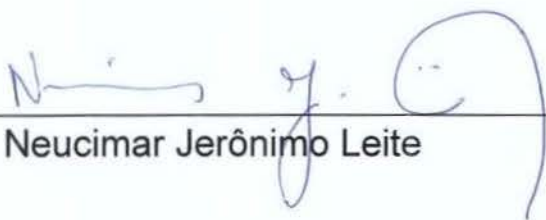
Tese de Mestrado defendida e aprovada em 26 de junho de 1998
pela Banca Examinadora composta pelos Professores Doutores



Prof. Dr. Arnaldo de Araújo Albuquerque



Prof. Dr. Ricardo de Oliveira Anido



Prof. Dr. Neucimar Jerônimo Leite

Prefácio

Esta dissertação aborda o problema de acompanhamento (tracking) do movimento humano numa sequência de imagens. O interesse pelo acompanhamento do movimento humano está relacionado com a análise e reconhecimento do movimento utilizado para diferentes aplicações, tais como, análise clínica, análise do movimento no esporte, ergonomia, coreografia de dança, etc.

Em particular, nós abordamos este problema utilizando marcadores (adesivos) atados ou colados ao corpo humano, diferentemente de outros métodos que utilizam modelos do corpo humano e seu movimento.

A utilização de marcadores permite obter resultados de maior precisão e que são de grande importância nas aplicações relacionadas com a biomecânica.

Neste trabalho apresentamos algoritmos para a perseguição automática dos marcadores numa sequência de imagens. O conjunto nestes algoritmos inclui a extração dos marcadores, a predição da posição nos próximos quadros, e sua identificação e classificação.

O método discutido aqui permite a implementação de sistemas de análise do movimento humano baseados em câmeras normais, ao contrário dos sistemas que utilizam câmeras especiais, em que os custos são elevados.

Um ponto importante deste trabalho é a introdução de ferramentas de processamento de imagens, em particular a morfologia matemática, para a extração dos marcadores.

Abstract

In this work we consider the problem of tracking human movements. The interest for this problem concerns the analysis and recognition of human motion used in a variety of applications, such as clinical gait analysis, improvement of the human performance in sports, ergonomics, choreography of dance, and so on.

In particular, we treat this problem using markers attached or fixed to human body, different from the methods based on human body and their motion models.

The use of markers allows us to get more precise results in applications, such as Biomechanics where this precision is very important.

We present some algorithms for tracking markers in a sequence of images, consisting in: extraction (segmentation) of the markers, prediction of their position in the next frames, and their recognition and classification.

The definition of these algorithms yields the implementation of systems for human motion analysis based on normal CCD cameras, unlike the systems based on special optoelectronic devices.

Agradecimentos

Gostaria de agradecer a todas as pessoas que de uma forma ou outra contribuíram para a finalização deste trabalho.

A meu pai e meus irmãos, por me darem todo o apoio e me convencerem que eu era capaz de superar todas as dificuldades.

A meu orientador, Prof. Neucimar J. Leite, pelo apoio constante no desenvolvimento deste trabalho e por sua paciência na correção da redação.

A meu co-orientador, Prof. Ricardo M. Barros, pelo apoio constante e por brindar todas as facilidades na utilização de equipamentos no laboratório de instrumentação para a biomecânica (LIB).

A meus amigos, pela alegria e o ânimo recebido nos momentos difíceis, em especial ao pessoal do LIB da Educação Física e ao pessoal do Instituto da Computação.

Ao povo brasileiro por fazer-me sentir em casa.

A CAPES e FAPESP pelo apoio financeiro

Conteúdo

Prefácio	v
Abstract	vi
Agradecimentos	vii
1 Introdução	1
2 Análise do Movimento Humano	4
2.1 Métodos baseados em modelos	4
2.2 Métodos baseados em extração de características	6
2.2.1 Marcadores ativos	7
2.2.2 Marcadores passivos	7
2.3 Definição de marcadores para sistemas baseados em câmeras CCD	8
2.4 Conclusão	10
3 Segmentação de Marcadores	13
3.1 Morfologia Matemática	14
3.1.1 Operações básicas: Dilatação e Erosão	15
3.1.2 Exemplos de outras operações morfológicas	16
3.1.3 Linha divisora de águas	17
3.1.4 Dinâmica em imagens numéricas	19
3.2 Detecção de contornos	20
3.3 Seleção de regiões	21
3.4 Algoritmos de segmentação para a extração de marcadores	23
3.4.1 Algoritmo baseado na intensidade dos marcadores	23
3.5 Conclusão	26
4 Predição da Trajetória	31
4.1 Extrapolação	31

4.2	Filtro de Kalman	32
4.2.1	Introdução	32
4.2.2	Aplicação para tracking do movimento humano	34
4.3	Análise dos resultados de predição	35
4.4	Conclusão	37
5	Casamento de Padrões	38
5.1	Técnicas de correlação	39
5.2	Método baseado em distância com tolerância	42
5.3	Método baseado nos valores maximais	42
5.4	Definição do modelo dos marcadores	42
5.5	Classificação das marcas e critérios de decisão	43
5.6	Conclusão	45
6	Implementação e Teste de Algoritmos para a Perseguição Automática de Marcadores	46
6.1	Características do sistema DVIDEOW	46
6.1.1	Entrada e saída de dados	48
6.1.2	Visualização dos dados	48
6.1.3	Ferramentas de processamento de imagens	49
6.1.4	Calibração e reconstrução tridimensional	49
6.1.5	Perseguição automática dos marcadores	50
6.2	Inicialização da perseguição automática de marcadores	50
6.3	Teste dos algoritmos de perseguição automática dos marcadores	58
6.4	Casos de oclusão ou perda de marcadores	59
6.5	Avaliação dos resultados	59
6.6	Conclusão	65
7	Conclusões	68
	Bibliografia	70

Lista de Tabelas

Lista de Figuras

2.1	exemplos de modelos do corpo humano.	5
2.2	exemplo de utilização de marcadores ativos	8
2.3	utilização de marcadores passivos	9
2.4	marcadores ativos (LEDs).	11
2.5	marcadores passivos (adesivos).	11
2.6	marcadores passivos (pinturas).	12
2.7	marcadores passivos (isopores).	12
3.1	elementos estruturantes típicos B_1 , B_2 e B_3 (o X indica a origem).	14
3.2	um exemplo de dilatação.	15
3.3	um exemplo de erosão.	15
3.4	exemplo de abertura e fechamento: a) a imagem original, b) o elemento estruturante, c) a abertura, d) o fechamento.	16
3.5	elementos estruturantes homotópicos L e M, em malhas 4- e 8-conexas. . .	17
3.6	uma superfície topográfica e definição intuitiva de LDA.	18
3.7	dinâmica de um máximo regional.	19
3.8	cálculo da dinâmica de um mínimo regional.	20
3.9	compacidade: (a) forma compacta, (b) forma não compacta.	21
3.10	excentricidade.	22
3.11	extensibilidade: (a) utilizando o envelope retangular, (b) o envelope retangular não representa corretamente a extensibilidade de curvas.	22
3.12	elementos estruturantes B , L e E	23
3.13	algoritmo para a segmentação de marcadores baseado em sua intensidade. .	26
3.14	seqüência de processamentos para segmentação dos marcadores (isopores). .	27
3.15	seqüência de processamentos para segmentação dos marcadores (LEDs). .	28
3.16	seqüência de processamentos para segmentação dos marcadores (adesivos). .	29
3.17	seqüência de processamentos para segmentação de um marcador (pintura). .	30
4.1	o ciclo do filtro de Kalman.	33
4.2	predição utilizando a extrapolação ponderada.	36

4.3	predição utilizando o filtro de Kalman.	37
5.1	um exemplo de correlação	40
5.2	subdivisão da região de busca.	44
6.1	o ambiente do DVIDEOW.	47
6.2	um cubo de referência utilizado para a calibração.	49
6.3	seqüência de procedimentos da perseguição automática.	51
6.4	janela de opções para a seleção de algoritmos da perseguição automática.	52
6.5	seleção de operações para a segmentação dos marcadores.	54
6.6	um exemplo de janela de parâmetros para a configuração de uma operação.	54
6.7	definição da máscara para as operações morfológicas.	55
6.8	janela das máscaras para afinamento e espessamento.	55
6.9	seleção de um limiar para operações de filtragem.	55
6.10	parâmetros do filtro de Kalman.	56
6.11	janela de parâmetros para o casamento de padrões.	57
6.12	trajetória dos marcadores para duas seqüências de imagens.	60
6.13	trajetória de marcadores do tipo LED.	61
6.14	trajetória de marcadores do tipo adesivo com casos de oclusão.	61
6.15	movimento na esteira.	62
6.16	trajetória de uma das marcas do movimento na esteira.	62
6.17	perseguição de marcadores adesivos para análise de respiração.	63
6.18	superfície do corpo reconstruída a partir de 14 pontos para un instante de respiração.	63
6.19	perseguição de marcadores para análise de corrida.	64
6.20	perseguição de um jogador de futebol.	64
6.21	Coordenadas X da trajetória dos marcadores obtida manualmente.	66
6.22	Coordenadas X da trajetória dos marcadores obtida automaticamente.	66
6.23	Coordenadas Y da trajetória dos marcadores obtida manualmente.	67
6.24	Coordenadas Y da trajetória dos marcadores obtida automaticamente.	67

Capítulo 1

Introdução

Visão computacional está relacionada com aplicações em diferentes áreas, com o objetivo de construir sistemas inteligentes capazes de interagir com o ser humano.

Nas diversas aplicações do reconhecimento do movimento humano, tais como realidade virtual, sistemas de vigilância, análise do movimento no esporte, coreografia de dança, tradução de linguagens de sinais e interfaces de usuários guiados por gestos, os objetivos das análises podem ser diferentes. Em alguns casos precisa-se determinar o tipo de movimento [10], em outros, reconhecer partes do corpo humano [20], reconhecer pessoas por sua forma de andar [21], determinar a postura do corpo em movimento [3, 10], analisar o movimento das articulações [11], etc.

De maneira geral, a abordagem destes problemas pode ser subdividida nos seguintes tópicos: captura de imagens com sensores ou vídeo-câmeras, digitalização e geração de sequência de imagens, extração do movimento, análise e reconhecimento do movimento.

A extração do movimento consiste em obter uma descrição ou representação deste, a partir de uma sequência de imagens. Esta fase engloba as etapas de extração (segmentação), reconhecimento e classificação dos objetos de interesse. Este processo de localizar e reconhecer os objetos, quadro a quadro, é denominado perseguição (*tracking*) de objetos.

A biomecânica, em particular, é uma das áreas em que a análise do movimento humano é de grande importância [7]. As aplicações nesta área podem ser exemplificadas pelo estudo do desempenho de pessoas submetidas a um treinamento físico, estudos relativos ao mecanismo de controle motor, avaliação de portadores de deficiência ou lesados cerebrais que apresentem seqüelas motoras, ou mesmo atletas e pacientes que necessitem avaliação dos progressos obtidos durante um processo recuperatório. Todos estes casos exigem a análise quantitativa, com um certo grau de exatidão, do movimento humano. Para tanto, são utilizados sistemas especiais, baseados em optoeletrônica e no uso de marcadores [2]. Estes sistemas se caracterizam pelo alto custo e o seu desempenho está fortemente

limitado às condições ambientais.

O fácil acesso às câmeras para aquisição de imagens, e do computador para simulação do complexo sistema visual humano, empregando técnicas modernas de processamento e análise de imagens, tem tornado crescente o interesse pela descrição automática do movimento do corpo humano. Este estudo permite a abordagem de problemas em diversas áreas, tais como medicina, esporte, realidade virtual, sistemas de vigilância, entre outros. Assim, foram propostos alguns métodos para a perseguição automática do movimento, a qual envolve a detecção e extração do movimento e sua representação. Entre estes podemos mencionar os métodos baseados em modelos predefinidos do corpo humano e seu movimento [1, 23, 3, 10]. A utilização destas técnicas exige uma definição dos modelos para se obter resultados corretos, o que não é uma tarefa fácil tendo em vista o grande número de segmentos e graus de liberdade no movimento do corpo humano.

O acompanhamento automático desse movimento é um problema complexo devido, entre outros fatores, à presença de articulações e à possibilidade de oclusão de algumas partes do corpo.

Recentemente, no trabalho apresentado em [7], foi proposto um sistema para a análise tridimensional do movimento humano, baseado em vídeo-câmeras CCD (*charged-coupled device*) e no uso de marcadores. Os marcadores, que representam componentes de interesse da seqüência de imagens, podem ser diodos emissores de luz, pinturas ou adesivos colados ou atados ao corpo, etc. Neste sistema, para não depender de equipamentos dedicados (câmeras e placas especiais), a extração do movimento é realizada por software. No entanto, o sistema não apresenta algoritmos sofisticados para a perseguição de marcadores, o que constitui o ponto de partida do presente trabalho.

Seguindo a idéia da utilização de marcadores para a extração do movimento, o objetivo principal deste trabalho é introduzir ferramentas de processamento de imagens, técnicas de predição e casamento de padrões, definindo algoritmos para a extração e perseguição automática dos marcadores. Estes algoritmos fazem parte de um sistema em fase de desenvolvimento, a ser utilizado no campo da biomecânica para a avaliação do sistema locomotor humano. Os algoritmos aqui definidos podem ser facilmente estendidos a outros tipos de movimento, como ilustram os exemplos do Capítulo 6.

A perseguição de marcadores pode ser subdividida em três etapas principais: segmentação, predição da posição e reconhecimento dos marcadores.

O processamento de imagens permite a extração dos objetos de interesse de uma cena através da segmentação. Em particular, a Morfologia Matemática contém ferramentas poderosas para esta finalidade, como será ilustrado posteriormente.

O filtro de Kalman [19] é uma técnica bastante empregada na perseguição de objetos, com o objetivo de predizer sua posição nos diferentes quadros da seqüência de imagens.

O reconhecimento de padrões [25, 29, 27] consiste em reconhecer e classificar os objetos

da imagem de acordo a sua forma, cor, textura, etc.

Este trabalho é o resultado de uma cooperação estabelecida entre a Faculdade de Educação Física e o Instituto de Computação da UNICAMP. O Capítulo 2 apresenta alguns conceitos relacionados com a análise do movimento humano, a seqüência de procedimentos e alguns métodos para a extração deste movimento. Também são citadas algumas características dos sistemas comerciais e o tipo de marcadores utilizados. No Capítulo 3 é abordado o problema de extração de marcadores. Inicialmente, são introduzidos alguns conceitos básicos de Morfologia Matemática e descritos alguns algoritmos propostos para a extração destes marcadores. O Capítulo 4 apresenta algumas técnicas de predição, para limitar a área de busca dos marcadores e a região de segmentação. O Filtro de Kalman é abordado detalhadamente. No Capítulo 5, são analisados alguns algoritmos para o casamento de padrões, apresentando alternativas para diferentes tipos de marcadores. O Capítulo 6 aborda as características do sistema em desenvolvimento, em que foram implementados e testados os algoritmos definidos nos capítulos anteriores. Finalmente, a conclusão é apresentada no Capítulo 7.

Capítulo 2

Análise do Movimento Humano

O interesse pelo estudo do movimento humano tem origem nos tempos remotos, segundo os recitos de Aristóteles e Leonardo da Vinci [8]. Apenas nestas últimas décadas, com a aparição das novas tecnologias, tais como as câmeras fotográficas rápidas, vídeo-câmeras e computadores digitais, foi possível explorar efetivamente novos horizontes nesta área.

O interesse crescente pelo estudo do movimento humano tem sido motivado pelo progresso na área de visão computacional, em que resultados consideráveis foram obtidos em relação à extração do movimento, e no desenvolvimento de sistemas optoeletrônicos para aplicações específicas na área da biomecânica.

De modo geral, para se determinar o movimento a partir de uma sequência de imagens, cada quadro da sequência é analisado e, de acordo com as aplicações e métodos desenvolvidos, são extraídos os objetos de interesse que podem ser marcas, contornos, figuras geométricas, modelos etc. Esses objetos são perseguidos ao longo de toda a sequência.

Sucintamente, os métodos relacionados com a análise do movimento humano podem ser estruturados em *top-down*, baseados na reconstrução geométrica do corpo (usando modelos), e *bottom-up*, baseados em características de baixo nível da imagem (pontos, vértices, contornos, regiões, etc). As próximas seções abordam estes dois tipos de métodos e apresentam os marcadores considerados neste trabalho.

2.1 Métodos baseados em modelos

Em geral, e para este estudo específico, o modelo humano pode ser considerado a partir de dois tipos de parâmetros [18]. Os parâmetros que definem a forma (corpo), os quais mudam para diferentes pessoas, e os parâmetros de controle (movimento), relacionados com a postura do corpo.

O acompanhamento do movimento através de modelos envolve a busca de parâmetros de controle que definam a postura ótima, tal que sua projeção coincida com os dados

da imagem original. O corpo humano possui uma estrutura bem definida que pode ser decomposta numa hierarquia natural de partes. Exemplos destes modelos são mostrados na Figura 2.1. Em seguida são apresentados alguns métodos de análise do movimento baseados em modelos.

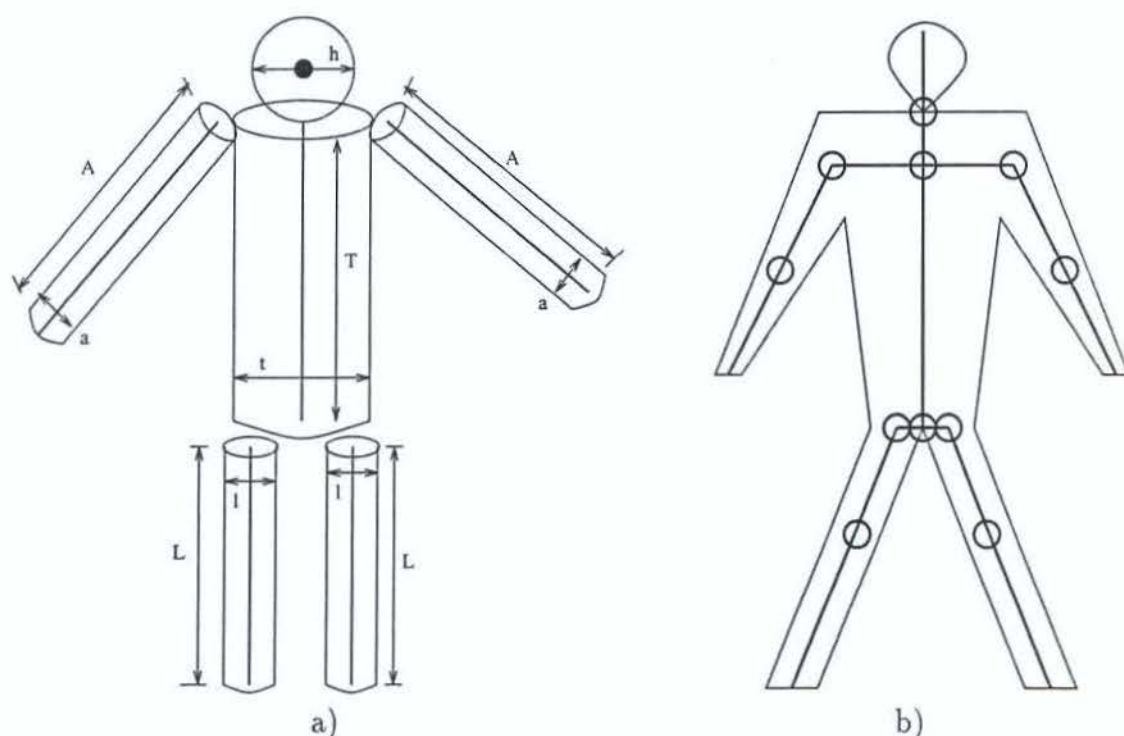


Figura 2.1: exemplos de modelos do corpo humano.

Akita [1] analisa o movimento humano utilizando um modelo em forma de 6 segmentos (Figura 2.1a): os braços, as pernas, o torso e a cabeça. Embora o corpo seja modelado usando cilindros, apenas sua projeção no plano da imagem é utilizada. Para prever a posição de uma parte do corpo com respeito a outra parte, na imagem atual, é utilizada uma sequência de chave de quadros. A sequência de chave de quadros consiste de uma sequência ordenada de figuras de segmentos lineares (modelos), previamente definidas, onde cada figura é diferente de sua predecessora e sucessora.

Para se determinar cada parte do corpo são extraídos os contornos da imagem. O reconhecimento das partes é realizado na seguinte ordem: pernas, cabeça, braços e torso, considerando que as pernas são as componentes mais “estáveis”, ou seja, fácil de reconhecer. O torso é menos estável porque muda de forma com maior frequência devido a oclusões.

Para se determinar a correspondência entre os quadros é utilizado um dos seguintes

métodos. Quando a mudança de posição de um segmento é muito pequena, a posição pode ser predita a partir de uma métrica específica (*window code distance*). Se através deste método não pode ser determinada a correspondência, então são utilizados a sequência de chave de quadros para determinar a postura atual do corpo.

Uma desvantagem deste método é que a sequência de chave de quadros fornece apenas informação sobre a posição relativa dos segmentos do corpo, e não é possível obter informação temporal.

A análise de movimento de um caminhante é tratada no trabalho de K. Rohr [23]. O autor utiliza 14 cilindros com corte transversal elíptico para modelar o corpo humano. Para o respectivo modelo são usados dados do estudo anterior sobre o movimento. A análise da imagem consiste da segmentação das diferentes regiões, utilizando algoritmos de detecção de mudanças. Para uma melhor definição dos parâmetros, o modelo do contorno é comparado com o contorno da imagem original. Os parâmetros do modelo, nos próximos quadros, são obtidos aplicando-se uma análise de regressão linear a partir do filtro de Kalman.

Uma outra forma de representar a projeção do corpo humano no plano da imagem é modelar o corpo como uma coleção de linhas antiparalelas denominados *apars*. Na verdade, o corpo é modelado pelas regiões fechadas (Figura 2.1b). No trabalho de Leung e Yang [20], o corpo é representado como uma coleção de seis *apars*. A posição relativa de cada segmento é menos importante, maior importância é dada à razão entre os segmentos representando o modelo do corpo. Inicialmente, a imagem de cada quadro é segmentada aplicando-se diferenças de imagens baseada nas informações dos quadros anteriores. Posteriormente, são extraídas as linhas antiparalelas a partir da imagem segmentada.

A identificação das partes é realizada utilizando considerações, tais como: a razão entre os braços, pernas, cabeça e torso, a razão largura/comprimento da cabeça etc., valores estes determinados de forma experimental.

Outros métodos de perseguição do movimento humano por modelos podem ser vistos em [16, 10, 21].

2.2 Métodos baseados em extração de características

Os métodos baseados diretamente nas características dos objetos da imagem, sem definição prévia de modelos, consistem basicamente das seguintes três fases: extração de características, correspondência de características, e agrupamento/partição de características em grupos de partes articuladas. Pontos específicos da imagem, contornos ou regiões são comumente utilizados como características.

Considerando o corpo humano como um modelo de segmentos rígidos, unidos pelas articulações, a extração do movimento do corpo humano pode ser reduzida à extração do

movimento das articulações [2]. A partir destas considerações foram implementados sistemas optoeletrônicos para a análise tridimensional do movimento humano. Estes sistemas comerciais, tais como OPTOTRACK, ELITE, WATSMART, SELSPOT etc., utilizam marcadores e câmeras especiais e são baseados em dispositivos sensíveis à posição (PSD), em *scanners* optomecânicos, em câmeras CCD, etc. Em [9, 28] são comparadas as principais características destes sistemas, cuja particularidade é o alto custo e o emprego de marcadores ativos ou passivos mencionados a seguir.

2.2.1 Marcadores ativos

Os marcadores ativos são dispositivos que emitem algum tipo de energia, geralmente diodos emissores de luz (LED). Estes precisam de energia externa e podem ser ativados através de uma estroboscopia para facilitar a identificação automática e individual das marcas. A maior parte dos sistemas optoeletrônicos trabalham com LEDs que emitem luz infra-vermelha.

