

UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS EXATAS - ICEX
DEPARTAMENTO DE CIÊNCIA DA COMPUTAÇÃO

**RASTREAMENTO E CONTAGEM DE PEIXES
UTILIZANDO FILTRO PREDITIVO**

ERIKSON FREITAS DE MORAIS

Belo Horizonte

29 de abril de 2005

ERIKSON FREITAS DE MORAIS

**RASTREAMENTO E CONTAGEM DE PEIXES
UTILIZANDO FILTRO PREDITIVO**

Dissertação apresentada ao Curso de Pós-Graduação em Ciência da Computação da Universidade Federal de Minas Gerais como requisito parcial para a obtenção do grau de Mestre em Ciência da Computação.

Belo Horizonte

29 de abril de 2005

UNIVERSIDADE FEDERAL DE MINAS GERAIS

FOLHA DE APROVAÇÃO

Rastreamento e contagem de peixes utilizando filtro preditivo

ERIKSON FREITAS DE MORAIS

Dissertação defendida e aprovada pela banca examinadora constituída pelos Senhores:

Prof. Dr. MÁRIO FERNANDO MONTENEGRO CAMPOS - Orientador
Departamento de Ciência da Computação – ICEX – UFMG

Prof. Dr. RODRIGO LIMA CARCERONI
Departamento de Ciência da Computação – ICEX – UFMG

Prof. Dr. ALEXEI MANSO CORREIA MACHADO
Departamento de Ciência da Computação – PUCMG

Belo Horizonte, 29 de abril de 2005.

Resumo

Este trabalho aborda o problema de se estimar o número de peixes que nadam através de mecanismos de transposição, também conhecidos como escadas de peixes. Este mecanismo são construídos com o objetivo de auxiliar os peixes em seus processos migratórios para desova, fenômeno este conhecido no Brasil como *Piracema*. O sistema de contagem de peixes proposto neste trabalho juntamente com um módulo complementar para classificação das espécies de peixes, compõe um sistema mais amplo cujo objetivo é realizar o monitoramento e a avaliação da eficácia de escadas de peixe. Devido a rica biodiversidade do ecossistema brasileiro, este trabalho representa uma tarefa de grande relevância para o estudo e quantificação dos volumes de espécies de peixes que se encontram em extinção ou ainda que tenham um significativo valor econômico. Diferentemente dos métodos existentes, os quais não fornecem meios adequados para aquisição de informações relevantes sobre espécies diferentes de peixes, tal como suas habilidades de nado, tempos de migração e picos de fluxo, este trabalho apresenta uma abordagem baseada em análise de sequência de imagens e no uso de um rastreador *bayesiano* de múltiplos objetos (BraMBLe), a qual permite a realização de um rastreamento mais confiável de peixes mesmo diante de variações ambientais bruscas, como variações de luminosidade e propriedades da água. A direção de migração dos peixes é estimada dividindo a cena em regiões de interesse que são constantemente monitoradas pelo sistema. Nossa metodologia foi validada com sucesso utilizando-se vídeos do mundo real, alcançando uma exatidão de contagem de cerca de 80%.

Abstract

This work addresses the problem of estimating the number of fish that swim by an observation window installed in a fish transposition mechanism, also known as a fish ladder, which are built to assist them in their upriver journey to spawning grounds. Our system and a fish classifier module compose a complex computer vision based system for monitoring and evaluating the adequacy and efficiency of fish ladders. Due to the opulent biodiversity of Brazilian ecosystem, this represents a formidable task, whose main concern is with species that are either in extinction or economically relevant. The target site is the hydroelectric generation plant of Igarapava, on the Rio Grande river in Minas Gerais, Brazil. Our method compares positively to existing ones in that it is able, besides counting, to acquire relevant information. Characteristics of different fish species that may be obtained are swimming ability, time of migration and peak flow rates. Our approach is based on image sequence analysis and on the use of a bayesian multiple-blob tracker (BraMBLe). This integration enables tracking fishes reliably even under severe environmental changes, such as variations in illumination parameters and water characteristics. Also, fish trajectories in the field of view and other necessary information for their accurate counting are achieved. Fish migration are estimated by dividing the scene in regions of interest that are constantly monitored by the system. Our methodology was successfully validated with video streams acquired at the fish ladder, and it was able to attain overall accuracy as high as 80.95%.

A meu amor, Claudiana, que soube exercitar o dom da paciência e se fez fonte de incentivos em nome do sonho de estarmos juntos.

Agradecimentos

Agradecimentos Em primeiro lugar devo agradecer a Deus que me deu forças para chagar onde cheguei. “Esperei com paciência ao SENHOR, e ele se inclinou para mim e ouviu o meu clamor”(SL 40:1).

Agradeço ao Professor Mário Campos pela paciência, aprendizado e amadurecimento científico. Agradeço também a meus colegas de laboratório que participaram direta ou indiretamente no trabalho. Em especial aos amigos Flávio Cardeal pelo apoio e incentivo constates, Pedro Shiroma pela paciência inabalável, Wagner Barros pela descontração, José Pio pelas revisões e interesse demonstrado, José Pinheiro, Vilar Fiuza, Patrícia Martins, Ligiane Alves, Fabiano Botelho e tantos outros amigos que sempre estiveram presentes, dentro ou fora do VeRLab.

Agradeço a meus pais que me ensinaram a ser quem sou, sempre com carinho, atenção e confiança. Mesmo sendo conhecedores de minhas limitações, nunca perderam a certeza de meu sucesso. A minha futura esposa, Claudiana, que me deu atenção quando precisei, compreendeu as severas restrições que passamos e aceitou paciente-mente a distância de 340 Km imposta entre nós.

Devo agradecer também a Kristin Branson e Serge Belongie por disponibilizarem sua implementação do algoritmo BraMBLe, assim como Michael Isard por sua ajuda na busca por uma implementação.

Por fim, agradeço ao Professor Ulisses Azevedo Leitão, que viabilizou a importante ajuda financeira fornecida pelo Instituto Doctum de Educação e Tecnologia, sem a qual o mestrado não seria possível.

A todos o meu muito obrigado.

Sumário

Lista de Figuras	x
Lista de Tabelas	xi
1 Introdução	1
1.1 Motivação	3
1.2 Abordagem	5
1.3 Contribuições	9
1.4 Organização do texto	9
2 Trabalhos Relacionados	10
2.1 Métodos convencionais de contagem de peixes	10
2.1.1 Métodos acústicos	10
2.1.2 Métodos de enfileiramento	13
2.1.3 Análise de imagens	15
2.2 Rastreamento Visual	17
2.2.1 Filtro de Kalman	18
2.2.2 CONDENSATION	21
2.2.3 BraMBLe – A Bayesian Multiple Blob Tracker	23
3 Metodologia	26
3.1 Descrição da metodologia	26
3.1.1 Pré-processamento	28
3.1.2 Treinamento	31
3.1.3 Modelo de Aparência	36
3.1.4 Rastreamento	40
3.1.4.1 Propagação temporal	41
3.1.4.2 Factored Sampling	41
3.1.4.3 Algoritmo CONDENSATION	43
3.1.5 Predição	44
3.1.6 Estimador de objetos válidos	48
3.1.7 Contagem de peixes	49

4 Experimentos	53
4.1 Planejamento dos experimentos	53
4.2 Primeiro experimento - Rastreamento de <i>blobs</i> coloridos	55
4.2.1 Configuração e resultados	55
4.2.2 Discussão	61
4.3 Segundo experimento - Rastreamento de peixe em laboratório	62
4.3.1 Configuração e resultados	62
4.3.2 Discussão	67
4.4 Terceiro experimento - Rastreamento de peixes no mecanismo de trans- posição	67
4.4.1 Configuração e resultados	67
4.4.2 Discussão	77
4.5 Contagem automática	78
5 Conclusões e direções futuras	81
5.1 Direções futuras	82
Referências Bibliográficas	87

Lista de Figuras

1.1	Exemplos de mecanismo de transposição de peixes.	4
1.2	Sala de observação do mecanismo de transposição instalado em Igarapava.	5
1.3	Esquemático do sistema de contagem e classificação de peixes.	6
1.4	Exemplo de recuperação de imagens usando o resultado de um algoritmo de visão estéreo.	7
1.5	Exemplo de identificação do objeto de interesse.	8
2.1	Esquemático da estrutura de PVC proposta no trabalho de [Menin, 1984].	12
2.2	Esquemático do túnel para contagem de organismos aquáticos.	14
2.3	Contagem de peixes usando imagens lineares paralelas.	16
2.4	Arquitetura do sistema de rastreamento de pessoas e ciclistas.	19
2.5	Rastreamento de formigas.	23
2.6	Pessoa identificada pela observação do método BraMBLe.	24
3.1	Exemplo de Segmentação e Rotulação de objetos.	27
3.2	Esquemático do funcionamento da metodologia proposta.	28
3.3	Exemplos de <i>kernel</i> discreto finito.	29
3.4	Pré-Processamento das imagens.	30
3.5	Aspecto de uma <i>fdp</i> modelada por combinação de Gaussianas.	32
3.6	<i>fdp</i> de uma imagem representada por combinação de Gaussianas.	33
3.7	Exemplo de mapa de probabilidades.	35
3.8	Factored Sampling.	43
3.9	Propagação de $P(X_t Z_t)$	45
3.10	Probabilidade total do peixe representado por s_t^r	48
3.11	Imagens dividida em regiões para a contagem.	51
4.1	Escala de cores adotada para a exibição das imagens de saída.	54
4.2	Esquemático do primeiro experimento.	55
4.3	Primeiro experimento.	57
4.4	Trajetórias encontradas no experimento 1.	60
4.5	Esquemático do segundo experimento.	62
4.6	Segundo experimento.	64
4.7	Trajetórias encontradas no experimento 2.	66
4.8	Esquemático do terceiro experimento.	68

4.9	Terceiro experimento A.	69
4.10	Terceiro experimento B.	71
4.11	Terceiro experimento C.	73
4.12	Trajetórias encontradas no experimento 3.	76
4.13	Exemplo de falso negativo.	79
4.14	Resultado da contagem automática.	80

Lista de Tabelas

4.1	Erros e desvios padrão encontrados no experimento 1.	61
4.2	Erros e desvios encontrados no experimento 2.	66
4.3	Número de imagens utilizadas para o treinamento e rastreamento das seqüências adquiridas em Igarapava	68
4.4	Erros encontrados no experimento 3.	77
4.5	Desvios padrão encontrados no experimento 2.	77

Lista de Algoritmos

1	Algoritmo de treinamento em um <i>pixel</i> (i, j).	36
2	Calculo da probabilidade logarítmica.	39
3	Algoritmo CONDENSATION.	46
4	Algoritmo de predição.	47
5	Algoritmo de estimação.	49
6	Algoritmo de contagem.	51

Capítulo 1

Introdução

A necessidade de contar objetos é antiga na história humana. O homem primitivo contava as ovelhas de seu rebanho mesmo sem conhecer um sistema de numeração. Cada ovelha correspondia a uma das pedras guardadas dentro de um saco, daí a origem da palavra *cálculo* usada nos tempos atuais, derivada da palavra latina *calculus* que significa pedrinha. A correspondência, unidade a unidade, era feita também usando nós em cordas, marcas em paredes, desenhos, talhes em ossos e outros tipos de marcação. Os sistemas de numeração surgiram, as civilizações se desenvolveram e se modernizaram sempre com a necessidade de contar coisas.

A tarefa de contar objetos é parte do cotidiano das pessoas. Em exames biológicos, a contagem possui significado importante, quantificando níveis de contaminação, presença de bactérias, qualidade do sangue e muitas outras.

Em processos industriais, a produtividade é medida contando-se as unidades produzidas com qualidade. A aceitação de um produto é medida por seu número de vendas, sendo que muitas vezes o processo de contagem é feito automaticamente.

Em particular, a contagem automática de objetos que se movem independentemente pode ser aplicada em contextos diferenciados. A quantificação do fluxo de pessoas por exemplo, constitui uma informação importante para os sistemas modernos de tráfego urbano, útil tanto no projeto quanto no controle do fluxo, permitindo melhor organização dos grandes centros metropolitanos.

O mesmo tipo de contagem automática aplicada em áreas estratégicas de estabelecimentos comerciais, fornece informações valiosas à equipe gerencial, permitindo otimizar o atendimento em função dos horários de maior procura, estudo da disposição de produtos de forma a melhor atender às necessidades dos clientes, criar políticas internas de marketing valorizando espaços, avaliar as vitrines em função do número

de pessoas que param para observar, dimensionamento rápido do pessoal de segurança em áreas de grande fluxo, tanto comercial quanto públicas como estações de embarque e desembarque. A contagem automática de automóveis permite determinar mudanças na estruturação do tráfego, minimizando transtornos e aumentando a segurança tanto de condutores quanto de pedestres, ou ainda, distribuir o fluxo de veículos de modo a aumentar a vida útil das vias. Exames médicos baseados em contagem também possuem métodos automáticos para a contagem, dando maior confiabilidade nos resultados. Como exemplo podemos citar a contagem de plaquetas e o hemograma.

A visão computacional é composta por um conjunto de técnicas matemáticas e computacionais que objetivam, principalmente, estimar ou reconstruir as características geométricas e dinâmicas do mundo 3D por meio de uma ou mais imagens [Trucco and Verri, 1998]. Dentre elas, técnicas para detecção de objetos e análise de movimento. Tais técnicas podem ser aplicadas ao rastreamento visual, dimensionamento do fluxo de veículos, controle de qualidade, exames médicos, contagem automática de pessoas, entre outros, sem interferir no ambiente dos objetos, ou seja, de forma não intrusiva.

A contagem de objetos por meio da visão computacional permite a identificação visual dos objetos contados utilizando informações úteis como forma, cor e textura. Em geral, os sistemas de contagem automática baseados em visão computacional permitem a análise de outras variáveis como velocidade dos objetos de interesse e trajetórias percorridas dentro do campo visual.

Pesquisadores no campo da biologia têm desenvolvido pesquisas que acompanham a população de cardumes, com o objetivo de preservar a vida de peixes bastante valiosos para pescadores e amantes da fauna fluvial. Essas pesquisas se manifestam com maior interesse nas espécies de *piracema*, que sobem os rios em períodos determinados para realizarem a desova, mas são impedidos por barragens artificiais construídas entre as áreas de alimentação e desova. A importância dos peixes de *piracema* vem sendo confirmada pelo poder público por meio da criação de leis para tornar obrigatória a instalação de mecanismos de transposição de peixes nas barragens, proporcionando caminhos alternativos para que os peixes consigam vencer as barragens artificiais.

A idéia principal desta dissertação é aplicar técnicas de visão computacional à contagem de peixes que nadam em mecanismos de transposição construídos em barragens hidrelétricas. Atualmente, a estimativa do número de peixes que transpõem as barragens hidrelétricas durante o período da desova é feita de forma manual. Um sistema baseado em visão computacional capaz de contar as ocorrências de peixes no mecanismo

de transposição, permite um melhor aproveitamento da mão-de-obra especializada que efetua a contagem, além de uma contagem com maior precisão. Assim sendo, podemos definir nosso problema da seguinte forma:

Dada uma seqüência temporal de imagens adquiridas do interior de um mecanismo de transposição por meio de uma câmera estacionária, contar o número de peixes que passam diante do campo visual da câmera, considerando as direções de fluxo a jusante (parte baixa do rio) e a montante (parte alta do rio).

A contagem automática de peixes baseada em visão computacional é um problema composto de partes menores, cada uma com um tratamento específico. Nesse capítulo vamos definir o problema e suas sub-divisões evidenciando os objetivos e a abordagem que será adotada.

1.1 Motivação

Ao longo dos anos a construção de barragens nos rios tem dificultado a reprodução de diversas espécies de peixes. Os peixes de piracema viajam longas distâncias para chegar às cabeceiras dos rios onde realizam a desova, mas ficam impedidos de subir a correnteza e se reproduzem quando encontram esses obstáculos. Uma solução frequentemente adotada é a construção de caminhos alternativos que possibilitem aos peixes transpor as barragens. A esses mecanismos dá-se o nome de *mecanismo de transposição de peixes*.

Existem diferentes tipos de mecanismos de transposição, como elevadores e eclusas, mas os mais comuns são as escadas de peixes. Em [Martins, 2000] o autor mostra alternativas de construção juntamente com seu funcionamento. Alguns exemplos de escada de peixes estão ilustrados na Figura 1.1. Como pode ser observado na parte (a) da mesma figura, a escada fica posicionada de forma estratégica sobre a barragem; soleiras ou degraus são utilizados, como pode ser visto na parte (b).

O primeiro mecanismo de transposição de peixes construído em barragem artificial data do ano de 1828 na Escócia [Martins, 2000], com projeto e execução por parte da iniciativa privada.

Os mecanismos de transposição de peixes estão concentrados em países desenvolvidos, sendo a Noruega o país com maior número de dispositivos instalados. Em países do hemisfério sul a diversidade de peixes é maior e com características diferentes, necessitando de dispositivos específicos. Esse fato desperta em tais países a necessidade de realização de pesquisas destinadas a mecanismos de transposição de peixes.

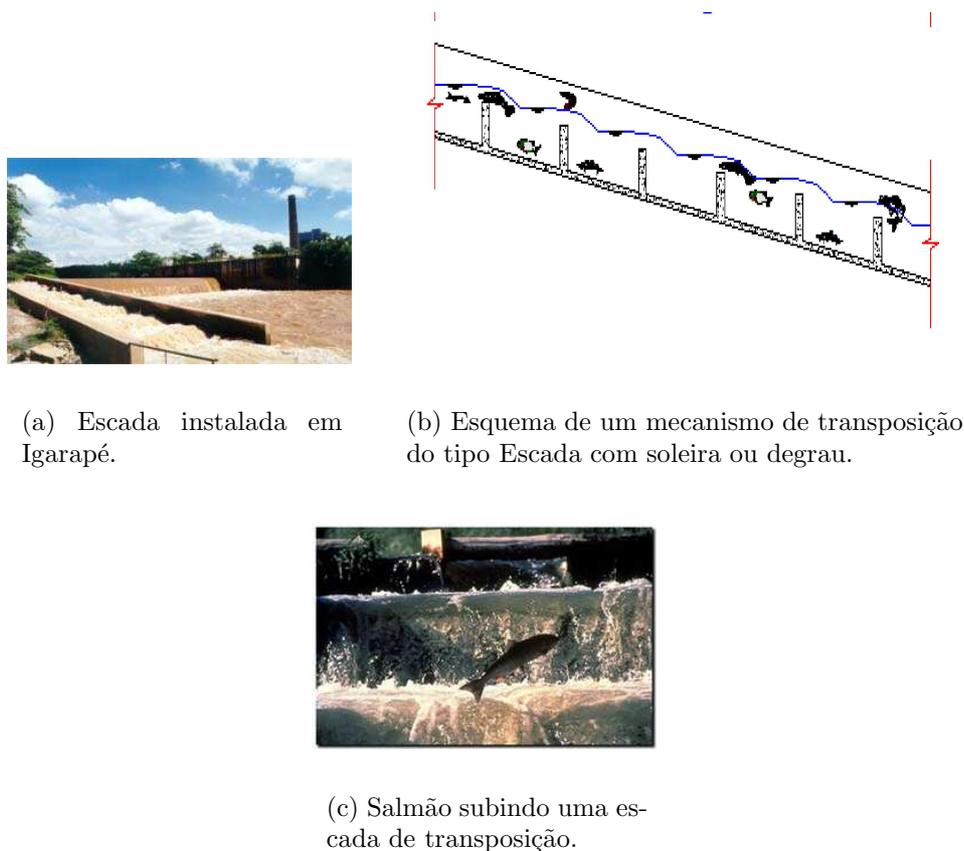


Figura 1.1: Exemplos de mecanismo de transposição de peixes. Mecanismo de Transposição de peixes do tipo Escada com soleira ou degrau.

O trabalho de contagem automática de peixes é parte integrante do projeto “Utilização Racional de Mecanismos de Transposição de Peixes”, *Transpeixe*, patrocinado pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico, CNPq-521259/2001-0, e pela Companhia Energética do Estado de Minas Gerais, CEMIG. Recentemente um grupo de pesquisadores do Centro de Estudos em Mecanismos de Transposição de Peixes da UFMG coordenou a construção de um mecanismo de transposição na Usina Hidrelétrica de Igarapava localizada no rio Grande. Em especial, esse mecanismo possui uma sala de observação construída ao seu lado com uma janela de vidro instalada na lateral do canal por onde os peixes nadam, permitindo aos pesquisadores visualizar os peixes como em um aquário (Figura 1.2).

Por meio de filmes adquiridos a partir da janela de observação instalada na usina hidrelétrica de Igarapava, os pesquisadores classificam e contam manualmente os peixes que utilizam o mecanismo de transposição da barragem.



Figura 1.2: Sala de observação do mecanismo de transposição instalado em Igarapava. Esta Sala de Observação foi construída abaixo do nível da água para possibilitar a observação dos peixes que utilizam o mecanismo de transposição.

A contagem automática dos peixes que nadam através da escada ajudará os pesquisadores na avaliação da eficácia dos mecanismos em desenvolvimento ou ainda na avaliação de mecanismos já instalados. Os dados da contagem automática em conjunto com a classificação automática de espécies, auxilia na elaboração de experimentos direcionados à verificação da influência dos mecanismos de transposição no tráfego dos peixes, fornecendo bases para elaborar protocolos de operação.

1.2 Abordagem

As tarefas de classificação e contagem dos peixes, que são realizadas manualmente pelos pesquisadores na usina de Igarapava, podem ser automatizadas com o auxílio da visão computacional segundo o diagrama da Figura 1.3, onde encontramos os processos principais de pré-processamento, contagem e classificação.

As imagens adquiridas pela câmera passam por um pré-processamento(1) que elimina efeitos espúrios das imagens causados pela água. Os peixes são detectados e rastreados nas imagens tratadas possibilitando a contagem(2). Com base nas trajetórias, são escolhidas sub-imagens contendo o peixe para a classificação(3).

Em virtude de características como a turbidez da água, refração e difração causadas pela propagação da luz através do vidro da janela de visualização instalada no mecanismo de transposição, as imagens adquiridas precisam passar por um processo

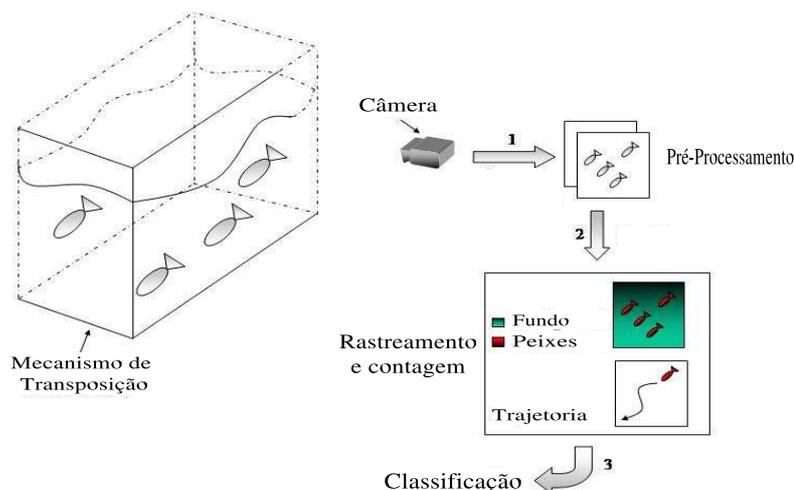
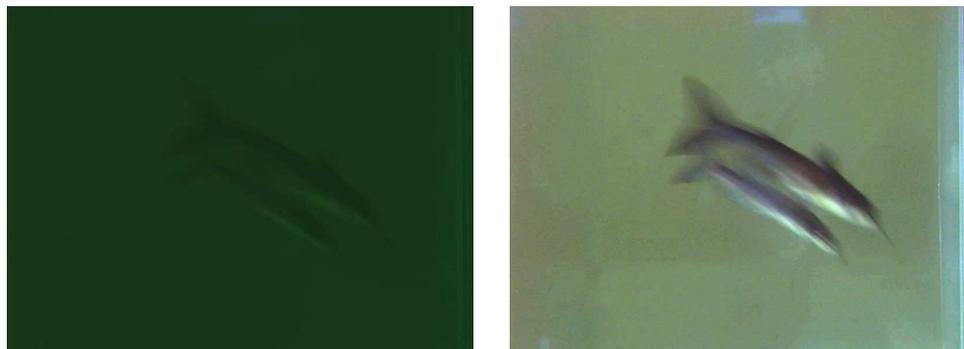


Figura 1.3: Esquemático do sistema de contagem e classificação de peixes. Uma câmera estacionária adquire imagens dos peixes que nadam diante da janela de observação do mecanismo de transposição. As imagens são pré-processadas para facilitar a detecção dos peixes que são rastreados e contados. Os dados do rastreamento alimenta outro sistema responsável pela classificação das espécies.

de tratamento que antecede a contagem. Esse processo visa melhorar a nitidez das imagens, permitindo a localização automática dos objetos de interesse. A Figura 1.4 mostra um exemplo de recuperação de imagem adquirida em um aquário que simula a janela de observação. O trabalho aqui tratado assume que as imagens de entrada do sistema foram tratadas e estão nítidas o suficiente para permitir a detecção automática dos peixes.

A cena monitorada pelo sistema é representada por um trecho do canal de passagem dos peixes. Portanto, os objetos de interesse entram e saem da cena, com tempo de permanência indeterminado. Esse comportamento resulta em detecções múltiplas de um mesmo objeto durante seu tempo de permanência diante da janela de observação. O rastreamento visual dos objetos de interesse busca relacionar as múltiplas ocorrências de um mesmo objeto ao longo do tempo e com isso eliminar as redundâncias da detecção e assim contar os objetos rastreados.

Além de eliminar as redundâncias da detecção, o rastreamento visual permite recuperar também as trajetórias percorridas pelos objetos de interesse. Usando uma câmera calibrada, podemos rastrear objetos de dimensões conhecidas, recuperando inclusive a informação de profundidade. Conhecendo a trajetória dos peixes na cena, é possível recuperar um conjunto de sub-imagens contendo um dado peixe e submeter o conjunto a outro sistema capaz de identificar a espécie e classificar a contagem [Nery, 2004].



(a) Imagem Original.

(b) Imagem recuperada.

Figura 1.4: Exemplo de recuperação de imagens usando o resultado de um algoritmo de visão estéreo [Queiroz-Neto et al., 2004]. As informações sobre profundidade obtidas da visão estéreo são utilizadas para remover os efeitos de dispersão e absorção da luz devido a distância em coluna d'água. Esse trabalho ainda se encontra em fase de desenvolvimento.

Uma etapa importante no processo de rastreamento é a localização dos objetos de interesse. Conhecendo os objetos de interesse e a cena onde serão encontrados, é possível medir a probabilidade de cada ponto da imagem ser, ou não, parte de um objeto de interesse (*segmentação probabilística*). Dessa forma, o objeto de interesse fica destacado em uma nova imagem mantendo a imagem de entrada original inalterada. A Figura 1.5 mostra uma imagem de entrada que foi segmentada medindo-se a probabilidade de cada ponto ser parte do plano de fundo ou parte de um objeto de interesse. O resultado é uma imagem final que evidencia o peixe atribuindo valores conhecidos para o plano de fundo e para o objeto de interesse.

Com base na segmentação obtida em todos os quadros da seqüência, os peixes podem ser rastreados relacionando suas múltiplas ocorrências ao longo de um trecho de filme. Isso impede que o mesmo peixe seja contado em todos os quadros em que é encontrado. A segmentação probabilística é usada como um modelo de observação, que dá ao rastreador condições de encontrar os objetos de interesse nas imagens de entrada.

As informações a respeito dos objetos contidos em uma imagem, como localização e forma, compõem o *estado* de uma imagem. Durante o rastreamento, o sistema observa cada quadro do filme separadamente e propaga no tempo uma distribuição de probabilidades a respeito do *estado* da imagem. A partir dessa distribuição de probabilidade e a dinâmica do objeto, é possível estimar a localização de um dado objeto no próximo

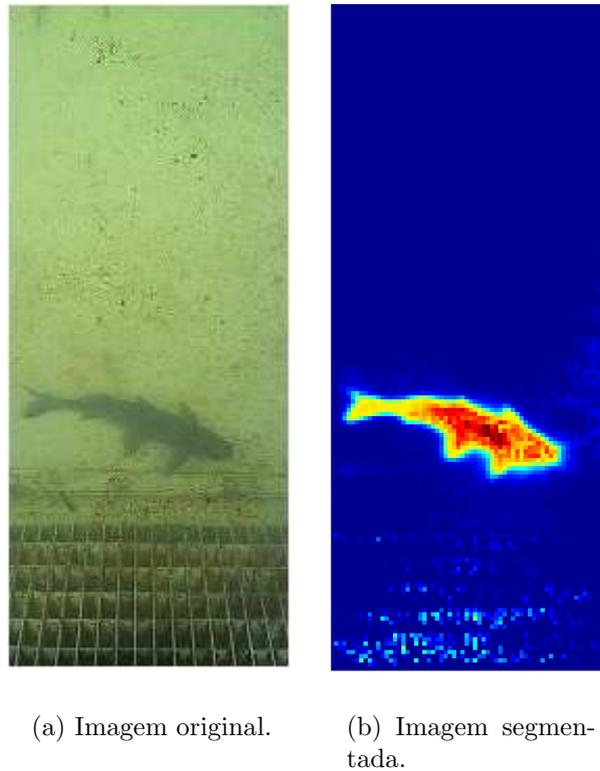


Figura 1.5: Exemplo de identificação do objeto de interesse. A Figura 1.5(a) mostra a imagem que deu origem à Figura 1.5(b). Essa transformação é feita com base em um conhecimento *a priori* da cena e do objeto de interesse.

quadro da seqüência. Comparando com a medição que será feita quando a imagem estiver disponível, têm-se a correspondência do objeto entre os quadros juntamente com a atualização da função de probabilidades para a nova predição.