A vantagem no uso destes marcadores reside na facilidade de se identificar automaticamente (e por *hardware*) cada uma das marcas, e de se poder aumentar a frequência de captura quando poucos marcadores são considerados. Uma desvantagem imediata refere-se ao grau de liberdade do corpo, limitado pelo uso de fios de conexão. Os LEDs têm um ângulo restrito de emissão de luz (Figuras 2.2 e 2.4), o que complica a detecção da marca em pontos onde a intensidade é muito baixa. Por outro lado, pontos onde esta intensidade é muito forte acabam comprometendo a precisão da detecção.

É importante ressaltar que o tempo indispensável aos sensores, para detecção das marcas individuais limita o número de marcadores utilizados pelo sistema, que tem de considerar, ainda, a frequência de captura em caso de movimentos rápidos.

2.2.2 Marcadores passivos

Os marcadores passivos são objetos que refletem luzes projetadas e que aparecem na imagem como manchas claras. Estas marcas são, por exemplo, objetos cobertos com papel refletor, atados ao sujeito ou, ainda, pinturas ou adesivos colados à pele (Figura 2.3).

Os marcadores passivos não precisam de fios de conexão, o que confere maior liberdade na realização de movimentos rápidos. A possibilidade de captura simultânea desses marcadores pelos sensores é outra vantagem que, teoricamente, permite que um número ilimitado de marcadores seja considerado. Aqui, a identificação das marcas é feita por *software* e problemas ligados a sua etiquetagem, tais como sobreposição, oclusão etc. precisam ser eficientemente abordados.

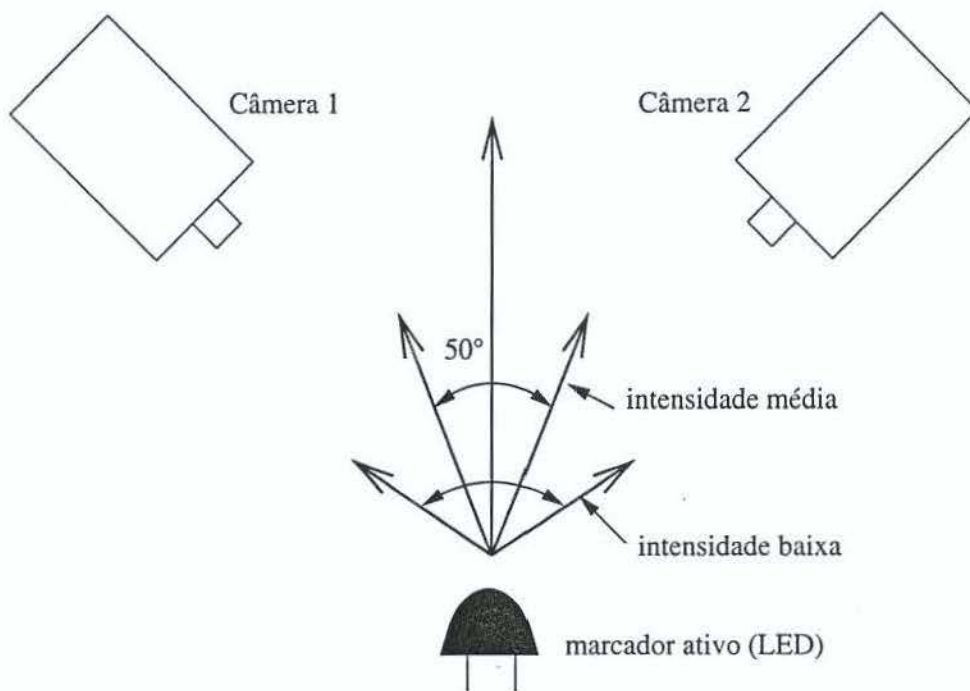


Figura 2.2: exemplo de utilização de marcadores ativos

O fato dos marcadores passivos poderem ser capturados simultaneamente pelos sensores é uma vantagem do ponto de vista de flexibilidade do sistema que necessita de *software* adequado para diferenciar e etiquetar cada um desses marcadores.

2.3 Definição de marcadores para sistemas baseados em câmeras CCD

O alto custo dos sistemas comerciais disponíveis deve-se, principalmente, aos sensores especiais utilizados na captura de imagens e ao *hardware* dedicado à extração e reconhecimento dos marcadores.

Com o desenvolvimento considerável dos computadores, tanto na capacidade de cálculo como no armazenamento de grandes quantidades de informação, a tendência atual é resolver estes problemas por *software*, conduzindo a sistemas de baixo custo baseados apenas em câmeras CCD normais. Desta forma espera-se que, num futuro próximo, tais sistemas possam estar disponíveis a um maior número de pesquisadores e laboratórios.

No trabalho apresentado em [8] é analisado o sistema ELITE em que parte da detecção dos marcadores é executada por *hardware*, deixando para o computador resolver o problema da correspondência quadro a quadro dos marcadores, a partir destes dados, sem outras informações adicionais.

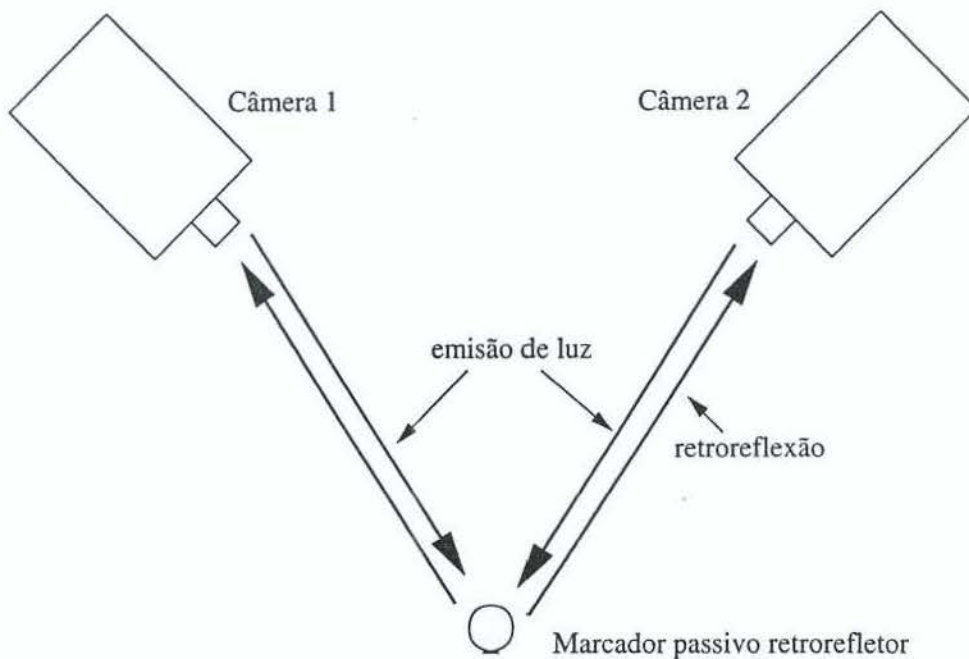


Figura 2.3: utilização de marcadores passivos

Neste sistema, são utilizados marcadores passivos. As imagens são capturadas por sensores do tipo câmeras CCD, com uma frequência de 50 ou 100 Hz. As imagens são digitalizadas com 16 níveis de cinza, numa matriz de 256x256 pixels.

O processador encarregado da detecção dos marcadores calcula a correlação entre a imagem digitalizada e um padrão predefinido. O padrão é uma matriz de tamanho 6x6 configurada de forma a obter uma alta correlação com o marcador e uma baixa correlação com o fundo da imagem. Para a correspondência dos marcadores, é levada em consideração a posição relativa dos pontos da imagem em relação ao dados do modelo.

O trabalho apresentado em [7] propõe um sistema de análise tridimensional do movimento baseado apenas em câmeras CCD e num computador para a extração e reconhecimento dos marcadores. Um dos problemas principais na implementação de tais sistemas refere-se à definição dos marcadores e dos algoritmos para a sua detecção e reconhecimento.

A utilização de câmeras CCD permite maior flexibilidade na seleção dos movimentos a serem analisados, diminui as restrições ao ambiente de coleta, permitindo a utilização tanto de marcadores ativos (LEDs) quanto de marcadores passivos (isopores esféricos, pinturas, adesivos etc.) O objetivo inicial é definir marcadores que possam ser diferenciados em relação ao restante da imagem, facilitando o reconhecimento de cada um deles através de *software*.

Neste trabalho, foram considerados alguns destes marcadores e possíveis algoritmos para sua extração e reconhecimento. O primeiro tipo de marcadores são os LEDs (Figura 2.4) que apesar das desvantagens mencionadas anteriormente podem ser utilizados em ambientes controlados. Estes LEDs, que não podem ser controlados à distância, não precisam de cabos mas de pilhas pequenas e podem ser empregados em número de filmagens limitado.

Outros tipos de marcadores considerados são os adesivos, pinturas ou isopores (Figuras 2.5, 2.6, 2.7, respectivamente). Os adesivos têm a vantagem de serem de fácil manipulação, mas com a possibilidade de se descolarem durante o movimento, ao contrário das pinturas cuja maior dificuldade é a própria realização das marcas. Os isopores são atados ou colados ao sujeito, o que restringe um pouco o movimento, como no caso dos adesivos ou LEDs, mas conservam a projeção circular que pode ser explorada na sua detecção.

2.4 Conclusão

Neste capítulo foram abordados dois métodos relacionados com a análise do movimento humano: métodos baseados em modelos e em extração de características. Foram abordados alguns tipos de marcadores e, finalmente, ilustrados os marcadores a serem considerados neste trabalho. O próximo capítulo define algumas formas de extração dos marcadores mencionados nesta seção.



Figura 2.4: marcadores ativos (LEDs).



Figura 2.5: marcadores passivos (adesivos).

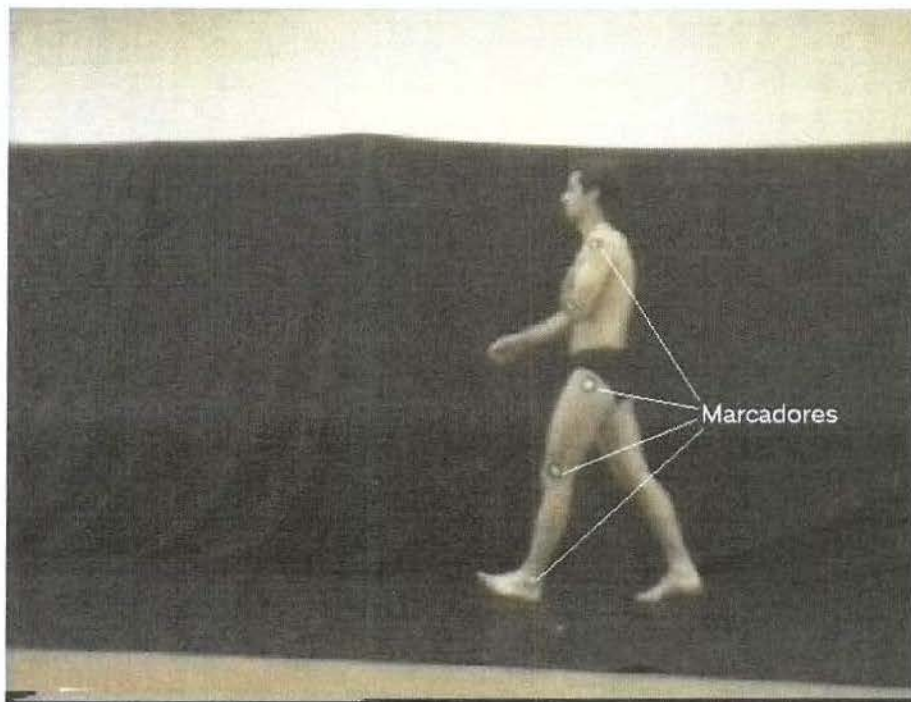


Figura 2.6: marcadores passivos (pinturas).



Figura 2.7: marcadores passivos (isopores).

Capítulo 3

Segmentação de Marcadores

Um dos pontos importantes deste trabalho é a introdução de técnicas de processamento de imagens, mais precisamente de segmentação, na perseguição automática de marcadores. Informalmente, a segmentação consiste em processar os dados da imagem original de modo a obter uma imagem particionada com algum significado relevante (regiões ou segmentos) [29].

A segmentação facilita a extração de características da imagem, muito importante para o reconhecimento de padrões empregados em visão computacional, por exemplo.

Os métodos de segmentação podem ser subdivididos em três grupos, de acordo com a característica dominante considerada: segmentação baseada num conhecimento global da imagem ou de suas partes, segmentação baseada em contornos e segmentação baseada em regiões [27].

A binarização da imagem através de um limiar é uma das formas mais direta de segmentação que em alguns casos pode ser usada para separar facilmente os objetos do fundo da imagem [24].

A segmentação baseada nos contornos da imagem é uma técnica amplamente empregada na área de visão computacional, já que muitas das características da forma dos objetos podem ser determinadas a partir de seus contornos [17].

Os métodos baseados em regiões definem zonas homogêneas da imagem de acordo com algum critério de similaridade dos pixels. Por exemplo, o método de crescimento de regiões (*region merging*) agrupa regiões ou pixels vizinhos em regiões maiores. Um outro método é o de subdivisão de regiões (*region splitting*) que, partindo da imagem original, a subdivide em regiões menores, de acordo com um grau de similaridade. Estas duas técnicas podem ser combinadas (*split-and-merge*) visando compensar erros de decisão oriundos do emprego de um único método [27].

O algoritmo que define as linhas divisoras de água - LDA (*watershed*) de uma função, e que constitui a base da segmentação morfológica considerada neste trabalho, é um outro

exemplo de segmentação baseada em regiões [5].

Para a extração dos marcadores foi definida uma sequência de processamentos utilizando ferramentas de Morfologia Matemática. O pré-processamento das imagens permite ressaltar os marcadores e definir melhor seus contornos, facilitando a sua extração.

Etiquetar regiões definidas pelos contornos e eliminar aquelas não obedecendo a determinados critérios é o passo final do processo de segmentação.

A próxima seção aborda informalmente algumas funções e conceitos básicos de morfologia matemática, entre elas são introduzidas duas operações importantes para este trabalho: a dinâmica, e a linha divisora de águas (LDA). As duas seções seguintes abordam a detecção de contornos e algumas formas de seleção de regiões de interesse da imagem a partir destes contornos. A última seção reúne todas estas operações na definição de algoritmos para a extração dos marcadores mencionados anteriormente.

3.1 Morfologia Matemática

A Morfologia Matemática [25, 13], inicialmente formalizada para imagens binárias e posteriormente estendida para imagens em níveis de cinza, define ferramentas para processamento de imagens baseada na geometria e forma de seus objetos. As operações morfológicas simplificam as imagens, mantendo as principais características de seus componentes [27].

Seja $f(x)$ uma imagem binária e discreta, $\{f(x) \in \{0, 1\} : x \in Z^2\}$, onde Z denota o conjunto de números inteiros. Esta imagem pode ser representada por um conjunto X definido como $X = \{x \in Z^2 : f(x) = 1\}$. A transformação morfológica ϕ é dada pela relação da imagem (conjunto de pontos X) com outro conjunto de pontos B , denominado elemento estruturante. Elementos estruturantes típicos são mostrados na Figura 3.1.

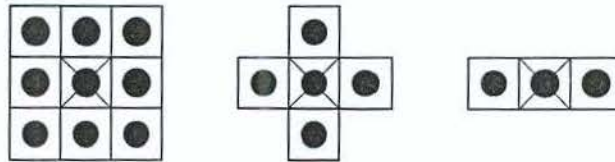


Figura 3.1: elementos estruturantes típicos B_1 , B_2 e B_3 (o X indica a origem).

O complemento X^c de um conjunto X é definido como $X^c = \{x \in Z^2 : f(x) = 0\}$. Um elemento estruturante B é denominado homotópico se ele não altera a conexidade e o número de componentes conexos da imagem após uma operação com este elemento.

3.1.1 Operações básicas: Dilatação e Erosão

As duas operações básicas da morfologia matemática são a dilatação e a erosão. A dilatação de uma imagem binária X com um elemento estruturante B pode ser definida a partir da adição vetorial dos elementos dos conjuntos X e B :

$$(X \oplus B) = \{x + b \in Z^2 : \forall x \in X \wedge b \in B\}. \quad (3.1)$$

A Figura 3.2 apresenta um exemplo de dilatação.

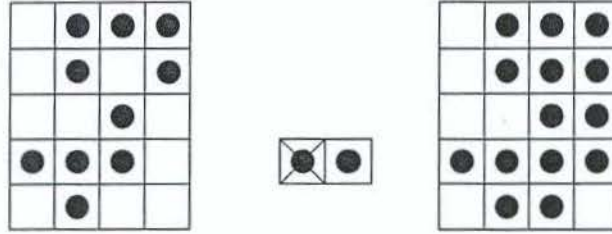


Figura 3.2: um exemplo de dilatação.

De forma análoga, a erosão de uma imagem binária X com um elemento estruturante B pode ser definida a partir da subtração vetorial dos conjuntos X e B :

$$(X \ominus B) = \{d \in Z^2 : d + b \in X, \forall b \in B\}. \quad (3.2)$$

Um exemplo de erosão é mostrado na Figura 3.3

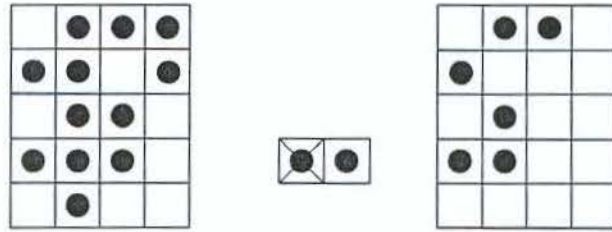


Figura 3.3: um exemplo de erosão.

As duas operações acima podem ser estendidas para uma imagem numérica f (em níveis de cinza) [13]. Sejam $F, G \subseteq Z^2$ e duas funções $f : F \rightarrow Z$ e $g : G \rightarrow Z$, representando imagens em níveis de cinza. Assim, a dilatação, $f \oplus g$, e a erosão, $f \ominus g$, podem ser definidas por:

$$(f \oplus g)(x) = \max_{z \in G, x-z \in F} \{f(x-z) + g(z)\}. \quad (3.3)$$

$$(f \ominus g)(x) = \min_{z \in G, x+z \in F} \{f(x+z) - g(z)\}. \quad (3.4)$$

Para um elemento estruturante planar (de representação binária, fig. 3.1) a dilatação e a erosão podem ser representadas por:

$$(f \oplus B)(x) = \max_{z \in B, x-z \in F} \{f(x-z)\}. \quad (3.5)$$

$$(f \ominus B)(x) = \min_{z \in B, x+z \in F} \{f(x+z)\}. \quad (3.6)$$

3.1.2 Exemplos de outras operações morfológicas

As operações morfológicas mais complexas são definidas a partir de combinações das operações de dilatação e erosão. Nesta subseção são abordadas algumas destas operações basicamente para imagens numéricas. As definições correspondentes para imagens binárias podem ser encontradas em [4, 25, 30].

- *abertura e fechamento*. Aplicando-se iterativamente a dilatação e a erosão, é possível eliminar alguns detalhes específicos da imagem, menores que o elemento estruturante, sem destruir a geometria global das partes não suprimidas. A *abertura*, por exemplo, suaviza os contornos, elimina picos agudos e pequenas ilhas (Figura 3.4). Esta operação é definida por:

$$f \circ B = (f \ominus B) \oplus B \quad (3.7)$$

Por sua vez, o *fechamento* fecha aberturas agudas e golfos longos e finos, elimina pequenos buracos e espaços entre contornos (fig. 3.4). Esta operação é definida por:

$$f \bullet B = (f \oplus B) \ominus B \quad (3.8)$$

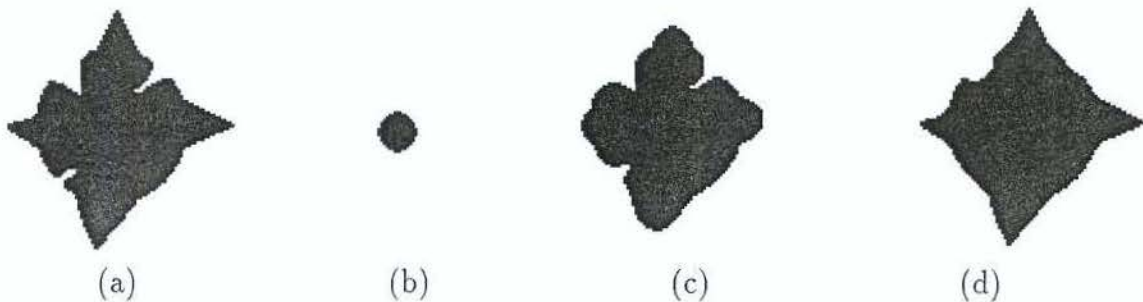


Figura 3.4: exemplo de abertura e fechamento: a) a imagem original, b) o elemento estruturante, c) a abertura, d) o fechamento.

- *afinamento e espessamento*. Estas operações morfológicas utilizam um elemento estruturante homotópico composto B . Este é denotado por um par de conjuntos

disjuntos, $B = (B_1, B_2)$, e pode ser representado por uma única matriz em que B_1 contém os elementos de valor "1", e B_2 os elementos de valor "0", como indicado na Figura 3.5. Um asterisco "*" na matriz denota um elemento que pode assumir valor "0" ou "1" indiferentemente.

Na Figura 3.5, são mostrados alguns elementos estruturantes homotópicos utilizados para o afinamento em malhas 4- e 8-conexas.

A função $g(x)$ correspondente ao afinamento da imagem f por B , denotada por $f \oslash B$, é dada por:

$$g(x) = \begin{cases} \max_{y \in B_2} [f(y)], & \text{se } \max_{y \in B_2} [f(y)] < f(x) \leq \min_{y \in B_1} [f(y)] \\ f(x), & \text{em caso contrário} \end{cases} \quad (3.9)$$

Analogamente, o espessamento, denotado por $f \odot B$, é definido por:

$$g(x) = \begin{cases} \min_{y \in B_2} [f(y)], & \text{se } \max_{y \in B_2} [f(y)] \leq f(x) < \min_{y \in B_1} [f(y)] \\ f(x), & \text{em caso contrário} \end{cases} \quad (3.10)$$

Na prática, o afinamento e o espessamento são aplicados de forma iterativa, utilizando uma família de configurações homotópicas. Uma família de configurações homotópicas é obtida ao girar um elemento estruturante homotópico B de um ângulo dado (por exemplo, 90 e/ou 45 graus para malhas 4- e 8-conexas) de tal forma que a mudança de estado do ponto central, nestas configurações, não altera localmente o número de componentes conexos da vizinhança.

$$L_4 = \begin{bmatrix} 0 & 0 & 0 \\ * & 1 & * \\ 1 & 1 & 1 \end{bmatrix} \quad M_4 = \begin{bmatrix} * & 0 & * \\ * & 1 & * \\ 1 & 1 & 1 \end{bmatrix} \quad L_8 = \begin{bmatrix} * & 0 & * \\ 1 & 1 & * \\ * & 1 & * \end{bmatrix} \quad M_8 = \begin{bmatrix} * & 0 & * \\ 1 & 1 & 1 \\ * & 1 & * \end{bmatrix}$$

Figura 3.5: elementos estruturantes homotópicos L e M, em malhas 4- e 8-conexas.

3.1.3 Linha divisora de águas

Uma das ferramentas morfológicas mais importantes em segmentação de imagens refere-se à transformação de *watershed* ou linha divisora de águas (LDA) de uma função [5].