A contagem propriamente dita utiliza os dados obtidos com o rastreamento para quantificar os peixes que atravessaram a cena. Os detalhes serão tratados no Capítulo 3.

Dessa forma, a abordagem adotada pode ser subdividida em cinco etapas principais:

- Aquisição de imagens;
- Pré-processamento;
- Segmentação;
- Rastreamento;
- Contagem.

1.3 Contribuições

Este trabalho estuda o uso de técnicas de visão computacional para rastrear e contar peixes que nadam através de um mecanismo de transposição. A metodologia é baseada no uso de um filtro bayesiano para realizar os peixes, cujo número de indivíduos presente na cena pode variar com o tempo.

Diferente de outros métodos de contagem automática de peixes, essa abordagem proporciona meios adequados para a aquisição de informações relevantes sobre características de diferentes espécies de peixes, tais como habilidade de nado, tempo de migração e picos de fluxo.

A metodologia usada é capaz de estimar a trajetória ao longo do tempo, permitindo o estudo de comportamentos das espécies quando nadam em regiões de interesse.

Trabalhando juntamente com um sistema de classificação de espécies, a metodologia proposta o rastreamento e contagem de diferentes espécies simultaneamente. Devido à variedade do ecossistema brasileiro, essa tarefa tem grande relevância, principalmente quando são consideradas as espécies em extinção ou aquelas de grande valor econômico.

1.4 Organização do texto

O trabalho está organizado da seguinte forma: O Capítulo 2 mostra uma revisão das principais aplicações e métodos encontrados na literatura para a solução do problema de contagem automática de objetos que se movem independentemente, especialmente peixes, e trabalhos relacionados com o rastreamento de múltiplos objetos. No Capítulo 3 descreve-se a metodologia adotada para a solução do problema com as implementações e resultados compondo o Capítulo 4. Finalmente, as conclusões são mostradas no Capítulo 5.

Capítulo 2

Trabalhos Relacionados

Este capítulo apresenta trabalhos que se relacionam à contagem automática de peixes. Em um primeiro momento é apresentado o estado da arte na contagem e em seguida métodos usados para o rastreamento visual de objetos. Os trabalhos estão organizados em uma evolução cronológica, partindo de modelos específicos como o mostrado em [Menin and Paulus, 1974] para métodos com aplicações mais gerais.

2.1 Métodos convencionais de contagem de peixes

A necessidade de estimar populações de peixes levou pesquisadores a investigar formas de efetuar a contagem automaticamente. Nesta seção, agrupamos alguns métodos encontrados na literatura em três categorias principais de acordo com o princípio de funcionamento. Dentre os trabalhos encontrados estão alguns que utilizam princípios acústicos, métodos que exigem a passagem de um peixe por vez (e por esse motivo foram classificados como de *enfileiramento*) e métodos que utilizam imagens. Cada grupo será tratado separadamente.

2.1.1 Métodos acústicos

Esses métodos utilizam basicamente um sistema de sonares para emitir um sinal e processar o eco retornado pelo ambiente. Medindo-se, o tempo de retorno do eco pode-se calcular a distância que objeto se encontra.

Em [Menin and Paulus, 1974] os autores propõem uma forma de contagem que detecta a bolsa natatória (*swim bladder*) do salmão utilizando um conjunto de sonares dispostos em linha. Cada sonar emite pulsos de $100\mu s$ a uma frequência de $100KHz$.

Os peixes que passam pela faixa de cobertura dos sensores retornam um eco característico em função de sua bolsa natatória e cada ocorrência é contada. O sistema foi testado cuidadosamente em rios de águas claras no estado do Alaska, e a confiabilidade pôde ser avaliada por observação visual. Nos rios em que o sistema foi testado, a contagem de salmão era realizada por pessoas posicionadas em pequenas torres instaladas à margem do rio. O sistema de [Menin and Paulus, 1974] foi comparado com a contagem visual, obtendo uma precisão acima de 90%, sendo posteriormente aplicado em rios com a superfície congelada no estado do Alaska. Esse trabalho não mostra uma forma de eliminar falso-positivos em função pro exemplo, da baixa velocidade de nado dos peixes que pode causar detecções múltiplas. Também não mostra como são tratadas as possíveis oclusões que ocorrem durante a migração de cardumes.

Um dispositivo móvel para quantificação de cardumes, com o objetivo de dar suporte a estudos biológicos no Canadá, é descrito em [White, 1976]. Esse sistema consiste de um dispositivo acústico instalado em uma embarcação que se movimenta em velocidade constante, registrando o eco retornado pelo fundo e conseqüentemente os peixes que estiverem na faixa de atuação do sonar. O sistema também possui um conjunto de 6 a 8 termômetros dispostos em linha, cada um a uma profundidade maior que seu antecessor. Esse sistema possibilita relacionar a localização dos peixes e a temperatura da água. Assim como na abordagem anterior, o sistema não possui artifícios para a classificação dos peixes monitorados e restringe-se a ambientes navegáveis.

Outra abordagem destinada à contagem de peixes em rios durante o período de migração utiliza o efeito Doppler para eliminar os falsos positivos durante a detecção dos alvos (peixes) [Pincock and Easton, 1978]. A metodologia inicialmente proposta por [Braithwaite, 1974] teve sua viabilidade avaliada nesse novo trabalho, onde foi estudada a natureza do sinal recebido pelo sistema para mostrar quando é possível utilizar o efeito Doppler. A dificuldade que Braithwaite teve em separar o eco e o efeito Doppler desencorajou o uso da técnica, mas o trabalho de [Pincock and Easton, 1978] atesta a utilização do sistema com frequências contínuas entre 30 e 100KHz. Quando utilizado o sistema de pulsos em alta resolução, altas frequências são necessárias, o que restringe a utilização para a detecção de alvos a poucos metros.

Algumas abordagens baseadas em sonares foram patenteadas, como por exemplo [Menin, 1984], que desenvolveu um contador de salmão baseado em sonar capaz de separar a ocorrência de *king salmon*, uma espécie de salmão maior que as demais. Esse mecanismo é composto por um conjunto de emissores e receptores de ondas sonoras na faixa de 515KHz, emitidas em pulsos com duração de 100μs. A resposta

encontrada no receptor tende a ser alongada, sendo o eco que representa o *king salmon* alongado a $300\mu s$ ou mais. A resposta das outras espécies de salmão fica abaixo de $275\mu s$. Nessa abordagem, os peixes são obrigados a saltar sobre uma estrutura linear, instalada bem próxima à superfície da água sustentando uma tela fina até o fundo do canal de migração. O sistema de sonares é instalado paralelamente à estrutura de PVC de maneira a detectar a passagem dos peixes que saltam a estrutura. O sistema possui um funcionamento simples mas requer uma instalação não trivial, como mostra a Figura 2.1. A tela que força a passagem dos peixe sobre a estrutura, também retém os corpos estranhos trazidos pela correnteza. Logo, a manutenção periódica é necessária.

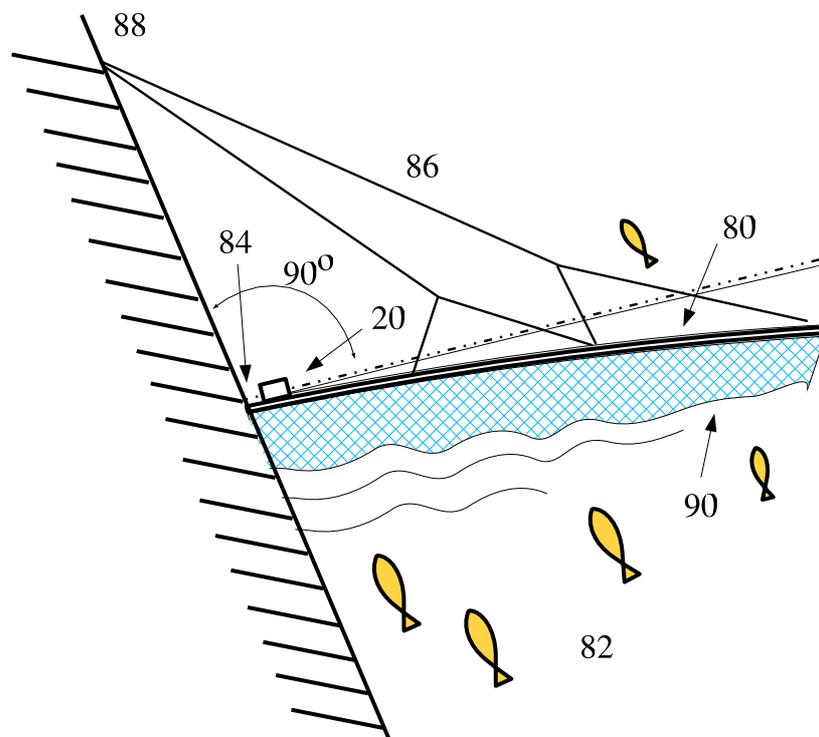


Figura 2.1: Esquemático da estrutura de PVC proposta no trabalho de [Menin, 1984]. A estrutura de PVC (80), sustentada por cabos (86), fixados a um ponto de apoio (88), possui uma tela fina (90) cuja extensão preenche a coluna d'água do rio (82) até o fundo. O conjunto de sonares (20) é instalado em uma das extremidades da estrutura de forma a detectar os peixes que saltam.

Outra abordagem patentada tem como alvo principal a contagem de peixes pequenos diminuindo erros de contagem causados por objetos ou lama presente na água durante a contagem visual [Takagi et al., 1997]. O sistema é projetado para ser implantado em passagens de peixes instaladas em rios, lagos ou tanques. Consiste de um conjunto de emissores e receptores de ultra-som direcionados para a passagem dos

peixes. O eco retornado forma uma imagem de *blobs* que será processada por um *software* especialista para a identificação de quais conjuntos de pixels representam peixes. O sistema assume uma distância mínima entre os *blobs* e portanto não é capaz de tratar oclusões, mas consegue eliminar as detecções do mesmo indivíduo em imagens adjacentes.

As abordagens baseadas em meios acústicos, geralmente são utilizadas para contar apenas indivíduos de determinada espécie. Em rios brasileiros a diversidade de espécies desperta o interesse da contagem classificada, porém com sistemas acústicos a classificação torna-se difícil senão impossível. A única informação que podemos utilizar para esse fim é a forma, recuperando o contorno dos *blobs* detectados. Outras informações como cor e textura são perdidas. O tratamento de oclusões parciais também fica muito prejudicado, já que os *blobs* coalescem quando estão sobrepostos.

Uma forma de evitar a contagem de dois ou mais indivíduos como apenas uma ocorrência é fazer com que os peixes sejam “forçados” a manter uma fila. A próxima seção mostra trabalhos que se baseiam nesse conceito.

2.1.2 Métodos de enfileiramento

Os sistemas de contagem de peixes baseados em meios acústicos têm a capacidade de detectar vários possíveis objetos de interesse no mesmo ciclo de varredura, o que traz a possibilidade de erros durante a contagem causados por oclusões, por exemplo.

Tais problemas são contornados em algumas metodologias que obrigam os peixes a passarem por caminhos estreitos onde o espaço físico disponível não permite que muitos peixes passem ao mesmo tempo. Por esse motivo, denominamos esses métodos como métodos de enfileiramento.

Um sistema baseado em infra-vermelho foi desenvolvido para contar peixes em passagens artificiais [Cadieux et al., 2000]. O sistema possui um conjunto de emissores e receptores, ambos dispostos em linha vertical, formando um plano perpendicular à direção de nado dos peixes. Esses sensores são instalados nas laterais de uma passagem única obrigando os peixes a atravessar o plano de luz. Ao passar pelo sensor o peixe obstrui feixe infra-vermelho tendo a silhueta escaneada.

Depois da formação da imagem, o sistema faz a extração das bordas e identifica a espécie utilizando um conjunto de classificadores, fundindo os resultados para se obter a identificação final, e então fazer a contagem. O sistema possui vantagens como baixo consumo e facilidade de instalação, mas não leva em consideração possíveis erros de detecção que podem ocorrer. Quando os peixes nadarem muito próximos, podem ser

escaneados como uma única ocorrência, pois o sistema assume uma distância mínima entre os peixes que são obrigados a passar pelo sensor um por vez.

Há um sistema comercial para contagem de organismos aquáticos que utiliza túneis de contagem submersos [Smith-Root, 2005]. A Figura 2.2 mostra um diagrama desse sistema. Esses túneis percebem a passagem de um peixe, ou outro animal que passe em seu interior, através de um conjunto de sensores capazes de perceber flutuações na condutividade da água durante a passagem. Cada túnel possui três sensores o que possibilita determinar a direção de passagem do peixe. Para a contagem de grandes quantidades simultaneamente, o fornecedor sugere a utilização de múltiplos túneis que são montados em uma placa para múltiplos túneis. Esse tipo de sistema não possui uma metodologia de identificação de espécies para a contagem classificada e nem mesmo considera a possibilidade de contar animais com dimensões diferentes durante o mesmo período de operação, considerando que os indivíduos possuem dimensões similares.

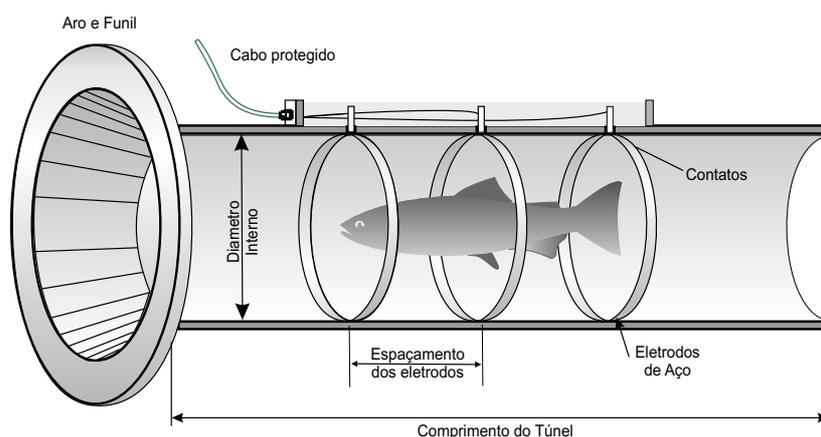


Figura 2.2: Esquemático do túnel para contagem de organismos aquáticos. Esse sistema possui três sensores que detectam flutuações na condutividade da água quando um peixe, por exemplo, passa em seu interior. A ordem de detecção nos sensores determina a direção de passagem do organismo [Smith-Root, 2005].

As duas metodologias expostas nesta seção tentam estruturar o ambiente para evitar alguns problemas possíveis como a contagem única de múltiplos indivíduos causada pela passagem de vários peixes pelo sensor. Entretanto, o trabalho mostrado em [Cadieux et al., 2000] dispõe os sensores em linha vertical, possibilitando escanear mais de um peixe na mesma imagem. Considerando que espécies de peixe como o salmão migram em cardumes, o sistema apresentaria contagem incorreta quando os peixes estivessem muito próximos entre si durante a passagem pelo sensor, sendo escaneados como se fossem um único peixe. A segunda abordagem [Smith-Root, 2005]

restringe ainda mais a passagem dos peixes, diminuindo a possibilidade da entrada de mais de um indivíduo no sensor, mas não possui nenhum mecanismo para classificação de espécies.

A técnica proposta em [Cadieux et al., 2000] realiza a contagem classificada de peixes utilizando para classificar, apenas informação de contorno onde estão implícitas a forma e o perímetro. Algumas vezes essas informações se mostram insuficientes para uma boa classificação, já que espécies diferentes podem apresentar semelhanças.

A próxima seção apresenta métodos que preservam informações visuais como cor e textura, que têm grande importância na classificação automática de espécies.

2.1.3 Análise de imagens

Informações visuais como cor e textura não são adquiridas em métodos baseados em meios acústicos ou enfileiramento apresentados nas seções anteriores. De acordo com estudos de biólogos especialistas, a coloração dorsal e ventral dos peixes constituem características importantes para a discriminação de espécies de peixes. Nesta seção serão apresentadas abordagens que utilizam câmeras para adquirir imagens da cena observada e obter estas informações.

O trabalho mostrado em [Terazono et al., 2000] utiliza câmeras digitais lineares (*line scene*) para a contagem de peixes em mecanismos de transposição ou passagens artificiais de peixes. A idéia é posicionar paralelamente duas câmeras lineares de forma que a linha de captura esteja perpendicular à direção de passagem dos peixes e paralelas entre si, como mostrado na Figura 2.3. De forma similar ao trabalho de Cadieux [Cadieux et al., 2000], as imagens de cada câmera são integradas e processadas para identificar o tamanho dos peixes, medindo quantos quadros existem entre a primeira e a última detecção do peixe. As duas câmeras fazem as amostragens ao mesmo tempo e em posições diferentes possibilitando determinar a direção de passagem, velocidade e tamanho dos peixes. Essa abordagem supõe uma distância mínima entre os peixes para que consiga separar os indivíduos.

As câmeras CCD comuns possuem um conjunto de elementos foto-sensíveis dispostos em um plano onde a imagem da cena observada será projetada. Apesar desse tipo de câmera adquirir imagens bidimensionais, o trabalho de [Terazono et al., 2000] utiliza duas câmeras unidimensionais instaladas paralelamente a uma distância conhecida entre elas. Considerando que após o processo de calibração de uma câmera conhecemos as dimensões de cada CCD, apenas uma câmera poderia ter sido utilizada, sem a preocupação de sincronismo entre câmeras distintas.

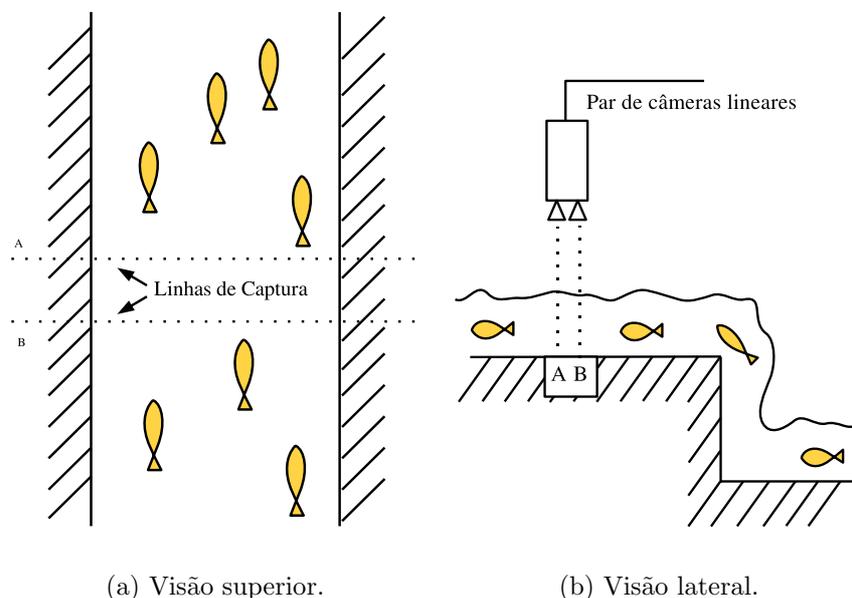


Figura 2.3: Contagem de peixes usando imagens lineares paralelas. As imagens são adquiridas simultaneamente permitindo identificar a velocidade de nado. O número de imagens adquiridas desde a primeira detecção do peixe até a última, determina o comprimento do peixe [Terazono et al., 2000].

A abordagem desta dissertação, utiliza apenas uma câmera equipada com um conjunto de CCD's agrupados em um plano, instalada diante de uma janela de observação vertical. Esta janela possibilita observar abaixo do nível da água, os peixes que nadam dentro do canal de passagem e portanto, evitando as variações causadas pelo movimento da superfície da água. No caso de [Terazono et al., 2000], as câmeras são posicionadas acima do nível da água apontadas verticalmente para baixo, ficando sujeitas a essas variações.

A contagem de peixes em imagens digitais pode ser considerada como uma instância do problema mais geral de contagem de múltiplos objetos que se movem independentemente, como pessoas.

A contagem de pessoas em tempo real para ambientes internos ou externos, utilizando uma câmera CCD é implementada em [Pádua, 2002]. Essa abordagem não necessita de infraestrutura especial para a instalação, e portanto não interfere nas trajetórias realizadas pelas pessoas. As pessoas são detectadas utilizando técnicas de estimação do fluxo óptico e as dimensões corpóreas aproximadas. A mesma idéia usada em [Terazono et al., 2000] para estimar a direção do movimento é utilizada, mas nesse caso adotaram-se duas faixas contíguas de largura suficiente para delimitar uma pessoa

observada de cima para baixo. A direção do movimento é dada pela ordem das faixas em que as pessoas são detectadas. No caso dos peixes, o tamanho dos indivíduos não é conhecido, a forma observada se altera durante a seqüência de imagens, devido aos movimentos de nado, e o movimento das pessoas é bem mais comportado que o dos peixes.

O método de [Rossi and Bopzzoli, 1994] mostra a contagem de pessoas utilizando toda a área das imagens adquiridas. Uma câmera é instalada com a direção de visualização apontando verticalmente para o piso, evitando oclusões. As imagens são divididas em três partes: Alerta (*Alerting*), determina quando um objeto de interesse entrou na cena; Rastreamento (*tracking*), rastreia o movimento do objeto até a linha de contagem; Interpretação (*interpretation*), identifica quantas pessoas correspondem aos objetos rastreados. A metodologia foi testada com seqüências de imagens reais de pessoas andando por um corredor. Como era de se esperar, o desempenho do sistema diminui quando o número de pessoas na cena aumenta. Mesmo assim, a precisão foi por volta de 90% com muitas pessoas transitando ao mesmo tempo. Mais uma vez, as pessoas estão caminhando em um mesmo plano e estão sendo observadas por uma câmera instalada acima delas e direcionada para o piso. Essa configuração evita a ocorrência de possíveis oclusões que o método não consegue tratar. No caso dos peixes, os indivíduos não estão presos a um plano único e se movem independentemente, o que exclui a possibilidade do uso desse método para rastrear e contar peixes.

As pessoas que entram no campo visual da câmera serão detectadas em diferentes *frames* e em diferentes posições ao longo da seqüência de imagens. Essa multiplicidade temporal de ocorrências do mesmo objeto de interesse precisa ser resolvida relacionando ocorrências ao mesmo objeto, ou ao rótulo que identifica um objeto. Com o rastreamento visual podemos determinar a correspondência entre essas várias ocorrências e impedir que o mesmo objeto de interesse seja contado múltiplas vezes.

Na próxima seção, são mostradas algumas formas de rastreamento visual que podem ser utilizadas.

2.2 Rastreamento Visual

O problema de rastreamento visual consiste em processar uma seqüência de imagens para descrever o movimento de um ou mais objetos em uma cena [Gevarter, 1984], ou ainda, gerar inferência sobre o movimento de um objeto dada uma seqüência de imagens [Forsyth and Ponce, 2002].

A idéia fundamental é associar características, como cor ou segmentos, quadro-a-quadro na seqüência de imagens. Técnicas mais recentes utilizam filtros preditivos capazes de estimar o estado de um alvo no próximo quadro da seqüência. A estimativa permite determinar qual objeto detectado no próximo quadro, melhor corresponde ao objeto rastreado. Com a correspondência quadro-a-quadro, é possível extrair parâmetros do movimento realizado como trajetória e velocidade. Maiores detalhes a respeito de filtros preditivos podem ser encontrados em [Goldenstein, 2004].

O *Filtro de Kalman* [Kalman, 1960], [Grewal and Andrews, 1993], [Pereira, 1999] é um dos filtros preditivos utilizados para estimar a posição em que o objeto de interesse irá ocorrer no próximo quadro e com isso diminuir o custo computacional da detecção. Outra técnica desenvolvida para o rastreamento visual de objetos, denominada CONDENSATION, é tipicamente usada para o rastreamento de contornos [Nummiaro et al., 2003]. Essas técnicas são discutidas nas próximas seções.

2.2.1 Filtro de Kalman

O filtro de Kalman foi desenvolvido inicialmente descrevendo uma solução recursiva para o problema de filtragem linear de dados discretos [Kalman, 1960]. A primeira implementação completa do filtro foi parte integrante do projeto Apollo da NASA cujo problema era estimação de trajetória e controle. O objetivo do projeto era levar uma nave à lua e trazê-la de volta [Grewal and Andrews, 1993]. Teoricamente, o filtro é um estimador para o problema Gaussiano-linear-quadrático, que é o problema da estimação dos estados instantâneos de um sistema linear dinâmico perturbado por um ruído Gaussiano branco usando-se medições linearmente relacionadas aos estados e também corrompidas por ruído branco [Pereira, 1999]. Quando o sistema é linear e variáveis aleatórias Gaussianas são representações apropriadas para os estados do sistema, o Filtro de Kalman é uma ótima escolha para a estimação dos estados.

As principais aplicações do filtro de Kalman estão nos sistemas de controle modernos, navegação e rastreamento de todos os tipos de veículos [Pereira, 1999]. Em [Trucco and Verri, 1998] encontramos um exemplo ilustrativo dessa técnica mostrando o rastreamento do centróide de um carro em movimento. Nesse exemplo, podemos perceber como o nível de incerteza na estimativa diminui com a evolução temporal do rastreamento. O exemplo utiliza um marcador em forma de cruz cujo tamanho é proporcional à incerteza na localização ao longo do movimento.

O filtro de Kalman foi utilizado por [Heikkilä and Silvén, 2004] para rastrear pedestres e ciclistas no trânsito. As imagens são adquiridas por câmeras CCD instaladas

em postes e processadas para a determinação de objetos que se movem na cena. A forma e a trajetória dos objetos são armazenadas em um cartão de memória para serem analisadas posteriormente.

Depois de alguns dias, o cartão de memória é transferido para um computador equipado com um software de análise. Nesse passo, os objetos são classificados entre pessoas e ciclistas com base nas características armazenadas durante a fase de rastreamento, desprezando aqueles com trajetória muito pequenas. A Figura 2.4 mostra a arquitetura básica do sistema.

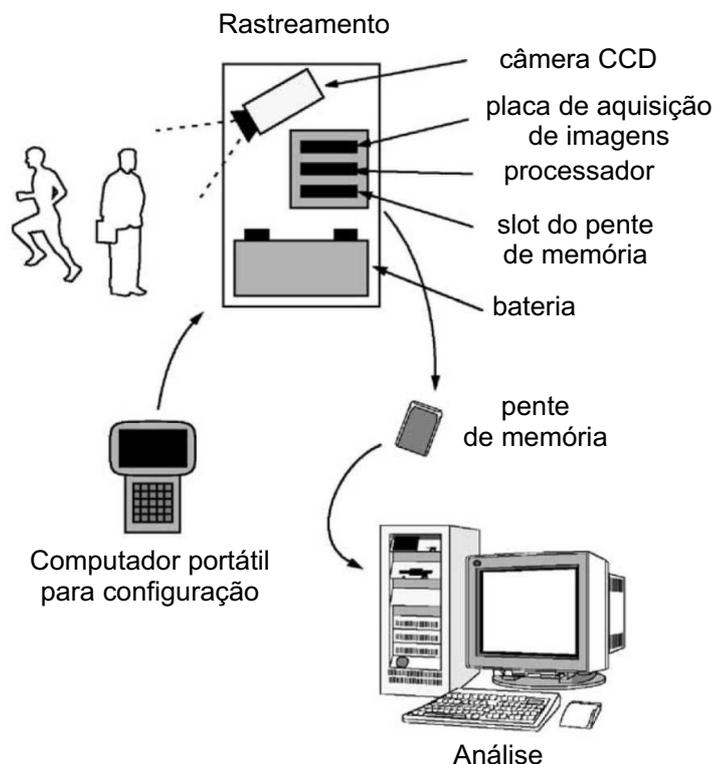


Figura 2.4: Arquitetura do sistema de rastreamento de pessoas e ciclistas [Heikkilä and Silvén, 2004]. Os objetos de interesse são detectados e rastreados armazenando a silhueta em um cartão de memória para serem analisadas em uma etapa futura.

A unidade de rastreamento detecta movimento por meio de subtração dos quadros por um *fundo* adaptativo que é comparada a um limiar [Heikkilä and Silvén, 2004]. Depois de passar por um filtro, o centróide dos *blobs* de interesse são utilizados para o rastreamento. Toda vez que um novo objeto é detectado na cena uma rotina de rastreamento é iniciada. Essa rotina compara a ocorrência desse suposto objeto com o próximo frame. Se a estrutura de características for semelhante, um rastreador

baseado em filtro de Kalman é iniciado para essa ocorrência. Um objeto é tido como ausente depois de não ser detectado durante um certo período de tempo, sendo que sua trajetória, um vetor de características e a silhueta rastreada, são armazenados no cartão de memória para análise futura.

Esse sistema foi testado por um período de 2 a 3 dias em função de limitação de bateria e conseguiu taxas de acerto entre 80 e 90% na contagem e classificação. Foram utilizados ambientes com pouca densidade de pessoas e ciclistas. Em seu experimento de maior duração (3 horas) ocorreram 107 pessoas, o equivalente a 0.5944 pessoas por minuto. A utilização de várias instâncias do filtro de Kalman aumenta a carga do processador, o que pode inviabilizar a utilização desse sistema de baixo custo em ambientes com alta densidade demográfica.

Assistentes de trânsito baseados em visão computacional precisam estimar a posição do veículo em relação às bordas da pista e reconhecer a topografia da mesma. O trabalho mostrado em [Paetzold and Franke, 2000] apresenta uma metodologia baseado em filtro de Kalman capaz de identificar as linhas estruturantes de vias de tráfego urbano como as guias laterais. O filtro é utilizado para minimizar o esforço computacional da localização de margens, beira de calçada, faixas de pedestres e linhas de parada em imagens de bordas. Com o filtro, a posição de uma estrutura pode ser estimada na próxima imagem para que seja encontrada mais facilmente. O sistema foi testado em uma máquina Power PC de 200MHz consumindo menos de 70ms para a localização global dos objetos estruturantes da cena e menos de 10ms para aproximar localmente usando o filtro.

Embora tenha sido originalmente aplicado em sistemas lineares, o filtro de Kalman pode ser estendido para aplicações de sistemas não lineares por meio de uma expansão da série de Taylor. A expansão lineariza as funções em torno da estimação corrente usando as derivadas parciais do processo e das funções de medição para calcular as estimações. O filtro resultante é chamado de *Filtro de Kalman Estendido*.

O filtro de Kalman estendido foi utilizado para rastrear objetos no espaço de Hough [Mills et al., 2003]. Usando a estimativa do filtro de Kalman o custo computacional da execução da transformada de Hough para linhas baixou de $O(n^3)$ para $O(n)$. A metodologia de [Mills et al., 2003] se mostrou robusta a ruídos e oclusões parciais, mas falha quando a aresta rastreada é ocluída e durante a oclusão surge na cena, outra aresta com orientação e localização próximas à aresta ocluída. O sistema assume que a nova aresta corresponde à aresta ocluída e passa a rastreá-la como parte do objeto original. Um inconveniente do método é a necessidade da escolha das linhas que serão

rastreadas no primeiro quadro. Desse ponto em diante, o filtro consegue trabalhar automaticamente.

O *Filtro de Kalman* é uma técnica ususal para o rastreamento baseado em características [Gordon et al., 1993]. O filtro é um estimador linear recursivo que serve de base para várias aplicações para o rastreamento visual, com a premissa de que as distribuições do ruído e o modelo de observação são Gaussianas com uma dinâmica do sistema linear. Porém, para os casos gerais de modelos não lineares, e não Gaussianos, o rastreamento torna-se analiticamente intratável e nesses casos, técnicas de aproximação podem ser utilizadas.

Os Filtros de Partículas, ou filtragem *bootstrap*, podem ser usados no problema de rastreamento. Um filtro de partículas pode, em teoria, representar outros tipos de modelos dinâmicos e de observação que não sejam Gaussianos [Bar-Shalom, 1992]. O algoritmo CONDENSATION, um filtro de partículas desenvolvido para o rastreamento visual [Isard and Blake, 1998], será exibido na próxima seção.

2.2.2 Condensation

O algoritmo CONDENSATION (*CONDitional dENSity propaGATION*) é uma abordagem probabilística para rastrear objetos em cenas ambíguas, como por exemplo objetos camuflados em que o objeto de interesse imita características do plano de fundo. Algoritmos anteriores possuem limitações na representação de distribuições de probabilidades como o Filtro de Kalman que assume distribuição Gaussiana para os eventos.

O algoritmo CONDENSATION consegue representar múltiplas hipóteses simultâneas propagando Funções de Distribuição de Probabilidades (*fdp*) não Gaussianas e potencialmente multimodais [Isard, 1998b] [Isard and Blake, 1998].

Para conseguir representações completamente gerais de funções de probabilidade, o algoritmo CONDENSATION baseia-se na técnica chamada *factored sampling* [Li et al., 2003] utilizada para localizar objetos conhecidos em imagens estáticas. Essa técnica gera um conjunto aleatório de N possíveis configurações da imagem chamadas de partículas, onde são descritos os parâmetros de objetos hipotéticos como localização, forma e orientação. Cada partícula possui uma probabilidade de representar a imagem, tornando o conjunto uma aproximação da distribuição *a posterior* $P(x|z)$, sendo x uma possível configuração e z a imagem observada [Isard and Blake, 1998]. A estratégia é utilizada de forma semelhante, sendo que para cada quadro da seqüência, ou seja, cada instante de tempo t , um novo conjunto de N partículas é criado com base em amostras do conjunto do instante anterior e levando-se em conta o movimento dos objetos

rastreados.

Essa técnica é tipicamente utilizada associada ao rastreamento de bordas [Isard and Blake, 1996], [Isard and Blake, 1998], [MacCormick and Blake, 2000], mas Nummiaro propôs usar o CONDENSATION para rastrear características baseadas em cores e identificar, nos histogramas de cores, vantagens para o rastreamento de objetos não rígidos como robustez a oclusões parciais, invariância a rotação e escala e cálculo eficiente [Nummiaro et al., 2003]. Para demonstrar o funcionamento da metodologia diante de oclusões e movimentos rápidos, usou-se uma seqüência de imagens correspondente a uma partida de futebol, onde um jogador foi rastreado em 438 quadros. Nessa seqüência, o jogador foi completamente ocluído por um instante mas o filtro conseguiu recuperar sua trajetória. Outros testes foram realizados, como por exemplo o rastreamento de faces de pessoas em uma escada rolante. O sistema funcionou comparativamente em tempo real executando em um PC $800MHz$ a uma taxa entre 10 e 30 quadros por segundo.

Originalmente, o algoritmo CONDENSATION não é capaz de detectar novos objetos que apareçam na cena, rastreando apenas aqueles presentes no passo de inicialização. Esse fato ocorre porque o passo de medição é utilizado apenas para calcular as probabilidades de cada partícula e não afeta os objetos rastreados diretamente. Então, se um novo objeto entrar na cena, ele não poderá ser rastreado. O trabalho apresentado em [Meier and Ade, 1999] alterou o passo de amostragem selecionando M partículas a menos e inserindo M novas baseadas na observação. Com as novas representações inseridas baseadas no passo de observação, o sistema adquire a capacidade de iniciar novos objetos na cena e com isso tornar o rastreamento dinâmico. O método foi testado em uma seqüência de 100 quadros de 16×64 pixels com uma taxa de 25 quadros por segundo e dois carros. O sistema teve sucesso no rastreamento dos dois objetos, sendo que o segundo surge na cena depois do passo de inicialização.

De forma simultânea e independente, uma estratégia baseada em filtro de partículas tem sido desenvolvida para aplicações não lineares e não Gaussianas. A abordagem é conhecida como *Seqüencial Monte Carlo Filters* e seu desenvolvimento independente mostra que o CONDENSATION pertence a um conjunto de variações de métodos com a mesma base teórica [Li et al., 2003].

Recentemente em [Khan et al., 2004] os autores rastream formigas em um ambiente estruturado, precisamente em uma pequena arena visualizada perpendicularmente ao plano de movimentação das formigas. Tradicionalmente os filtros de partículas possuem complexidade exponencial em função do número de objetos rastreados. Como

solução, essa abordagem substitui o passo tradicional de reamostragem do filtro de partículas pela reamostragem usada em *Markov Chain Monte Carlo* – MCMC, gerando uma seqüência de estados que se aproxima da distribuição de probabilidade de um estado, dada uma medição. O resultado é o rastreamento simultâneo de 20 formigas em 10.400 frames de 720 por 480 pixels com uma taxa de aquisição de 30 quadros por segundo. A Figura 2.5 ilustra o funcionamento do método.

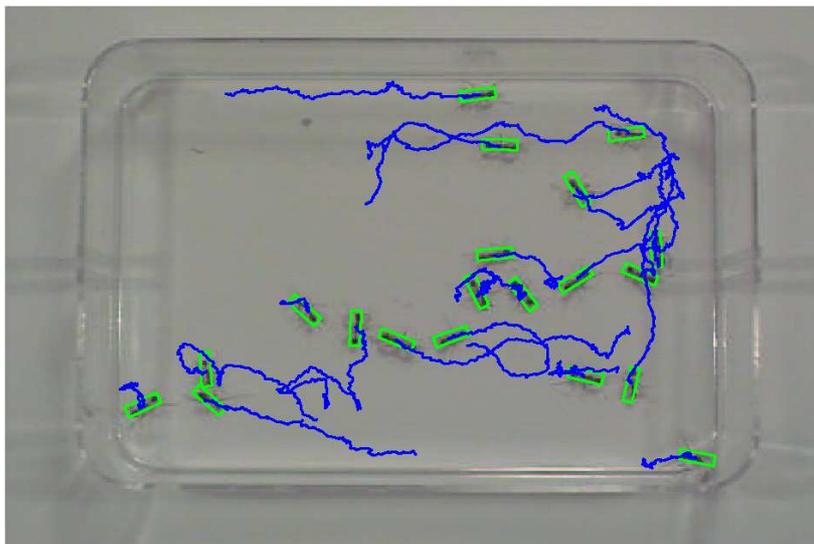


Figura 2.5: Rastreamento de formigas [Khan et al., 2004].

O objetivo do trabalho em [Khan et al., 2004] é descrever um filtro de partículas baseado em *Markov Chain Monte Carlo* para tratar alvos que se interagem. As interações entre os objetos de interesse causam problemas para os métodos tradicionais no tocante à associação dos dados. Essa abordagem utiliza um modelo de dinâmica sofisticado para manter a identidade dos objetos, diminuindo as falhas do filtro. O filtro resultante consegue tratar iterações complicadas quando os objetos se aproximam, minimizando as falhas no rastreamento de insetos quando as sobreposições não são detectadas.

2.2.3 BraMBLe – A Bayesian Multiple Blob Tracker

O rastreamento de múltiplos objetos que se movem independentemente tem despertado grande interesse da comunidade de visão computacional, especialmente aqueles métodos que apresentam robustez à deformações do plano de fundo e separação efetiva do mesmo.

O método conhecido como BraMBLe [Isard and MacCormick, 2001], utiliza o CONDENSATION [Isard and Blake, 1998] em conjunto com um modelo de observação probabilístico e robusto, capaz de separar o plano de fundo dos objetos de interesse. O método possui uma fase de treinamento na qual o modelo de observação é criado combinando funções Gaussianas, aprendidas com o uso do algoritmo de agrupamento por *k*-médias. O modelo de observação possui uma componente que mede a probabilidade de um *pixel* da imagem de entrada ter sido gerado pelo plano de fundo e outra componente para medir a probabilidade do *pixel* ser parte de um objeto de interesse. A separação do plano de fundo é feita com base nesses dois valores.

Após a fase de treinamento realizada off-line, segue a fase de rastreamento. Essa fase utiliza o CONDENSATION adaptado para detectar novos objetos na cena e também retirar os objetos que deixaram a cena. Um exemplo de adaptação para o mesmo fim é mostrado em [Meier and Ade, 1999], que no passo de reamostragem, cria o novo conjunto de partículas com M partículas a menos e insere M novas baseadas na observação do sistema. De forma similar ao trabalho de [Meier and Ade, 1999], a detecção de novos objetos consiste em modificar o conjunto de partículas do sistema, mas no caso de BraMBLe, não são inseridas partículas completas e sim, ocorre uma escolha aleatória de partículas para receberem um novo vetor de estado com parâmetros de um objeto inicial. Durante o passo de predição do filtro, objetos são escolhidos para deixar o conjunto de representações. Essa retirada aleatória de objetos torna o conjunto de partículas suficientemente dinâmico para que os objetos que deixarem a cena sejam eliminados naturalmente pela reamostragem.

O método é apresentado rastreando pessoas em um corredor, se mostrando bastante eficiente. A câmera que adquire as imagens foi instalada em uma altura superior à das cabeças das pessoas, com um ângulo de observação que possibilita a visualização dos indivíduos em todo o corredor. O sistema funcionou em tempo real em uma CPU Pentium II de 447MHz, consumindo 85% do processamento, com um máximo de sete pessoas sendo rastreadas.

O CONDENSATION original representa os objetos de interesse como um vetor de parâmetros para a forma, ou o contorno dos objetos que será propagado ao longo do tempo. O BraMBLe generaliza as formas das pessoas em troncos de cone com quatro raios característicos em função das dimensões corpóreas aproximadas das pessoas, como ilustrado na Figura 2.6.

A técnica BraMBLe é referenciada como um importante avanço no rastreamento de múltiplos objetos independentes [Okuma et al., 2004b], capaz de rastrear e iniciar au-

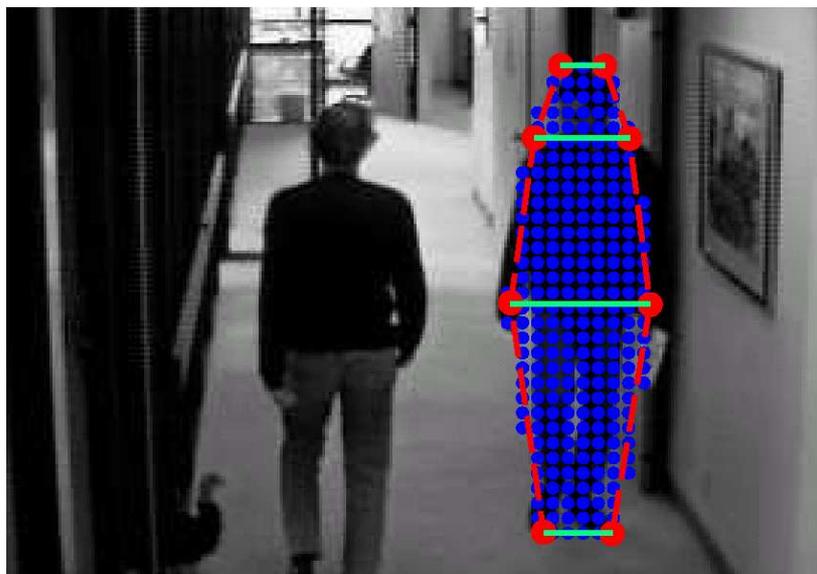


Figura 2.6: Pessoa identificada pela observação do método BraMBLe [Isard and MacCormick, 2001].

tomáticamente múltiplos objetos [Khan et al., 2004] [Branson et al., 2003]. Tendo em vista a combinação de características que o método possui, principalmente o rastreamento robusto provido pelo filtro de partículas CONDENSATION e a separação efetiva do plano de fundo, o método se mostra eficiente para o rastreamento de múltiplos objetos funcionando em tempo real em uma CPU modesta, exceto quando as trajetórias de duas pessoas se cruzam em frente de uma terceira. Neste caso, os rótulos das duas pessoas que passaram diante de uma terceira foram trocados, apesar de o filtro continuar rastreando três objetos. Com treinamentos separados para cada objeto, gerando um modelo de observação para cada um, a distinção dos objetos na cena pode ser mais precisa [Isard and MacCormick, 2001].

Neste capítulo foram apresentados trabalhos desenvolvidos para a contagem de peixes. Também foram mostrados alguns métodos que serão fundamentais para o desenvolvimento da metodologia. O problema de contar peixes que passam diante de uma janela de observação, remete problemas semelhantes ao problema de rastrear pessoas em um corredor mostrado em [Isard and MacCormick, 2001]. Por se tratarem de problemas com características semelhantes, nossa abordagem será baseada no algoritmo BraMBLe e terá detalhes de seu funcionamento no próximo capítulo.

Capítulo 3

Metodologia

Este capítulo mostra a metodologia proposta baseada no método BraMBLe para detectar, rastrear e posteriormente contar as ocorrências dos objetos de interesse, preservando informações visuais importantes como cor e textura, úteis para sistemas especialistas capazes de identificar a espécie dos peixes contados usando imagens, como o trabalho de [Nery, 2004].

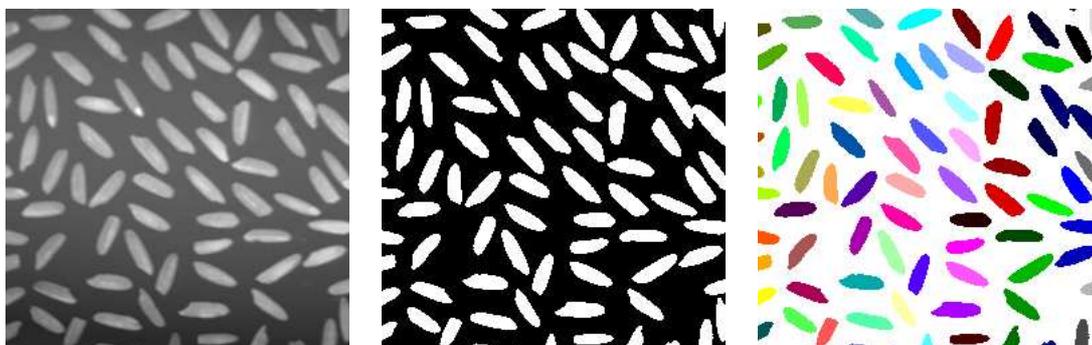
Esta metodologia utiliza um filtro de partículas para fazer o rastreamento dos objetos de interesse e um modelo de observação probabilístico para a segmentação dos objetos. Em uma fase de treinamento *a priori*, o modelo de observação encontra um conjunto de funções de probabilidades que lhe permite decidir quais pontos têm maior probabilidade de representar um objeto de interesse e quais pontos representam melhor a cena observada. O funcionamento de todos esses passos são mostrados nas próximas seções, onde será proposto um algoritmo para a contagem dos objetos rastreados.

3.1 Descrição da metodologia

O rastreamento visual de uma característica é o problema de identificar a mesma característica, quadro-a-quadro em uma longa seqüência de imagens [Trucco and Verri, 1998]. Podemos rastrear por exemplo, o centróide de um objeto, um ponto com cor especial que identifique o objeto de interesse, o contorno do objeto, uma fonte de calor, etc. Essa metodologia propõe rastrear conjuntos de pontos que correspondem a um mesmo objeto de interesse, conhecidos como *blobs*.

Cada *blob* possui informações intrínsecas como posição no sistema de coordenadas da imagem e área ocupada. Os *blobs* são detectados através de um processo de segmentação responsável por dividir a imagem em porções de interesse da aplicação.

Uma das formas de encontrar essas porções é a subtração de fundo. Na Figura 3.1 temos um processo de segmentação por subtração de fundo no qual os objetos são individualizados recebendo uma cor como rótulo ao final do processo.



(a) Imagem Original.

(b) Imagem Binária.

(c) Imagem Rotulada.

Figura 3.1: Exemplo de Segmentação e Rotulação de objetos. A subtração do fundo utilizou a operação de Abertura Morfológica (*Morphological Opening*) para estimar o plano de fundo. Essa operação retira os pequenos objetos de uma imagem preenchendo seu espaço com valores semelhantes àqueles que se encontram ao seu redor. Após uma binarização os *blobs* são rotulados com cores.

Analogamente, essa abordagem precisa detectar os *blobs* a cada quadro da seqüência associando rótulos de identificação. Para tanto, utiliza um algoritmo elaborado que separa o plano de fundo por meio de funções de probabilidades capazes de estimar quais *pixels* fazem parte do plano de fundo e quais representam os objetos de interesse. Esse conjunto de funções precisa ser aprendido em uma fase de treinamento *a priori* e constituem o Modelo de Aparência, utilizado no passo de medição do rastreador para encontrar os peixes. A Figura 3.2 mostra um esquema contendo esses passos.

Cada objeto observado possui uma representação com posição e forma permitindo estabelecer a *configuração* de uma imagem e assim saber o conjunto de objetos que a compõe e suas respectivas localizações. Os objetos são rastreados atualizando, a cada instante de tempo, a função de probabilidades $P(X_t | Z_t)$, que representa a probabilidade da configuração X_t dado que a imagem Z_t ocorreu. Essa função é representada por um conjunto de hipóteses de configuração chamadas de *partículas*, que são associadas a pesos obtidos, observando cada novo quadro do filme.

Um *Filtro de Partículas* é usado para atualizar, quadro-a-quadro, a descrição da função de probabilidades com base na dinâmica de movimento dos objetos de interesse e, por fim, estimar as posições dos objetos.

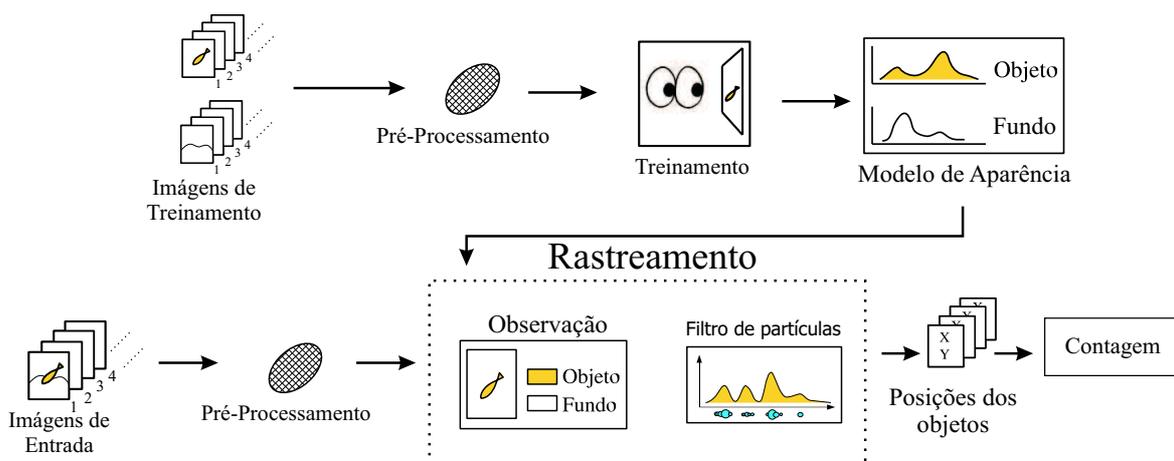


Figura 3.2: Esquemático do funcionamento da metodologia proposta. O treinamento baseia-se em dois conjuntos de imagens para determinar o modelo de aparência do plano de fundo e do plano dos objetos. Esse modelo é usado no passo de observação do rastreador.

A Figura 3.2 mostra dois fluxos principais: treinamento e rastreamento. O treinamento é executado *a priori* para gerar o modelo de aparência. O rastreamento usa o modelo de aparência para observar as imagens de entrada e detectar os objetos de interesse. Um filtro de partículas, no caso o CONDENSATION, é usado para propagar a função $P(X_t | Z_t)$ e estimar as posições dos objetos de interesse.

O passo de préprocessamento, comum tanto ao treinamento quanto ao rastreamento, tem o objetivo de eliminar ruídos da aquisição e evidenciar características consideradas importantes. As próximas seções mostram em maiores detalhes cada parte da Figura 3.2.

3.1.1 Pré-processamento

Em processamento digital de imagens, os filtros são usados principalmente para suprimir ruídos de baixa frequência, isto é, atenuar a imagem, ou evidenciar as altas frequências, isto é, detectar as bordas.

Normalmente a função de filtragem é aproximada por um *kernel* discreto e finito representado por uma pequena matriz contendo a estrutura do *kernel*. Dois exemplos de *kernel* são mostrados na Figura 3.3. Existem vários padrões de *kernel* para diferentes aplicações, nas quais a forma e o tamanho determinam características da operação.

Define-se a operação de convolução entre uma imagem e um *kernel* discreto como

0.1667	0.6667	0.1667
0.6667	-3.3333	0.6667
0.1667	0.6667	0.1667

(a) *Kernel* laplaciano.

0.0277	0.1110	0.0277
0.1110	0.4452	0.1110
0.0227	0.1110	0.0277

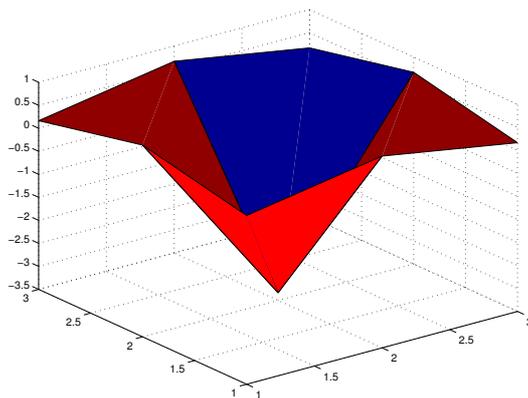
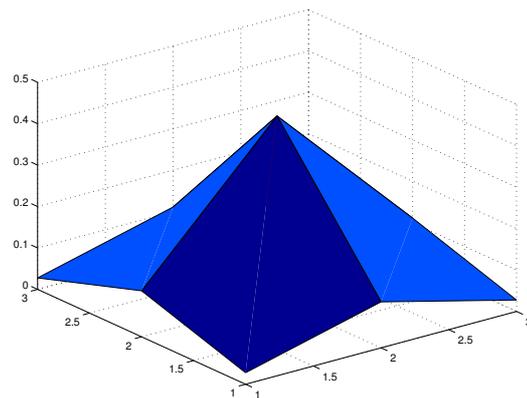
(b) *Kernel* gaussiano.(c) Visualização gráfica do *kernel* laplaciano.(d) Visualização gráfica do *kernel* gaussiano.

Figura 3.3: Exemplos de *kernel* discreto finito. As Figuras 3.3(a) e 3.3(b) mostram exemplos de *kernel* discreto finito. As Figuras 3.3(c) e 3.3(d) mostram as respectivas representações gráficas.

um conjunto de deslocamentos e somatórios de multiplicações, na qual deslocamos o *kernel* sobre a imagem e multiplicamos seus valores pelos *pixels* correspondentes da imagem. Para um *kernel* de tamanho $m \times m$, podemos calcular a convolução discreta usando a seguinte equação:

$$I_f(i, j) = \sum_{h=-\frac{m}{2}}^{\frac{m}{2}} \sum_{k=-\frac{m}{2}}^{\frac{m}{2}} A(h, k)I(i - h, j - k). \quad (3.1)$$

O resultado é a substituição do valor $I(i, j)$ pela soma ponderada dos valores de I na vizinhança (i, j) , sendo que os pesos da soma são os valores contidos no *kernel*.

Após passar por um filtro, a imagem de entrada dá origem a outra com o mesmo número de bandas da imagem original, já que cada banda é filtrada separadamente. O sistema proposto utiliza um banco de filtros e faz com que as bandas resultantes em cada filtro sejam concatenadas. Dessa forma, cada imagem de x bandas processada pelo banco de N filtros, resulta em uma imagem composta por Nx bandas.

Foram utilizadas imagens no espaço de cor YUV, desprezando a componente de luminância, e um banco de dois filtros. Esse fato faz com que as imagens de entrada sejam compostas de duas bandas e o resultado do Pré processamento em cada uma tenha quatro bandas. A Figura 3.4 ilustra o processo de Pré-processamento.

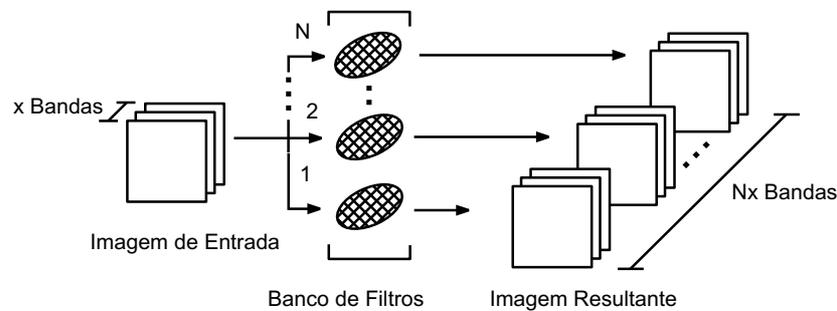


Figura 3.4: Pré-processamento das imagens. As x bandas de uma imagem são processadas separadamente por cada um dos filtros contidos no banco de N filtros. O resultado é uma imagem composta por Nx bandas

O filtro Gaussiano compõe o banco de filtros, sendo usado para atenuar detalhes e ruídos das imagens. O outro é o laplaciano da Gaussiana que calcula a segunda derivada espacial em uma imagem. Isso significa que em áreas onde a imagem possui intensidades constantes a resposta do filtro será zero, enquanto que na vizinhança de áreas com intensidades diferentes, a resposta será positiva do lado de valores maiores e negativa do lado de valores menores, evidenciando as bordas.

Tanto as imagens analisadas no treinamento quanto no rastreamento passam pelo processo de Pré-Processamento descrito. As seções seguintes mostram o treinamento e como será usado para o rastreamento.

3.1.2 Treinamento

Para rastrear os objetos de interesse o sistema precisa ser capaz de localizá-los nas imagens de entrada. Para isso utiliza-se um modelo de observação composto por um conjunto de funções que dizem a probabilidade de um ponto representar o fundo da cena ou o objeto de interesse.

Assumindo independência entre os *pixels* e os ruídos oriundos da aquisição das imagens, podemos considerar cada ponto como uma variável aleatória. Uma variável aleatória é uma função que atribui um número real a cada resposta de um experimento randômico presente no espaço amostral [Leon-Garcia, 1994]. Em nosso caso, o domínio da função, ou o espaço amostral, é o conjunto de cores que cada pixel pode assumir. Essas cores são mapeadas para um valor v que é sua representação no domínio dos números inteiros.