Uma definição intuitiva de LDA pode ser dada considerando a imagem f como uma superfície topográfica. Esta superfície apresenta mínimos regionais, como mostra a Figura 3.6. Imagine que esses mínimos são perfurados e que os submergimos progressivamente em água de tal modo que o nível continue o mesmo em toda a superfície. Durante o

processo de inundação são construídos diques para evitar a união de águas provenientes de dois mínimos diferentes. Ao final, o conjunto dos diques representando fronteiras entre os mínimos da superfície constitui a LDA da função imagem.

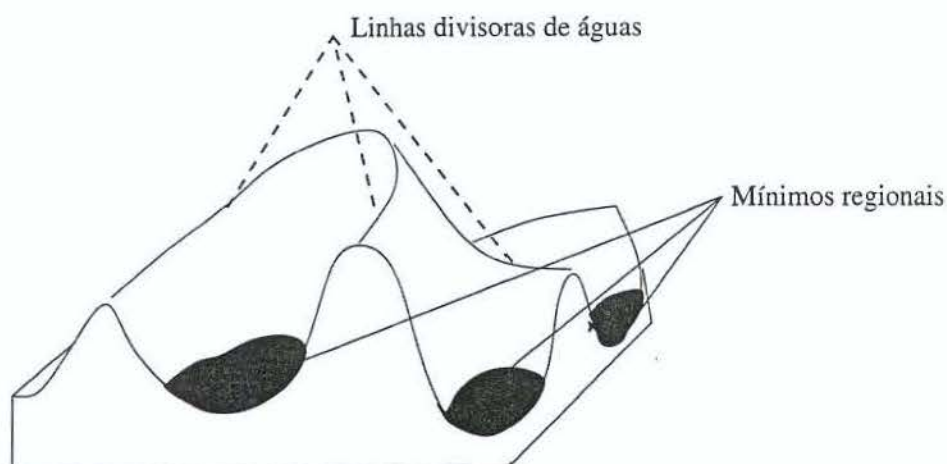


Figura 3.6: uma superfície topográfica e definição intuitiva de LDA.

Uma forma direta de se implementar a LDA é aplicar o afinamento com uma família homotópica de elementos estruturantes até a idempotência ou seja até não obter mais mudanças na imagem. Este método, além de constituir uma representação aproximada da LDA, é lento em termos de execução. Métodos mais eficientes que simulam o processo de imersão da superfície topográfica foram propostos.

Por exemplo, em [5] é apresentado um algoritmo baseado no uso de marcadores. Os marcadores são utilizados para evitar o problema da supersegmentação oriunda dos mínimos regionais insignificantes ou gerados por ruídos na imagem. Desta forma, no processo de inundação, são utilizados apenas aqueles mínimos escolhidos como marcadores. Neste algoritmo, é criada inicialmente uma fila com todos os níveis de prioridade representados pelo valor dos níveis de cinza da imagem, e todos os pontos que se encontram na fronteira dos marcadores são introduzidos nesta fila. Além disso, é criada uma imagem de etiqueta dos marcadores. Posteriormente, cada ponto é retirado da fila e analisado. Se o ponto extraído tem apenas uma região vizinha etiquetada, este é inserido a essa região (recebe a etiqueta correspondente). Se o ponto é vizinho a duas regiões com diferentes etiquetas, o ponto é considerado ponto de LDA. A dificuldade deste método está na determinação inteligente dos marcadores. Outro método baseado na simulação de imersões pode ser encontrado em [31].

3.1.4 Dinâmica em imagens numéricas

Os extremos (máximo ou mínimo regionais) de uma superfície topográfica contêm informações importantes para sua análise. Eles podem servir, entre outros, para filtrar a imagem original ou definir marcadores a serem utilizados no algoritmo de LDA.

A dinâmica [12] é uma das características que permite avaliar os extremos de uma imagem a partir de seus contrastes. A dinâmica de um máximo regional M , de uma imagem numérica f , é definida como o desnivelamento mínimo a ser atingido quando, partindo-se de M , se deseja alcançar um ponto de maior altitude (Figura 3.7). Assim:

$$\text{dyn}(M) = f(M) - \sup_{C=(p_0, p_1, \dots, p_n), p_0 \in M \wedge f(p_n) > f(p_0)} \{\inf\{f(x), x \in C\}\} \quad (3.11)$$

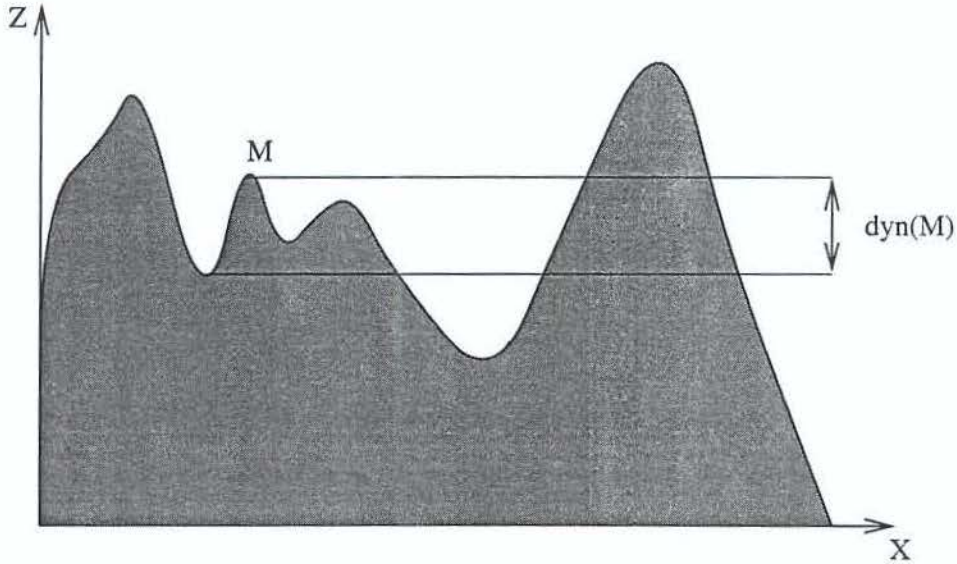


Figura 3.7: dinâmica de um máximo regional.

No trabalho em [12], é apresentado um algoritmo para a implementação da dinâmica que consiste em inverter a imagem e determinar a dinâmica dos mínimos regionais, considerando (Figura 3.8):

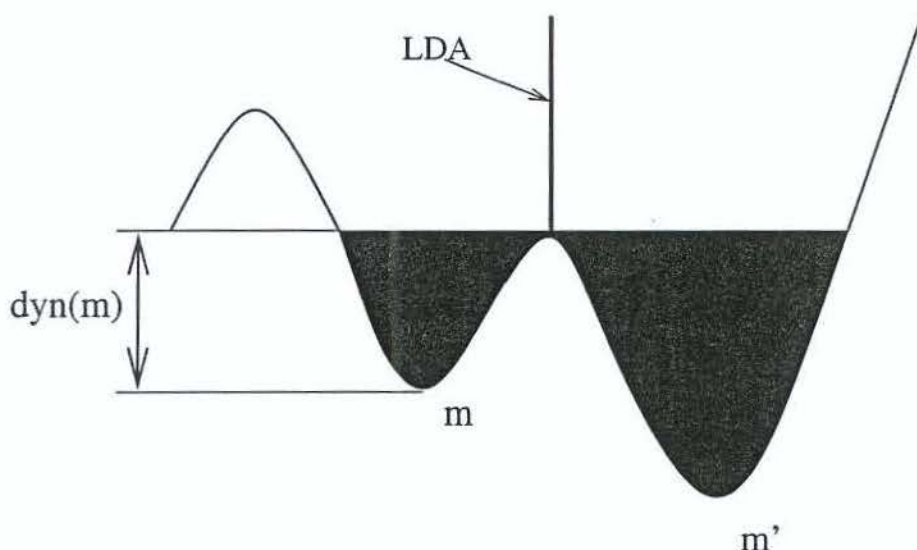
$$\text{dyn}(m) = h - f(m), \quad (3.12)$$

em que h é o nível mínimo tal que o “lago” proveniente da fonte m encontre um outro lago cuja fonte m' seja mais profunda (veja Figura 3.8).

Para se calcular a dinâmica de todos os mínimos, pode-se considerar o princípio de implementação da LDA. Durante o processo de inundação, cada mínimo possui uma etiqueta diferente. No momento em que duas águas de diferentes mínimos se encontram, ao invés de se construir diques, é calculada a dinâmica do menor mínimo, segundo a

equação 3.12, e toda a região é etiquetada com o valor da etiqueta do mínimo mais profundo (um lago absorve o outro).

Ao final do algoritmo, tem-se uma imagem de única etiqueta associada à região mais profunda. A dinâmica desta região recebe um valor máximo global (por exemplo 255 para imagens em nível de cinza). Caso o nível dos dois mínimos seja igual, no momento do encontro das duas águas, pode-se escolher um deles para absorver o outro.



3.2 Detecção de contornos

Os contornos de uma imagem são o resultado de mudanças em alguma propriedade física ou espacial de superfícies iluminadas. A idéia básica da detecção de contornos consiste em localizar a descontinuidade da intensidade na imagem [22]. Na literatura, pode-se encontrar uma grande variedade de métodos para a detecção de contornos.

Um exemplo bastante simples é representado pelo gradiente morfológico que pode ser definido por:

$$g(x) = (f(x) \oplus B) - (f(x) \ominus B) \quad (3.13)$$

ou ainda :

$$g(x) = (f(x) \oplus B) - f(x) \quad (3.14)$$

e

$$g(x) = f(x) - (f(x) \ominus B) \quad (3.15)$$

Outras técnicas de detecção de contornos empregam operadores diferenciais de primeira ou de segunda ordem. Estes ressaltam os contornos mas também amplificam o ruído da cena. Alguns trabalhos discutem comparativamente alguns destes operadores mais conhecidos [17, 26, 27].

3.3 Seleção de regiões

A representação e a descrição de regiões segmentadas é bastante utilizada na etapa de reconhecimento de padrões, por exemplo.

A análise das características particulares dos objetos permite a sua seleção ou classificação no conjunto dos objetos de uma cena. Neste sentido, faz-se necessária uma descrição de regiões através de alguns parâmetros de forma. Entre estes podemos citar [27]:

- **área.** Uma propriedade importante de uma região é sua área, definida pelo número de pixels que a constitui.
- **perímetro.** Outra característica de interesse de uma região é o perímetro, dado pelo número de pixels que compõem o seu contorno.
- **Compacidade.** Um índice de compacidade é dado pela seguinte relação:

$$\text{compacidade} = \frac{\text{área}}{l}$$

em que l representa o perímetro do contorno. Um exemplo de uma região compacta e outra não compacta é ilustrado na Figura 3.9.

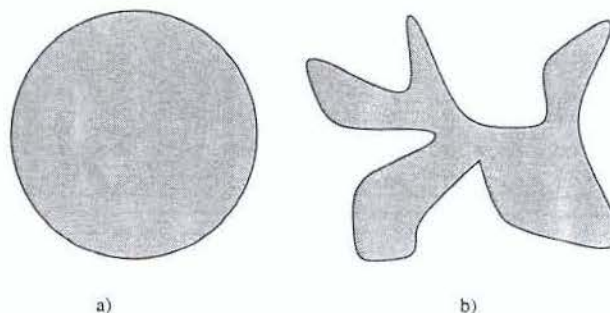


Figura 3.9: compacidade: (a) forma compacta, (b) forma não compacta.

- **Excentricidade.** Este parâmetro de forma é dado pela relação entre o comprimento da máxima corda A e da máxima corda B, perpendicular a A, como indicado na

Figura 3.10:

$$excentricidade = \frac{A}{B}$$

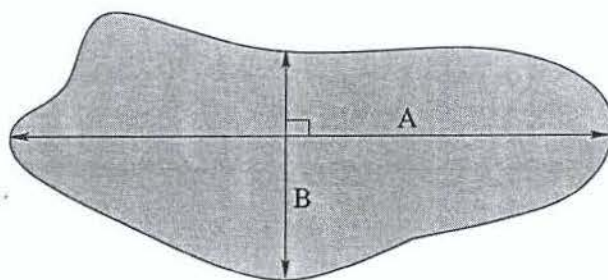


Figura 3.10: excentricidade.

- **Extensibilidade.** Definida como a razão entre a largura e o comprimento do mínimo envoltório retangular de uma região (Figura 3.11a):

$$excentricidade = \frac{a}{b}$$

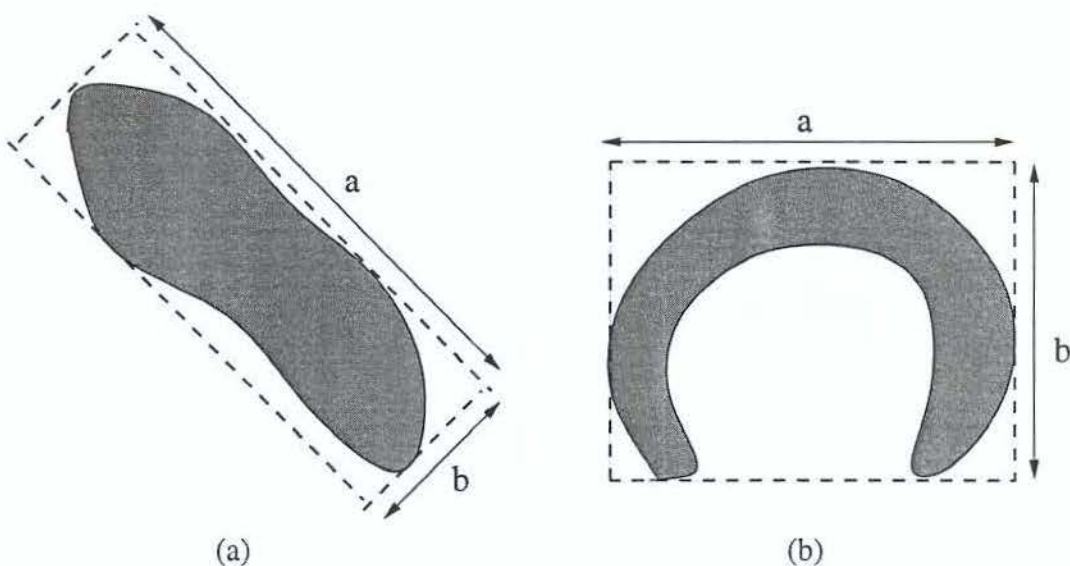


Figura 3.11: extensibilidade: (a) utilizando o envelope retangular, (b) o envelope retangular não representa corretamente a extensibilidade de curvas.

Como este critério não pode ser utilizado para regiões curvas (Figura 3.11b), a extensibilidade pode ser definida, ainda, como a razão entre a área da região e o

número de erosões necessárias à sua eliminação:

$$extensibilidade = \frac{\text{área}}{(2d)^2},$$

em que d é o número de erosões.

3.4 Algoritmos de segmentação para a extração de marcadores

A seguir, são apresentados algumas operações de segmentação/extração dos marcadores definidos no Capítulo 2, a partir de informações dos seus contornos.

Inicialmente, a região f da imagem original a ser processada e de tamanho $(2M + 1) \times (2N + 1)$ é dada por:

$$f = I(x, y) \quad \forall x, y : x_c - M \leq x \leq x_c + M \wedge y_c - N \leq y \leq y_c + N \quad (3.16)$$

I é a imagem original em níveis de cinza; x_c, y_c são as coordenadas do centro da região. Esta região f é denominada região de busca, cujo centro é definido na etapa de predição a ser discutida no próximo capítulo. Os contornos extraídos pela segmentação serão validados, posteriormente, na etapa de casamento de padrões para determinar a posição correta do marcador. Os elementos estruturantes utilizados na sequência de processamento são indicados na Figura 3.12.

1	1	1	0	0	0	0	0	0
1	1	1	*	1	*	0	1	0
1	1	1	1	1	1	*	1	*

Figura 3.12: elementos estruturantes B , L e E .

3.4.1 Algoritmo baseado na intensidade dos marcadores

O princípio geral do algoritmo de segmentação considerado aqui é baseado na “topografia” definida pelas marcas na região de busca. Esta topografia é dada pelo contraste das diferentes marcas, contraste este explorado pelos operadores morfológicos de segmentação. O algoritmo geral de segmentação (Figura 3.13) é representado pelas seguintes etapas (a topografia a ser analisada corresponde a vales ou mínimos regionais da imagem. Um algoritmo dual pode ser definido para os picos ou máximos regionais):

- i) A imagem original em níveis de cinza é invertida (para marcas de alta intensidade). O objetivo é obter um vale profundo (baixa intensidade) na região dos pixels correspondentes às marcas (Figura 3.14a).
- b) Uma operação de pré-processamento visando um maior realce das marcas. Este realce pode ser obtido simplesmente a partir de uma erosão ou abertura, por exemplo.
- c) Detecção de contornos aplicando a função gradiente definida na Seção 3.2
- d) A partir da imagem erodida (etapa b) é calculada a dinâmica dos mínimos regionais e são escolhidas as regiões com maior dinâmica. O número de regiões a ser selecionado depende do número de marcas procuradas. Como nem sempre se pode garantir que os maiores valores de dinâmica pertencem às marcas, este número deve ser ligeiramente maior que o número de marcas presentes numa região de busca. Isto aumenta a probabilidade de o conjunto das dinâmicas escolhidas representarem realmente as marcas nessa região (Figura 3.14d).
- e) A partir da imagem na etapa c) é calculada a LDA utilizando os mínimos regionais de maior dinâmica, definidos na etapa anterior, como marcadores para a segmentação (Figura 3.14e).
- f) Finalmente, as regiões são selecionadas de acordo com alguns critérios definidos na Seção 3.3, levando em consideração a forma e o tamanho das marcas.

O algoritmo acima descrito pode ser resumido em pré-processamento da imagem, segmentação/extração de contornos e seleção de regiões.

A imagem 3.14 ilustra a seqüência de processamentos para a extração de marcas do tipo isopor (Figura 2.7). O reconhecimento de cada das marcas é abordado no Capítulo 5. A Figura 3.15 apresenta outro exemplo de processamento para um grupo de marcas, neste caso são considerados os LEDs da Figura 2.4.

A Figura 3.16 ilustra uma segmentação bastante simples em que todas as marcas (adesivos pretos) constituem vales facilmente identificáveis no processo de segmentação.

De modo geral, o pré-processamento da região de busca pode variar de acordo com o tipo de marca utilizado. Por exemplo, a segmentação das marcas definidas por um círculo branco de contorno preto (Figura 2.6) pode ser obtida aplicando o algoritmo acima e considerando o seguinte pré-processamento (Figura 3.17):

- b1) Afinamento da imagem visando dilatar a marca e ressaltar o contorno. A operação é aplicada com todas as configurações (rotações) possíveis dos elementos estruturantes homotópicos L e E (Figura 3.12), e a iteração pode ser repetida até a idempotência. O elemento estruturante E é utilizado para eliminar os ramos parasitas.

$$f_1 = ((f \odot L)_n \odot E)_\infty$$

- b2) Fechamento da imagem, visando eliminar os pontos agudos gerados pelo afinamento e manter a forma circular das marcas.

$$f_2 = f_1 \bullet B$$

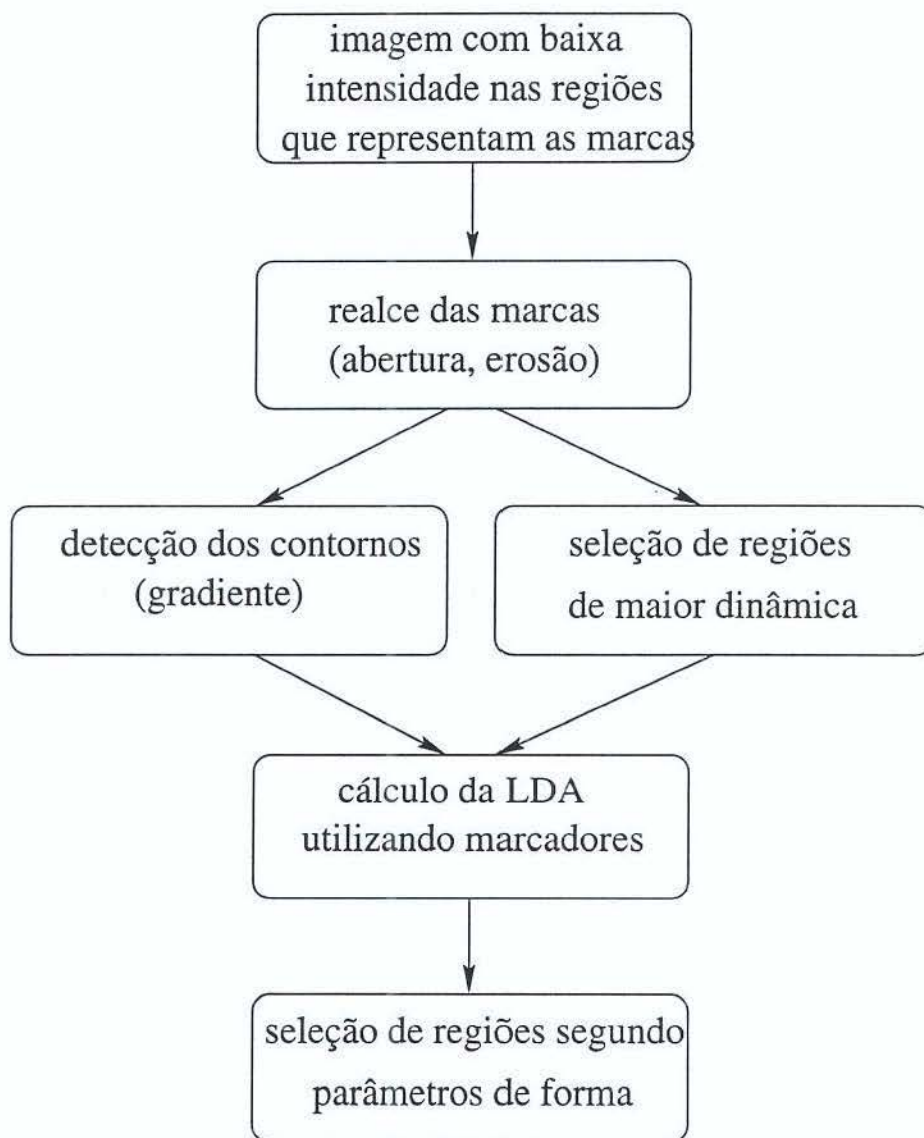


Figura 3.13: algoritmo para a segmentação de marcadores baseado em sua intensidade.

3.5 Conclusão

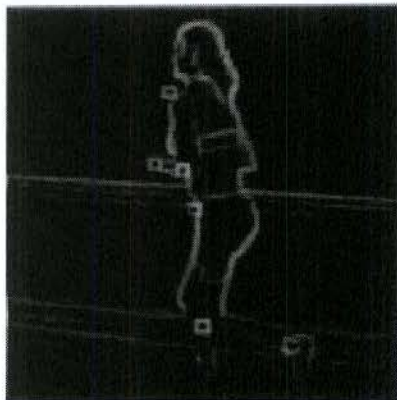
Na primeira parte deste capítulo, foram abordadas algumas ferramentas de processamento de imagens com o objetivo de definir um algoritmo genérico para a segmentação dos diferentes tipos de marcas. As regiões segmentadas/extraídas nesta fase serão utilizadas no Capítulo 5 na obtenção da posição correta das marcas em cada uma das imagens de uma sequência de quadros.



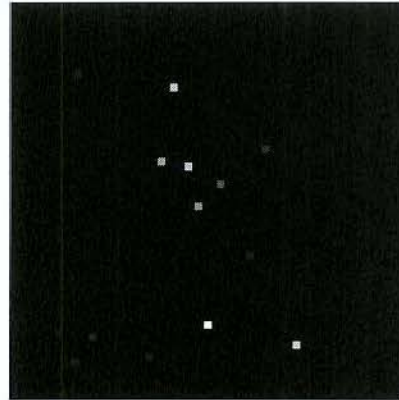
a) Imagem em níveis de cinza invertida.