O sistema assume que as imagens I são adquiridas por uma câmera estática. Portanto, se apenas o $pixel(i, j)$ for observado em uma seqüência de n imagens sucessivas referentes a uma cena também estática, serão encontrados valores v_i com pequena variação entre eles. Isso significa que se o $pixels(i, j)$ estiver localizado em uma região vermelha da imagem, os n valores observados serão variações de vermelho, em conseqüência dos ruídos existentes no processo.

A variação das cores em um *pixel* ao longo de uma coleção de n imagens, pode ser explicada por uma *fdp* Gaussiana, parametrizada pela média dos valores observados e por sua variância, de forma que cada possível valor v terá uma probabilidade de ocorrer no $pixel(i, j)$. Conhecendo essa probabilidade, podemos inferir variações na região da cena representada por aquele *pixel*. Por exemplo, um valor que resulte em uma probabilidade muito distante do valor esperado, significa grande variação da cena naquele ponto, podendo representar a passagem de algum objeto naquela região.

O objetivo do treinamento é identificar uma *fdp* capaz de explicar as ocorrências dos valores v no $pixel(i, j)$ para uma seqüência de imagens.

Uma escolha natural para a estimativa dessas funções é a função Gaussiana parametrizada pela média e variância dos valores v do $pixel(i, j)$. Apesar do cálculo simplificado, essa função não é apropriada para situações que as distribuições são multimodais, mas podem servir para compor distribuições mais complicadas.

Organizando-se como um vetor de dados, os valores v encontrados em um *pixel* durante uma seqüência de imagens, podemos estimar a *fdp* que dá a probabilidade de um valor v ocorrer naquele *pixel*. Para representar funções mais complicadas que a função Gaussiana, os dados do vetor são agrupados em k grupos, cujas *fdp*'s Gaussianas são combinadas em uma *fdp* mais geral, capaz de explicar as ocorrências dos valores v no *pixel* observado. A formação de agrupamentos pressupõe que os dados são gerados a partir de uma *pdf* P , gerada por uma mistura que possua k componentes, cada uma sendo por si só uma *pdf* [Russell and Norvig, 2004]. Seja a variável aleatória $C \in \{1, \dots, k\}$ que denota o componente. A *pdf* gerada por uma mistura de *pdf*'s é dada por:

$$P(v) = \sum_{i=1}^k P(C = i) P(v | C = i), \quad (3.2)$$

onde v se refere aos valores de dados do vetor, $P(C = i)$ é o coeficiente de mistura, isto é, o peso de cada componente na geração de cada probabilidade e $P(v | C = i)$ é a *pdf* da componente i . A Figura 3.5 mostra que uma densidade multimodal pode ser modelada combinando apenas três densidades Gaussianas. A Figura 3.6 mostra um exemplo prático da representação de *pdf* por combinação de Gaussianas. Nesse caso, a *pdf* resultante mostra uma distribuição que considera todos os *pixels* de uma imagem em escala de cinza.

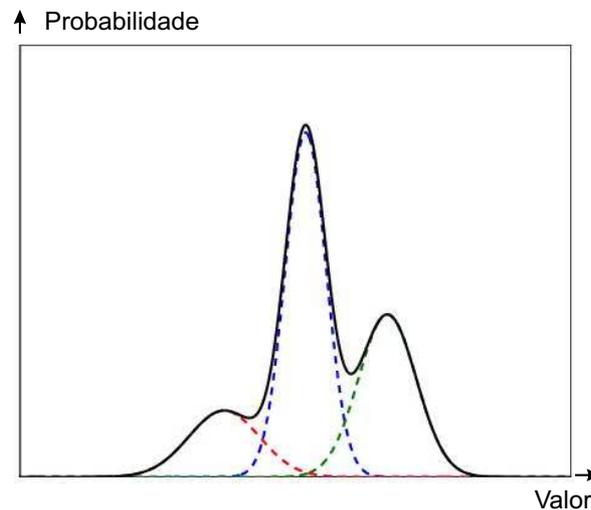


Figura 3.5: Aspecto de uma *fdp* modelada por combinação de Gaussianas.

O algoritmo de k -médias foi utilizado para encontrar os agrupamentos de valores em cada vetor de dados. Esse algoritmo agrupa os valores de um vetor de dados em k grupos, minimizando a distância de cada elemento do vetor ao centro do grupo

que está inserido. O k -médias é um algoritmo iterativo que recalcula os centros dos agrupamentos a cada passo, convergindo o resultado para um bom particionamento. O único parâmetro requerido além do vetor de dados é o número de agrupamentos que devem ser criados.

Sabendo em qual grupo cada valor v foi alocado, é possível calcular os parâmetros da função Gaussiana de cada agrupamento. O coeficiente de mistura $P(C = i), i \in \{1 \dots k\}$ pode ser definido como a frequência relativa do agrupamento i no vetor de dados, caracterizando a influência das componentes no vetor de dados.

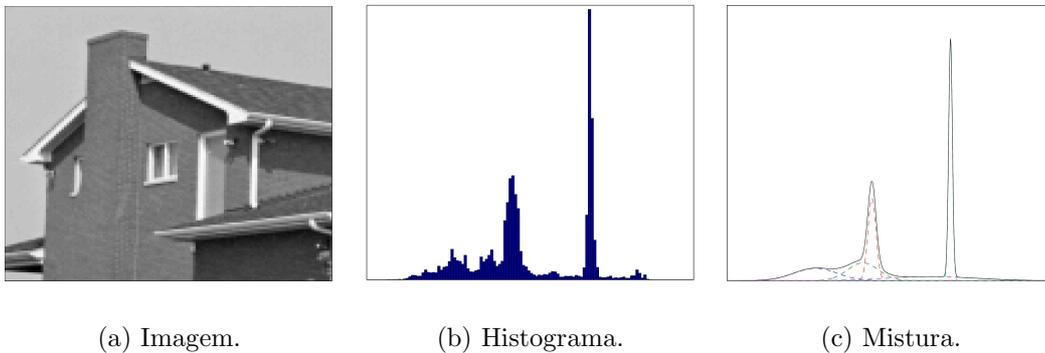


Figura 3.6: fdp de uma imagem representada por combinação de Gaussianas. A Figura 3.6(b) mostra a fdp da imagem na Figura 3.6(a). A Figura 3.6(c) mostra a representação da fdp por combinação de Gaussianas.

Da Seção 3.1.1 percebe-se que são utilizadas imagens compostas por mais de uma banda de cor no processamento que sucede o Pré-processamento. Com isso, o valor v de cada $pixel(i, j)$ em um conjunto de imagens sucessivas passa a ser considerado como um vetor $V = [v_1, \dots, v_b]^T$ de b componentes, sendo b o número de bandas das imagens de entrada, cada uma constituindo uma variável aleatória.

Para v unidimensional, a fdp de cada componente da mistura de Gaussianas é dada pela forma simples da função Gaussiana

$$P(v | C = i) = \frac{1}{\sqrt{2\pi}\sigma_i} e^{-(v-m_i)^2/2\sigma_i^2},$$

com v sendo o valor do $pixel$, m_i a média dos valores do componente i e σ_i^2 a variância. Em nosso caso, cada fdp é dada pela densidade de probabilidades Gaussiana conjunta

[Leon-Garcia, 1994]

$$P_V(V | C = i) = P_{X_{1\dots b}}(v_{1\dots b}) = \frac{1}{(2\pi)^{\frac{b}{2}} |K_i|^{\frac{1}{2}}} e^{-\frac{1}{2}(V-M_i)^T K_i^{-1}(V-M_i)}, \quad (3.3)$$

onde $V = [v_1, \dots, v_b]^T$ é o vetor de dados, $M_i = [m_1, \dots, m_b]^T$ o vetor das médias de cada variável aleatória e K_i a matriz de covariâncias definida como:

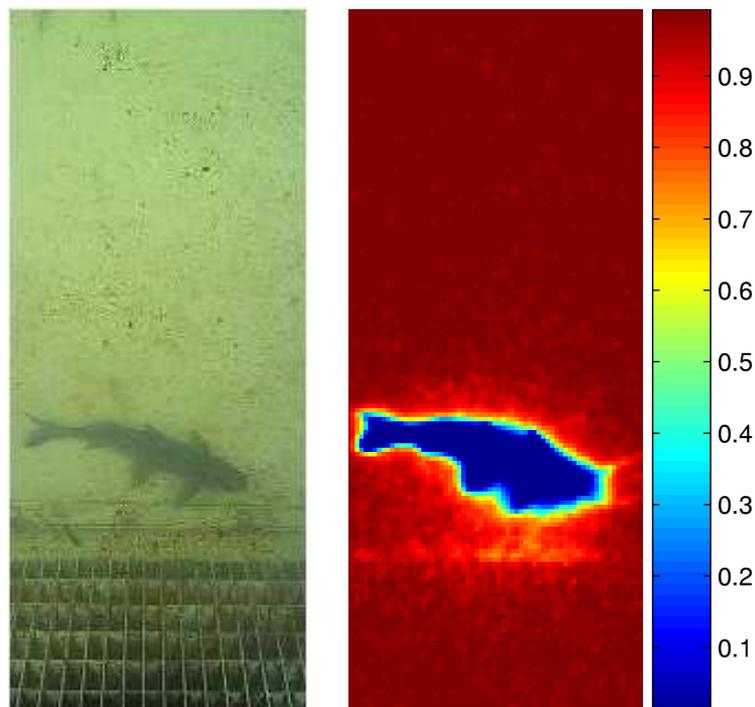
$$V = \begin{bmatrix} VAR(X_1) & COV(X_1, X_2) & \dots & COV(X_1, X_b) \\ COV(X_2, X_1) & VAR(X_2) & \dots & COV(X_2, X_b) \\ \vdots & & \ddots & \\ COV(X_b, X_1) & COV(X_b, X_2) & \dots & VAR(X_b) \end{bmatrix}.$$

Após a execução do algoritmo de agrupamento k-médias para os valores V do *pixel*(i, j) ao longo da coleção de imagens de treinamento, temos a representação de cada componente da densidade de probabilidades sub-entendida por seus parâmetros: média, covariâncias e peso na composição da mistura. Esses parâmetros são o resultado do treinamento e permitem o cálculo da probabilidade do valor V ocorrer no *pixel*(i, j) em imagens que não estão na coleção de treinamento, mas que sejam da mesma cena e mantém as mesmas características de iluminação. A *fdp* do valor V ocorrer no *pixel*(i, j) é uma expansão da Equação 3.2 como segue, onde V é o valor do *pixel* na nova imagem:

$$P(V) = \sum_{i=1}^k P(C = i) P_V(V | C = i), \quad (3.4)$$

Estendendo esse conceito para toda a área da imagem teremos uma estimacão de *fdp* para cada *pixel*. Conseqüentemente obtemos um mapa de probabilidades nas mesmas dimensões da imagem que diz a probabilidade de um *pixel* representar o espaço de treinamento. Supondo que todas as imagens do espaço de treinamento representam o plano de fundo da cena, podemos dizer que cada ponto do mapa terá a probabilidade do *pixel* correspondente representar o plano de fundo. A Figura 3.7 mostra um exemplo desse mapa treinado apenas com imagens do plano de fundo. Os *pixels* de baixa probabilidade mostram a presença de algo na cena que não representa o plano de fundo.

Depois de conhecidas as *fdp*'s relativas ao plano de fundo, uma região com probabilidade baixa indica a existência de algum objeto estranho à cena. É preciso medir a probabilidade de cada *pixel* deste corpo estranho representar um objeto de interesse



(a) Imagem observada.

(b) Mapa de probabilidades.

Figura 3.7: Exemplo de mapa de probabilidades. A figura mostra uma imagem observada(3.7(a)) e um mapa de probabilidades (3.7(b)) obtido computando as probabilidades de cada *pixel* após o passo de Pré-processamento.

por meio de uma *fdp* específica. O plano de fundo ocupa toda a área da cena e pode conter características muito diferentes nas diversas regiões, daí a necessidade de dividir o plano de fundo em pequenas regiões e treinar uma *fdp* em cada região. Os objetos de interesse podem ocorrer em qualquer região da imagem e mantém características como cor e textura. Isso permite representá-los com apenas uma *fdp*, treinada com *pixels* que correspondem aos objetos em diversas posições.

Em uma coleção de imagens contendo o objeto de interesse e devidamente filtradas de acordo com o banco de filtros, todos os pontos referentes a objetos de interesse devem ser concatenados em um único vetor para o treinamento da *fdp*. O número de agrupamentos para o plano de fundo não precisa ser o mesmo para o treinamento dos objetos de interesse e normalmente não é. Como existe apenas uma *fdp* para explicar as ocorrências dos objetos e esses valores estão sujeitos a variações ao longo da cena, é razoável que a *fdp* do plano dos objetos seja gerada por um conjunto maior de agrupamentos que as *fdp*'s do plano de fundo.

Usando o treinamento descrito nesta seção conseguimos detectar ocorrências de objetos estranhos ao treinamento que surgem na cena e detectar pontos com potencial de representar algum objeto de interesse. A Figura 3.7 mostra uma imagem e seu mapa de probabilidades obtido a partir de um treinamento feito apenas com imagens do fundo da cena. A combinação das probabilidades do fundo da cena e dos objetos de interesse será explorada pelo modelo de aparência para detectar os objetos de interesse nas imagens de entrada, medindo a probabilidade de um ponto ser fundo e a probabilidade do mesmo ponto representar algum objeto de interesse. O Algoritmo 1 sumariza o processo de treinamento em um $pixel(i, j)$.

```

1:  $c \leftarrow$  o número de componentes em cada  $pixel(i, j)$ ;
2:  $col \leftarrow$  o conjunto dos valores  $V = [v_1, \dots, v_c]$  de cada  $pixel(i, j)$ ;
3:  $k \leftarrow$  no agrupamentos;
4:  $[centros, rotulos] \leftarrow k$ -médias( $col, k$ );
5: for  $g = 1$  to  $k$  do
6:    $A_{m \times c} \leftarrow$  a matriz dos valores  $V = [v_1, \dots, v_c]$  de  $col$  onde  $rotulos = g$ ;
7:    $covars(g) \leftarrow$  Covariâncias entre as  $c$  colunas de  $A$ ;
8:    $priors \leftarrow$  freqüência relativa de  $g$  em  $rotulos$ ;
9: end for

```

Algoritmo 1: Algoritmo de treinamento em um $pixel(i, j)$.

3.1.3 Modelo de Aparência

O passo de treinamento baseia-se em uma coleção de imagens contendo apenas o plano de fundo para determinar um conjunto de fdp 's que caracterizam a cena. Analisando todos os $pixels$ de uma só vez teremos áreas com probabilidade baixa quando um objeto diferente entrar na cena (Figura 3.7). Um mapa de probabilidades desse tipo é suficiente apenas para verificar a probabilidade de uma área representar ou não o fundo da cena. O fato de ter baixa probabilidade de representar o plano de fundo não significa que irá representar bem o objeto de interesse.

É preciso duas densidades de probabilidades para encontrar o objeto de interesse na imagem. Na primeira, verifica-se a probabilidade de um $pixel$ representar o plano de fundo e, na segunda, a probabilidade de representar o objeto de interesse, ou o plano dos objetos. O modelo de aparência une as duas probabilidades em cada $pixel$ dando origem a um valor de contribuição do plano dos objetos em relação ao plano de fundo. Isso permite calcular a *Função de Probabilidade multi-blob* $P(Z | X)$, que representa a probabilidade de uma imagem Z dada a hipótese de configuração X .

Uma hipótese de configuração X é constituída por um conjunto de possíveis objetos. Cada objeto x^i é representado por um par de vetores $x^i = (P^i, F^i)$ que representam posição e forma respectivamente. Em nosso caso particular, os peixes são representados por elipses parametrizadas por semi-eixo maior, semi-eixo menor e ângulo medido no sentido anti-horário a partir do eixo das abscissas. Uma hipótese de configuração dada por

$$X = (m, x^1, x^2, \dots, x^m), \quad (3.5)$$

possui m representações de objetos, caracterizando múltiplos *blobs* simultâneos.

Normalmente as imagens são formadas por uma grande quantidade de *pixels*. Dessa forma, computar o treinamento das *fdp*'s em cada *pixel* torna-se um trabalho desnecessariamente extenso. Quando um objeto entra na cena, a variação de probabilidades ocorre na área correspondente ao objeto. Isso permite treinar as *pdf*'s em *pixels* espaçados de um pequeno intervalo. Semelhante a [Sullivan et al., 1999], após o passo de pré-processamento, cada imagem Z é casada a uma grade retangular de G pontos fixos, cada um rotulado como z_g . Tanto nas imagens de treinamento quanto nas imagens de rastreamento, são considerados apenas os pontos z_g tal que $g = \{1, \dots, G\}$. Essa amostragem de pontos das imagens permite diminuir consideravelmente a computação das imagens de entrada sem comprometer o modelo de aparência nem o treinamento.

Os pontos z_g de uma imagem de entrada Z são condicionalmente independentes. Logo, a probabilidade da imagem Z dada uma configuração X é obtida por

$$P(Z | X) = \prod_{g=1}^G P(z_g | X).$$

Como visto anteriormente, uma configuração X é composta por um conjunto de elipses que englobam um conjunto de pontos z_g . Cada ponto possui um valor V com uma probabilidade de ocorrer. Para que um ponto não seja considerado em mais de uma elipse na mesma configuração, cada ponto z_g recebe um rótulo $r_g \in \{0, \dots, m\}$ que obedece a ordem de profundidade dos objetos de acordo com a configuração X . Dessa forma, cada ponto possui a informação de ser parte do plano de fundo ($r_g = 0$), ou estar contido em qualquer uma das m elipses de representação ($r_g = i$ para o objeto i), transformando as probabilidades parciais $P(z_g | X)$ em $P(z_g | r_g)$. Agora a *fdp*

original pode ser escrita como

$$P(Z | X) = \prod_{g=1}^G P(z_g | X) = \prod_{g=1}^G P(z_g | r_g). \quad (3.6)$$

Usando os modelos de plano de fundo e do plano dos objetos, treinados *off-line*, para calcular $p(z_g | r)$ tal que z_g corresponda ao vetor de coordenadas de cada *pixel* e r seja a escolha do modelo treinado, o produtório da Equação 3.6 pode ser simplificado para uma seqüência de somas usando probabilidades logarítmicas.

Sabendo que $\log AB = \log A + \log B$, calculamos previamente os mapas de probabilidades logarítmicas $y_g^r = \log(P(z_g | r))$ para $r \in \{0, \dots, m\}$ e usamos como tabelas de consulta para realizar as somas. O filtro de partículas utilizado no rastreamento dos objetos de interesse utiliza a resposta do modelo de aparência apenas como uma constante multiplicativa o que nos permite definir

$$y_g^r = \log(P(z_g | r \neq 0)) - \log(P(z_g | r = 0)). \quad (3.7)$$

Dessa forma, o modelo de aparência une as duas informações de probabilidade logarítmica dando como resposta em cada *pixel*, uma taxa de logaritmos comparando o plano do objeto e o plano de fundo. Podemos notar que, nas regiões em que o rótulo r seja igual a zero, temos $y_g^r = 0$, garantindo que apenas o plano do objeto contribui para a probabilidade logarítmica dada por:

$$L = \sum_{g=1}^G y_g^l. \quad (3.8)$$

Sumarizamos o cálculo da probabilidade logarítmica no Algoritmo 2. O primeiro passo que compreende as linhas de 3 a 7, assume que dois objetos estão posicionados a uma distância mínima, descartando as hipóteses em desacordo. Após iniciadas as variáveis, a probabilidade dos pontos de cada objeto é acumulada em L de acordo com a Equação 3.8. A função $G_{obj}^{\psi(i)}$ retorna os pontos de y_g^r referentes ao objeto $\psi(i)$. A rotulação dos pontos no comando da linha 19 garante que a probabilidade de um ponto não será acumulada mais de uma vez.

Como visto, o modelo de aparência combina duas probabilidades em cada *pixel* z_g referentes ao plano dos objetos ($r \neq 0$) e plano de fundo ($r = 0$). A *fdp* do plano de fundo é aprendida para cada seqüência de testes, utilizando um conjunto de imagens que contenha apenas o plano de fundo, gravadas com as mesmas condições ambientais

```

1:  $X \leftarrow$  Configuração  $(m, x^1, x^2, \dots, x^m)$ ;
2:  $\delta_d \leftarrow$  distância mínima entre blobs;
3: for all  $i \neq j$  do
4:   if  $dist(x^i, x^j) < \delta_d$  then
5:      $L \leftarrow -\infty$  e termine;
6:   end if
7: end for
8:  $L \leftarrow 0$ ;
9: for  $g = 1$  to  $G$  do
10:   $l_g \leftarrow 0$ ;
11: end for
12: Ordene os objetos de maneira que  $\psi(i), i = 1 \dots m$ 
    enumera os objetos em ordem de profundidade;
13: for  $i = 1$  to  $m$  do
14:  Compute  $G_{obj}^{\psi(i)}$ ;
15:  if  $G_{obj}^{\psi(i)} = \emptyset$  then
16:     $L \leftarrow -\infty$  e termine;
17:    for all  $g \in G_{obj}^{\psi(i)}$  tal que  $l_g = 0$  do
18:       $L \leftarrow L + \psi(i)$ ;
19:       $l_g \leftarrow \psi(i)$ ;
20:    end for
21:  end if
22: end for

```

Algoritmo 2: Calculo da probabilidade logarítmica.

da seqüência de testes. Essa exigência garante que a *fdp* conseguirá representar o plano de fundo nas mesmas condições do aprendizado.

Cada ponto z_g terá uma *fdp* de plano de fundo associada, mas o plano dos objetos será representado por apenas uma *fdp* em toda a imagem. O plano de fundo ocorre em toda a imagem e pode ser bastante complexo, com cores, objetos e texturas diferenciadas. O plano dos objetos é formado por objetos específicos que se deslocam pela imagem. O treinamento do plano dos objetos é realizado concatenando-se, em um único vetor de dados, todos os pontos referentes a objetos de interesse em todas as imagens de treinamento. Isso significa dizer que para cada imagem de entrada existem G representações de *fdp* para o plano de fundo e apenas uma para o plano dos objetos.

O modelo de aparência pode responder a sombras como sendo parte do plano dos objetos. Para evitar essas falsas respostas, um pequeno múltiplo da matriz identidade $\Delta_B = \delta_b I$ é acrescentado a cada matriz de covariâncias do modelo de plano de fundo aprendido. Quando aumentamos a variância de uma função Gaussiana, estamos au-

mentando a largura de sua base. Conseqüentemente aumentamos a probabilidade dos valores correspondentes a sombras serem parte do plano de fundo. Com isso, ocorre uma diminuição do peso final L dado pelo modelo de aparência naquele ponto, significando que o valor V do *pixel* tem maior chance de ser parte do fundo que parte de um objeto de interesse.

As amostras não observadas na fase de treinamento poderão ter probabilidades $P(z_g | r) = 0$, causando inconsistência no cálculo da probabilidade logarítmica na Equação 3.7. Para dar robustez a valores não observados durante o treinamento, adicionamos uma função uniforme de pequena amplitude à mistura de Gaussianas, dando uma probabilidade mínima para cada modelo. Isso equivale a somar uma constante τ à *fdp*. Representando a Equação 3.3 na forma parametrizada $\mathcal{N}(\mu, \Sigma)$, sendo μ o vetor de médias e Σ a matriz de covariâncias, as densidades de probabilidades dadas pela Equação 3.4 terão a seguinte forma:

$$P(z_g | r_g = 0) = \sum_{i=1}^{k_B} P(C = i) \mathcal{N}(\mu_g^i, \Sigma_g^i + \Delta_B) + \tau_B$$

$$P(z_g | r_g \neq 0) = \sum_{i=1}^{k_F} P(C = i) \mathcal{N}(\mu_g^i, \Sigma_g^i) + \tau_F$$

As próximas seções são dedicadas ao rastreamento dos objetos de interesse. Detalhamos o funcionamento do algoritmo CONDENSATION e a importância do modelo de aparência para esse algoritmo.

3.1.4 Rastreamento

O algoritmo CONDENSATION foi desenvolvido sob o desafio de rastrear curvas em ambientes densos e desordenados [Isard and Blake, 1998]. O objetivo era rastrear contornos de objetos que compõem o plano dos objetos de uma cena substancialmente confusa. Nesse tipo de cena, o plano dos objetos e plano de fundo possuem características muito semelhantes que confundem o observador. Em cenas com casos de camuflagem, os objetos de interesse imitam as características do plano de fundo para se tornar mais difícil de ser percebido.

A cada quadro o sistema precisa estimar um estado de configuração \hat{X}_t propagando no tempo a probabilidade $P(X_t | Z_{1,\dots,t})$. A estimação do espaço temporal, ou o rastreamento de uma característica ao longo do tempo, pode ser feita usando-se o filtro

de Kalman nos casos em que $P(X_t)$ é modelada como uma função Gaussiana. Esta solução não é adequada quando a propagação de X_t resulta em uma densidade de probabilidade multimodal, difícil de ser modelada [Isard and Blake, 1998].

O Algoritmo CONDENSATION baseia-se na idéia de *Factored sampling* usado para representar as distribuições de probabilidades em imagens isoladas, para propagar a probabilidade $P(X_t | Z_{1,\dots,t})$ ao longo dos quadros. Esse método utiliza um conjunto de amostras de X , chamadas de partículas, cada uma ponderada por um peso que representa sua probabilidade. Os pesos das partículas são atribuídos pelo passo de observação utilizando o modelo de aparência.

3.1.4.1 Propagação temporal

De acordo com [Isard, 1998a], a regra para a propagação temporal da densidade de probabilidade dos estados é dada por

$$P(X_t | Z_t) = k_t P(Z_t | X_t) P(X_t | Z_{t-1}), \quad (3.9)$$

onde

$$P(X_t | Z_{t-1}) = \int_{X_{t-1}} P(X_t | X_{t-1}) P(X_{t-1} | Z_{t-1}) dX_{t-1} \quad (3.10)$$

e k_t uma constante de normalização que independe de X_t . A Equação 3.9 pode ser interpretada como uma equivalente da regra de Bayes (Equação 3.11) para inferir a densidade posterior de estados a partir dos dados de entrada, ao longo do tempo [Isard, 1998a].

A propagação temporal da *fdp* de estados é um processo recursivo que leva em consideração a *fdp* do passo anterior. O cálculo da *fdp* $P(X_t | Z_{t-1})$ une à *fdp* do passo anterior a probabilidade de um estado no instante atual dado o estado anterior. Essa probabilidade se refere à dinâmica dos objetos de interesse.

O rastreamento possui um passo de observação que analisa as imagens de entrada. Nesse passo, o sistema mede a probabilidade da imagem de entrada atual dado o estado de configuração no instante atual, afim de acrescentar à propagação da *fdp* de estados um fator reativo em função da observação. A probabilidade $P(Z_t | X_t)$ correspondente ao passo de observação, é computada pelo modelo de aparência da Seção 3.1.3.

A propagação da *fdp* de estados pode ser realizada de forma mais simplificada por meio de um conjunto de amostras que represente a *fdp*. A cada passo o conjunto é criado escolhendo-se elementos do conjunto anterior e considerando sua probabilidade. Essa idéia é baseada no algoritmo *Factored Sampling*.

3.1.4.2 Factored Sampling

Encontrar um objeto parametrizado por X , com uma probabilidade *a priori* $P(X)$, dada uma imagem Z , é um problema de reconhecimento estatístico de padrões, normalmente instanciado como a busca pelo melhor valor estimado de \hat{X} . A localização do objeto é expressa pela probabilidade *a posteriori* $P(X | Z)$, que em geral é difícil de calcular por não seguir um comportamento unimodal [Isard, 1998a]. Nesse caso a regra de Bayes é aplicada:

$$P(X | Z) = \frac{P(Z | X) P(X)}{P(Z)}, \quad (3.11)$$

onde $P(X)$ é a probabilidade *a priori* de X , que representa o conhecimento do objeto e $P(Z)$ uma constante que independe de X . A probabilidade $P(Z | X)$ dada pelo modelo de observação mede a probabilidade de uma imagem Z existir com uma configuração X .

Quando o modelo de observação $P(Z | X)$ ou a descrição do objeto $P(X)$ representam funções difíceis de serem calculadas analiticamente, $P(X | Z)$ não pode ser calculada diretamente. Nesse caso precisamos de uma aproximação para essa função. O algoritmo chamado *Factored Sampling* aproxima a distribuição de probabilidades $P(X | Z)$ por um conjunto randômico e discreto de possíveis realizações de X , ponderadas pela probabilidade dada pelo modelo de observação $P(Z | X)$ da Seção 3.1.3 calculado de acordo com a Equação 3.6.

Inicialmente o conjunto de realizações, ou partículas $\{S^1, \dots, S^N\}$ é gerado a partir da função de probabilidades *a priori* $P(X)$. Em seguida, cada realização S_i recebe uma probabilidade π_i de representar a imagem dada pelo modelo de observação, como ilustra a Figura 3.8, sendo $i \in \{1, \dots, N\}$ e

$$\pi_i = \frac{P(Z | X = S^i)}{\sum_{j=1}^N P(Z | X = S^j)}.$$

Cada partícula representa uma hipótese de configuração X . O conjunto de partículas ponderadas pelo modelo de observação $\{(S^1, \pi^1), \dots, (S^N, \pi^N)\}$, aproxima-se à função de probabilidades *a posteriori* $P(X | Z)$ com uma precisão que aumenta proporcionalmente ao número de partículas N .

Com a representação aproximada da *fdp* das hipóteses de configuração, podemos estimar a configuração \hat{X} que corresponde à localização do objeto de interesse, compu-

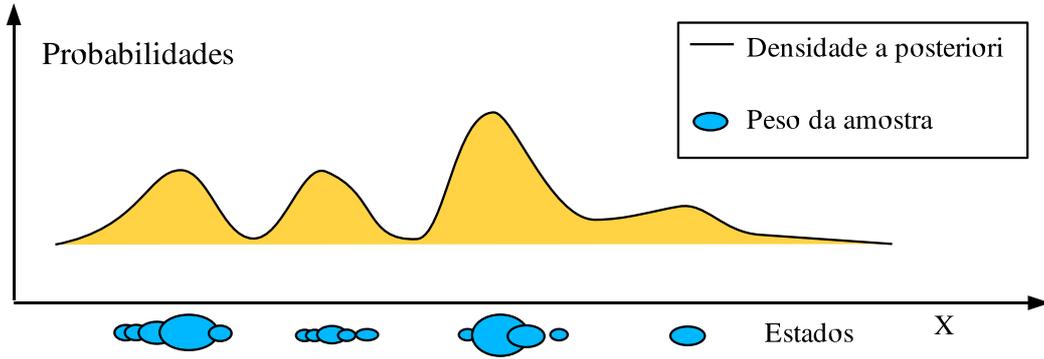


Figura 3.8: Factored Sampling. A figura mostra um conjunto de partículas amostradas randomicamente com probabilidade π_i proporcional a $P(Z | X = S^i)$. As amostras são representadas pelos centros dos *blobs* e a área corresponde a probabilidade π_i associada.

tando a média *a posteriori* das partículas do conjunto, como segue em [Isard, 1998a].

$$\mathcal{E}[X | Z] = \frac{\sum_{n=1}^N S^n P_Z(S^n)}{\sum_{n=1}^N P_Z(S^n)}$$

A forma apresentada para representar a função $P(X | Z)$ por um conjunto discreto de realizações, utiliza a probabilidade $P(Z | X = S^i)$ apenas como um peso para a partícula S^i e com isso, determinar as partículas que melhor representam a imagem observada Z . Dessa forma, o rastreamento está livre para utilizar como peso, a resposta dada pela Equação 3.8, que otimiza o cálculo de $P(Z | X = S^i)$ e mantém a ordem de relevância entre as partículas.

3.1.4.3 Algoritmo Condensation

Baseado em um conhecimento *a priori* do objeto de interesse, o algoritmo *factored sampling* usa um modelo de observação e um conjunto de hipóteses de configuração aleatórias para calcular a probabilidade *a posteriori* $P(X | Z)$. O algoritmo CONDENSATION estende a idéia para sucessivas imagens de uma seqüência criando um conjunto de partículas $\{(S_t^n, \pi_t^n), n = 1, \dots, N\}$ para a imagem do instante de tempo t , tendo como base o conjunto de partículas do instante anterior. Dessa forma a PDF condicional $P(X_t | Z_t)$ é propagada no tempo, permitindo realizar o rastreamento.

Pelas Equações 3.9 e 3.10, a propagação da *fdp* $P(X_t | Z_t)$ necessita conhecer a dinâmica de movimento dos objetos de interesse para calcular a probabilidade $P(X_t | X_{t-1})$, usada no cálculo de probabilidade *a priori* de um estado no instante atual.

Usando a idéia de *factored sampling* para representar funções multimodais, a propagação da PDF no instante $t - 1$ para o instante t é feita criando um novo conjunto de amostras derivado da representação no instante anterior, cada uma acrescida de um movimento e ponderada por uma nova observação. O novo conjunto de partículas é constituído por N sorteios de partículas do conjunto anterior, permitindo repetição. A cada sorteio escolhemos um número aleatório r , a partícula do instante $t - 1$ que possui a menor probabilidade π_{t-1}^n maior que r , constituindo o novo conjunto $\{S_t^n, n = 1, \dots, N\}$. Esse processo descarta as partículas de baixa probabilidade, já que, para cada escolha, existe um novo valor de r permitindo ocorrências múltiplas da mesma partícula.

Para facilitar a eliminação das partículas de baixo peso, elas são mantidas ordenadas por peso com um acumulador de probabilidades C associado, de forma que

$$\begin{aligned} C_t^0 &= 0, \\ C_t^n &= C_t^{n-1} + \pi_t^n, \quad (n = 1, \dots, N). \end{aligned}$$

A partícula escolhida será aquela correspondente ao menor j tal que $C_{t-1}^j \geq r$ com $r \in [0, 1]$.

O conjunto de partículas escolhidas $\{S_t^n\}$ passa por um passo de predição que acrescenta movimento às partículas baseado em uma dinâmica de movimento previamente estabelecida, como mostra a próxima seção. Juntamente a esse movimento conhecido, o sistema acrescenta um movimento aleatório que causa o espalhamento das partículas idênticas escolhidas. Finalmente o passo de observação de *factored sampling* encontra os pesos para cada partícula a partir da distribuição de probabilidades $P(Z_t | X_t)$, obtendo a representação $\{(S_t^n, \pi_t^n, C_t^n)\}$ no tempo t . Esses passos podem ser observados na Figura 3.9. Dado o conjunto de partículas representando a distribuição de probabilidades $P(X_t | Z_t)$, a estimativa do estado \hat{X}_t pode ser encontrada através da média ponderada das partículas. O Algoritmo 3 resume todos esses passos.

Considerando que a coleção de imagens possui apenas um objeto de interesse e a representação de estados da Equação 3.5, a estimativa do estado \hat{X}_t corresponde à estimativa do objeto rastreado. O método BraMBLe faz o rastreamento de múltiplos objetos simultaneamente por meio de partículas compostas pelas representações dos m objetos de interesse detectados. Nesse caso a estimativa de \hat{X}_t precisa ser substituída pelas estimativas dos respectivos objetos. O dinamismo do sistema que permite a entrada de novos objetos na cena e a saída dos objetos rastreados, desperta a necessidade de um preditor capaz de identificar a saída de objetos conhecidos e a entrada de outros

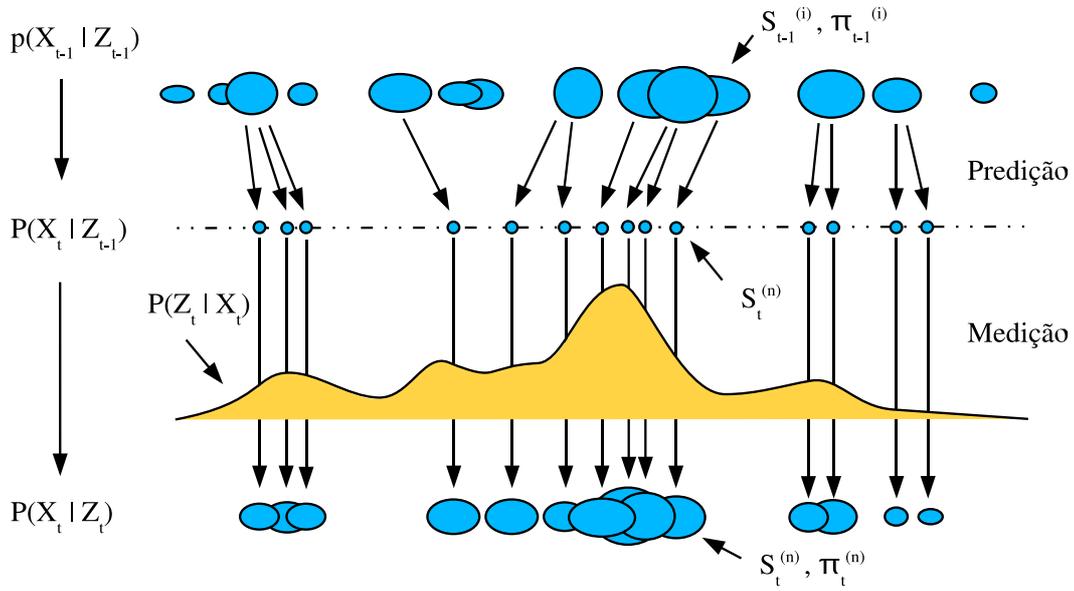


Figura 3.9: Propagação de $P(X_t | Z_t)$. Baseado na distribuição do passo anterior são escolhidas N partículas que recebem movimento de acordo com um algoritmo de predição. Em seguida o modelo de observação dá a cada partícula um peso representado na figura pela área das partículas.

novos. Os passos de predição e estimação são mostrados separadamente nas seções seguintes.

3.1.5 Predição

O passo de predição do algoritmo CONDENSATION é responsável por estimar a próxima configuração de um estado dada a configuração atual. Essa predição juntamente com o modelo de observação, permitem propagar no tempo a *fdp* de uma hipótese de configuração dada a imagem de entrada.

Um estado é representado como um conjunto X de m objetos como mostra a Seção 3.1.3. Para o preditor, cada objeto x^i precisa de duas novas componentes para guardar um identificador único ϕ^i e a velocidade V^i , passando a ser representado da seguinte forma:

$$x^i = (\phi^i, P^i, F^i, V^i)^T.$$

A função de predição f prediz novos valores de P^i , F^i e V^i em cada objeto da coleção, considerando a dinâmica dos objetos de interesse e preservando o identificador ϕ^i .

Neste caso, a dinâmica adotada considera que os peixes executam movimento uniforme nas direções x e y, corrompidos por ruído branco. Em consequência do ruído

- 1: **for all** $n \in \{1, \dots, N\}$ **do**
- 2: Escolher S_t^m em $\{(S_t^n, \pi_t^n, C_t^n)\}$ da seguinte forma:
 - a) Gere um número randômico $r \in [0, 1]$, uniformemente distribuído;
 - b) Encontre o menor j tal que $C_{t-1}^j \geq r$;
 - c) faça $S_t^m \leftarrow S_{t-1}^j$;
- 3: Predizer $S_t^n = P(X_t | X_{t-1} = S_t^m)$
- 4: Medir a probabilidade $\pi_t^n = P(Z_t | X_t = S_t^n)$ da partícula;
- 5: **end for**
- 6: Normalizar as probabilidades para que $\sum_n \pi_t^n = 1$;
- 7: construa $\{(S_t^n, \pi_t^n, C_t^n)\}$, sendo que

$$\begin{aligned} C_t^0 &= 0, \\ C_t^n &= C_t^{n-1} + \pi_t^n, \quad (n = 1, \dots, N) \end{aligned}$$

- 8: Estime o estado \hat{X}_t

$$\hat{X}_t = \mathcal{E}[X_t] = \sum_{n=1}^N \pi_t^n S_t^n$$

Algoritmo 3: Algoritmo CONDENSATION.

inserido, novas velocidades precisam ser calculadas e para isso, são consideradas variação de tempo unitárias entre os quadros da seqüência. As dimensões das elipses e o ângulo dados respectivamente por a , b e θ , apenas recebem ruído branco.

Com isso, a predição da função f pode ser separada em duas partes, uma determinística e outra estocástica. A parte determinística faz com que as partículas idênticas, escolhidas repetidamente na reamostragem, se mantenham idênticas depois da predição. Já a parte estocástica, causa a separação dos objetos idênticos tornando as partículas diferentes umas das outras. A função de predição adotada é mostrada a seguir:

$$\begin{aligned}
f(\phi, (P, F, V)^T) &= (\phi, (P', F', V')^T) \\
(\phi, (P', F', V')^T) &= (\phi, ((x', y'), (a', b'), (v'_x, v'_y))^T) \\
x' &= x + v_x + \omega_P \\
y' &= y + v_y + \omega_P \\
v'_x &= x - x' \\
v'_y &= y - y' \\
a' &= a + \omega_F \\
b' &= b + \omega_F \\
\theta' &= \theta + \omega_F,
\end{aligned}$$

onde ω_P e ω_F são variáveis aleatórias Gaussianas de média zero e desvio padrão de 1.2 *pixels*.

A cada nova imagem existe a probabilidade de um novo objeto entrar na cena, bem como a probabilidade de um dos objetos presentes deixar a cena. Esse fato exige uma função de inicialização $g(t)$ dos novos objetos que surgem e uma maneira de detectar a saída de outros.

A função de inicialização $g(t)$ é uma função simples que gera um objeto com forma e velocidades padronizadas. A posição do novo objeto é encontrada monitorando as bordas da imagem. Na posição em que ocorrer uma região com probabilidade de ser um objeto de interesse maior que um dado limiar, será iniciado uma nova hipótese de objeto. O Algoritmo 4 resume os passos da predição.

No Algoritmo 4 nota-se que o passo de predição é realizado separadamente para cada partícula S_t^n . Na primeira linha o número de elementos da partícula predita S_t^n é inicializado com zero. Isso porque alguns objetos podem ser considerados como ausentes na cena ou outros podem entrar. Para cada representação de objeto da partícula é gerado um número aleatório r que dirá a probabilidade de um objeto continuar presente. Em caso positivo, o objeto recebe o devido movimento por meio da função de predição $f(x_t^n)$. Ao final do cálculo de movimento de todas as representações de objetos em S_t^n , um novo r é escolhido aleatoriamente para decidir se um novo objeto será ou não iniciado. Esse processo é consistente com uma distribuição de Poisson explicando a entrada de objetos e uma distribuição exponencial para sua permanência [Isard and MacCormick, 2001].

```

1:  $m_t^n \leftarrow 0$ ;
2: for  $i \leftarrow 1$  to  $m_{t-1}^n$  do
3:   Gere um número randômico  $r \in [0, 1)$  uniformemente distribuído;
4:   if  $r < \lambda_r$  then
5:      $m_t^n \leftarrow m_t^n + 1$ ;
6:      $x_t^{n,m_t^n} \leftarrow f(x_{t-1}^{m,i})$ ;
7:   end if
8: end for
9: Gere um número randômico  $r \in [0, 1)$  uniformemente distribuído;
10: if  $r < \lambda_i$  then
11:    $m_t^n \leftarrow m_t^n + 1$ ;
12:    $x_t^{n,m_t^n} \leftarrow g(\phi)$ ;
13: end if

```

Algoritmo 4: Algoritmo de predição.

3.1.6 Estimador de objetos válidos

O algoritmo CONDENSATION não faz distinção dos objetos contidos nas partículas. Sua tarefa é propagar a distribuição de probabilidades $P(X_t | Z_t)$ no tempo usando um modelo de observação para computar as probabilidades de suas hipóteses de configuração, sem a preocupação de como é a estrutura das partículas. O método precisa estimar as representações dos objetos separadamente. Isso é feito usando a probabilidade total de um objeto estar presente na cena no instante de tempo t .

Assumindo que o conjunto de partículas S_t^1, \dots, S_t^N são eventos mutuamente exclusivos representados por áreas específicas, cujo tamanho corresponde ao peso ϕ_t^i de cada partícula, como mostra a Figura 3.10. Seja um peixe específico s_t^r representado pela elipse cinza na Figura 3.10, o tamanho da área da elipse, corresponde à probabilidade do peixe representado por s_t^r ocorrer na imagem observada no instante t .

Então, por meio do teorema da probabilidade total, sabe-se que a probabilidade do peixe s_t^r é dada por:

$$\begin{aligned}
P(\mathbf{s}_t^r) &= P(\mathbf{s}_t^r \cap S_t^1) + \dots + P(\mathbf{s}_t^r \cap S_t^N) \\
&= P(\mathbf{s}_t^r | S_t^1)P(S_t^1) + \dots + P(\mathbf{s}_t^r | S_t^N)P(S_t^N) \\
&= P(\mathbf{s}_t^r | S_t^1)\pi_t^1 + \dots + P(\mathbf{s}_t^r | S_t^N)\pi_t^N.
\end{aligned} \tag{3.12}$$

Por meio de um limiar λ_d pré-estabelecido, encontramos os objetos que possuem a probabilidade total maior que λ_d para serem estimados. Esse processo elimina inicializações mal sucedidas durante o passo de predição.

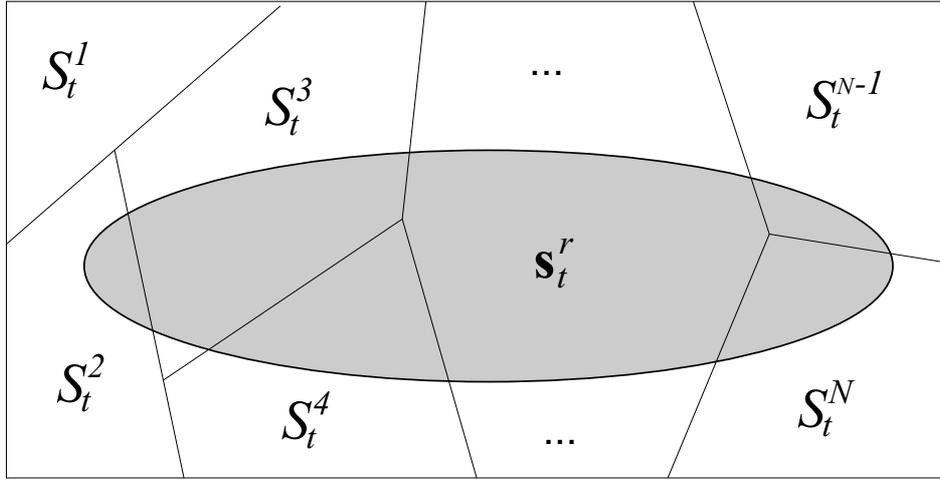


Figura 3.10: Probabilidade total do peixe representado por s_t^r . Considerando cada partícula um evento mutuamente exclusivos, a probabilidade de um peixe s_t^r será dada pela área da elipse cinza.

Cada objeto identificado por Φ^i precisa de uma estimativa no instante t . Sabendo o maior número de objetos M_t representado por uma partícula, podemos criar um conjunto $\mathcal{M}_t^{\Phi_i}$ dos objetos rotulados por Φ_i , sendo que cada elemento do conjunto é composto pelo par ordenado que identifica o j -ésimo objeto da n -ésima partícula. Com o conjunto $\mathcal{M}_t^{\Phi_i}$, sabe-se em quais partículas do conjunto S_t^1, \dots, S_t^N ocorrem os objetos e portanto podemos calcular a probabilidade total $\Pi_t^{\Phi_i}$ do objeto Φ_i .

```

1: for  $i = 1$  to  $M_t$  do
2:   Compute  $\mathcal{M}_t^{\Phi_i} = \{(n, j) : \phi_t^{n,j} = \Phi_i\}$ ;
3:   Compute  $\Pi_t^{\Phi_i} = \sum_{(n,j) \in \mathcal{M}_t^{\Phi_i}} p(\phi_t^{n,j})$ 
4:   if  $\Pi_t^{\Phi_i} > \lambda_d$  then
5:      $\hat{x}_t^{\Phi_i} = \sum_{(n,j) \in \mathcal{M}_t^{\Phi_i}} p(\phi_t^{n,j}) \phi_t^{n,j} / \Pi_t^{\Phi_i}$ 
6:   end if
7: end for

```

Algoritmo 5: Algoritmo de estimação.

A estimativa dos objetos que possuem probabilidade total maior que λ_d , será dada pela média das representações. O Algoritmo 5 resume os passos da estimação de objetos válidos.

A cada quadro, as estimativas $\hat{x}_t^{\Phi_i}$ dos objetos são armazenadas, o que possibilita determinar, por exemplo, a trajetória do objeto ao longo de uma seqüência de ima-

gens. Os dados da saída do rastreador formarão uma lista das estimativas dos objetos detectados e rastreados.

3.1.7 Contagem de peixes

Com a técnica BraMBLe, conseguimos rastrear os objetos de interesse de uma cena, dada uma seqüência de imagens sucessivas. Ou seja, conseguimos detectar a presença dos objetos de interesse e relacionar as ocorrências do mesmo objeto ao longo dos quadros, atribuindo a elas o mesmo rótulo de identificação.

Em uma imagem isolada, poderíamos contar os rótulos associados aos objetos detectados e saberíamos quantos objetos estão na imagem. Em nosso caso, queremos contar objetos que se movem em uma cena, especificamente peixes logo, não podendo contar todas as ocorrências encontradas, já que o mesmo objeto é detectado muitas vezes. Para eliminar as redundâncias da contagem, proveniente das ocorrências múltiplas ao longo da coleção, usamos um rastreador de múltiplos objetos.

Nesse novo contexto, os objetos são identificados por suas estimativas de trajetória. Dessa forma, contar os objetos significa contar as trajetórias encontradas no rastreamento. Quando um novo objeto é detectado na cena, uma nova trajetória é iniciada e só terminará quando a posição do objeto não for mais estimada. Isso significa que o objeto saiu da cena ou que o sistema perdeu o objeto.

Em nosso contexto, os peixes estão nadando em um dispositivo de transposição e em algum momento passam diante de uma janela de observação. Como estamos interessados em quantificar o fluxo de peixes que sobem ou descem o referido dispositivo, vamos contar os peixes que atravessam a imagem de um lado a outro. Por esse motivo adotamos a metodologia de [Rossi and Bopzzoli, 1994] que divide a área das imagens em detecção, rastreamento e contagem para contar pessoas em um ambiente específico. Naquele caso, as pessoas eram contadas apenas em uma direção. Em nosso problema contamos os peixes que nadam em duas direções e por isso temos duas regiões de detecção e duas de contagem como mostra a Figura 3.11. O sistema detecta novos peixes nas áreas laterais e então inicia o rastreamento em toda a área das imagens até chegar à outra área lateral onde o objeto será contado.

Um objeto só é contado quando sua trajetória não é mais estimada. Quando isso ocorre, o sistema precisa verificar se o objeto deixou a imagem ou foi perdido pelo rastreador. A verificação é feita, executando uma rotina de contagem. A rotina de contagem verifica se o objeto não estimado atravessou a imagem e, nesse caso, efetua a contagem. Observe no Algoritmo 6 que a contagem é direcionada, separando os objetos

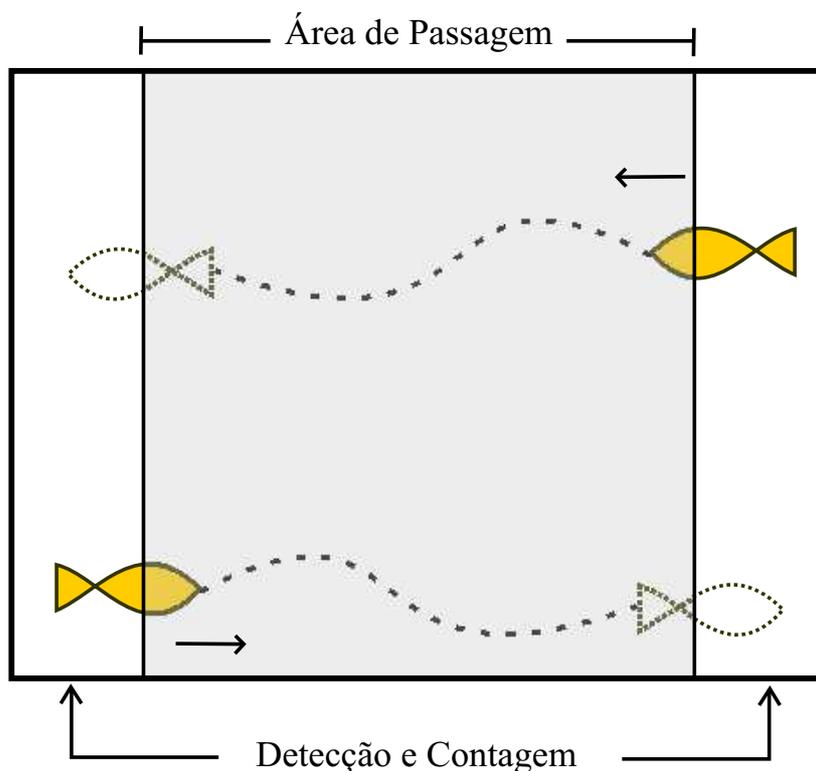


Figura 3.11: Imagens dividida em regiões para a contagem. Partes de observação nas imagens, baseado no trabalho de [Rossi and Bopzzoli, 1994]. Os objetos são detectados nas laterais e rastreados até a outra extremidade, onde são contados.

que se movem da direita para a esquerda daqueles que se movem da esquerda para a direita.

O algoritmo de contagem mostrado no Algoritmo 6 faz a contagem apenas dos *blobs* que atravessam a imagem. O sistema sabe que um objeto atravessou a área da imagem quando as coordenadas da primeira e última estimativas estão dentro das áreas de detecção e contagem em lados opostos da imagem. Nos casos em que um peixe entra na cena por um dos lados e deixa a cena pelo mesmo lado, não existe a contagem. Apesar de esse comportamento gerar uma trajetória em nosso conjunto, ela não será contada porque nosso interesse é contar os peixes que sobem o dispositivo de transposição e os peixes que descem. Consideramos que um peixe subiu ou desceu quando entra por um lado da image e sai pelo outro.

Neste capítulo foi apresentada uma metodologia baseada em visão computacional para contar peixes que nadam através de um mecanismo de transposição, analisando seqüência de imagens digitais. Como foi visto, o rastreamento dos peixes no plano das imagens tem como objetivo relacionar as ocorrências do mesmo objeto ao longo

```

1: est_anterior ← vazio;
2: for all Quadros da seqüência do
3:   estimativas ← resultado do estimador;
4:   Encontre os objetos não estimados comparando est_anterior e estimativas;
5:   for all Objeto não estimado do
6:     if (origem( $\hat{x}_t^{\Phi_i}$ ) ∈ AreaEsquerda) AND (final( $\hat{x}_t^{\Phi_i}$ ) ∈ AreaDireita) then
7:        $\text{CONTAGEM}_{ed} \leftarrow \text{CONTAGEM}_{ed} + 1$ ;
8:       Marca como contado;
9:     end if
10:    if (origem( $\hat{x}_t^{\Phi_i}$ ) ∈ AreaDireita) AND (final( $\hat{x}_t^{\Phi_i}$ ) ∈ AreaEsquerda) then
11:       $\text{CONTAGEM}_{de} \leftarrow \text{CONTAGEM}_{de} + 1$ ;
12:      Marca como contado;
13:    end if
14:  end for
15:  est_anterior ← estimativas;
16: end for

```

Algoritmo 6: Algoritmo de contagem.

da seqüência e contar uma única vez. No próximo capítulo são mostrados testes e resultados obtidos aplicando uma implementação da metodologia a imagens reais.

Capítulo 4

Experimentos

Este capítulo mostra resultados obtidos aplicando-se a metodologia adotada a imagens reais, obtidas tanto de forma controlada em laboratório, como em situações reais sem interferir nas condições do ambiente no qual foram obtidas. Todas as seqüências foram gravadas para serem processadas em laboratório, utilizando-se um computador equipado com processador Pentium 4 de 3.0GHz, 512KB de memória cache e 1GB de memória RAM, executando a versão FEDORA 2 do sistema operacional GNU LINUX.

A metodologia foi implementada usando o software MATLAB V7 como ambiente de programação fazendo uso do sistema operacional sem se preocupar com performance.

Cada vídeo processado pelo sistema foi previamente fragmentado em uma coleção de imagens do tipo PPM. As imagens utilizadas foram nomeadas com o número do quadro correspondente e armazenadas em diretórios cujo nome identifica o vídeo.

4.1 Planejamento dos experimentos

O objetivo dos experimentos é medir a precisão das trajetórias e verificar a contagem dos peixes nas seqüências de vídeo adquiridas em Igarapava. Para isto, foram projetados três grupos de experimentos que iniciam por filmes adquiridos em laboratório. Nesse ambiente, a iluminação foi mantida constante e a cena observada não possuía objetos que pudessem interferir no modelo de observação.

O primeiro experimento mostra o funcionamento do rastreador sem a preocupação de contar os objetos rastreados. O segundo caso, apresenta o rastreamento onde a detecção do objeto de interesse, é dificultada por características da cena que interferem na cor do objeto rastreado. O terceiro experimento contou com imagens reais adquiridas da janela de observação instalada na escada de transposição da Usina Hidrelétrica de

Igarapava. Este experimento foi dividido em duas etapas: rastreamento e contagem, sendo a contagem mostrada em uma seção específica.

Em cada um dos experimentos as imagens de saída mostram o resultado obtido pelo modelo de observação em cada quadro. Esse resultado é exibido por meio de imagens cujos valores dos *pixels* representam a taxa de contribuição do plano dos objetos em relação ao plano de fundo, ou seja, quão maior é a probabilidade do valor de um *pixel* ser parte de algum objeto da cena, comparando com a probabilidade do mesmo *pixel* representar o plano de fundo. Essa taxa de participação do plano dos objetos em relação ao plano de fundo, é calculada da seguinte forma, tal como mostrado na Equação 3.7 da Seção 3.1.3:

$$Z(i, j) = \log (P_{frente} (I(i, j))) - \log (P_{fundo} (I(i, j))). \quad (4.1)$$

Para facilitar a observação e avaliação das imagens de saída, criou-se uma paleta de cores específica que mapeia os valores de cada *pixel* $Z(i, j)$ para uma cor no espaço RGB. Com base na Equação 4.1, observa-se que os valores abaixo de zero são aqueles com maior probabilidade de representar o plano de fundo e os valores positivos são os que melhor representam os objetos de interesse. Esses valores são representados respectivamente por cores frias e quentes, como mostra a Figura 4.1.