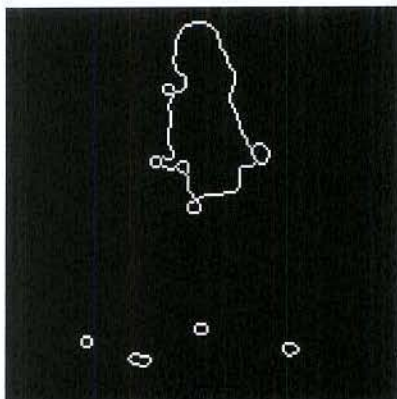
b) A imagem após uma erosão.



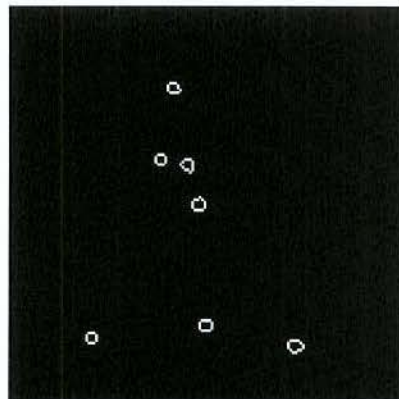
c) Gradiente da imagem.



d) Dinâmica dos mínimos regionais.



e) Regiões obtidas aplicando LDA, utilizando como marcadores a imagem d.



f) Seleção de algumas regiões.

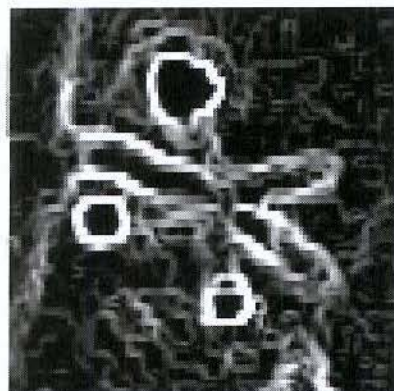
Figura 3.14: sequência de processamentos para segmentação dos marcadores (isopores).



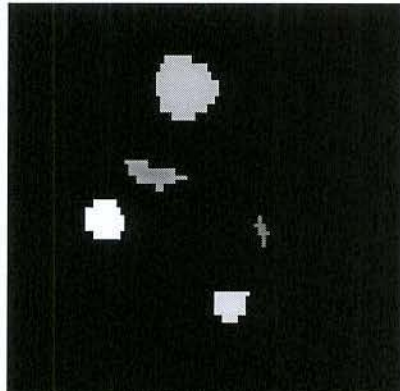
a) Imagem em níveis de cinza invertida.



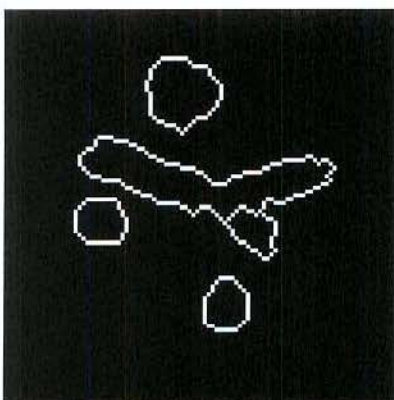
b) A imagem após uma abertura.



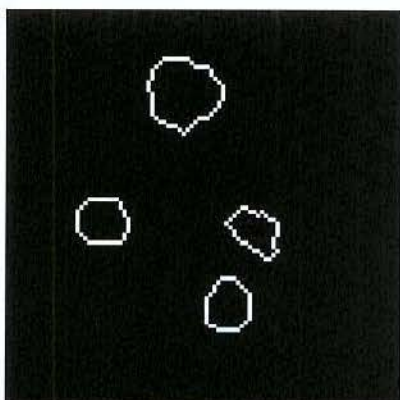
c) Gradiente da imagem.



d) Dinâmica dos mínimos regionais.

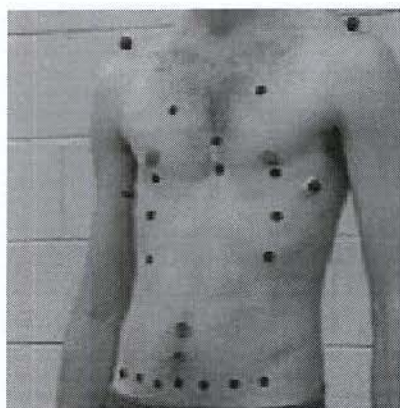


e) Contornos obtidos aplicando LDA, utilizando como marcadores a imagem d.

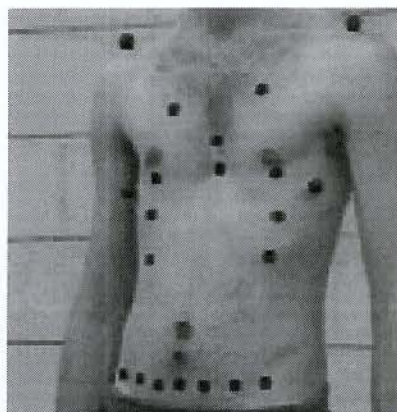


f) Seleção de alguns contornos.

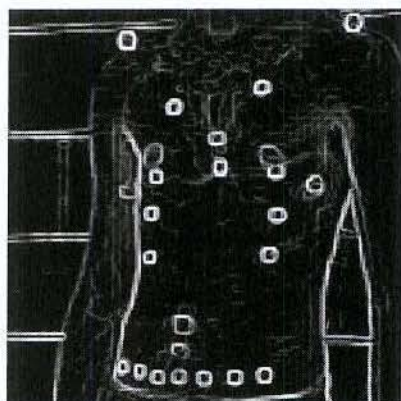
Figura 3.15: sequência de processamentos para segmentação dos marcadores (LEDS).



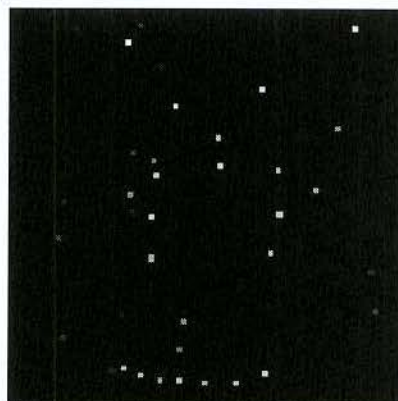
a) Imagem em níveis de cinza.



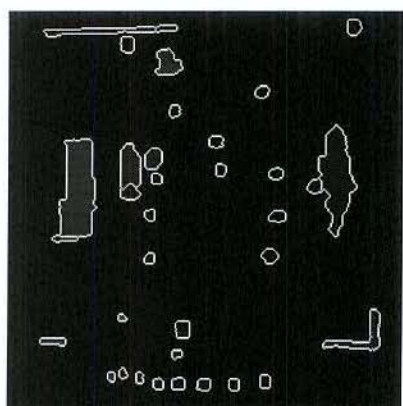
b) A imagem após uma erosão.



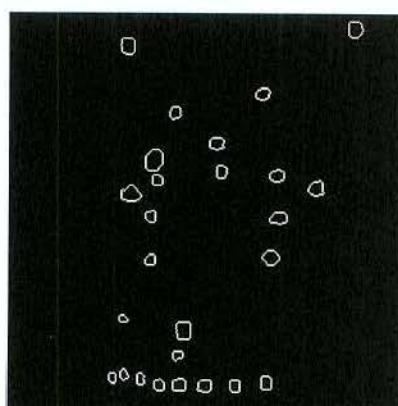
c) Gradiente da imagem.



d) Dinâmica dos mínimos regionais.

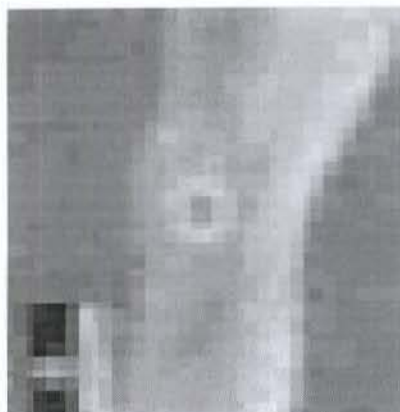


e) Contornos obtidos aplicando LDA, utilizando como marcadores a imagem d.



f) Seleção de alguns contornos.

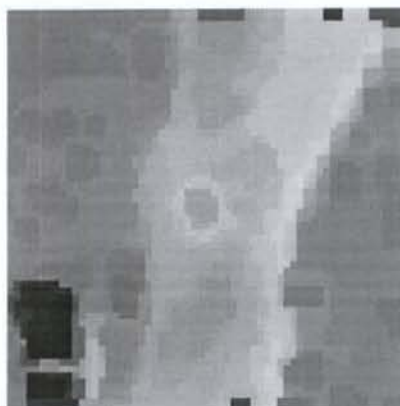
Figura 3.16: seqüência de processamentos para segmentação dos marcadores (adesivos).



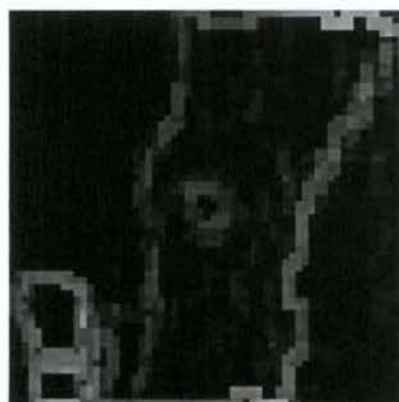
a) Imagem em níveis de cinza invertida.



b1) Afinamento



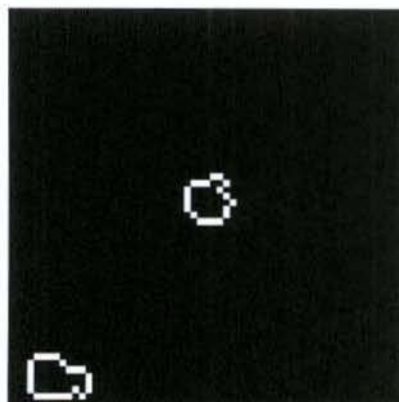
b2) Fechamento



d) Gradiente.



e) Contornos obtidos aplicando LDA, utilizando como marcadores a dinâmica da imagem b2.



f) Seleção de alguns contornos.

Figura 3.17: sequência de processamentos para segmentação de um marcador (pintura).

Capítulo 4

Predição da Trajetória

O problema de perseguição está relacionado com a correspondência entre os diversos quadros que compõem a sequência de imagens. Para diminuir a complexidade, um dos passos importantes é limitar o espaço de busca, introduzindo restrições quanto à uniformidade, velocidade máxima, movimento comum, rigidez etc., ou podem ser usadas técnicas de predição a partir do conhecimento da forma e do modelo do movimento. Outro motivo para se reduzir a área de busca refere-se à diminuição do tempo de processamento, fator este importante devido ao grande número de quadros contidos numa sequência de imagens.

Neste capítulo, são abordadas duas técnicas de predição: a extrapolação e o filtro de Kalman. Como será visto, esta predição servirá a delimitar a área de segmentação e busca dos marcadores nas imagens sucessivas.

4.1 Extrapolação

A trajetória descrita pelo movimento humano é de difícil descrição analítica devido a fatores tais como presença de articulações e elevado grau de liberdade nos movimentos. Como o corpo está submetido à influência da gravidade, a derivada segunda da variável posição é contínua, descrevendo uma trajetória suave. Embora esta informação seja válida apenas no espaço, é possível restringir a descrição da trajetória se considerar a posição da câmera ortogonal à direção do movimento. Isto permite definir uma função de extrapolação usando os parâmetros considerados em quadros anteriores. Um exemplo de uma função de extrapolação, com coeficientes definidos experimentalmente [7], é dada por:

$$x_c = \frac{1}{3}(7x_k - 5x_{k-1} + x_{k-2}) \quad (4.1)$$

$$y_c = \frac{1}{3}(7y_k - 5y_{k-1} + y_{k-2}) \quad (4.2)$$

em que x_k and y_k são as coordenadas dos pontos no quadro k . As coordenadas x_c e y_c formam o centro da região de busca que se procura definir no quadro $k + 1$.

Esta função realiza uma predição a partir do quarto quadro e, portanto, os valores nos primeiros quadros devem ser obtidos através de uma medição manual ou a região de busca deve ser ampliada e ser utilizado o seguinte cálculo para definir os valores no terceiro quadro:

$$x_c = \frac{1}{3}(7x_2 - 4x_1) \quad y_c = \frac{1}{3}(7y_2 - 4y_1)$$

Neste caso, os valores do segundo quadro assumem os mesmos valores do primeiro.

4.2 Filtro de Kalman

4.2.1 Introdução

O filtro de Kalman [19, 14] é uma técnica recursiva de atualização usada para determinar os parâmetros corretos do modelo de um processo. Dados alguns valores iniciais, pode-se prever e ajustar os parâmetros do modelo através de cada nova medição, obtendo a estimativa do erro em cada atualização. A possibilidade de incorporar efeitos de erros e sua estrutura computacional fizeram com que o filtro de Kalman tivesse um amplo campo de aplicações, especialmente no que se refere à análise de trajetória em visão computacional [6, 15, 23].

A propriedade do filtro de Kalman, não presente em outras técnicas de predição, refere-se ao ajuste de seus próprios parâmetros de acordo com a estatística das medições de cada imagem. Em outros tipos de filtros, a porcentagem de erro é mantida como constante durante toda a sequência. O filtro de Kalman permite mudar essa porcentagem automaticamente de acordo a precisão do estado de seus parâmetros.

O modelo discreto deste filtro contém os seguintes elementos:

- cálculo da matriz de ganho

$$K_k = P_k^- H_k^T (H_k P_k^- H_k^T + R_k)^{-1} \quad (4.3)$$

em que P_k^- é a matriz de covariância de erros dos valores estimados, H_k é a matriz de conexão entre as medições e o estado do vetor atual, R_k denota a matriz de covariância de erros das medições, a qual deve ser conhecida com antecipação.

- cálculo do vetor de parâmetros:

$$\hat{x}_k = \hat{x}_k^- + K_k(z_k - H_k \hat{x}_k^-) \quad (4.4)$$

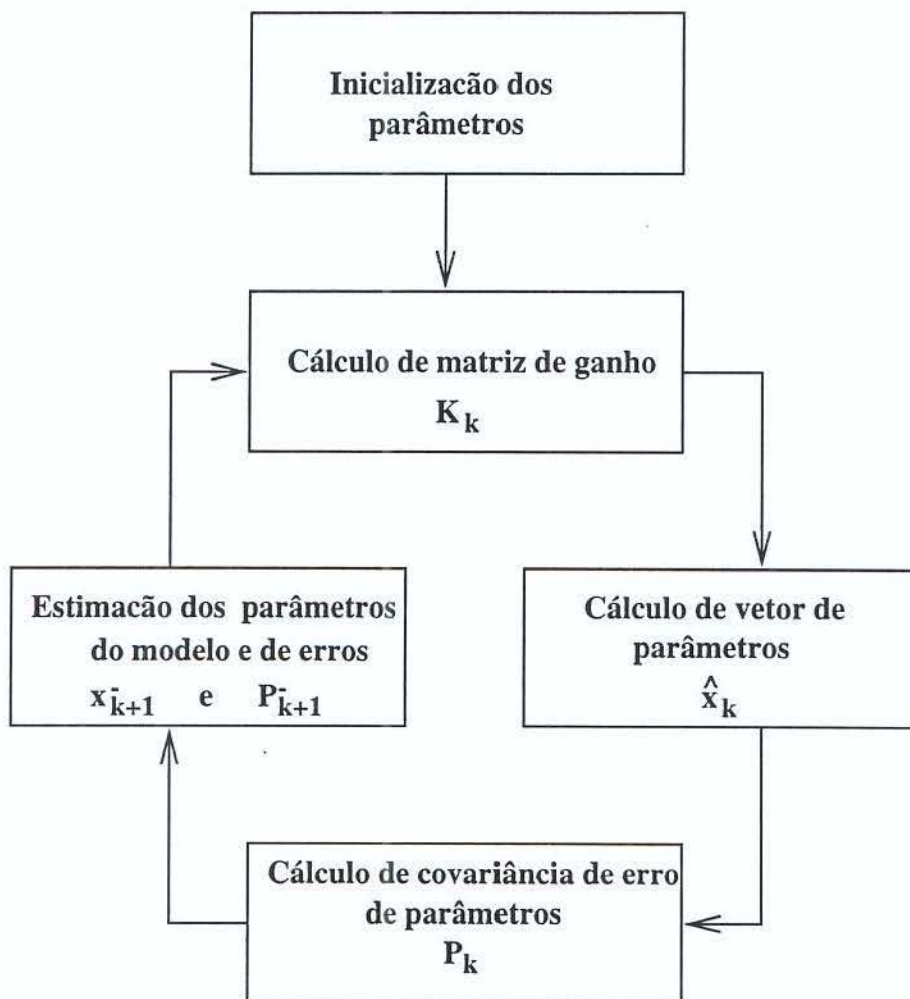


Figura 4.1: o ciclo do filtro de Kalman.

em que x_k^- é o vetor que contém os parâmetros estimados, z_k é o vetor referente aos valores medidos. Um exemplo de parâmetros pode ser $\hat{x}_k^T = \{x, y, \dot{x}, \dot{y}\}$, onde x, y indica a posição do objeto, $\dot{x}, \dot{y}$ as velocidades.

- cálculo da matriz de covariância do erro de parâmetros:

$$P_k = (I - K_k H_k) P_k^- \quad (4.5)$$

em que I é a matriz identidade. A matriz K_k determina a proporção de erro a ser usada nos valores dos parâmetros no estado atual.

- estimação dos parâmetros e de erros:

$$\hat{x}_{k+1}^- = \Phi_k \hat{x}_k \quad (4.6)$$

$$P_{k+1}^- = \Phi_k P_k \Phi_k^T + Q_k \quad (4.7)$$

A predição dos parâmetros do modelo e do erro, na instância $k+1$, é calculada pelas equações acima. Φ_k é a matriz de transição determinada a partir de um modelo do movimento. Q_k representa a matriz de covariância do erro do modelo que permite ajustar os valores de P_k . O ciclo do filtro de Kalman é mostrado na figura 4.1.

4.2.2 Aplicação para tracking do movimento humano

Os elementos que formam parte do filtro de Kalman para esta aplicação são definidos da seguinte forma: x_k é o vetor que contém os parâmetros no tempo t_k , ou seja $x_k^T = [x, \dot{x}, y, \dot{y}]$, onde x e y denotam as coordenadas da posição do marcador, e $\dot{x}$ e $\dot{y}$ são a representação das velocidades nas respectivas direções. z_k é o vetor que contém as medições no quadro k . Nesta aplicação, são consideradas apenas as coordenadas da posição dos marcadores, isto é, $z_k^T = [x, y]$. Φ representa a matriz de transição de x_k para x_{k+1} .

$$\Phi_k = \begin{bmatrix} 1 & \delta t & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & \delta t \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

em que δt é o intervalo do tempo entre os quadros.

A matriz H_k que representa a conexão sem ruídos entre x_k e x_{k+1} , e é dada por:

$$H_k = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

A matriz Q_k representando o erro do modelo também é definida a priori:

$$Q_k = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \sigma_{Q\dot{X}}^2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_{Q\dot{Y}}^2 \end{bmatrix}$$

em que $\sigma_{Q\dot{X}}$ e $\sigma_{Q\dot{Y}}$ são as correlações das velocidades. Analogamente, a matriz de covariância de erros das medições é definida por:

$$R_k = \begin{bmatrix} \sigma_{RX}^2 & 0 \\ 0 & \sigma_{RY}^2 \end{bmatrix}$$

Existem duas possibilidades de inicialização da matriz de covariância de estimação dos valores P_k [6]. A primeira alternativa é aguardar os três primeiros quadros para se calcular a velocidade a partir das duas primeiras imagens, fazer uma predição na terceira imagem e assim determinar o erro de predição. Uma outra solução é inicializar as velocidades na matriz de covariância de parâmetros estimados P_k^- com zeros e, simultaneamente, inicializar a matriz de covariância dos parâmetros P_k com valores grandes.

$$P_k = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \sigma_{P\dot{X}}^2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & \sigma_{P\dot{Y}}^2 \end{bmatrix}$$

4.3 Análise dos resultados de predição

Na figura 4.2 e figura 4.3 são apresentados os gráficos dos valores de uma predição, utilizando a extrapolação e o filtro de Kalman, respectivamente. Os pontos de cor preta indicam as medições da coordenada Y (em pixels) obtidas no tracking automático de uma das marcas da figura 2.7.

Para uma melhor visualização do desvio dos valores de predição em relação às medições, os pontos foram conectados por segmentos retilíneos. Para se determinar a precisão dos algoritmos e comparar os valores de predição, pode-se utilizar o valor do resíduo, em especial o valor máximo ou o valor médio. O resíduo é dado por:

$$r(k) = |z(k) - y(k)| \quad (4.8)$$

onde $z(k)$ é o valor da medição e $y(k)$ o valor da predição no quadro k . Para o caso da figura 4.2, o resíduo máximo é $\max(r_e) = 7$ e $\bar{r}_e = 4$, e no caso de Filtro de Kalman $\max(r_k) = 5$ e $\bar{r}_k = 3$.

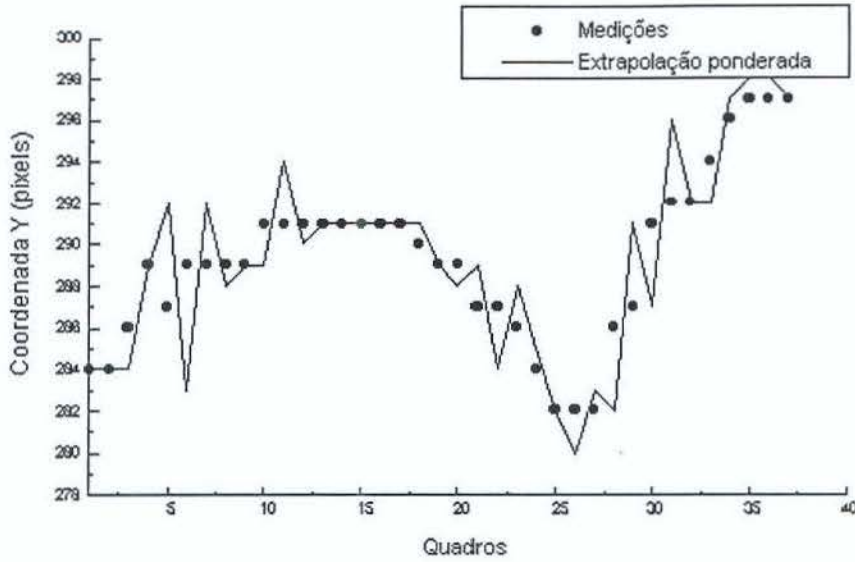


Figura 4.2: predição utilizando a extrapolação ponderada.

A frequência de captura das imagens foi de 25 quadros por segundo, isto é $\delta t = 0.04$. Os valores das covariâncias nas matrizes foram definidos a partir das considerações da seção anterior.

A inicialização dos parâmetros, ou seja, a posição do marcador e a velocidade, são obtidas das primeiras duas medições. A matriz que representa o erro do modelo Q_k é inicializada com valores elevados, tornando os valores da matriz significativos e os valores estimados sensíveis às medições mais recentes (equação 4.7). Assim, consideramos:

$$\sigma_{Q\dot{x}} = 1000; \quad \sigma_{Q\dot{y}} = 1000;$$

Analogamente, R_k é inicializado com a covariância do erro das medições. Este é modelado como um ruído branco (distribuição Gaussiana) com um desvio padrão de um pixel. Assim:

$$\sigma_{RX} = 1; \quad \sigma_{RY} = 1;$$

A matriz P_k é inicializada considerando as velocidades iniciais, e com valores suficientemente grandes para permitir a influência das medições anteriores, possibilitando um ajuste automático do filtro de acordo com os novos valores de entrada.

$$\sigma_{P\dot{x}} = 100; \quad \sigma_{P\dot{y}} = 100$$

A partir dos gráficos pode-se verificar que o filtro de Kalman gera uma trajetória mais suave que a extrapolação. A dificuldade principal do filtro de Kalman refere-se à inicialização dos seus parâmetros.

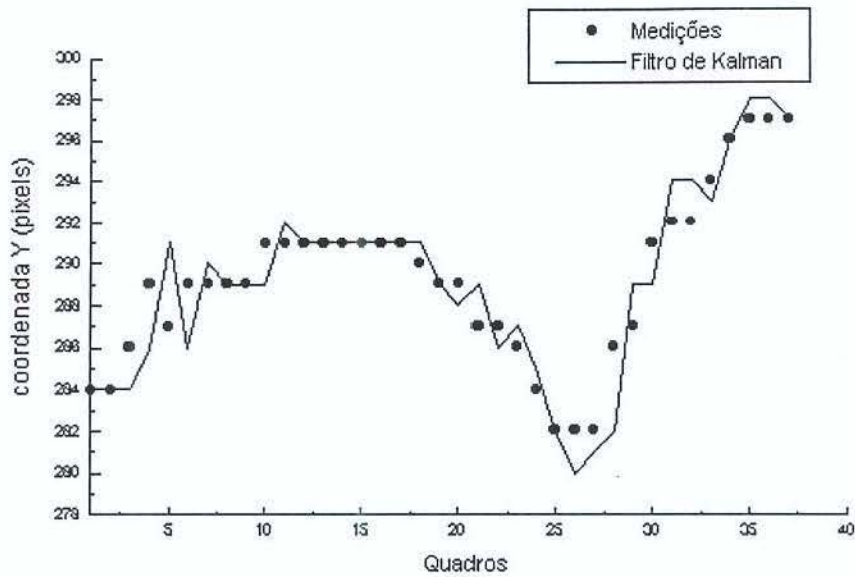


Figura 4.3: predição utilizando o filtro de Kalman.

4.4 Conclusão

Neste capítulo, foram abordados dois métodos de predição: a extrapolação e o filtro de Kalman. Estas técnicas são utilizadas neste trabalho para definir a região de busca, sendo sua precisão de pouca importância. Como será visto no próximo capítulo, esta predição poderá ser explorada, ainda, na etapa de casamento de padrões como critério de decisão no caso de múltiplos casamentos.

Capítulo 5

Casamento de Padrões

O objetivo desta parte do trabalho é determinar a posição dos marcadores no quadro atual, a partir da semelhança entre o padrão (modelo do marcador) definido na etapa de inicialização, ou no quadro anterior, e as respectivas regiões definidas na etapa de segmentação.

Determinar a similaridade entre dois padrões não é tarefa fácil, devido a fatores que influenciam a qualidade da imagem, como por exemplo, a mudança da intensidade da luz ao longo da seqüência, a mudança na forma dos objetos, a presença de ruídos, oclusão, etc.

Como não existe um critério que garanta a posição correta do marcador na imagem, é definido p_{ij} como a probabilidade de que o ponto (i, j) represente esta posição. Desta forma, os possíveis pontos que representam a posição do marcador podem ser definidos escolhendo os máximos p_{ij} , calculados aplicando alguns algoritmos de casamento de padrões.

Vale salientar que para o processo de segmentação foram utilizadas apenas imagens em níveis de cinza. No casamento de padrões a informação das cores, ou seja, das três bandas RGB (vermelho, verde e azul), pode ser muito importante na definição de semelhança entre os padrões. Assim a probabilidade do ponto (i, j) será definida como a média das probabilidades nas três bandas.

$$P_{ij} = \frac{1}{3}(p_{ij}^R + p_{ij}^G + p_{ij}^B) \quad (5.1)$$

O casamento de padrões pode ser definido de diferentes formas; a correlação entre um padrão e pequenas regiões dentro da área de busca é um critério geral [27, 29].

A semelhança entre dois padrões também pode ser determinada através da distância entre as cores, definindo regiões com intensidades próximas à região dada (modelo). Características particulares, tais como informações de regiões mais claras ou mais escuras, podem ainda ser empregadas. Geralmente, as cores preta ou branca são as que melhor

conservam suas características espectrais ao longo da sequência e, assim, critérios relacionados a valores de máximos ou mínimos regionais podem constituir bons parâmetros de casamento.

Estas técnicas são abordadas nas próximas seções deste capítulo, considerando as informações obtidas na etapa de segmentação. Os critérios considerados são aplicados apenas nas regiões extraídas. No final deste capítulo é considerado o problema da classificação no caso que duas ou mais marcas se encontram próximas umas das outras.

5.1 Técnicas de correlação

Seja $h(x - \xi, y - \eta)$ a imagem do objeto de interesse. Os parâmetros (ξ, η) definem a posição do objeto dentro da imagem. Uma imagem $f(x, y)$ pode ser considerada como uma composição de $h(x, y)$ e do fundo $\underline{n}(x, y)$, ou seja:

$$f(x, y) = h(x - \xi, y - \eta) + \underline{n}(x, y) \quad (5.2)$$

O objetivo é determinar se a cena contém o objeto e, se for o caso, localizar a sua posição na imagem. Uma forma de tratar este problema é deslocar o objeto ao longo da imagem, casar este padrão localmente e selecionar a posição $(\hat{\xi}, \hat{\eta})$ do padrão de melhor casamento [29]. Em particular, para este objetivo pode ser utilizada a *distância euclidiana*.

$$d^2(\xi, \eta) = \int \int_{x,y} (f(x, y) - h(x - \xi, y - \eta))^2 dx dy, \quad d(\hat{\xi}, \hat{\eta}) = \min\{d(\xi, \eta)\} \quad (5.3)$$

Substituindo as variáveis de integração, a equação acima pode ser escrita da seguinte forma:

$$d^2(\xi, \eta) = \int \int_{x,y} (f(x + \xi, y + \eta) - h(x, y))^2 dx dy \quad (5.4)$$

Para o caso discreto, a integração pode ser substituída por uma somatória finita.

$$d_{i,j}^2 = \sum_{k=-K}^K \sum_{l=-L}^L (f_{i+k,j+l} - h_{k,l})^2 \quad (5.5)$$

Onde K e L determinam o tamanho do objeto de interesse (padrão). Estes valores são escolhidos de tal forma que o padrão h cobre totalmente o objeto de interesse.

Na figura 5.1, é mostrado um exemplo de uma correlação aplicando a fórmula 5.5. A primeira imagem (a) é uma região extraída da figura 2.6. A imagem (c) mostra o resultado da correlação que representa a distância euclidiana $d_{i,j}$. Esta possui uma região bastante escura correspondendo à posição do padrão na imagem original, em que a distância é mínima.

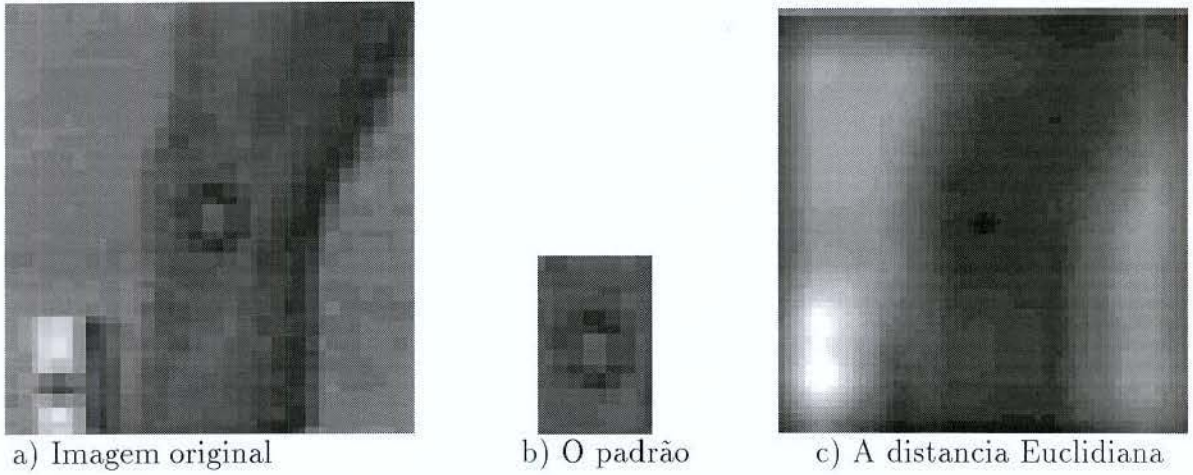


Figura 5.1: um exemplo de correlação

Descompondo a equação 5.5 em componentes tem-se:

$$d_{i,j}^2 = \sum_{k=-K}^K \sum_{l=-L}^L h_{k,l}^2 - 2 \sum_{k=-K}^K \sum_{l=-L}^L f_{i+k,j+l} h_{k,l} + \sum_{k=-K}^K \sum_{l=-L}^L f_{i+k,j+l}^2 \quad (5.6)$$

O primeiro componente (energia local do padrão) não depende da imagem observada e pode ser omitido, sem afetar a qualidade na detecção e localização. O último termo (energia local da imagem) pode ser igualmente omitido já que depende apenas da imagem observada. Desta forma temos a seguinte correlação cruzada:

$$C_{i,j} = \sum_{k=-K}^K \sum_{l=-L}^L f_{i+k,j+l} h_{k,l} \quad (5.7)$$

São vários os fatores que podem dificultar a detecção do padrão e sua localização, utilizando este método. Por exemplo, os fatores de iluminação, reflexão, tamanho, forma e orientação do objeto. No entanto a influência de alguns destes fatores pode ser neutralizada a partir de uma normalização. Para o caso da iluminação, por exemplo, se o fundo constante definido é desconhecido na fórmula $f(x, y) = h(x - \xi, y - \eta) + \underline{n}(x, y) + \underline{b}$, então a influência da iluminação $\underline{b}$ pode ser neutralizada utilizando a média dos níveis de cinza do padrão e da imagem.

$$\bar{h} = \frac{1}{(2K+1)(2L+1)} \sum_{k=-K}^K \sum_{l=-L}^L h_{k,l} \quad (5.8)$$

$$\bar{f}_{i,j} = \frac{1}{(2K+1)(2L+1)} \sum_{k=-K}^K \sum_{l=-L}^L f_{i+k,j+l} \quad (5.9)$$

Portanto, a correlação cruzada, considerando esta normalização, pode ser calculada da seguinte forma:

$$C_{m,n} = \sum_{k=-K}^K \sum_{l=-L}^L (f_{i+k,j+l} - \bar{f}_{i,j})(h_{k,l} - \bar{h}) \quad (5.10)$$

Outro parâmetro de ruído que pode complicar a localização do padrão na imagem é o ganho $\underline{a}$, na equação $f(x, y) = \underline{a}h(x - \xi, y - \eta) + \underline{n}(x, y)$. Para reduzir a influência deste parâmetro, deve ser considerada a normalização das energias locais.

$$\bar{h}^2 = \frac{1}{(2K+1)(2L+1)} \sum_{k=-K}^K \sum_{l=-L}^L h_{k,l}^2 \quad (5.11)$$

e

$$\bar{f}_{i,j}^2 = \frac{1}{(2K+1)(2L+1)} \sum_{k=-K}^K \sum_{l=-L}^L f_{i+k,j+l}^2 \quad (5.12)$$

Para este caso, a correlação cruzada pode ser dada por:

$$C_{i,j} = \frac{\sum_{k=-K}^K \sum_{l=-L}^L f_{i+k,j+l} h_{k,l}}{\sqrt{\bar{f}_{i,j}^2} \sqrt{\bar{h}^2}} \quad (5.13)$$

Se os parâmetros $\underline{a}$ e $\underline{b}$ são desconhecidos, então deve-se calcular a normalização considerando-se ambos os fatores:

$$C_{i,j} = \frac{\sum_{k=-K}^K \sum_{l=-L}^L (f_{i+k,j+l} - \bar{f}_{i,j})(h_{k,l} - \bar{h})}{\sqrt{\sum \sum (f_{i+k,j+l} - \bar{f}_{i,j})^2} \sqrt{\sum \sum (h_{k,l} - \bar{h})^2}} \quad (5.14)$$

Este último resultado é denominado coeficiente de correlação cruzada por ser análogo ao coeficiente de correlação normalizada entre duas variáveis aleatórias.

Para se obter o melhor casamento, é calculado o máximo valor das correlações.

$$\lambda = \max C_{i,j}; \quad i, j \in R$$

Onde R é a região de busca de tamanho $M \times N$. Na nossa aplicação, as coordenadas (i, j) para as quais λ é máximo são consideradas como o centro da marca. A Equação 5.14 gera valores entre -1 e 1, portanto as probabilidades das correlações podem ser dadas por:

$$p_{ij} = \begin{cases} C_{ij} & \text{if } C_{ij} > 0 \\ 0 & \text{else} \end{cases} \quad (5.15)$$

5.2 Método baseado em distância com tolerância

Nos casos de modelos homogêneos, por exemplo, se o objeto ou uma parte do objeto é de mesma cor, pode-se utilizar a distância média entre as intensidades dos pontos da cena e do modelo para a sua detecção. Seja h a intensidade de um ponto no modelo, e f_{ij} a intensidade do ponto na região de busca. A distância é determinada como a diferença entre as intensidades, i.e., $|f_{ij} - h|$. Para se determinar o critério de similaridade, é calculada a média entre as distâncias aos pontos vizinhos, de acordo com um determinado limiar Δ .

$$p_{ij} = \frac{1}{(2K+1)(2L+1)} \sum_{k=-K}^K \sum_{l=-L}^L |f_{i+k,j+l} - h| \quad \forall k, l \text{ em que } |f_{i+k,j+l} - h| < \Delta \quad (5.16)$$

Com o objetivo de considerar os pontos de menor distância média como pontos de maior probabilidade, estes valores são normalizados:

$$p_{ij} = \frac{1}{(2K+1)(2L+1)} \sum_{k=-K}^K \sum_{l=-L}^L \left(1 - \frac{|f_{i+k,j+l} - h|}{\Delta}\right) \quad (5.17)$$

Como no caso da correlação, p_{ij} é calculada para cada banda de cores e, portanto, P_{ij} é definida como a média das probabilidades nas três bandas (equação 5.1).

5.3 Método baseado nos valores maximais

Para alguns tipos de marcadores com alto contraste, pode-se levar em consideração as informações sobre os máximos valores da sua intensidade. Na região de tamanho definida pelo padrão é calculada a média das intensidades de todos os pontos que formam a região, e é escolhido o ponto que tem a maior média. Embora este método dependa da qualidade da imagem, ele é rápido e pode ser utilizado para casos em que o tempo de processamento seja importante.

Seguindo a idéia das probabilidades, definidos para os outros métodos, temos que:

$$p_{ij} = \frac{1}{(2K+1)(2L+1)} \sum_{k=-K}^K \sum_{l=-L}^L (f_{i+k,j+l} / \text{MAX_NIVEL}) \quad (5.18)$$

em que MAX_NIVEL é o máximo valor de cinza na região do padrão. Neste caso, não precisa uma definição do marcador, apenas é preciso definir o tamanho. O marcador pode ser entendido como uma matriz com todos seus valores iguais a MAX_NIVEL.

5.4 Definição do modelo dos marcadores

Geralmente, o tamanho do modelo do marcador é definido na etapa de inicialização (primeiro quadro da sequência), a partir da indicação de um primeiro ponto correspondente

ao centro da marca, e de mais um outro ponto para a obtenção do retângulo envolvente. O padrão ou modelo do marcador, para o quadro no instante t_l , pode ser definido a partir da última marca encontrada no instante $t_l - 1$, ou da primeira marca do quadro no instante t_0 . A primeira opção deve ser considerada quando a intensidade e a forma do marcador variam consideravelmente, quadro a quadro. Isto permite atualizar a informação sobre a marca em função das possíveis variações nas imagens. Neste caso, existe a possibilidade de após um eventual erro na definição do melhor marcador, no quadro atual, se continuar procurando uma marca “errada” nos quadros subsequentes. A segunda alternativa é melhor adaptada aos problemas de oclusão em que um valor limiar, associado à probabilidade P_{ij} , pode sinalizar a presença ou não de marcas nos diversos quadros. A desvantagem deste método refere-se, naturalmente, à baixa correlação obtida, à medida que a intensidade da imagem muda, dificultando, assim, a escolha do marcador na região de busca. Em ambos os casos, o padrão é extraído da imagem considerando o centro da marca obtido em cada quadro, e o tamanho inicial definido pelo usuário.

5.5 Classificação das marcas e critérios de decisão

O critério utilizado para a escolha da marca, em cada imagem da sequência, é baseado no valor máximo obtido no casamento de padrões. No entanto, este valor nem sempre garante o resultado esperado, já que em certas circunstâncias podem aparecer valores de P_{ij} bastante próximos. Assim, critérios de decisão devem ser associados ao casamento de padrões abordado anteriormente, visando uma maior precisão dos resultados.

Além do maior valor das probabilidades, outro critério de decisão considerado neste trabalho baseia-se na distância de possíveis centros de marcadores ao ponto predito (Capítulo 4). Para tanto, é criada uma lista ordenada de probabilidades.

Esta lista é gerada aplicando o casamento de padrões em regiões obtidas a partir da segmentação por LDA das imagens (Capítulo 3). Estas regiões foram etiquetadas, após segmentação, com valores maiores que 1. O casamento de padrões é realizado apenas nos pontos com tais etiquetas, e o máximo casamento para cada uma das regiões é calculado. Finalmente, a lista é ordenada.

Para casos em que o casamento de padrões tenha que ser aplicado em toda a região de busca, por exemplo, em caso de perda de marcadores por erros na segmentação (abordado no próximo capítulo), pode ser considerado o seguinte procedimento.

A região de busca é dividida em S pequenos retângulos, de tamanho igual ao tamanho do marcador (veja figura 5.2). Para cada uma destas regiões ou retângulos, R_s , é definido o valor p_s dado como o máximo casamento, isto é, $p_s = \max(P_{ij})$, $\forall s \in S$, $ij \in R_s$. Para evitar que na lista dois casamentos estejam referindo-se à mesma marca no caso em que o seu centro esteja na fronteira dos retângulos (por exemplo, p_1 , p_2 e p_3 na figura

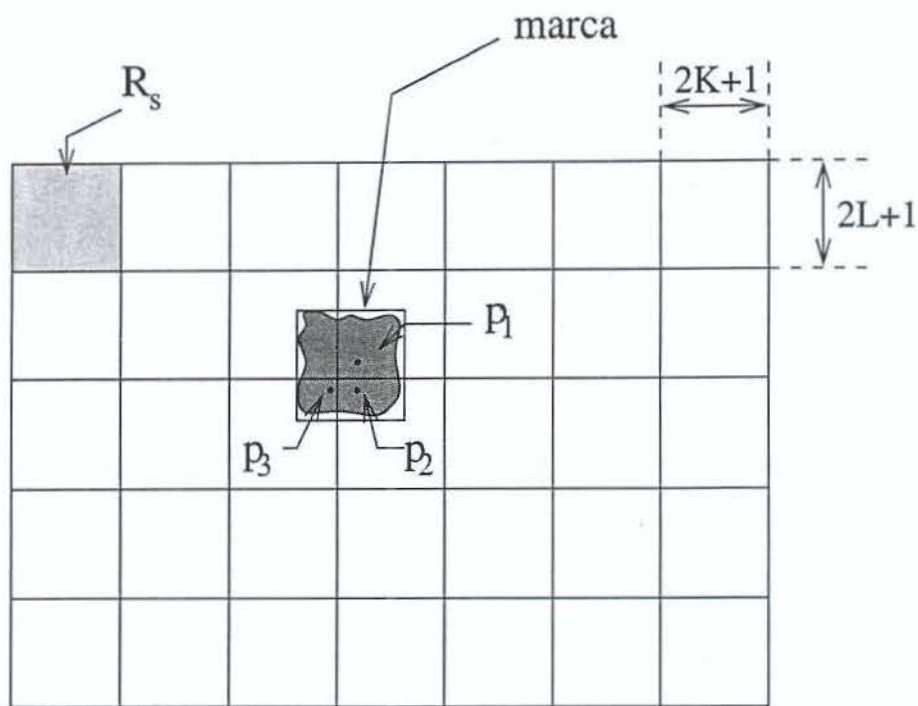


Figura 5.2: subdivisão da região de busca.

5.2), é calculada a distância entre os respectivos pontos. Caso esta distância seja menor que o diâmetro da marca, o ponto de menor valor é anulado. O passo final do processo é ordenar a lista.

Estes procedimentos permitem a introdução de novos critérios na definição da posição, sobretudo no caso em que pontos com probabilidades altas refiram-se a diferentes marcadores.

Como mencionado anteriormente, quando o casamento de padrões define pontos com valores de probabilidade muito próximos, a distância destes ao ponto predito (para o centro da região) pode ser utilizada na decisão da escolha do ponto correto. Este é um fator decisivo para a identificação de marcadores, por exemplo, no caso de se ter marcadores de mesma característica muito próximos. Uma probabilidade ponderada pode ser definida como:

$$P_{ij}^{pond} = \frac{1}{A+B} (A * P_{ij} + B * P_{ij}/d_{ij}) \quad d \geq 1 \quad (5.19)$$

onde d_{ij} é a distância 8-conexa do ponto (ij) ao ponto predito, A e B são duas constantes que definem o peso de cada critério (com ou sem informação de distância). Desta forma, dependendo do critério de maior peso, a decisão pode ser ponderada através dos valores de A e B .

Para um conjunto de marcadores próximos (figura 3.14) é preferível realizar um pré-

processamento em grupo, visando diminuir o tempo de segmentação. Neste caso, surgem os problemas relacionados com a classificação ou correspondência dos marcadores, especialmente quando os mesmos têm características similares. Como discutido anteriormente, estes problemas podem ser tratados considerando-se um maior peso B ao critério de distância da equação 5.19.

5.6 Conclusão

Neste capítulo, foram abordados alguns métodos de casamento de padrões para determinar a posição correta dos marcadores. Para casos de valores de casamento muito próximos, foi considerada a distância ao ponto predito como critério de decisão. Utilizando estes parâmetros foi definida uma probabilidade ponderada. No próximo capítulo, estas informações são consideradas na definição da perseguição automática dos marcadores.

Capítulo 6

Implementação e Teste de Algoritmos para a Perseguição Automática de Marcadores

Para a implementação da perseguição automática dos marcadores foram consideradas as três etapas de processamento definidas nos capítulos anteriores: segmentação, predição e casamento de padrões. Foi desenvolvido um sistema bastante flexível, permitindo ao usuário escolher uma sequência de algoritmos de processamento de acordo com o tipo de marcadores utilizados e o tipo de movimento analisado. Neste capítulo, são abordadas algumas características do sistema, as ferramentas implementadas e os testes para diferentes tipos de movimentos. Também são discutidos alguns problemas relacionados com a oclusão e a precisão das medições.

6.1 Características do sistema DVIDEOW

O sistema DVIDEOW(Digital Video for WIndows), em fase de desenvolvimento, é uma interface gráfica que permite ao usuário realizar tanto a medição manual como a medição automática de diferentes tipos do movimento humano. Este contém as ferramentas de processamento de imagens, algoritmos de predição e casamento de padrões, assim como as ferramentas de calibração e reconstrução 3D que foram obtidas de uma primeira versão do sistema [7]. O sistema DVIDEOW (Figura 6.1) foi programado na linguagem Visual C++, para Windows de 32 bits, trabalhando com uma sequência de imagens de 24 bits. Consegue carregar 4 janelas de calibração e 4 janelas de sequência de imagens (não existe um limite fixo para o número de quadros, nem para o número de marcadores, aspectos estes dependentes da memória disponível).