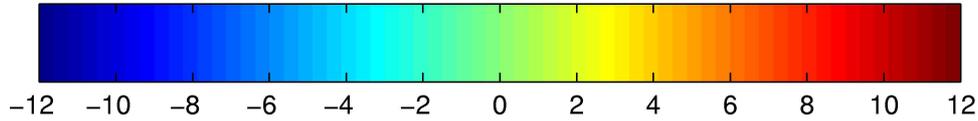


Figura 4.1: Escala de cores adotada para a exibição das imagens de saída. Com base na Equação ??, é possível concluir que os valores abaixo de zero, representam regiões do *pixels* de fundo e os valores acima de zero, são os que melhor representam o plano dos objetos.

Nas mesmas imagens, são desenhadas as elipses que representam o objeto rastreado naquele quadro, mostrando-se parcialmente o resultado do rastreamento. As trajetórias completas são exibidas em forma gráfica, comparando-se as trajetórias estimadas pelo sistema e as trajetórias reais obtidas manualmente.

Em cada experimento foram medidos os erros médios absolutos, dados por:

$$\sqrt{\frac{1}{n} \sum_{i=1}^n |\bar{v} - v_m|},$$

onde \bar{v} representa o valor estimado pelo rastreador e v_m o valor médio de várias medidas manuais obtidas independentemente, cujo desvio padrão médio encontrado foi de 1.9716 *pixels* no eixo x e 1.2370 *pixels* no eixo y .

A contagem automática dos peixes foi comparada com a contagem obtida pelo método tradicional, observando o vídeo. Os resultados são exibidos em forma de gráficos onde é possível perceber os falso negativos encontrados.

4.2 Primeiro experimento - Rastreamento de *blobs* coloridos

4.2.1 Configuração e resultados

Este experimento tem o objetivo de verificar o funcionamento da técnica de rastreamento adotada, sendo os alvos *blobs* de cores diferentes. Os *blobs* coloridos podem ser utilizados como marcações colocadas em objetos diversos para que o objeto possa ser detectado sem maiores problemas.

Para a realização do experimento foram utilizadas quatro bolinhas de cores diferentes (verde, amarelo, azul e vermelho) rolando sobre o piso do laboratório. A seqüência de imagens foi adquirida por uma câmera digital HITACHI modelo KP-D50 instalada no teto apontando para o piso com um ângulo não determinado. As bolinhas foram lançadas manualmente no campo visual da câmera e portanto, as velocidades não foram determinadas. A Figura 4.2 mostra um diagrama da configuração física estabelecida para a realização do experimento.



Figura 4.2: Esquemático do primeiro experimento. Uma câmera instalada no teto, adquire imagens de bolas coloridas que rolam no piso do laboratório.

O treinamento do plano de fundo contou com as 60 imagens iniciais do filme ad-

quirido, as quais contém apenas o cenário. Cada *fdp* do plano de fundo foi obtida agrupando os pontos correspondentes em quatro agrupamentos. Para o treinamento do plano dos objetos foram separadas 70 imagens da seqüência que continham todas as bolinhas para que fosse possível encontrar a *fdp* que abrangesse todas as quatro cores, segmentadas por subtração de fundo como na Figura 3.1 para que apenas os *pixels* de interesse fossem considerados. A *fdp* do plano dos objetos foi obtida por uma composição de vinte e cinco parciais, cada uma representando um grupamento.

A Figura 4.3 mostra alguns quadros de uma seqüência composta por 310 imagens processadas onde é possível visualizar a seguimentação dada pelo modelo de observação e as elipses desenhadas nas posições estimadas pelo rastreador.

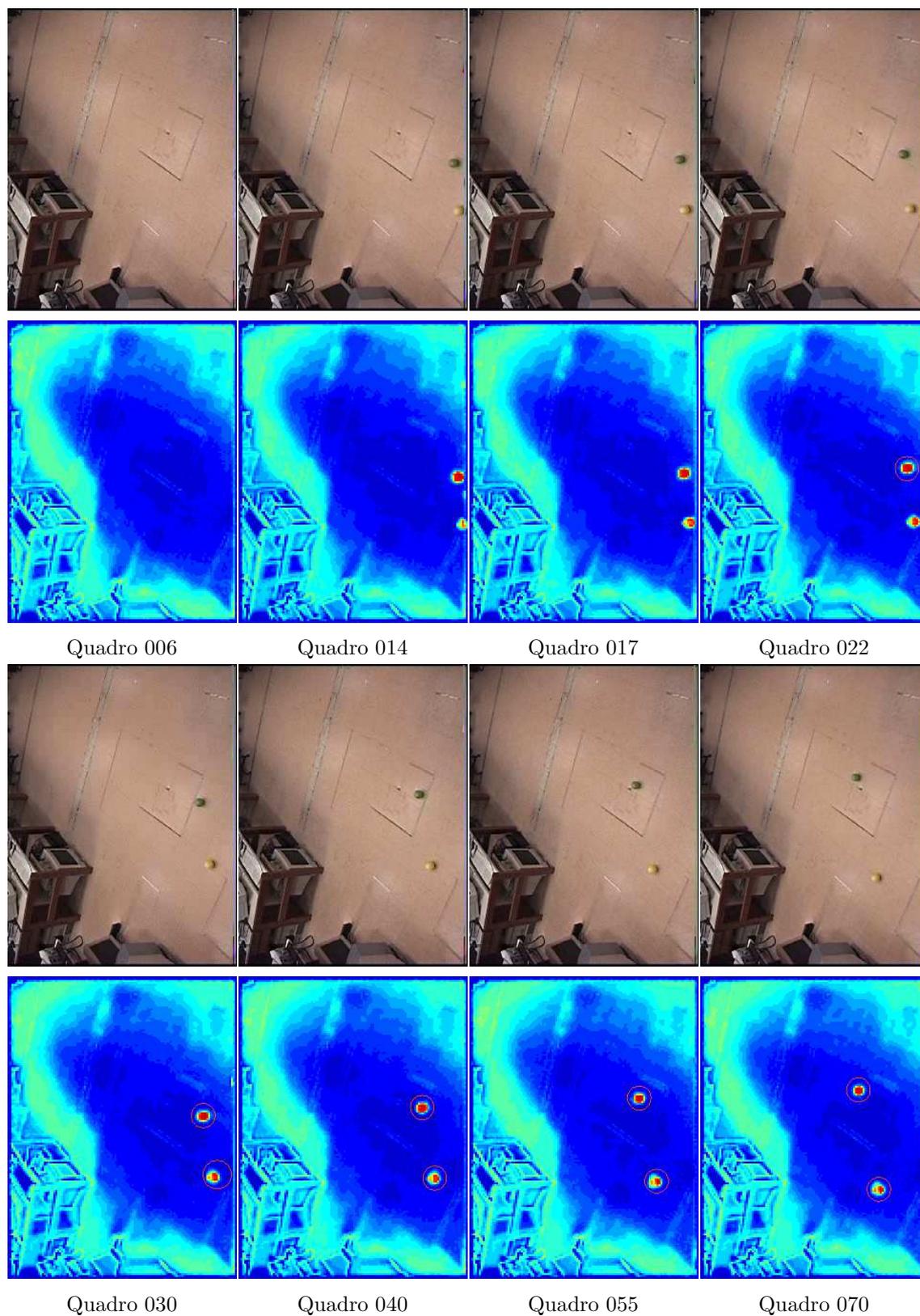


Figura 4.3: Primeiro experimento. Rastreamento de bolinhas rolando no piso do laboratório. Em cada quadro é mostrada a imagem adquirida pela câmera e a imagem observada pelo sistema. A presença de uma elipse sobre o objeto na imagem observada pelo sistema, caracteriza o rastreamento daquele objeto.

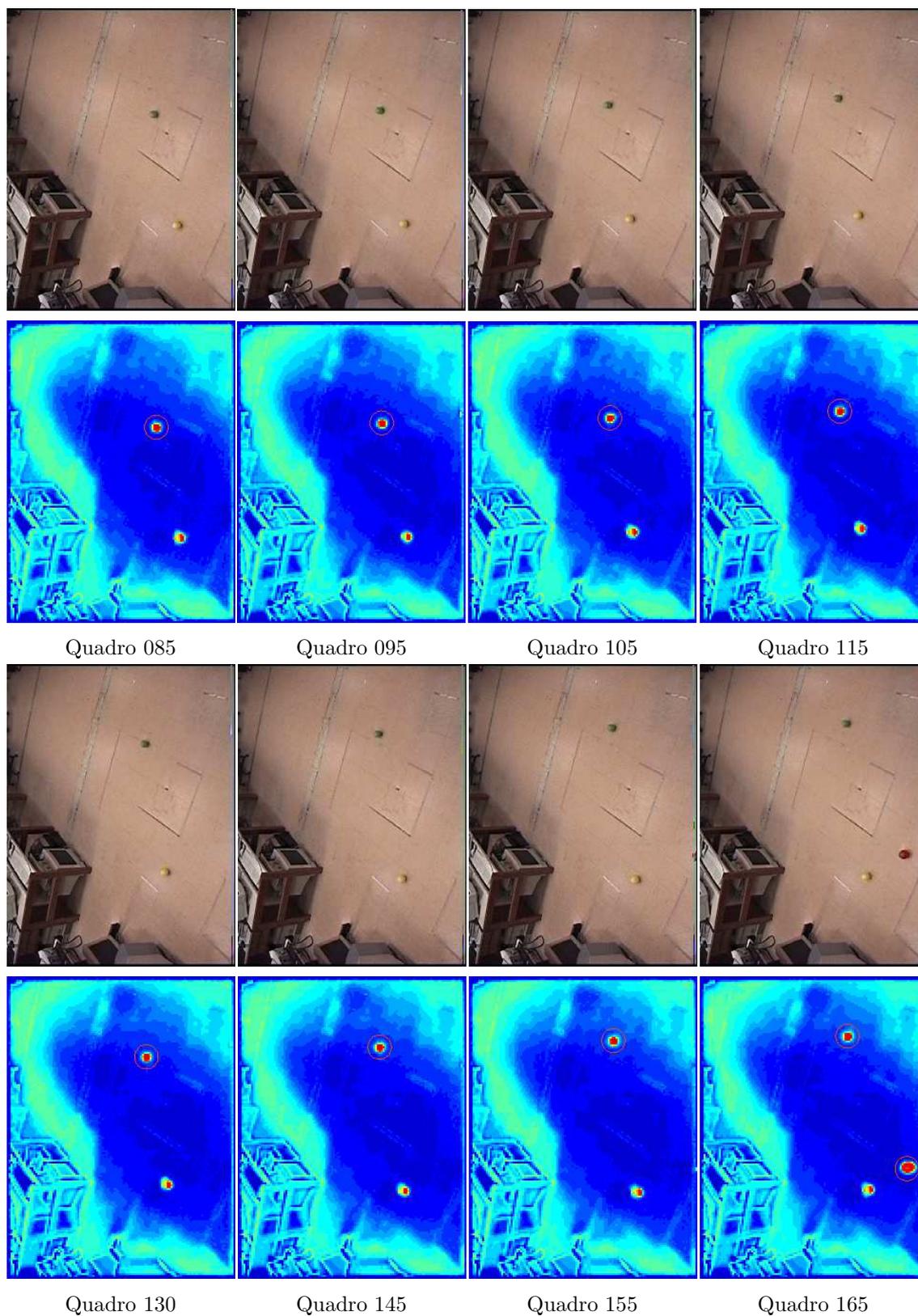


Figura 4.3: Primeiro experimento (Cont.).

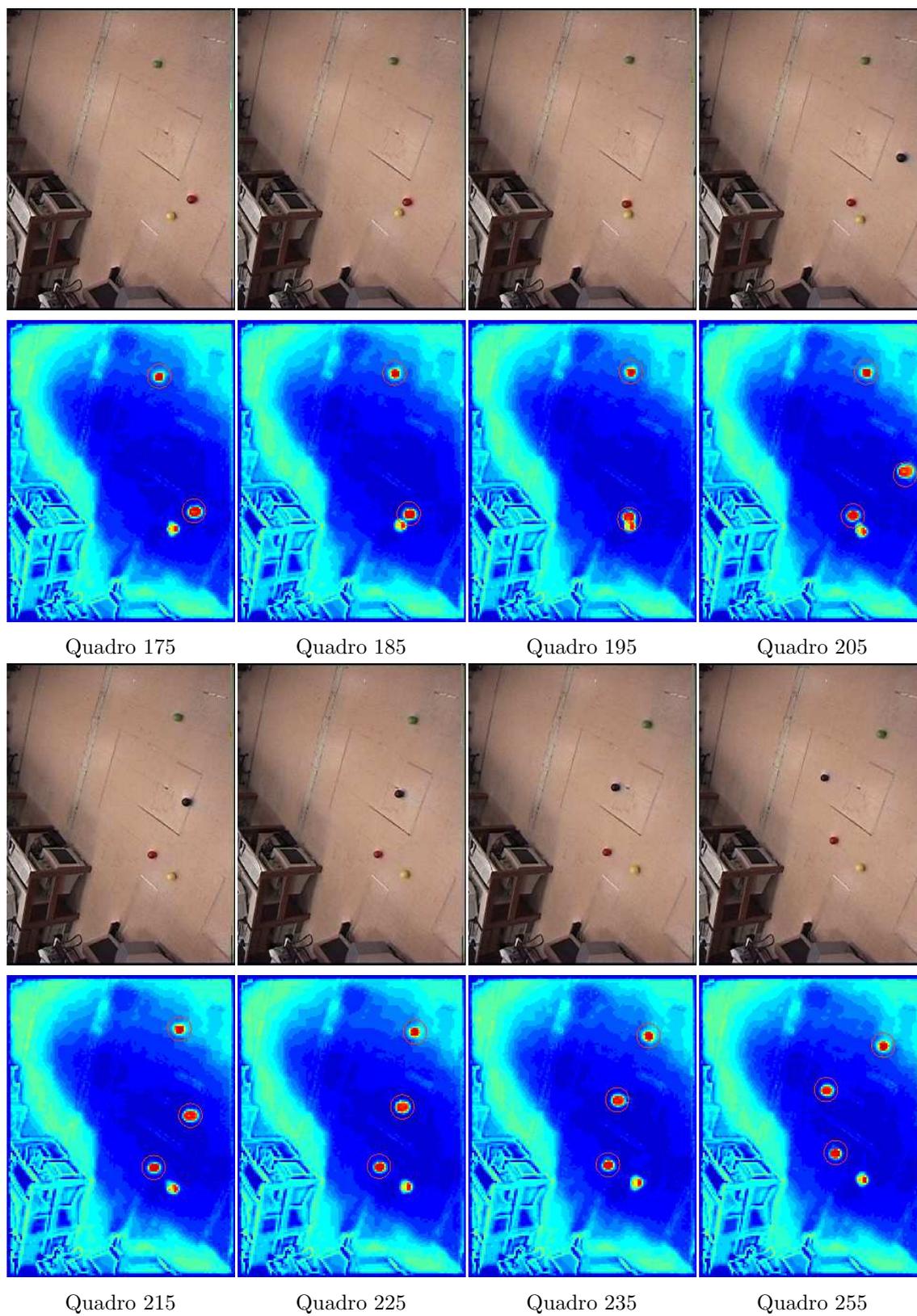


Figura 4.3: Primeiro experimento (Cont.).

Como em todos os testes, os objetos foram iniciados apenas nas laterais dos quadros. Nesta seqüência, utilizaram-se apenas 20 *pixels* de largura para as faixas de detecção. A detecção dos objetos que entram na cena é evidenciado pelas elipses que surgiram sobre os objetos. As trajetórias encontradas pelo rastreador são mostradas na Figura 4.4, sendo os erros encontrados exibidos na Tabela 4.1.

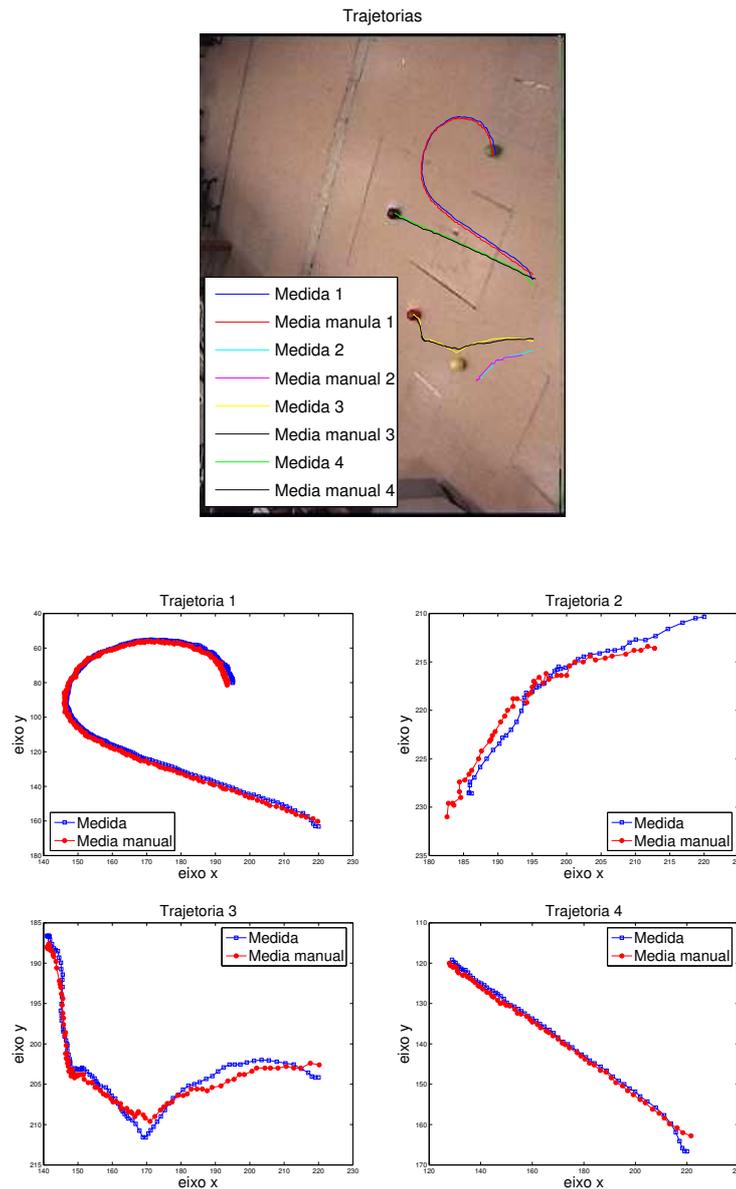


Figura 4.4: Trajetórias encontradas no experimento 1. As trajetórias encontradas pelo sistema são desenhadas juntamente com a média de trajetórias obtidas manualmente.

Tabela 4.1: Erros e desvios padrão encontrados no experimento 1.

	Eixo X	Eixo Y
Erro médio absoluto	1.5529	0.8987
Desvio padrão	1.2514	0.8284

4.2.2 Discussão

O experimento mostra a segmentação de quatro bolinhas com cores diferentes. As imagens, correspondentes ao resultado obtido pelo modelo de observação, mostram que apenas os pontos correspondentes a objetos de interesse receberam valores acima de zero, caracterizando a segmentação dos objetos. Esse fato pode ser verificado observando as imagens da Figura 4.3.

O *blob* referente à segmentação da bolinha amarela, recebeu valores próximos de zero em parte de sua área. O quadro 75 da Figura 4.3 mostra a representação destes valores em tons de amarelo tendendo a verde em uma posição que facilita a comparação visual do *blob* com seu vizinho. Esta segmentação em valores baixos fez com que o rastreamento do objeto fosse interrompido no quadro 75.

O rastreador do sistema baseia-se em um filtro de partículas para múltiplos objetos. A presença de um objeto com valores baixos de segmentação em uma imagem com outros objetos de valores mais altos, pode levar o filtro a perder o objeto de menor peso. O filtro tende a eliminar as partículas de menor peso no passo de reamostragem o que favorece aquelas de peso maior. Na seqüência mostrada pelo primeiro experimento o rastreador perde a bolinha amarela por estar com valores baixos na segmentação e mantém o rastreamento da outra bolinha presente na cena.

A Figura 4.4 mostra as trajetórias encontradas pelo rastreador sobrepostas às trajetórias medidas manualmente. Os erros médios quadráticos menores que um *pixel* no eixo *y* e menores que dois *pixels* no eixo *x*, mostram que o sistema pode ser utilizado para rastrear múltiplos objetos simultâneos e independentes com precisão.

4.3 Segundo experimento - Rastreamento de peixe em laboratório

4.3.1 Configuração e resultados

No segundo experimento foram utilizadas imagens adquiridas também em laboratório com objetivo de realizar o rastreamento em ambiente com características que se aproximam do real. Foram adquiridas imagens de um peixe nadando através de um protótipo de mecanismo de transposição instalado no Centro de Pesquisas Hidráulicas e Recursos Hídricos (CPH), da Universidade Federal de Minas Gerais (UFMG). Esse protótipo possui as laterais confeccionadas em material transparente possibilitando a observação dos peixes que nadam através do mecanismo. A água utilizada no experimento foi contaminada com impurezas para que apresentasse um aspecto semelhante à água de um rio comum.

As imagens foram adquiridas utilizando-se uma câmera digital Sony modelo DCR-TRV320 posicionada diante do protótipo com o apoio de um tripé. O trecho de vídeo tratado possui duração de 9 segundos armazenados em uma fita HI-8 que foram digitalizados posteriormente. A Figura 4.5 mostra um diagrama do *setup* físico adotado neste experimento.

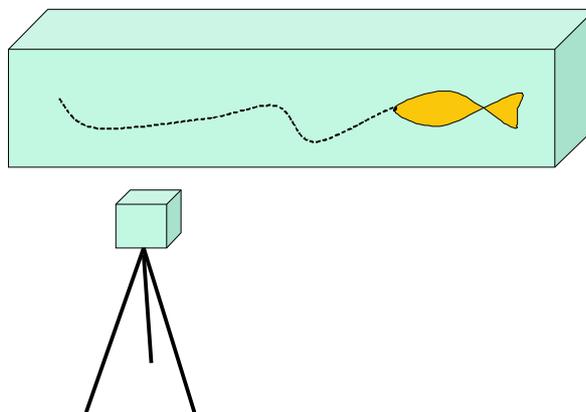


Figura 4.5: Esquemático do segundo experimento. Uma câmera estacionária adquire imagens de um peixe que nada através de protótipo de mecanismo de transposição.

Para o treinamento do plano de fundo, foram separadas 27 imagens consecutivas da seqüência capturada que não continham objetos de interesse (no caso o peixe) e 22 imagens contendo o objeto de interesse foram escolhidas para o treinamento do plano do objeto. A *fdp* relativa aos objetos de interesse foi encontrada, agrupando-se os *pixels*

correspondentes em 25 grupamentos para a composição de Gaussianas que dão origem à *fdp*. O conjunto de *fdp*'s do plano de fundo é composto por quatro agrupamentos cada um.

Na Figura 4.6 são exibidos alguns quadros da seqüência de 267 imagens processadas, mostrando-se o resultado obtido pelo modelo de observação que seguimenta os objetos de interesse e as elipses estimadas pelo rastreador.

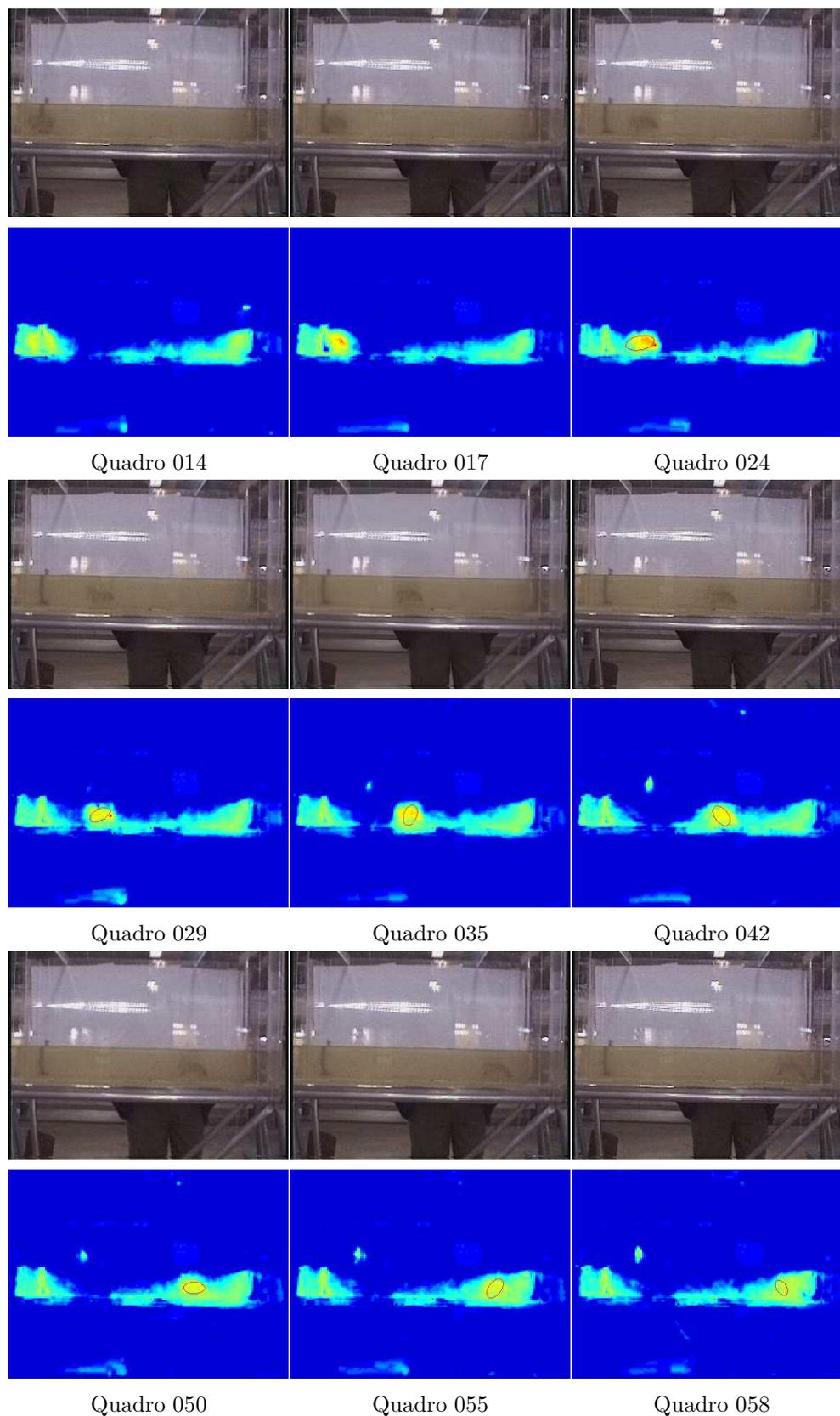


Figura 4.6: Segundo experimento. Rastreamento de um peixe nadando em um protótipo de mecanismo de transposição. Em cada quadro é mostrada a imagem adquirida pela câmera e a imagem observada pelo sistema. A presença de uma elipse sobre o objeto na imagem observada pelo sistema, caracteriza o rastreamento daquele objeto.