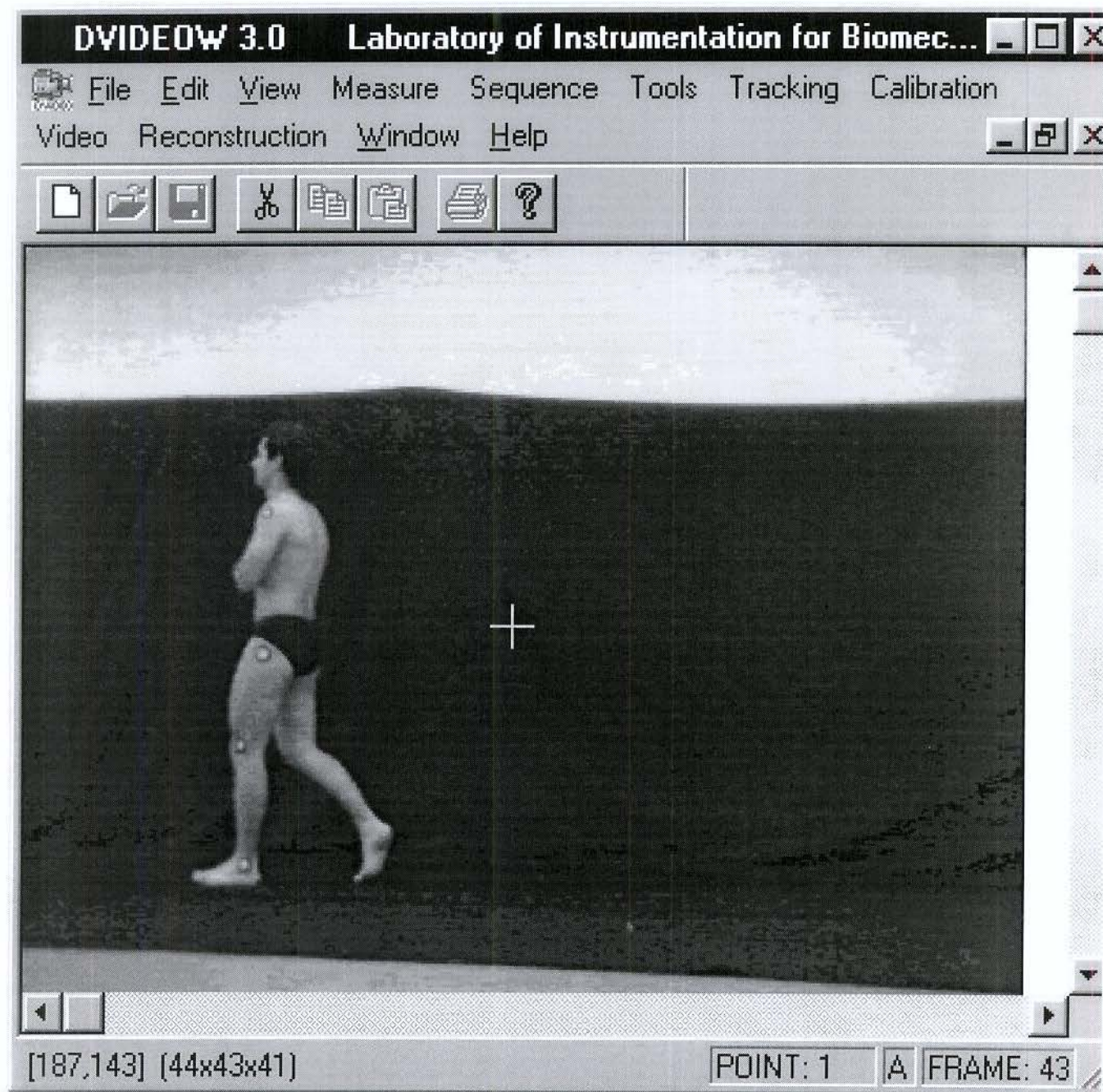


Figura 6.1: o ambiente do DVIDEOW.

6.1.1 Entrada e saída de dados

Para realizar tanto a medição manual quanto a medição automática, deve ser carregada a sequência de imagens. Uma sequência de imagens pode ser armazenada num arquivo (formato vídeo) ou num conjunto de arquivos (formato *Bitmap*), onde cada arquivo armazena uma imagem e algumas informações específicas.

O nome dos arquivos no formato *Bitmap* é composto de uma palavra comum e um índice diferente para cada arquivo que representa o respectivo número da imagem (quadro). Esta forma de organização facilita o reconhecimento pelo sistema dos arquivos pertencentes a uma mesma sequência. Inicialmente, o sistema carrega um arquivo da sequência e a partir deste são fornecidas as informações sobre o número de imagens, os índices da imagem inicial e final. Durante o *tracking*, cada um dos arquivos a ser processado é carregado individualmente.

O sistema trabalha com imagens no formato *bitmap* de 24 bits, ou seja para cada pixel possui três valores de cor no padrão RGB, com 8 bits destinados a cada canal. As imagens devem ter um tamanho padrão (640x480 ou 320x240 pixels).

Aproveitando a disponibilidade dos *drivers* de interface do *Windows* que permitem o acesso direto e rápido às imagens de vídeo, foi implementada a leitura de arquivos de vídeo. Desta forma, o sistema permite ler arquivos no formato AVI, utilizando o *driver* correspondente ao tipo de compressão (*hardware* ou *software*) do vídeo.

Finalmente, os dados de saída do sistema podem ser tanto medições em 2D como em 3D. As medições 2D são os valores das coordenadas x e y dos marcadores na imagem, para cada quadro. Por sua vez, as medições 3D [7] contêm as três coordenadas no espaço x, y e z da posição dos marcadores.

6.1.2 Visualização dos dados

Um fator importante nos sistemas de interface com o usuário é a visualização dos dados. O sistema desenvolvido permite mostrar as imagens em diferentes tamanhos, facilitando assim a visualização e indicação manual dos marcadores. Igualmente, pode ser ativada ou desativada a visualização do processamento, durante a perseguição automática. Isto permite avaliar a execução certa da sequência de processamento, como também determinar possíveis causas de erros na segmentação.

Os dados de saída também podem ser visualizados, permitindo assim uma avaliação aproximada dos resultados. A possibilidade de definir e mostrar os segmentos que unem os marcadores, assim com visualizar os valores das intensidades (R,G,B) da imagem, são algumas funções adicionais do sistema.

6.1.3 Ferramentas de processamento de imagens

As diferentes ferramentas de processamento de imagens, em especial as de morfologia matemática definidas no Capítulo 3, formam parte do sistema. Assim é possível experimentar de forma interativa cada uma destas ferramentas visando definir a sequência de operações para a segmentação utilizada na perseguição automática. Entre as opções existe a possibilidade de processar a imagem inteira ou apenas uma região de tamanho definido.

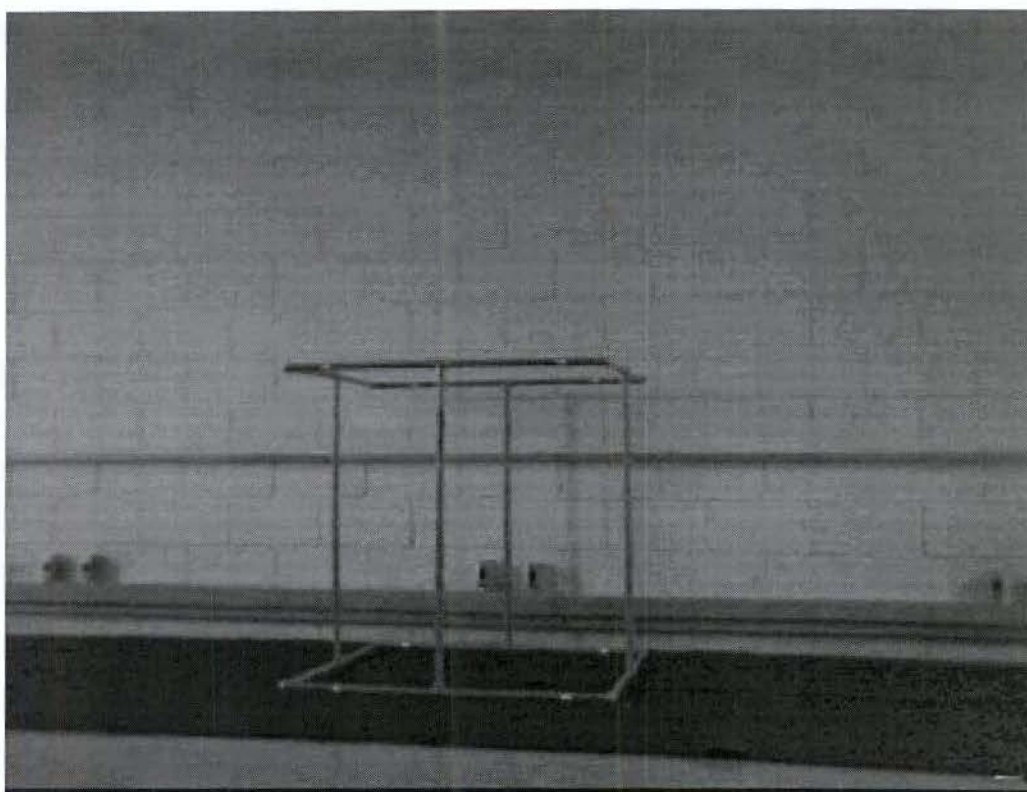


Figura 6.2: um cubo de referência utilizado para a calibração.

6.1.4 Calibração e reconstrução tridimensional

A calibração das câmeras é necessária para a reconstrução tridimensional das coordenadas no espaço. Esta calibração é obtida a partir da medição manual de pontos de referência na imagem, e aplicando o método DLT (*direct linear transformation*) [2], por exemplo. Os valores das coordenadas no espaço dos pontos de referência devem ser conhecidos (distância real com relação a algum ponto). Normalmente, para esta finalidade podem ser utilizados cubos de referência (Figura 6.2). A reconstrução tridimensional é considerada a partir das calibrações e medições de no mínimo duas câmeras [7, 24].

6.1.5 Perseguição automática dos marcadores

O módulo do sistema, relacionado com o principal objetivo deste trabalho, refere-se ao *tracking* ou perseguição automática dos marcadores. A implementação deste módulo foi realizada a partir das três etapas de processamento abordadas nos capítulos anteriores.

Para cada uma destas etapas foram implementadas as ferramentas que permitem ao usuário definir a sequência de procedimentos a ser utilizada na perseguição automática, com a possibilidade de salvar esta sequência como um algoritmo, no caso em que utilize o mesmo tipo de marcador e o mesmo tipo de movimento nas próximas medições. A sequência de etapas do algoritmo geral de uma perseguição de marcadores é apresentada na Figura 6.3. A visão da janela de opções é mostrada na Figura 6.4.

As próximas seções abordam mais detalhadamente a inicialização, teste dos algoritmos e avaliação dos resultados da perseguição automática.

6.2 Inicialização da perseguição automática de marcadores

A fase de inicialização consiste da definição de padrão do marcador, a partir do primeiro quadro, onde, através do mouse, são indicados o centro e o retângulo que engloba o marcador. Em alguns casos, quando o deslocamento é muito grande, é preciso marcar os primeiros 2 ou 3 quadros para definir uma predição acertada nos quadros subsequentes. Nesta etapa, são também definidos o tamanho da região de busca, dependente da velocidade do movimento, e o tamanho da região ocupada pelos marcadores (caso de *tracking* em grupo). A sequência de procedimentos a serem utilizados são também definidos na inicialização, assim como seus respectivos parâmetros. O *tracking* pode ser interrompido e reiniciado em qualquer ponto. A seguir, são detalhados as opções e os parâmetros mostrados na Figura 6.4.

- **Setup.** O tamanho do marcador (**Marker size**) define o raio horizontal (**dx**) e vertical (**dy**) em pixels. Desta forma, o marcador é definido como um retângulo de tamanho $(2dx + 1) \times (2dy + 1)$. Este tamanho é utilizado caso o marcador seja definido por apenas um ponto.

O tamanho da região de busca (**Area of search**) é definido igualmente através da definição de um raio horizontal e vertical em pixels.

O número de pontos que define o marcador (**Number of marker points**) pode ser um ou dois. Se for indicado apenas um ponto, então todos os marcadores utilizam o tamanho definido no campo *marker size*. Uma melhor definição é obtida utilizando

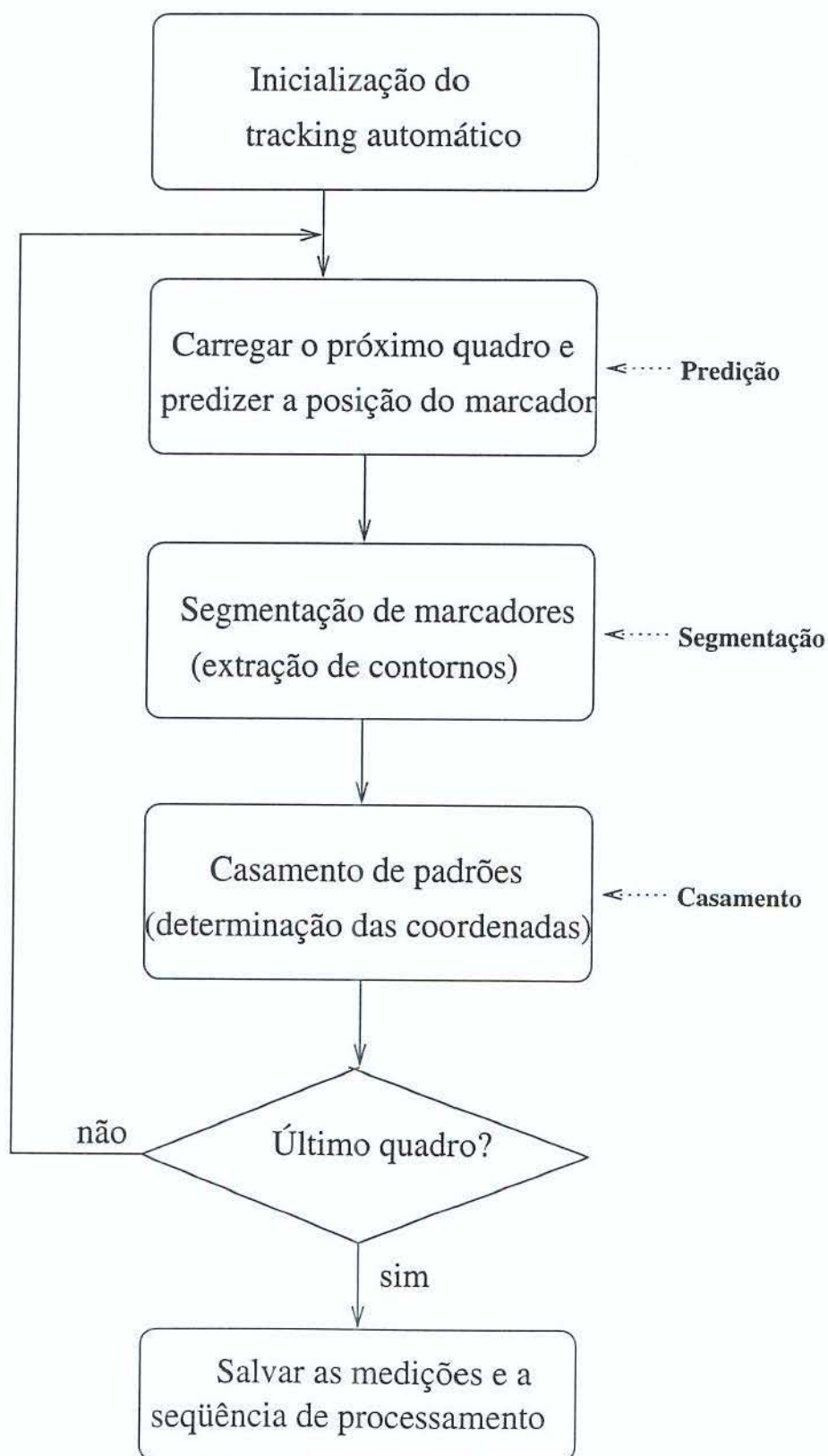


Figura 6.3: sequência de procedimentos da perseguição automática.

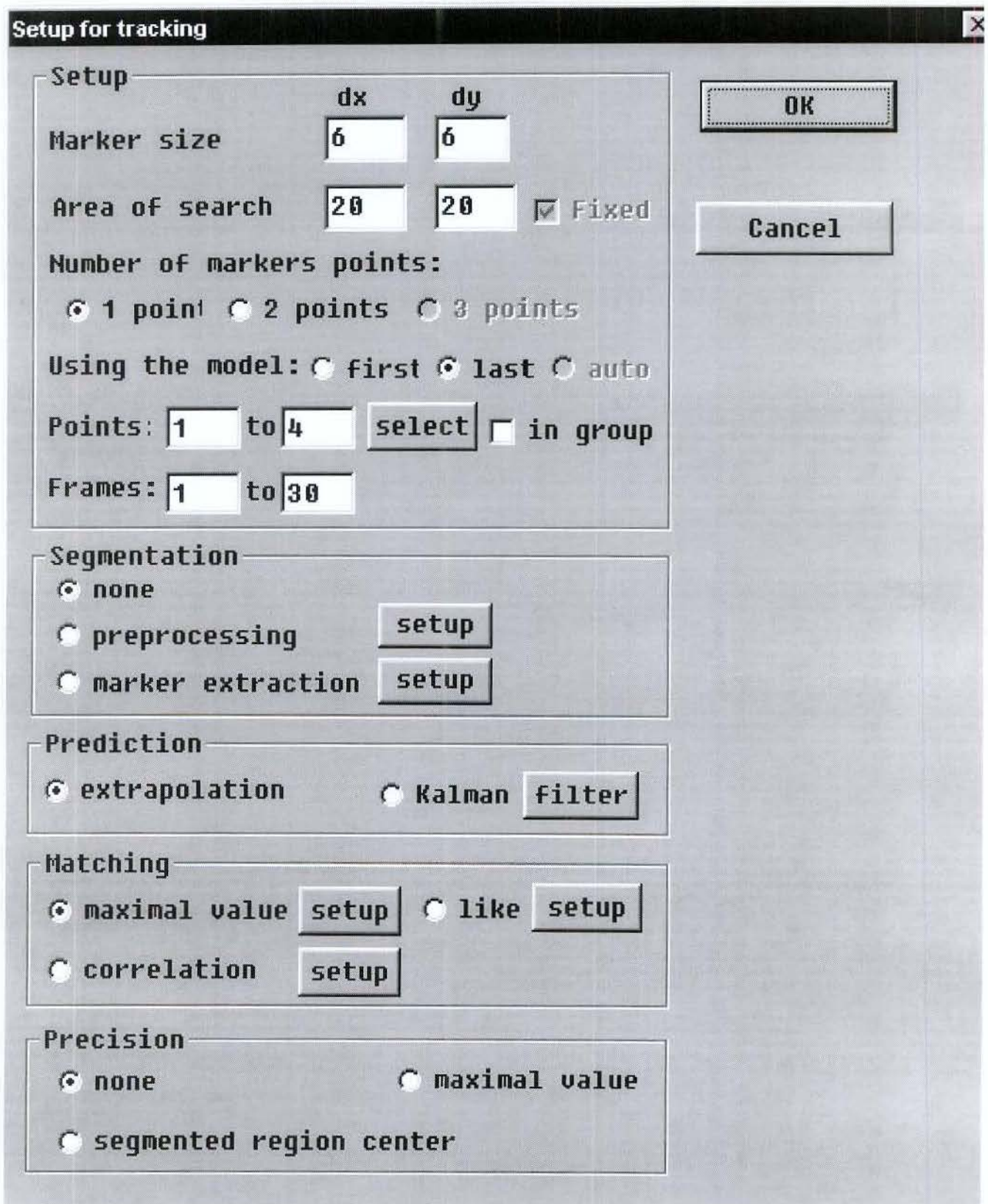


Figura 6.4: janela de opções para a seleção de algoritmos da perseguição automática.

dois pontos já que, neste caso, não se precisa conhecer o tamanho da marca em pixels, além disso, os marcadores podem ter diferentes tamanhos.

Para se escolher o modelo a ser utilizado (**Using the model**) no casamento de padrões, ao longo da seqüência, existem duas opções: o modelo definido a partir da primeira imagem (**first**) ou da última imagem (**last**).

O número de marcadores (**Points**) a serem perseguidos pode ser selecionado fornecendo o número inicial e final, com a possibilidade de selecionar (**select**) alguns deles que se encontram nesse intervalo. Uma opção (**in group**) permite a segmentação em grupo de marcadores que se encontram próximos.

O número de quadros (**Frames**) considerados também pode ser definido. Esta opção é importante na sincronização das câmeras, já que o primeiro quadro numa câmera não representa necessariamente o mesmo instante na outra câmera [7].

- **Segmentation.** Dependendo da aplicação, o que não é o caso da análise do movimento humano considerado aqui, pode ser possível a perseguição dos marcadores sem a necessidade de uma segmentação (**none**), ou apenas com um pré-processamento (**preprocessing**) das imagens. Estas opções podem ser úteis em aplicações em que a precisão é menos importante que o tempo de processamento.

A extração dos marcadores (**marker extraction**) é feita a partir da segmentação apresentada anteriormente, obtendo um *tracking* mais preciso. Tanto para o pré-processamento como para a extração pode ser definido um conjunto de funções de imagens a serem executadas seqüencialmente. Estas funções são agrupadas de acordo com o parâmetro utilizado. A janela que contém estas operações é mostrada na Figura 6.5. Exemplos de configurações de algumas destas operações são apresentadas nas Figuras 6.6 - 6.9.

Algumas destas operações e seus parâmetros são:

- Dilatação, erosão, abertura, fechamento e gradiente utilizam uma máscara (**mask**) planar de tamanho 3x3 (Figura 6.7).
- Afinamento e espessamento utilizam duas máscaras homotópicas (M1 e M2) e duas máscaras para eliminar ramos parasitas (E1 e E2). Pode ser escolhido o número de iterações (**times**) da operação (Figura 6.8).
- Para selecionar as regiões (**Select Contour**) podem ser utilizados os critérios de perímetro (**Length**), a área (**area**), a compacidade (**Length/Area**) e a excentricidade (**Width/Height**). Todas as regiões que cujos critérios têm valores maiores (**greater**) ou menores (**less**) que os valores definidos são selecionados (Figura 6.6).

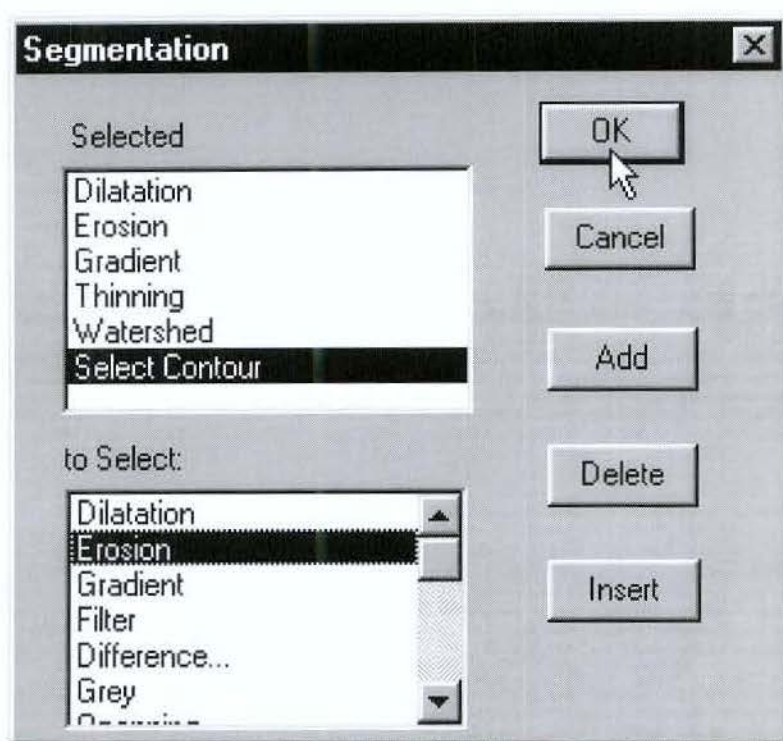


Figura 6.5: seleção de operações para a segmentação dos marcadores.

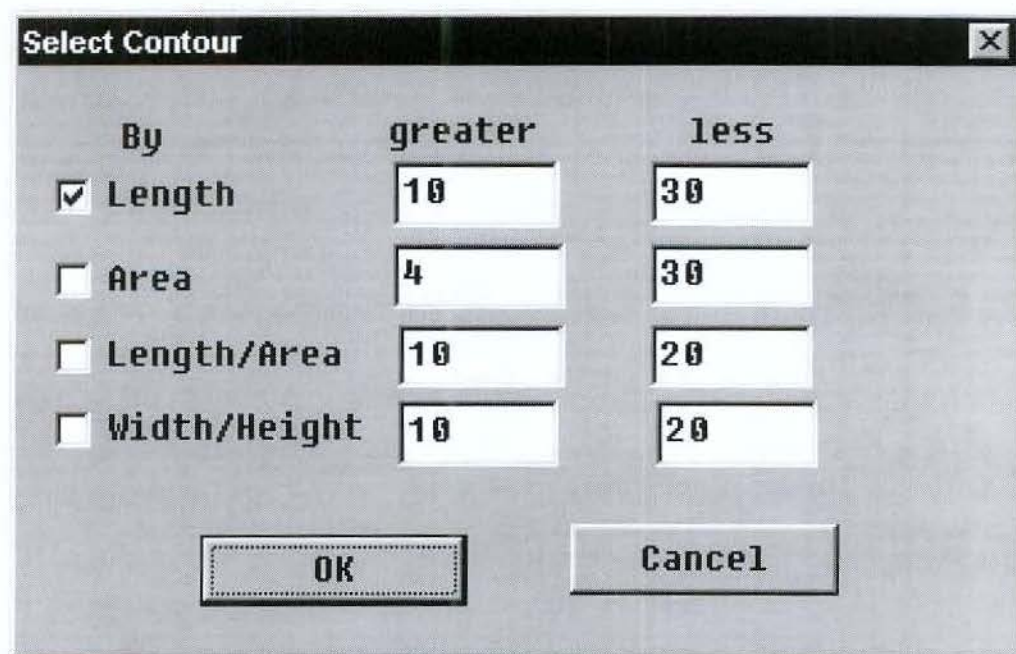


Figura 6.6: um exemplo de janela de parâmetros para a configuração de uma operação.

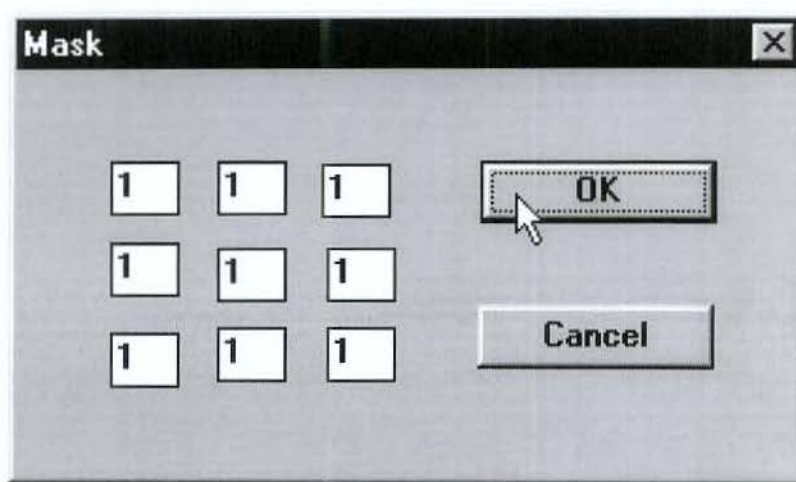


Figura 6.7: definição da máscara para as operações morfológicas.

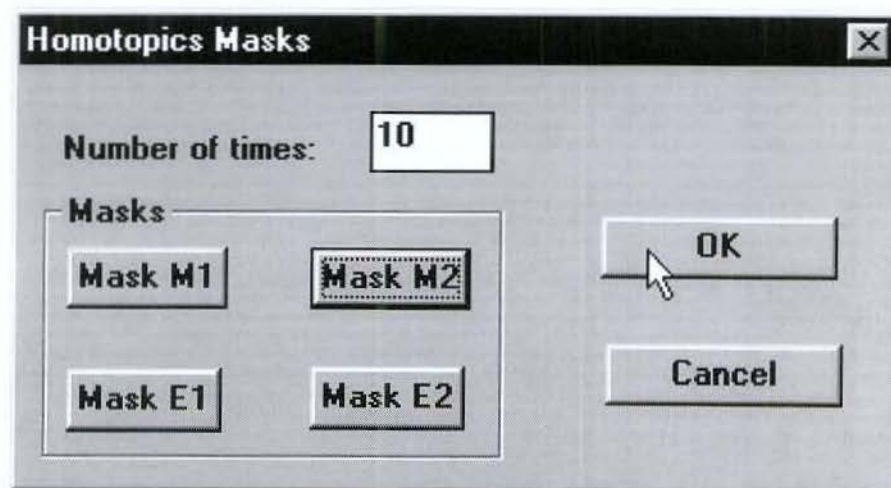


Figura 6.8: janela das máscaras para afinamento e espessamento.

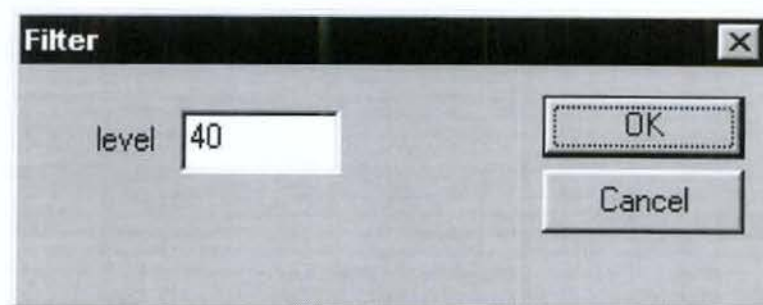
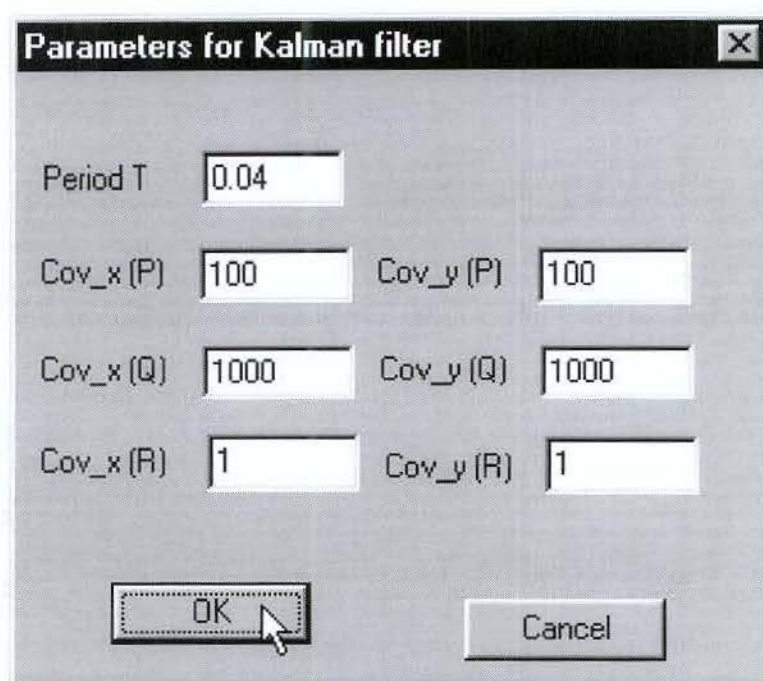


Figura 6.9: seleção de um limiar para operações de filtragem.

- **Prediction.** Com o objetivo de definir a região de busca, na etapa de predição, foram implementados dois métodos: a extrapolação e o filtro de Kalman. A extrapolação é de cálculo simples mas os valores determinados são menos exatos que o filtro de Kalman (Capítulo 4). O filtro de Kalman requer a definição de alguns parâmetros, de uma inicialização e portanto de definição de alguns parâmetros. A Figura 6.10 mostra a janela para a entrada dos parâmetros do filtro de Kalman.



Parameters for Kalman filter	
Period T	0.04
Cov_x (P)	100
Cov_y (P)	100
Cov_x (Q)	1000
Cov_y (Q)	1000
Cov_x (R)	1
Cov_y (R)	1

OK Cancel

Figura 6.10: parâmetros do filtro de Kalman.

- **Matching** No casamento de padrões foram implementados alguns métodos para a definição da posição dos marcadores (determinação das coordenadas) (Capítulo 5). De acordo com a opção selecionada na etapa de segmentação (com extração de contornos ou apenas pré-processamento), são aplicados os algoritmos concernentes. Assim, o casamento pode ser baseado dos seguintes parâmetros: máximo valor (**maximal value**), correlação (**Correlation**) ou na distância com tolerância (**like**). Para cada um destas opções deve ser definido o valor mínimo do casamento (**Minimal value**) a ser considerado. Caso não tenha nenhum ponto com probabilidade maior que este valor mínimo, a marca é considerada perdida.
- Os parâmetros **A** e **B** definem o peso dos critérios (casamento de padrões e distância ao ponto predito) na definição do marcador.

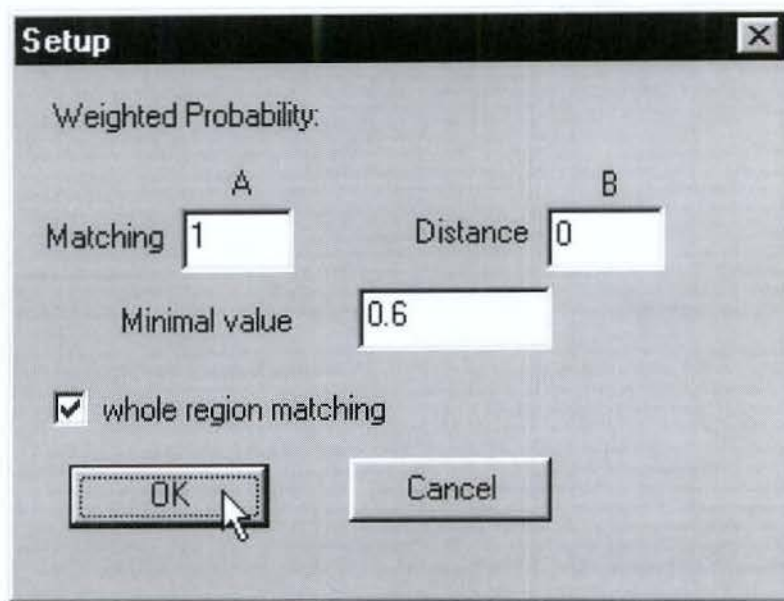


Figura 6.11: janela de parâmetros para o casamento de padrões.

- **Precisão.** Um fator importante para algumas aplicações, como a análise do movimento humano, é a precisão nas medições. A motivação na utilização dos marcadores deve-se principalmente a este fator. No entanto, apesar de melhorar a precisão restringindo o espaço com a ajuda dos marcadores, ainda persiste o problema de determinar com exatidão o centro de cada marcador.

Entre as opções para melhorar a precisão pode ser considerado o centro da região extraído na segmentação (**segmented region center**) ou o valor máximo (**maximal value**) do casamento limitado a esta região (seção 5.3).

Para determinar o centro da região podemos considerar:

$$x_c = \frac{1}{N} \sum_{i=1}^N x_i \quad y_c = \frac{1}{N} \sum_{i=1}^N y_i \quad (6.1)$$

em que N é o número de pontos e x_i e y_i são as coordenadas do ponto i na região.

Para evitar o problema de aproximação no espaço discreto e do resultado do baricentro em região pouco uniforme, consideramos a mediana de x_i e y_i no cálculo do centro da região. A avaliação dos resultados de precisão é abordado na última seção deste capítulo.

6.3 Teste dos algoritmos de perseguição automática dos marcadores

Nesta seção, são mostrados alguns resultados do *tracking* aplicado para diferentes tipos de movimentos, utilizando os algoritmos referentes à etapa de pré-processamento, segmentação e casamento de padrões discutidos anteriormente.

Um primeiro exemplo ilustra a aplicação para a análise de marcha. Neste caso, com o objetivo de obter métricas de quantificação da simetria na marcha, foi analisado o movimento de um sujeito operado do joelho e outro sujeito normal. Em ambos os casos foi possível obter as medições através do *tracking* automático. Estas duas seqüências são apresentadas na Figura 6.12 em que os marcadores são a pintura (círculo branco com contorno preto). Neste caso, foi utilizado o algoritmo de segmentação definido no Capítulo 3 (Figura 3.17), e a correlação para o casamento de padrões. A perseguição de outros tipos de marcadores é mostrada na Figuras 6.13 e 6.14. Para tanto, foi obtida aplicando a seqüência de procedimentos ilustrados nas Figuras 3.15 e 3.14, respectivamente, com o casamento de padrões podendo ser dado por qualquer um dos métodos descritos no Capítulo 5.

Igualmente, foi possível realizar a perseguição automática para outros tipos de movimentos, ou para outros tipos de aplicações, dentro dos projetos em desenvolvimento no Laboratório de instrumentação para a Biomecânica. Entre estes, a Figura 6.15 mostra a imagem em que o objetivo é calcular a energia mecânica durante o movimento na esteira relacionada com o consumo de oxigênio. Na Figura 6.16, é apresentada a coordenada X da trajetória da marca superior.

Outra seqüência de imagens (Figura 6.17) utilizada para a análise da respiração foi considerada com bons resultados. Esta aplicação foi útil para testar outras propriedades do sistema, por exemplo, a capacidade de processar uma maior quantidade de informação, neste caso 14 marcas e mais de 5000 quadros, embora a perseguição destes pontos não seja complicada devido ao pouco deslocamento entre cada quadro e ao tamanho maior da imagem que melhora bastante a qualidade da representação das marcas. O tempo de execução para as 14 marcas foi de aproximadamente 10 horas, sem a visualização do processamento, num computador de 133 MHz. Estes resultados nos levam a pensar na otimização dos algoritmos como trabalho futuro. A Figura 6.18 ilustra a superfície construída a partir destes 14 pontos.

A Figura 6.19 apresenta um outro exemplo de perseguição dos marcadores isopores para a análise da corrida.

O exemplo seguinte ilustra o caso geral em que os “marcadores” são definidos por alguma característica própria do objeto a perseguir. Este exemplo refere-se ao acompanhamento mais complexo do movimento dos jogadores de futebol durante um jogo (Figura

6.20). A abordagem específica dos problemas presentes nesta aplicação pode constituir uma extensão interessante a este trabalho. A trajetória mostrada na Figura 6.20 corresponde a 150 quadros, equivalentes a 5 segundos de jogo. No momento em que dois ou mais jogadores se aproximam começam as dificuldades no reconhecimento.

Estes exemplos ilustram a flexibilidade dos algoritmos e sua possível extensão a outras aplicações.

6.4 Casos de oclusão ou perda de marcadores

Um problema abordado no *tracking* do movimento humano é a possibilidade de oclusão de algumas partes do corpo durante o movimento. Esta oclusão deve-se ao fato de o corpo humano possuir pontos de articulação que permitem virar ou girar em inúmeras direções, de tal forma que alguns pontos não sejam mais visíveis.

Outro fator que pode gerar uma oclusão temporal é a sobreposição, causada entre partes do corpo. Por exemplo, na Figura 6.14 pode-se notar a oclusão da marca no trocante pela mão do paciente.

Os casos de oclusão não têm um tratamento geral e neste trabalho as oclusões são consideradas como perda temporal da marca. A dificuldade inicial de uma oclusão é a sua detecção. Aqui, a marca é considerada perdida quando o valor das probabilidades na etapa de casamento é muito baixo (observe que etapa de segmentação não se pode garantir a existência de oclusões já que a marca pode não ser detectada por um erro na segmentação).

Ao se detectar a perda da marca, o modelo do marcador empregado nos próximos quadros é o modelo anterior à oclusão. Este modelo só será atualizado, posteriormente, após a obtenção de valores altos de probabilidades. Em casos de oclusão, os valores de predição são utilizados para definir a posição do marcador no quadro atual. Por exemplo, na figura 6.14, em dois intervalos da seqüência houve oclusão do marcador no trocante e os valores obtidos nestes intervalos foram extrapolados.

6.5 Avaliação dos resultados

Os resultados obtidos no teste dos algoritmos mostram que na maioria dos casos foi possível fazer um *tracking* sem intervenção de um operador. As dificuldades iniciais estão relacionadas com a definição dos parâmetros dos algoritmos que envolve um treinamento com a primeira seqüência até a obtenção dos adequados para um determinado tipo de movimento e marcador. Os isopores mostraram ser o melhor tipo de marcador na análise de marcha. De forma similar, o *tracking* com as pinturas funcionaram corretamente em quase todas as seqüências, utilizando o algoritmo correspondente definido no Capítulo 3.

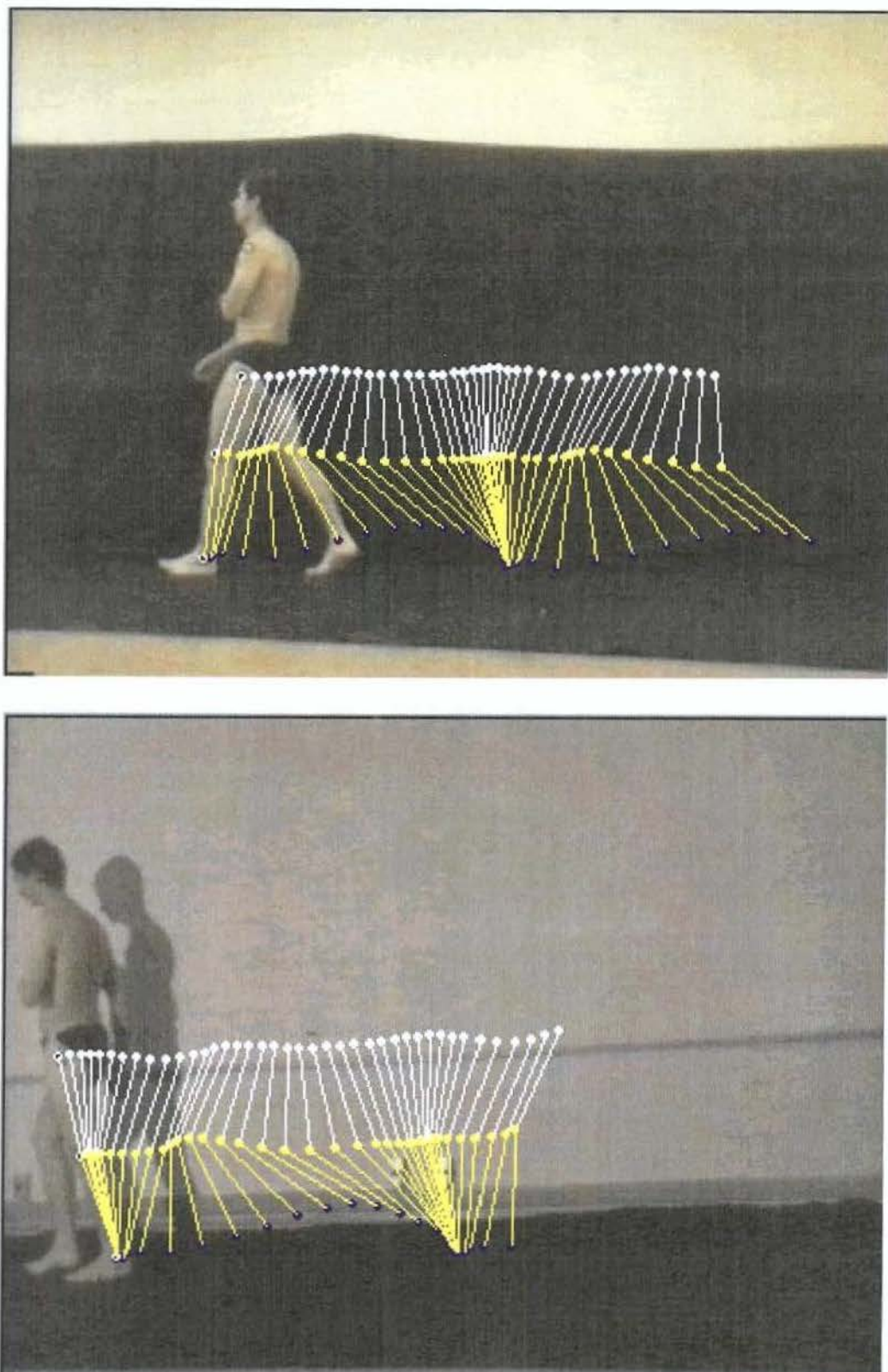


Figura 6.12: trajetória dos marcadores para duas seqüências de imagens.



Figura 6.13: trajetória de marcadores do tipo LED.

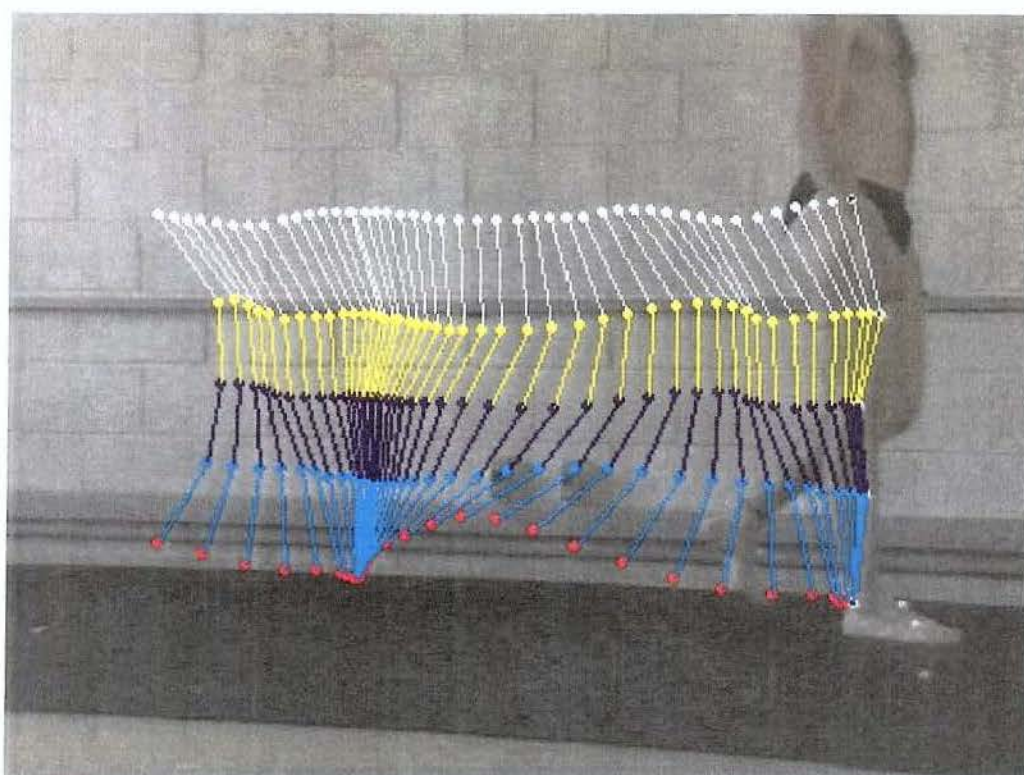


Figura 6.14: trajetória de marcadores do tipo adesivo com casos de oclusão.



Figura 6.15: movimento na esteira.