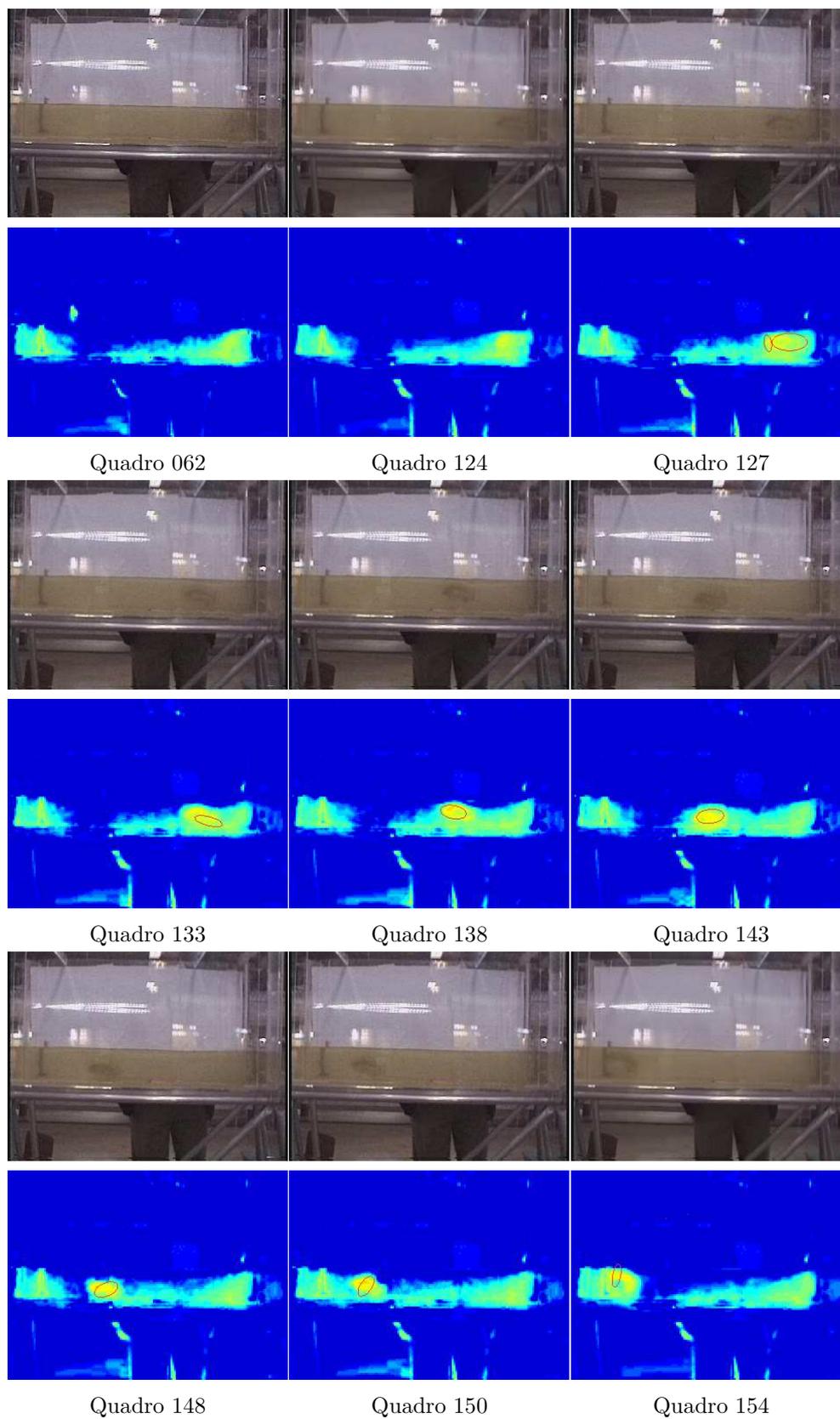
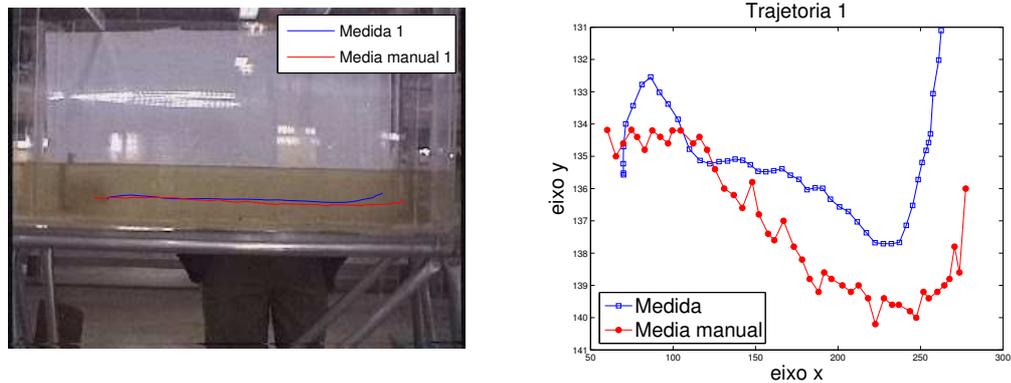
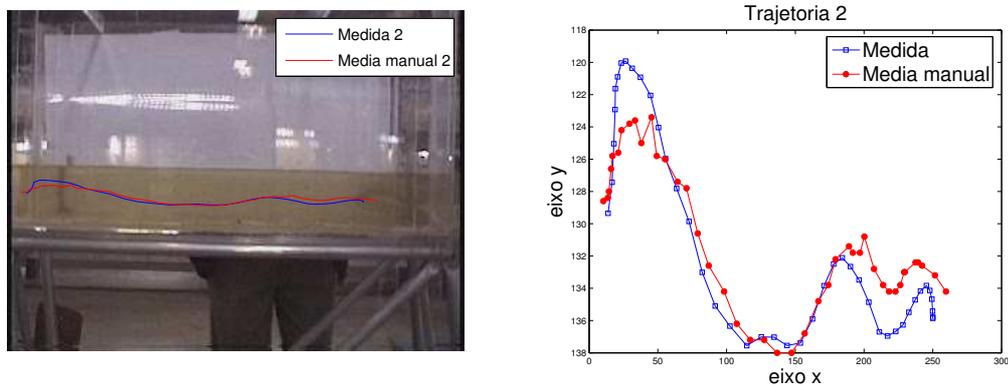


Figura 4.6: Segundo experimento (Cont.).

As trajetórias encontradas pelo rastreador foram comparadas com a médias de trajetórias medidas manualmente de forma independente. A Figura 4.7 exibe as duas trajetórias em cores diferentes e a Tabela 4.2 mostra os erros encontrados.



(a) Movimento da esquerda para a direita.



(b) Movimento da direita para a esquerda.

Figura 4.7: Trajetórias encontradas no experimento 2. As trajetórias encontradas pelo sistema são desenhadas juntamente com a média de trajetórias obtidas manualmente.

Tabela 4.2: Erros e desvios encontrados no experimento 2.

	Eixo X	Eixo Y
Erro médio absoluto	4.4201	2.0781
Desvio Padrão	3.9564	1.4315

4.3.2 Discussão

Neste experimento, foram observadas instabilidades na segmentação geradas pela semelhança das cores do plano de fundo na região onde o peixe é encontrado e as cores relativas ao peixe. O peixe se confunde com a cena por estar imerso em água propositalmente contaminada. As impurezas diluídas na água têm o objetivo de simular a turbidez de um rio com água barrenta. Por outro lado, dificulta a detecção do peixe e justifica a necessidade de um sistema capaz de retificar as imagens de entrada.

O experimento mostra as áreas de instabilidade concentradas nas extremidades da imagem. O protótipo de mecanismo de transposição utilizado possui alhetas verticais instaladas nessas posições que forçam a passagem do peixe no lado que fica mais distante da câmera.

As alhetas aumentam a concentração de impurezas no local e fazem com que o peixe passe em uma região com maior coluna de água. Esse fato dificulta a detecção porque a água interfere nas cores dos pontos observados que correspondem ao peixe.

Em consequência da instabilidade gerada nas áreas de entrada e saída do peixe na cena, as faixas de detecção na seqüência avaliada foram dilatadas para setenta *pixels*, permitindo que o peixe fosse detectado já saindo dessa região.

Apesar da dificuldade de detecção do peixe ao longo dos quadros da seqüência, o sistema conseguiu detectar e rastrear o peixe nadando nos dois sentidos de nado com um erro médio próximo de seis *pixels* no eixo x e próximo de três no eixo y . Esse fato mostra que o sistema pode ser aplicado para o rastreamento de peixes em águas parcialmente turvas.

4.4 Terceiro experimento - Rastreamento de peixes no mecanismo de transposição

4.4.1 Configuração e resultados

O terceiro experimento mostra o rastreamento de peixes que nadam através do mecanismo de transposição instalado na Usina Hidrelétrica de Igarapava. As imagens foram adquiridas por uma filmadora digital SONY modelo DCR-TRV950 instalada na sala de observação do mecanismo. A câmera foi posicionada diante da janela que permite aos pesquisadores observarem os peixes que nadam através do mecanismo, como ilustrado na Figura 4.8

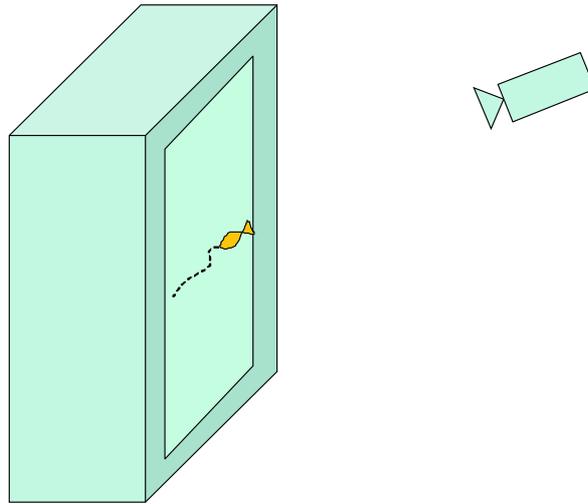


Figura 4.8: Esquemático do terceiro experimento. Uma câmera estacionária adquire imagens de peixes que nadam através do mecanismo de transposição instalado na Usina Hidrelétrica de Igarapava.

Para cada um dos 3 trechos apresentados, houve um treinamento específico cujos número de imagens utilizado é mostrado na Tabela 4.3. Em todos os casos, as imagens utilizadas para o treinamento não fazem parte da seqüência utilizada para o rastreamento. As *fdp*'s do plano de fundo foram encontradas combinando-se quatro funções Gaussianas que representam os agrupamentos em cada ponto. As *fdp*'s que representam os objetos de interesse foram determinadas agrupando os pontos em vinte e cinco grupos.

Tabela 4.3: Número de imagens utilizadas para o treinamento e rastreamento das seqüências adquiridas em Igarapava.

Filme	Rastreamento	<i>foreground</i>	<i>background</i>
1	204	26	40
2	95	35	65
3	240	56	50

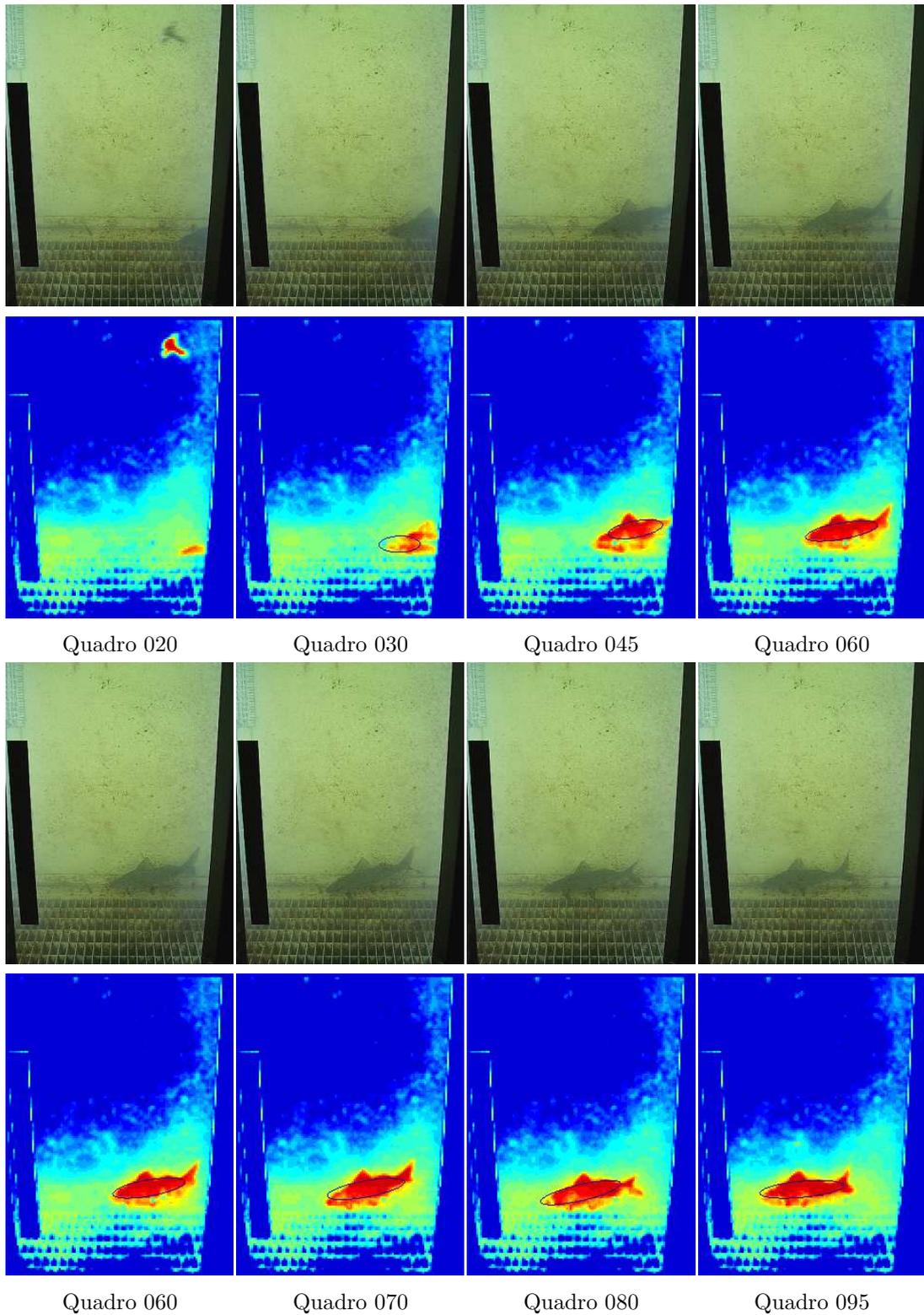


Figura 4.9: Terceiro experimento A. Rastreamento de peixes nadando em um mecanismo de transposição Instalado na Usina Hidrelétrica de Igarapava. Em cada quadro é mostrada a imagem adquirida pela câmera e a imagem observada pelo sistema. A presença de uma elipse sobre o peixe na imagem observada pelo sistema, caracteriza o rastreamento daquele peixe.

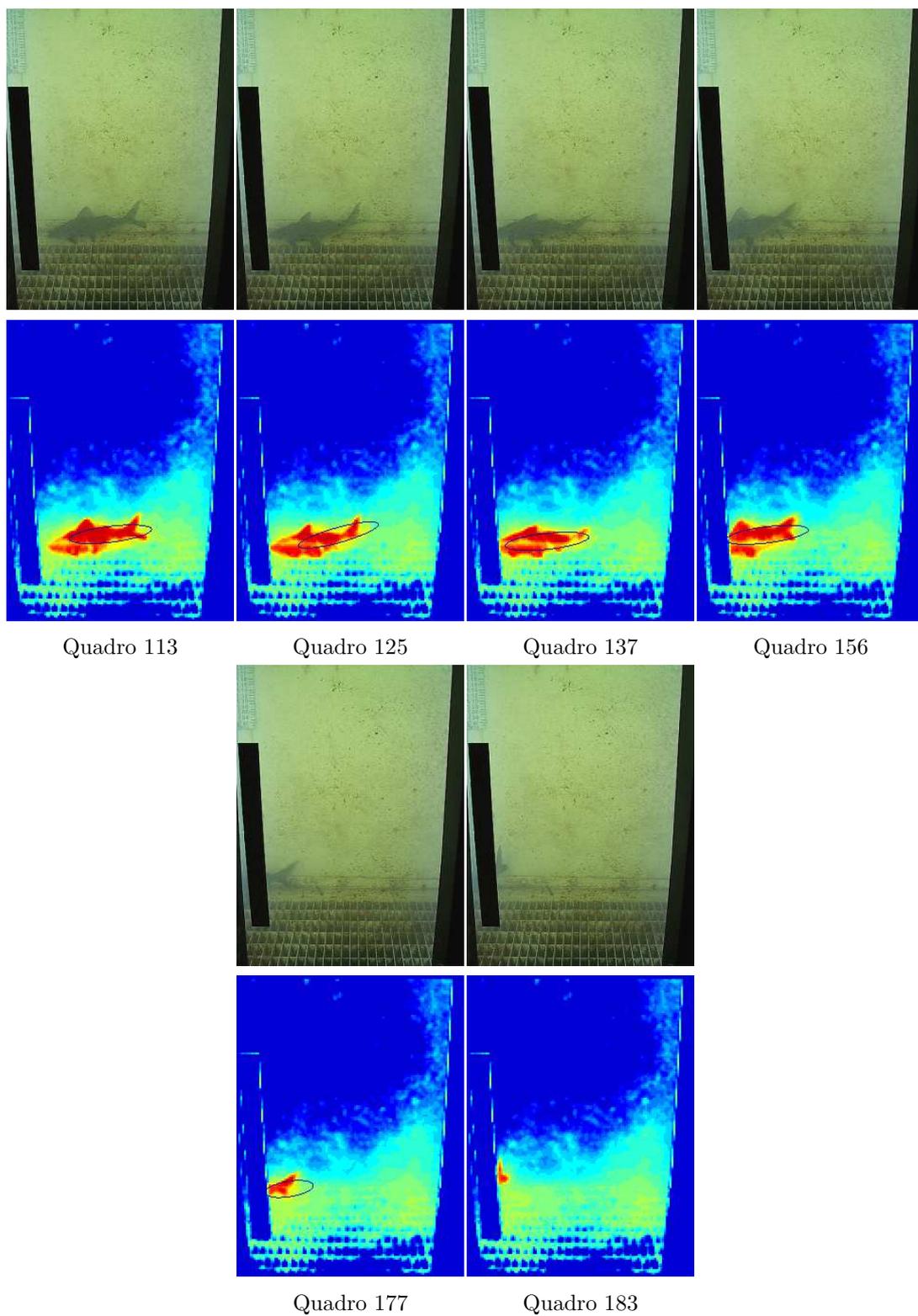


Figura 4.9: Terceiro experimento A.(Cont.)

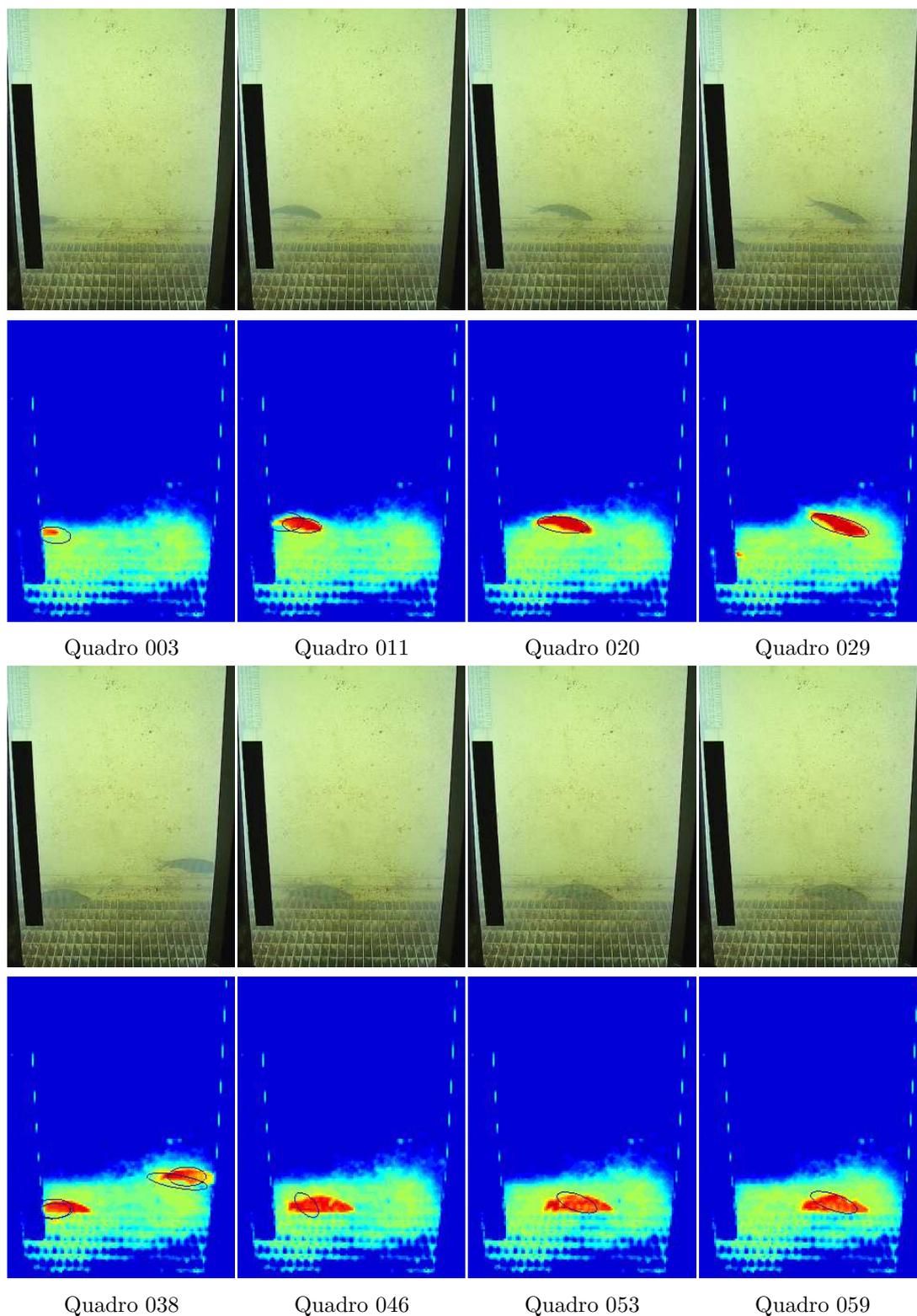


Figura 4.10: Terceiro experimento B. Rastreamento de peixes nadando em um mecanismo de transposição Instalado na Usina Hidrelétrica de Igarapava. Em cada quadro é mostrada a imagem adquirida pela câmera e a imagem observada pelo sistema. A presença de uma elipse sobre o peixe na imagem observada pelo sistema, caracteriza o rastreamento daquele peixe.

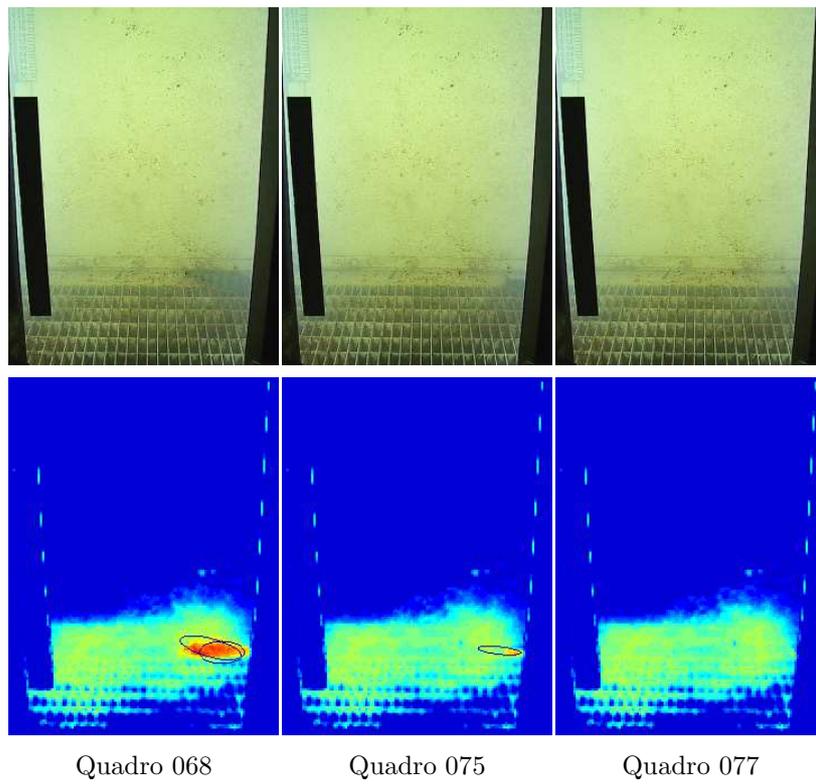


Figura 4.10: Terceiro experimento B.(Cont.)

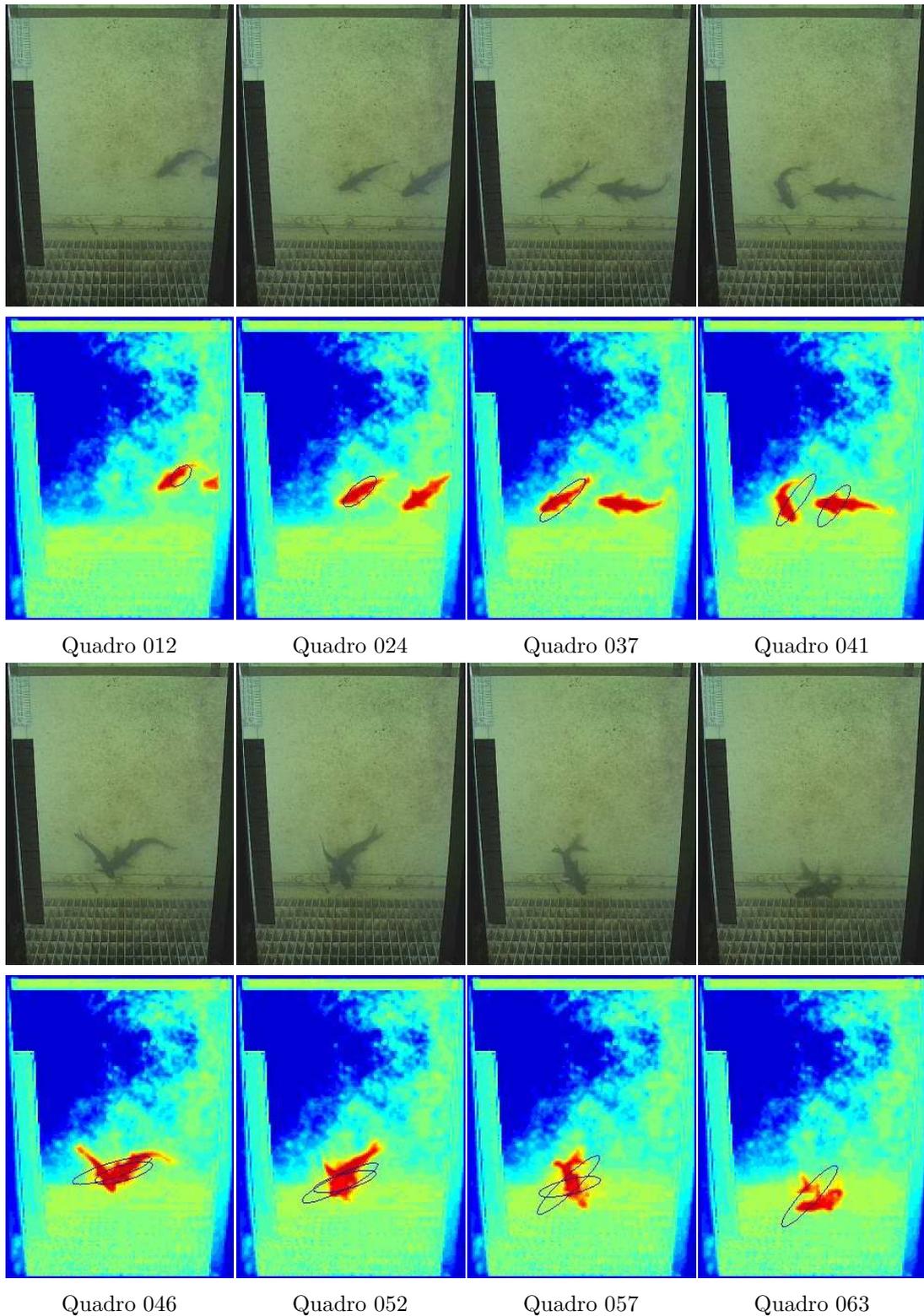


Figura 4.11: Terceiro experimento C. Rastreamento de peixes nadando em um mecanismo de transposição Instalado na Usina Hidrelétrica de Igarapava. Em cada quadro é mostrada a imagem adquirida pela câmera e a imagem observada pelo sistema. A presença de uma elipse sobre o peixe na imagem observada pelo sistema, caracteriza o rastreamento daquele peixe.

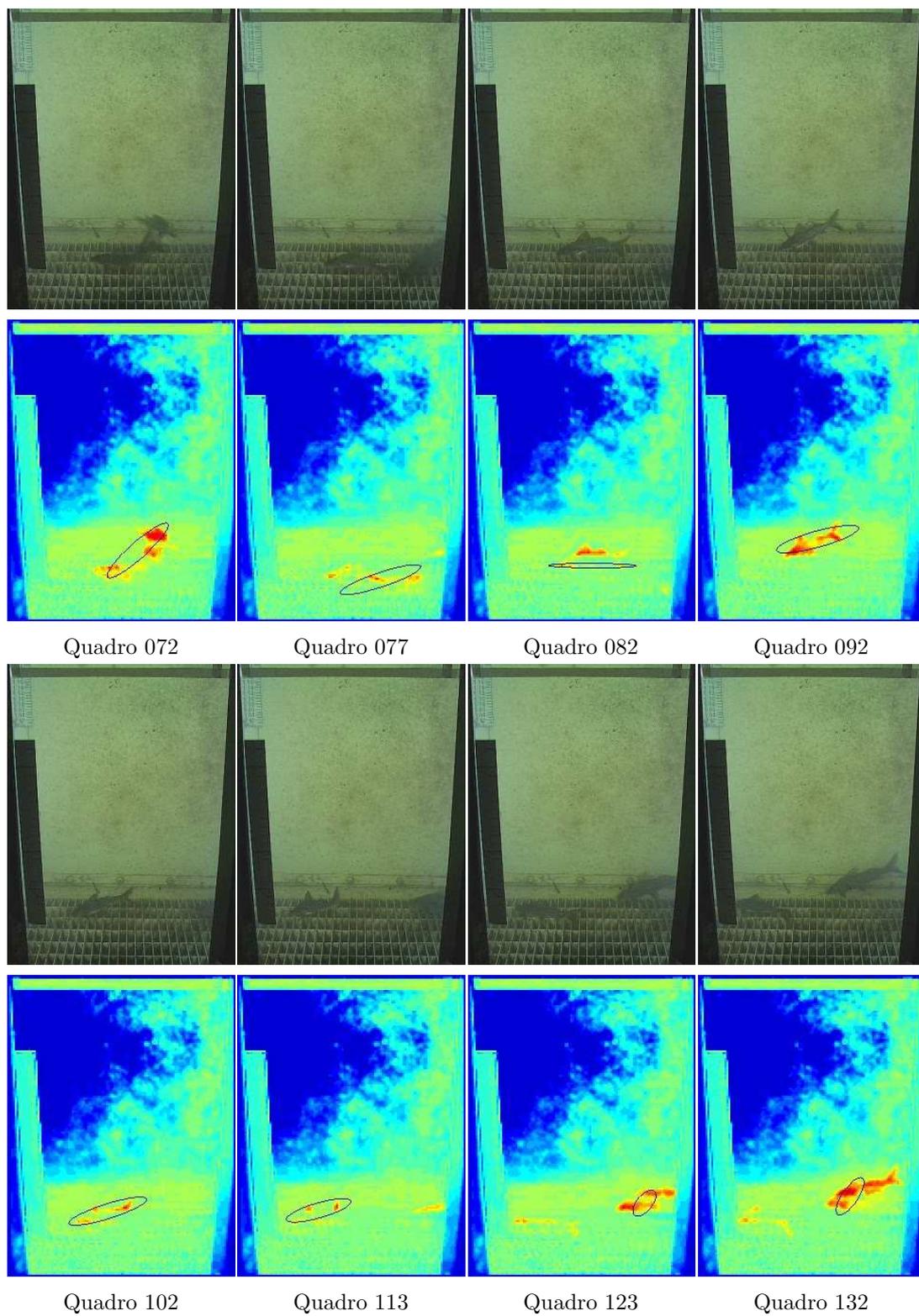


Figura 4.11: Terceiro experimento C. (Cont.)