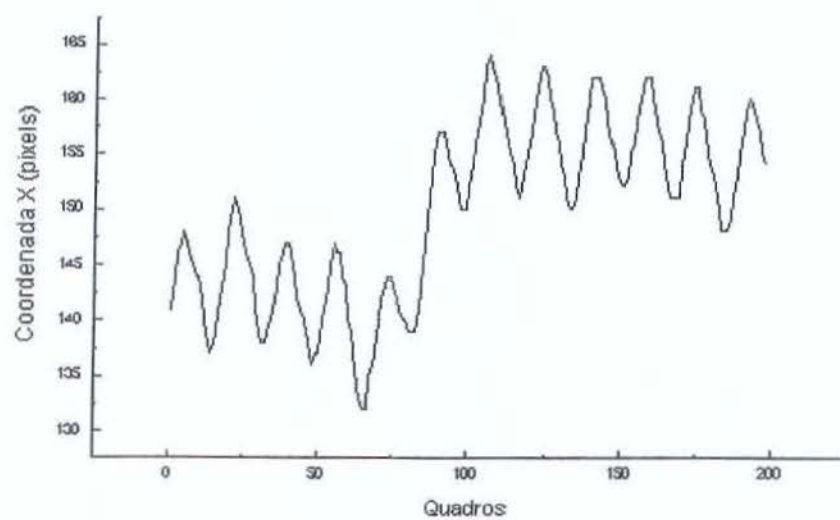


Figura 6.16: trajetória de uma das marcas do movimento na esteira.

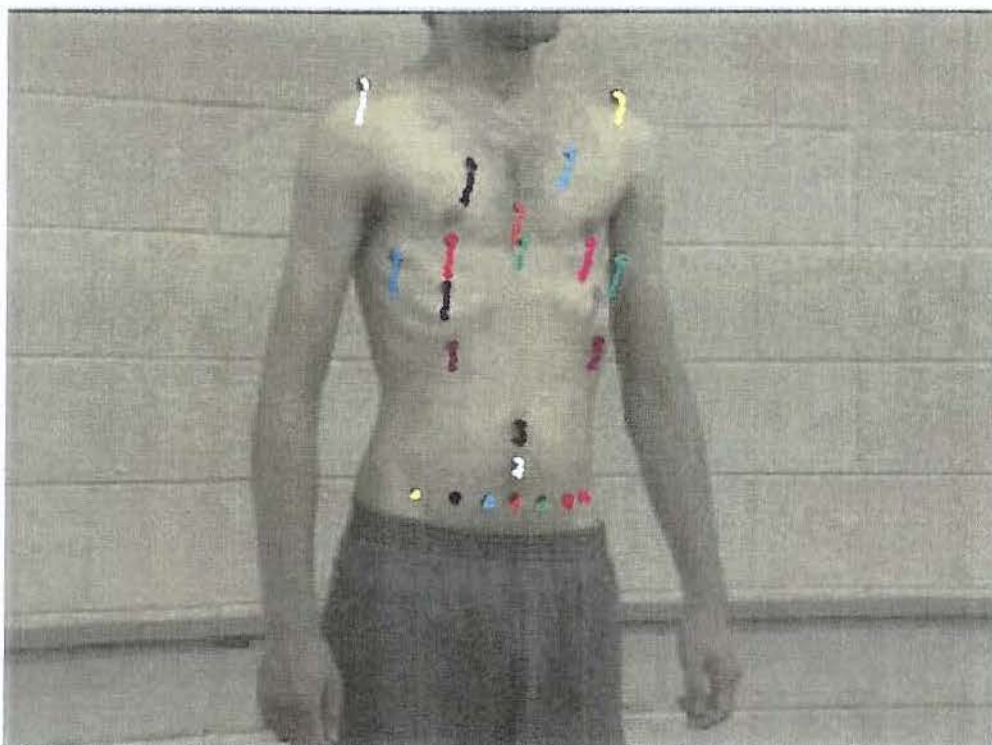


Figura 6.17: perseguição de marcadores adesivos para análise de respiração.

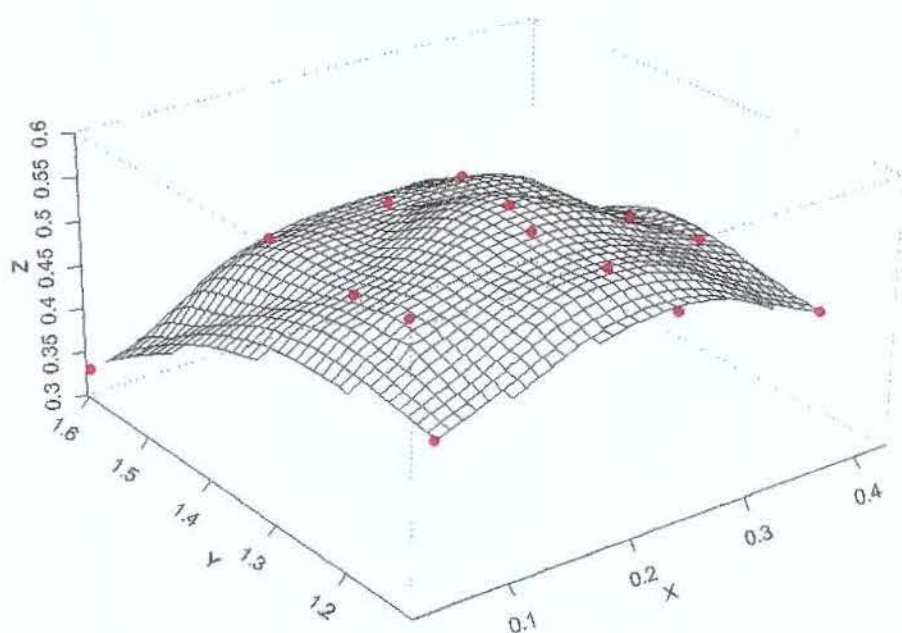


Figura 6.18: superfície do corpo reconstruída a partir de 14 pontos.

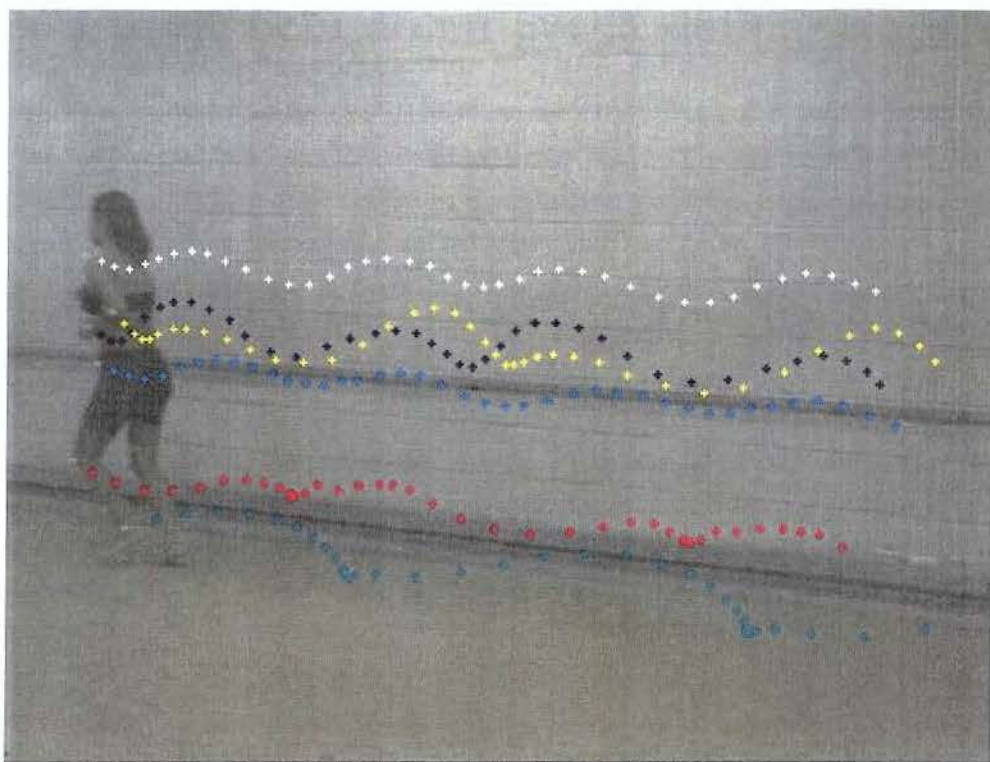


Figura 6.19: perseguição de marcadores para análise de corrida.



Figura 6.20: perseguição de um jogador de futebol.

Com o objetivo de avaliar a precisão dos dados da perseguição automática, podem ser utilizados os resultados de uma medição manual, embora a mesma não garanta uma medição exata, devido a erros aleatórios na marcação (a suavidade na trajetória pode ser um outro fator de avaliação).

Nas figuras 6.21 e 6.22 são mostradas as coordenadas X da trajetória obtida através de medição manual e medição automática, respectivamente, dos marcadores da Figura 6.12. Analogamente, as figuras 6.23 e 6.24 mostram as coordenadas Y das medições.

6.6 Conclusão

Neste capítulo, foram apresentadas as particularidades do sistema em desenvolvimento e a implementação dos algoritmos abordados nos capítulos anteriores na definição da perseguição automática de marcadores. Finalmente, foram mostrados alguns resultados obtidos em imagens reais.

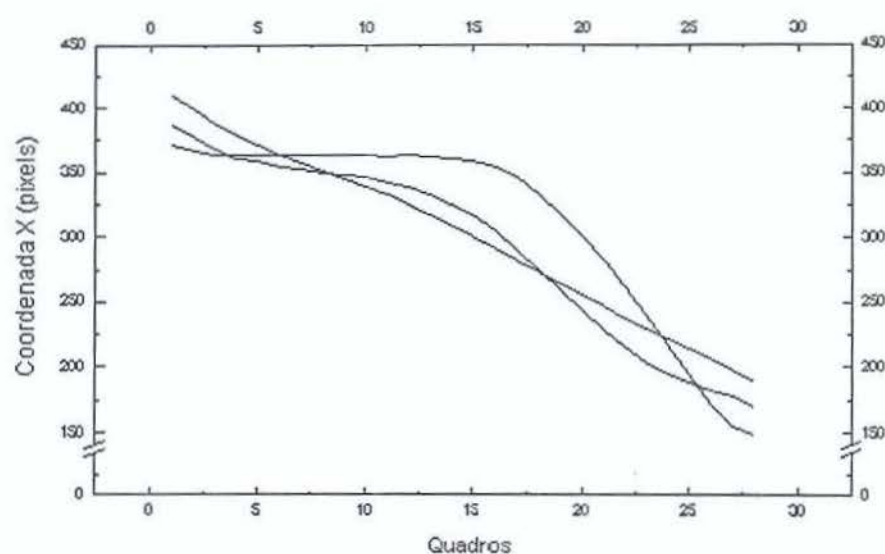


Figura 6.21: Coordenadas X da trajetória dos marcadores obtida manualmente.

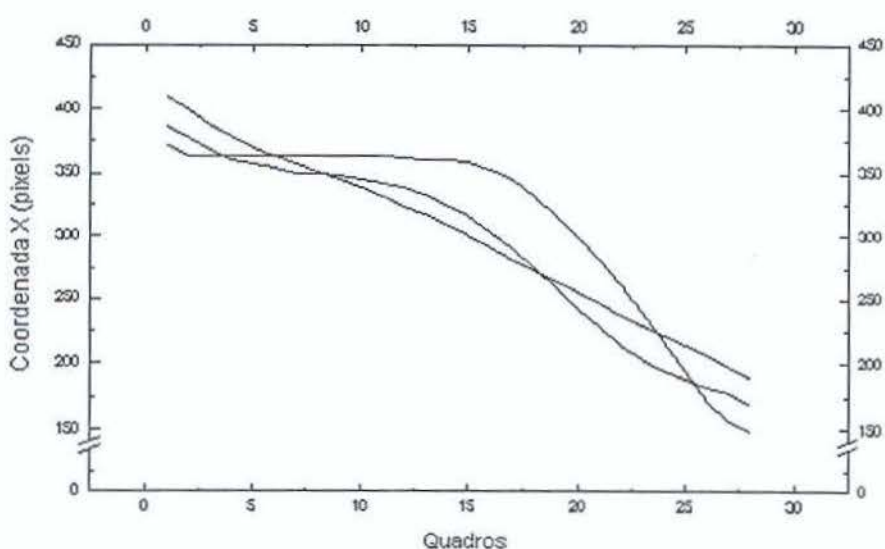


Figura 6.22: Coordenadas X da trajetória dos marcadores obtida automaticamente.

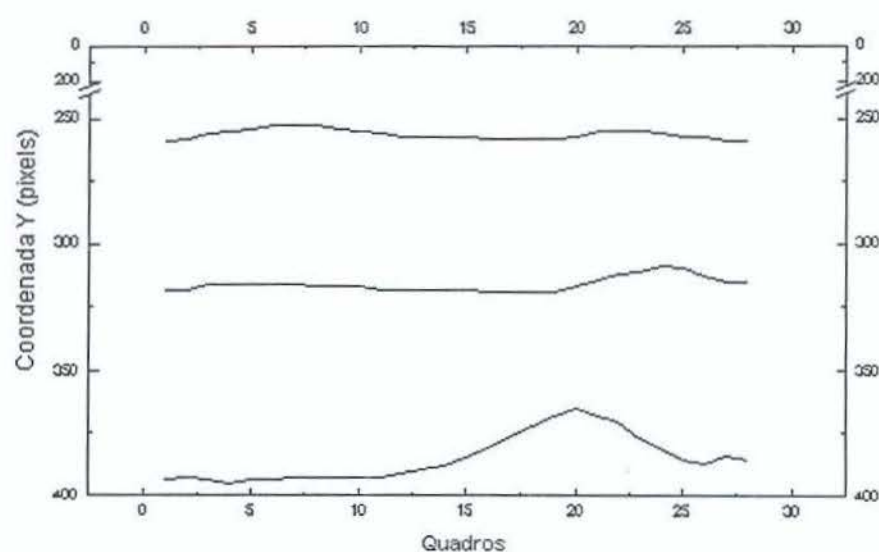


Figura 6.23: Coordenadas Y da trajetória dos marcadores obtida manualmente.

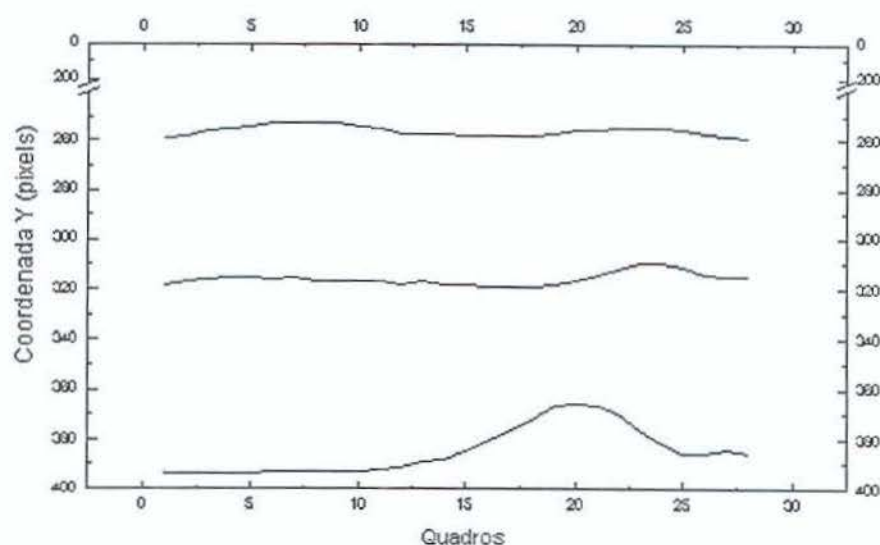


Figura 6.24: Coordenadas Y da trajetória dos marcadores obtida automaticamente.

Capítulo 7

Conclusões

Neste trabalho, é abordado o problema da perseguição automática de marcadores associada a análise do movimento humano. Nós apresentamos uma seqüência de técnicas para a perseguição de marcadores utilizando segmentação de imagens, predição da posição dos marcadores e casamento de padrões.

A possibilidade de utilizar câmeras normais para a filmagem do movimento e o computador para a detecção e perseguição destes marcadores facilita a implementação e utilização de tais sistemas em laboratórios, sem necessidade de equipamentos sofisticados. Outra vantagem é a flexibilidade de adaptar os algoritmos de acordo com o tipo de marcadores, condições ambientais e tipo de movimento analisado.

Utilizando os algoritmos definidos neste trabalho, foi desenvolvido um sistema que permite a perseguição automática dos marcadores e que pode ser utilizado na análise tridimensional do movimento humano. Esta análise consiste em determinar as coordenadas do marcador no espaço, utilizando técnicas de reconstrução, a partir de medições em imagens bidimensionais. As medições são obtidas aplicando o tracking automático para duas seqüências de imagens filmadas com duas câmeras estáticas, com diferentes ângulos de visão. Os parâmetros da reconstrução são obtidos a partir da calibração das câmeras empregando a DLT (*direct linear transformation*) [2], por exemplo, e alguns pontos de referência dos quais é possível obter informação das coordenadas reais e coordenadas na imagem.

Um ponto importante na definição dos algoritmos para o tracking automático é a introdução de ferramentas de processamento de imagens, em particular da morfologia matemática. A seqüência de ferramentas definida para a extração dos marcadores é baseada, principalmente, nas informações de forma e intensidade destes marcadores na imagem.

Os resultados obtidos a partir dos testes realizados com diferentes seqüências foram bastante satisfatórios. Para um grande número de diferentes seqüências foi possível apli-

car o tracking totalmente automático e em outros casos semi-automático, precisando de algumas marcações adicionais quando a marca foi perdida (caso de grande oclusões temporais).

São muitos os fatores que podem dificultar o tracking automático, um deles é a presença de outros marcadores de mesma característica na região vizinha. Outro fator refere-se à oclusão. Estes problemas são abordados neste trabalho, incluindo critérios de decisão como, por exemplo, a trajetória dos pontos (distância ao ponto predito).

Extensões a este trabalho incluem alternativas para resolver os problemas de decisão considerando, por exemplo, a trajetória no espaço para a extrapolação. Esta alternativa deve ser explorada em casos de aplicações que necessitem de reconstrução tridimensional. Um outro caminho para se abordar tal problema pode ser através de um modelo criado a partir dos mesmos marcadores, considerando a posição relativa entre eles e a forma do modelo do corpo (pontos unidos por segmentos).

Para a extensão do casamento de padrões, pode ser considerado o reconhecimento estatístico de padrões. Por se tratar de uma seqüência de imagens, a estatística das mudanças é um detalhe importante a ser levado em consideração.

Bibliografia

- [1] K. Akita. Image Sequence Analysis of Real World Human Motion. *Pattern Recognition*, 17(1):73–83, 1984.
- [2] P. Allard, I. Stokes, and J. Blanche. *Three-Dimensional Analysis of Human Movement*. Human Kinetics, 1995.
- [3] A. Azarbayejani, C. Wren, and A. Pentland. Real-Time 3-D Tracking. Technical Report TR96-374, Perceptual Computing Section MIT Media Laboratory, 20 Ames St., Cambridge, MA 02139, 1996.
- [4] G. J. Banon and J. Barrera. Bases da morfologia matemática para a análise de imagens binárias. IX Escola de Computação. 24 a 31 de Julho, Recife, PE, Brazil, 1994.
- [5] S. Beucher and F. Meyer. The morphological approach to segmentation: The watershed transformation. In E.R Dougherty, editor, *Mathematical Morphology in Image Processing*, chapter 12, pages 433–481. Marcel Decker, Inc, New York, 1992.
- [6] I. J. Cox and S. L. Hingorani. An efficient implementation of Reid’s multiple hypothesis tracking algorithm and its evaluation for purpose of visual tracking. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 18(2):138–150, 2 1996.
- [7] Ricardo M. Leite de Barros. *Concepção e Implementação de um Sistema para Análise Cinematica de Movimentos Humanos*. PhD thesis, Universidade Estadual de Campinas - Brasil, 1997.
- [8] G. Ferrigno, N. A. Borghese, and A. Pedotti. Pattern recognition in 3d automatic human motion analysis. *IPRS Journal of Photogrammetry and Remote Sensing*, 45(4):227–246, 8 1990.
- [9] E. H. Furnée. Opto-electronic movement measurement systems: Aspects of data acquisition, signal processing. In U. Boenick, M. Näder, and C. Mainka, editors, *Gangbildanalyse, Stand der Messtechnik und Bedeutung für die Orthopädie-Technik*,

- pages 112–129. Technische Universität Berlin, Mecke Druck und Verlag, Duderstad, 2 1990.
- [10] D.M. Gavrilu and L.S. Davis. Tracking of Human in Action: a 3-D Model-Based Approach. Technical report, Computer Vision Laboratory, CfAR, University of Maryland, College Park, MD 20742, USA, 1996.
- [11] Nigel H. Goddard. *The Perception of Articulated Motion: Recognizing Moving Light Display*. PhD thesis, University of Rochester, 1992.
- [12] Michael Grimand. A new measure of contrast: Dynamics. In *Proc. Image Algebra and Morphological Image Processing III*, pages 294–305, 1992.
- [13] R. M. Haralick, S. R. Sternberg, and X. Zhuang. Image analysis using mathematical morphology. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 9(4):532–549, 7 1987.
- [14] Simon Haukin. *Modern Filters*. Macmillan Publishing Company, 1989.
- [15] R. Hosie and G. West. Prediction ball trajectories in uncontrolled environment. In *Proceedings DICTA-95: Digital image Computing: Techniques and Applications*, Brisbane, 6-8 December 1995.
- [16] D.P. Huttenlocher, J.J. Noh, and W.J. Rucklidge. Tracking Non-Rigid Objects in Complex Scenes. Technical Report TR92-1320, Computer Science Department Cornell University, Ithaca NY 14853 USA, 1992.
- [17] R. P. Johnson. Contrast based edge detection. *Pattern Recognition*, 23(3/4):311–318, 1990.
- [18] Shanon Ju. Human motion estimation and recognition. Depth oral report, University of Toronto, 1996.
- [19] R. E. Kalman. A new approach to linear filtering and prediction problems. *ASME J. Basic Eng.*, pages 35–45, 3 1960.
- [20] M.K. Leung and Y.H. Yang. First Sight: A human body outline labeling system. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 17(4):359–377, 1995.
- [21] S. A. Niyogi and E. H. Adelson. Analyzing and Recognizing Walking Figure in XYT. Technical Report TR94-223, Perceptual Computing Section MIT Media Laboratory, 20 Ames St., Cambridge, MA 02139, 1992.

- [22] D. W. Paulus and J. Hornegger. *Pattern Recognition and Image Processing in C++*. Vieweg & Sohn Verlagsgesellschaft, 1995.
- [23] K. Rohr. Towards Model-Based Recognition of Human Movements in Image Sequences. *CVGIP: Image Understanding*, 59(1):94–115, 1994.
- [24] A. Rosenfeld. *Multiresolution Image Processing and Analysis*. Springer Verlag, Berlin, 1984.
- [25] J. Serra. *Image Analysis and Mathematical Morphology*. Academic Press, second printing, London, 1992.
- [26] Joseph S. Shu. One-pixel-wide edge detection. *Pattern Recognition*, 22(6):665–673, 1989.
- [27] M. Sonka, V. Hlavac, and R. Boyle. *Image Processing, Analysis and Machine Vision*. Chapman & Hall Computing, 1993.
- [28] E. Stüssi and R. Müller. Vergleichende bewertung kommerziell erhältlicher 3d-kinematik-systeme für die gangbildanalyse. In U. Boenick, M. Näder, and C. Mainka, editors, *Gangbildanalyse, Stand der Messtechnik und Bedeutung für die Orthopädie-Technik*, pages 86–97. Technische Universität Berlin, Mecke Druck und Verlag, Duderstad, 2 1990.
- [29] Ferdinand van der Heijden. *Image Based Measurement Systems*. John Wiley & Sons Ltd., 1994.
- [30] Luc Vicent. Morphological transformation of binary images with arbitrary structuring elements. *Singal Processing*, 22(1):3–23, 1 1991.
- [31] Luc Vicent and Pierre Soille. Watershed in digital spaces: An efficient algorithm based on immersion simulations. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 13(6):583–598, 6 1991.