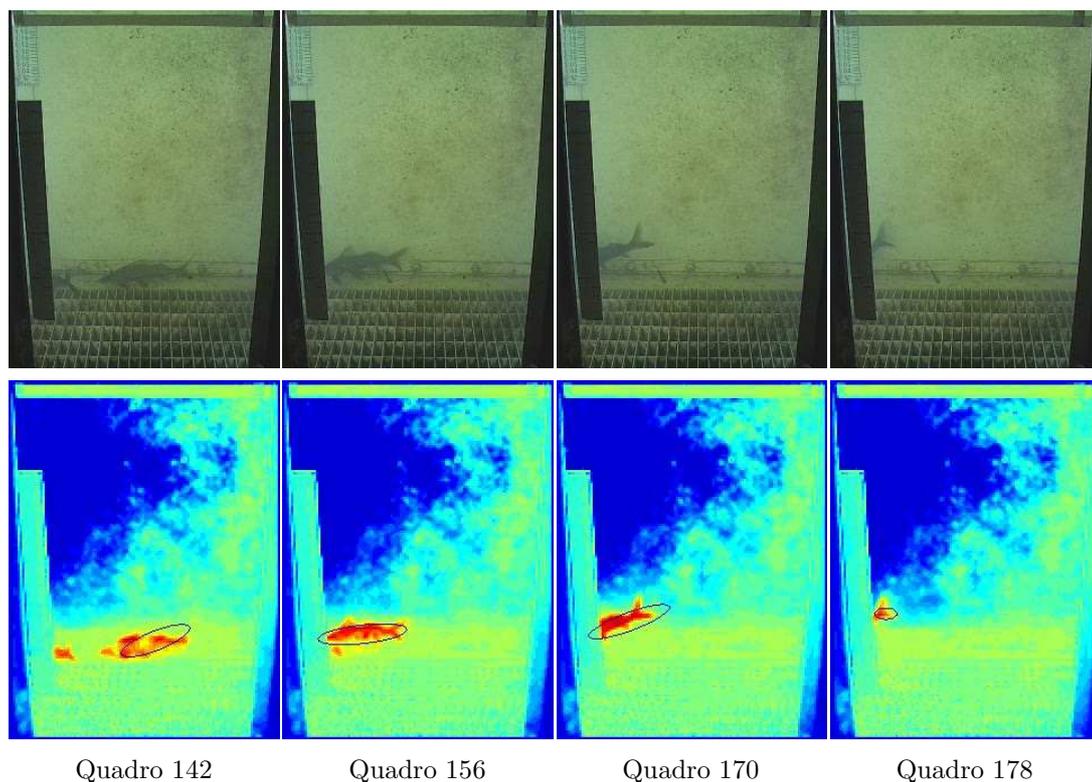
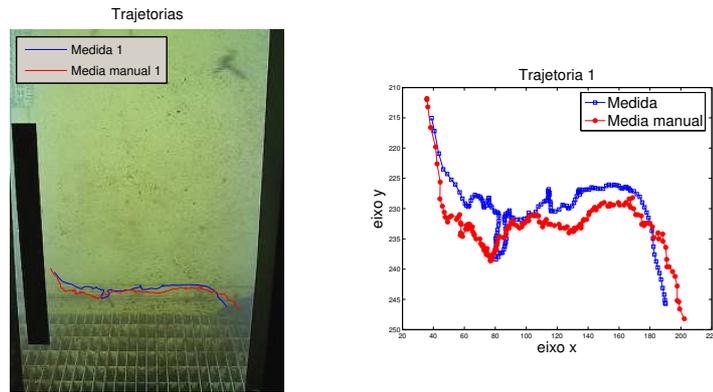
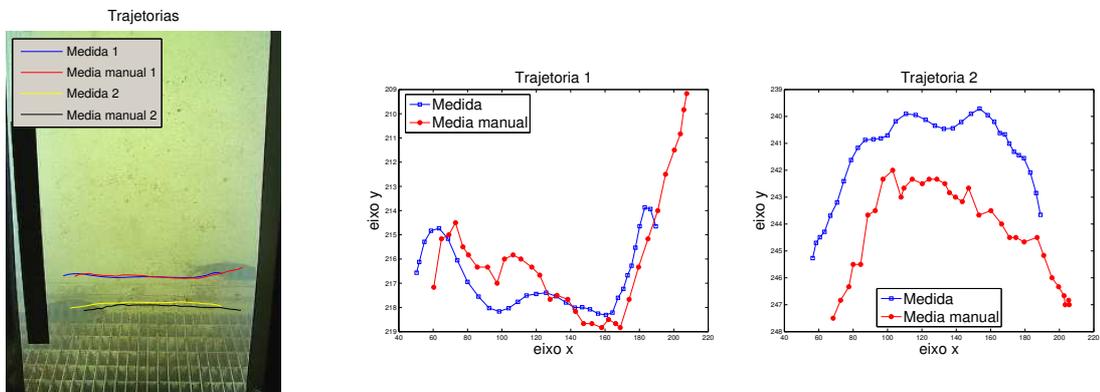


Figura 4.11: Terceiro experimento C. (Cont.)

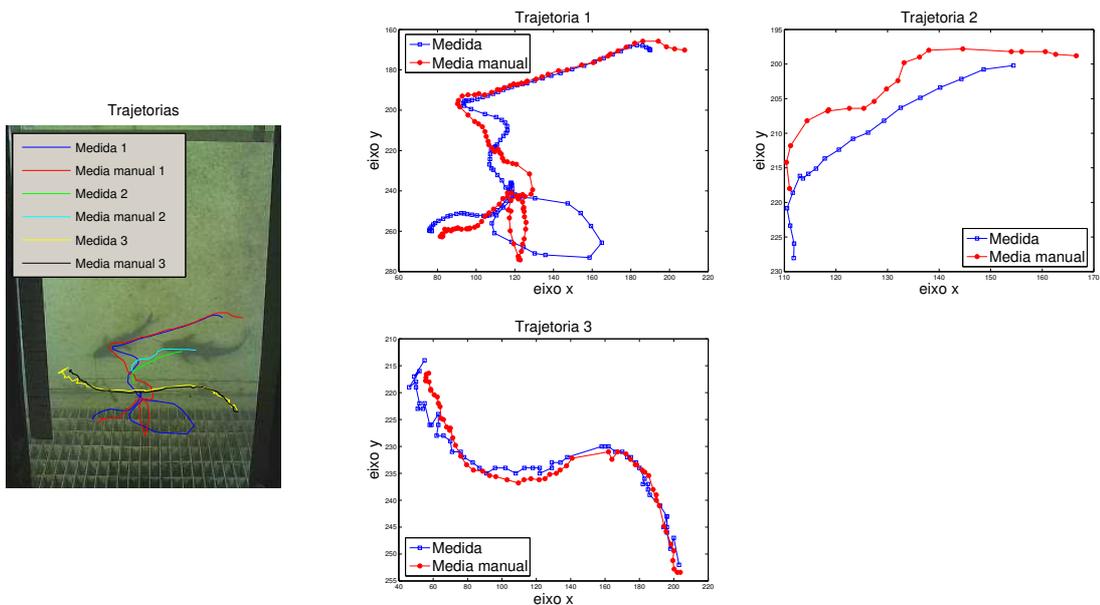
Em cada uma das seqüências analisadas o sistema encontrou um conjunto de trajetórias para os peixes rastreados. A Figura 4.12 mostra as trajetórias encontradas juntamente com as trajetórias medidas manualmente. A Tabela 4.4 mostra os erros encontrados para cada uma das seqüências analisadas e na Tabela 4.5 são encontrados os desvios.



(a) Primeira seqüência.



(b) Segunda Seqüência.



(c) Terceira Seqüência.

Figura 4.12: Trajetórias encontradas no experimento 3. As trajetórias encontradas pelo sistema são desenhadas juntamente com a média de trajetórias obtidas manualmente.

Tabela 4.4: Erros encontrados no experimento 3.

	Eixo X	Eixo Y
Seq. 1	6.8292	2.8807
Seq. 2	13.6881	2.0539
Seq. 3	6.3044	3.4888

Tabela 4.5: Desvios padrão encontrados no experimento 2.

	Eixo X	Eixo Y
Seq. 1	4.2060	1.4410
Seq. 2	5.2479	1.5303
Seq. 3	6.8186	2.9868

4.4.2 Discussão

Neste experimento também são encontradas as áreas de instabilidade. Neste caso elas ocorrem devido a uma grade presente na parte inferior da cena que possui pontos com cores parecidas com as cores dos peixes rastreados. A Figura 4.11 mostra alguns quadros da seqüência onde o peixe rastreado entra em uma área de instabilidade como por exemplo o quadro 102.

Essa seqüência de imagens utilizou a característica de instabilidade a favor do sistema. O peixe seria perdido pelo sistema quando passasse pela grade, já que o peso de cada objeto rastreado é calculado pelo somatório dos pontos envolvidos pela elipse correspondente (Seção 3.1.3).

A estratégia foi acrescentar um pequeno múltiplo da matriz identidade à matriz de covariâncias da *fdp* referente aos objetos de interesse. O incremento das variâncias causa alargamento nas bases das funções Gaussianas que compõem a *fdp* do plano dos objetos. Com isso, os valores próximos daqueles que representam o plano dos objetos e que possuíam probabilidade baixa de serem considerados como parte de um objeto, passam a ter probabilidade de ser objeto de interesse um pouco maior, se aproximando dos valores obtidos pela *fdp* do plano de fundo.

Dessa forma, quando o valor z de cada *pixel* é calculado, as probabilidades de fundo e objeto estarão mais próximas, resultando em valores de z que tendem a zero. Com isso, os pontos de fundo causarão menor interferência no peso do objeto, permitindo que os pontos de valores altos sobressaiam e mantenham o objeto sendo rastreado.

Sem essa estratégia, o peixe do quadro 102 da Figura 4.11 não estaria sendo rastreado naquele ponto.

As trajetórias medidas comparadas com a média das trajetórias obtidas manualmente, foram exibidas na Figuras 4.12. As trajetórias referentes à última seqüência, mostram interferência na trajetória do primeiro peixe. Quando o segundo peixe surge na cena e se aproxima do primeiro peixe, ocorre um deslocamento da primeira trajetória na direção da segunda. O deslocamento ocorre novamente quando a medição do mesmo peixe é novamente influenciada por pontos referentes ao segundo peixe.

Mesmo já não sendo mais rastreado, o segundo peixe possui alguns pontos observados com valores positivos que influenciam a medição do primeiro. No passo de predição das partículas o sistema acrescenta uma parcela de movimento aleatória fazendo com que os pontos do outro peixe sejam considerados na medição de algumas partículas. Dessa forma os pontos do segundo peixe são incluídos na medição do primeiro, causando os desvios de trajetória.

Mesmo com os desvios causados por deficiência da segmentação e da observação, o sistema se manteve com erros em níveis razoáveis. A Tabela 4.4 mostra os erros médios quadráticos encontrados nos eixos x e y . Na terceira seqüência, o erro no eixo x se aproxima de nove *pixels* e no eixo y fica próximo de três.

4.5 Contagem automática

A contagem automática dos peixes foi parte do experimento mostrado na Seção 4.4. Para testar o processo de contagem, 15 vídeos adquiridos da janela de observação em Igarapava foram processados. 17 peixes foram contados automaticamente de um total de 21, o que resulta em 80,95% de acerto.

Em nenhum dos casos houve a ocorrência de falso positivo, ou seja, o sistema não contou o mesmo peixe mais de uma vez ou considerou como peixe algum outro objeto que por ventura estivesse presente na cena. Esses dados são exibidos em forma de gráfico na Figura 4.14.

As ocorrências de falso negativos se deram em três vídeos. Todas as três situações tiveram a influências da grade ao fundo da cena. A Figura 4.13 mostra quadros de uma das seqüências que tiveram falso negativo. A influência da grade na seguimentação dos peixes está bem representada na figura.

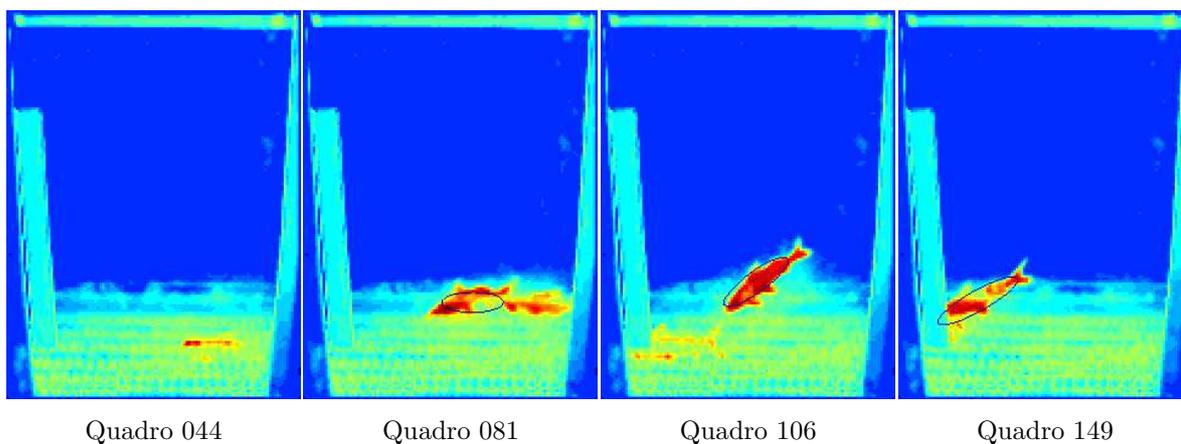
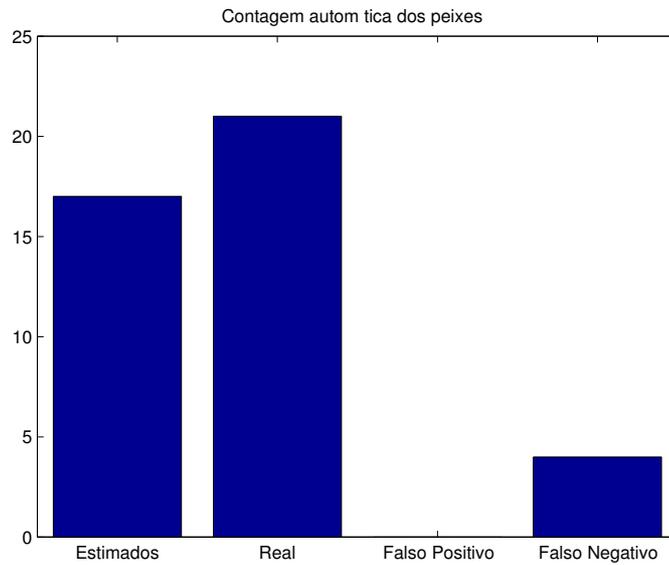


Figura 4.13: Exemplo de falso negativo. Na figura percebemos a presença de dois peixes, mas o sistema detectou apenas um

Outro fator que causou o falso negativo em uma das seqüências, foi o surgimento dos peixes com oclusão parcial. Nesse caso dois peixes entram na cena sendo parcialmente ocluído pelo primeiro. Esse fato fez com que o sistema considerasse os dois peixes como apenas um.

Nos experimentos mostrados na Seção 4.4 foi percebida a influência causada pela proximidade de dois peixes. Aqui os pontos relativos ao segundo peixe foram considerados como parte do primeiro, já que não havia uma representação para o segundo peixe, parcialmente ocluído.



	Estimado	Real	Falso Positivo	Falso Negativo
Totais	17	21	0	4

Figura 4.14: Resultado da contagem automática. Foram processados 15 vídeos onde haviam um total de 21 peixes. O sistema de contagem totalizou 17 peixes.

Neste capítulo foram apresentados experimentos realizados para a validação da metodologia proposta. O rastreamento foi testado com objetos diversos sendo que as trajetórias estimadas foram comparadas qualitativamente a trajetórias obtidas manualmente. Foram obtidos erros médios abaixo de dois *pixels* rastreando múltiplos objetos simultâneos, em ambientes bem controlados. Em situações em que o objeto rastreado se confundia com o fundo, o erro médio quadrático chegou próximo de 12 *pixels*.

Capítulo 5

Conclusões e direções futuras

A proposta inicial deste trabalho foi a implementação de um sistema de contagem automática de peixes que nadam através de um mecanismo de transposição usando imagens sucessivas adquiridas por uma câmera de vídeo. As múltiplas ocorrências do mesmo objeto de interesse em várias imagens sucessivas traz a necessidade de se aplicar um sistema de rastreamento para relacionar as ocorrências do mesmo objeto a um rótulo que o identifica.

A metodologia adotada foi inspirada no trabalho de Isard e MacCormick [Isard and MacCormick, 2001] dedicado ao rastreamento de pessoas em ambiente controlado. Assim como as pessoas os peixes são objetos que se movem independentemente, por esse motivo existe a necessidade de um rastreador capaz de tratar múltiplos *blobs* independentes.

Os objetos de interesse foram representados como elipses no plano das imagens. Cada elipse envolve os pontos que serão considerados na medição daquele objeto. A medição do objeto utilizou um modelo de observação que conseguiu separar o plano de fundo do plano dos objetos. Em áreas do plano de fundo em que existem pontos com cores semelhantes às cores dos pontos correspondentes a objetos de interesse, surgem áreas de instabilidade que podem causar a perda do objeto de interesse.

Durante a execução da contagem automática, o sistema processou 15 seqüências de imagens adquiridas da janela de observação instalada na usina Hidrelétrica de Igarapava. A grade existente na parte inferior da janela, possui pontos que causaram instabilidade no resultado do modelo de observação. Essa instabilidade é gerada quando as probabilidades do plano de fundo e do plano do objeto possuem valores próximos, resultando em valores que tendem a zero e sensibilizam a medição do sistema, quando existem outros objetos na cena cuja segmentação não possui esse problema.

Os testes de contagem realizados tiveram 80,95% de acerto sendo que os falso positivos encontrados foram conseqüência da instabilidade gerada pela grade na parte inferior da cena.

Nas situações em que as cores do fundo da cena e dos objetos de interesse se contrastam, o sistema consegue segmentar e rastrear os objetos de interesse sem problemas. O primeiro experimento e as partes do terceiro experimento em que o peixe rastreado nada com a parede ao fundo são exemplos dessa situação.

Por outro lado, as situações em que o peixe nada por regiões onde as cores do fundo são semelhantes, como nas posições que coincidem com a grade da cena, a segmentação deficiente prejudica o rastreamento no tocante à medição do sistema. Os objetos medidos sob essas condições possuem pesos que são insuficientes para se manter no conjunto de partículas juntamente com outros objetos, resultando na perda do objeto mal segmentado.

5.1 Direções futuras

A continuidade deste trabalho visa principalmente uma implementação capaz de rastrear e conseqüentemente contar objetos em tempo real. A versão em MATLAB utilizada não tinha como objetivo executar as rotinas analisando os quadros de uma seqüência vinda diretamente de uma filmadora digital.

Uma implementação em tempo real acoplada a um sistema de classificação que trabalhasse em tempo compatível, possibilitaria totalizações instantâneas que poderiam ser disponibilizadas de diversas formas. Por exemplo, uma página WEB disponibilizando relatórios instantâneos, ou alimentar um banco de dados que registre a passagem de indivíduos para estudos futuros, dentre outras utilidades.

A precisão do sistema poderia ser melhorada trabalhando-se com trajetórias tridimensionais. A instância adotada se restringe ao rastreamento dos peixes no plano da imagem. Essa restrição foi imposta ao sistema pelo fato do não conhecimento das dimensões dos objetos de interesse. Os peixes que utilizam o mecanismo de transposição não possuem dimensões fixas e portanto impedem a inferência de profundidade por meio de uma câmera calibrada.

A utilização de um par estéreo permitiria obter a distância em que os objetos estão do conjunto de aquisição. A inferência de profundidade em um sistema de rastreamento de múltiplos objetos independentes que se movem no espaço como os peixes, constitui um dado importante para o tratamento de possíveis oclusões.

A técnica na qual esse trabalho foi inspirada consegue rastrear seus objetos de interesse no espaço 3- D por considerar que todos possuem dimensões aproximadas. Se uma pessoa com estatura fora dos padrões entrar na cena, o sistema não dará a trajetória correta ou a pessoa poderá ser desconsiderada.

Como foi implementada, a metodologia utiliza um filtro para rastrear todos os objetos de uma só vez. Isso faz com que a confiança de um objeto seja afetada quando for inserido um novo objeto no conjunto.

Com um sistema capaz de gerenciar um filtro para cada objeto rastreado e fundir as *fdp*'s [Okuma et al., 2004a], a precisão e confiança no rastreamento poderia ser ainda melhor. A utilização de múltiplos filtros altera o modo de gerenciamento de entrada e saída de objetos. Um rastreador desse tipo precisa conseguir decidir a ordem de profundidade dos objetos para não considerar o mesmo *pixels* em duas *fdp*'s, precisa decidir quando um novo objeto entrou na cena e iniciar um novo filtro para ele, além de saber quando um objeto foi perdido para eliminar o seu filtro.

Referências Bibliográficas

- [Bar-Shalom, 1992] Bar-Shalom, Y. (1992). *Multitarget-Multisensor Tracking: Applications and Advances*, volume III. Artech House.
- [Braithwaite, 1974] Braithwaite, H. (1974). Sonar fish counting. In *European Inland Fisheries Advisory Commission Symp.*
- [Branson et al., 2003] Branson, K., Rabaud, V., and Belongie, S. (2003). Tree brown mice: See how they run. In *In Proc. IEEE International Workshop on VS-PETS*, pages 78 – 85, Nice, France.
- [Cadieux et al., 2000] Cadieux, S., Michaud, F., and Lalonde, F. (2000). Intelligent system for automated fish sorting and counting. volume 2, pages 1279–1284.
- [Forsyth and Ponce, 2002] Forsyth, D. A. and Ponce, J. (2002). *Computer Vision: A Modern Approach*.
- [Gevarter, 1984] Gevarter, W. B. (1984). Robotics and artificial intelligence applications series. In *Computer Vision: An Overview*, volume 3. Business/Technology Books.
- [Goldenstein, 2004] Goldenstein, S. (2004). A gentle introduction to predictive filters. *Revista de Informatica Teórica e Aplicada (RITA)*, xi(1):61 – 89.
- [Gordon et al., 1993] Gordon, N., Salmond, D., and Smith., A. (1993). A novel approach to nonlinear/non-gaussian baesian state stimation. In *IEE Proceedings on RADAR and signal processing*, pages 107–133.
- [Grewal and Andrews, 1993] Grewal, M. S. and Andrews, A. P. (1993). *Kalman Filtering and prediction problems*. Prentice Hall.
- [Heikkilä and Silvén, 2004] Heikkilä, J. and Silvén, O. (2004). A real-time system for monitoring of cyclists and pedestrians. *Image and Vision Computing*, 22:563 – 570.
- [Isard, 1998a] Isard, M. (1998a). *Visual Motion Analysis Probabilistic Propagation of Conditional Density*. PhD thesis, Oxford University.
- [Isard and Blake, 1996] Isard, M. and Blake, A. (1996). Contour tracking by stochastic propagation of conditional density. In *Proc. IEEE European Conference on Computer Vision*, volume 1, pages 343 – 356, Cambridge UK.

- [Isard and Blake, 1998] Isard, M. and Blake, A. (1998). Condensation – conditional density propagation for visual tracking. In *International Journal of Computer Vision*, 29(1), pages 5–28.
- [Isard and MacCormick, 2001] Isard, M. and MacCormick, J. (2001). Bramble: A bayesian multiple-blob tracker. In *Proc. IEEE International Conference on Computer Vision*, vol 2, pages 34–41.
- [Isard, 1998b] Isard, M. A. (1998b). The condensation algorithm.
- [Kalman, 1960] Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. In *Transactions of ASME-Journal of Basic Engineering*, 82(D), pages 35–45.
- [Khan et al., 2004] Khan, Z., Balch, T., and Dellaert, F. (2004). An mcmc-based particle filter for tracking multiple interacting targets. In *Proc. IEEE European Conference on Computer Vision (ECCV04)*.
- [Leon-Garcia, 1994] Leon-Garcia, A. (1994). *Probability and Random Processes for Electrical Engineering*. Addison-Wesley Publishing Company, Inc, second edition.
- [Li et al., 2003] Li, P., Zhang, T., and Pece, A. E. C. (2003). Visual contour tracking based on particle filters. *Image and Vision Computing*, 21:111 – 123.
- [MacCormick and Blake, 2000] MacCormick, J. and Blake, A. (2000). A probabilistic exclusion principle for tracking multiple objects. In *International Journal of Computer Vision*, 39(1), pages 57–71.
- [Martins, 2000] Martins, S. L. (2000). *Sistemas para a transposição de peixes*. PhD thesis, Escola Politécnica da Universidade de São Paulo.
- [Meier and Ade, 1999] Meier, E. B. and Ade, F. (1999). Tracking cars in range images using the condensation algorithm. In *Proc. IEEE/IEEEJ/JSAI International Conference on Intelligent Transportation Systems*, pages 129 – 134.
- [Menin and Paulus, 1974] Menin, A. and Paulus, R. D. (1974). Fish counting by acoustic means. In *IEEE Oceans*, volume 6, pages 166 – 168.
- [Menin, 1984] Menin, A. S. (1984). Salmon counter with separate king salmon tabulator. United States Patent.
- [Mills et al., 2003] Mills, S., Pridmore, T., and Hills, M. (2003). Tracking in a hough space with the extended kalman filter. In *Proc. British Machine Vision Conference*.
- [Nery, 2004] Nery, M. S. (2004). Classificação de espécies de peixes da bacia do rio grande baseada em visão computacional. Master’s thesis, Federal University of Minas Gerais, Computer Science Department, Belo Horizonte, MG, Brasil.

- [Nummiaro et al., 2003] Nummiaro, K., Koller-Meier, E., and Goll, L. V. (2003). An adaptative color-based particle filter. *Image and Vision Computing*, 21:99 – 110.
- [Okuma et al., 2004a] Okuma, K., Taleghani, A., de Freitas, N., Little, J. J., and Lowe, D. G. (2004a). A boosted particle filter: Multitarget detection and tracking. *Proc. IEEE European Conference on Computer Vision (ECCV04)*.
- [Okuma et al., 2004b] Okuma, K., Taleghani, A., de Freitas, N., Little, J. J., and Lowe, D. G. (2004b). A bosted particle filtre: Multitarget detection and tracking. In *In Proc. European Conference on Computer Vision (ECCV)*.
- [Paetzold and Franke, 2000] Paetzold, F. and Franke, U. (2000). Road recognition in urban enviroment. *Image and Vision Computing*, 18.
- [Pereira, 1999] Pereira, G. A. S. (1999). Filtro de kalman: Teoria e aplicações. <http://www.dcc.ufmg.br/~gpereira/kalman/>.
- [Pincock and Easton, 1978] Pincock, D. G. and Easton, N. W. (1978). The feasibility of doppler sonar fish counting. *Proc. IEEE Journal of Oceanic Engineering*, OE–3(2):37 – 49.
- [Pádua, 2002] Pádua, F. L. C. (2002). Real-time pedestrian counting based on computer vision. Master’s thesis, Federal Uninversity of Minas Gerais, Computer Science Department, Belo Horizonte, MG, Brasil. no. 401.
- [Queiroz-Neto et al., 2004] Queiroz-Neto, J. P., Carceroni, R. L., Barros, W. F., and Campos, M. F. M. (2004). Underwater stereo. In *SIBGRAPI Conference Proceedings, IEEE Computer Society Press*, pages 170–177.
- [Rossi and Bopzzoli, 1994] Rossi, M. and Bopzzoli, A. (1994). Tracking and counting moving people. In *IEEE Internertional Conference on Image Processing*, volume 3, pages 212 – 216.
- [Russell and Norvig, 2004] Russell, S. and Norvig, P. (2004). *Inteligência Artificial*. Campus.
- [Smith-Root, 2005] Smith-Root, I. (2005). *Fish counting systems*.
- [Sullivan et al., 1999] Sullivan, J., Blake, A., Isard, M., and MacCormick, J. (1999). Object localization by bayesian correlation. In *7th Int. Conf. on Computer Vision*, volume 2, pages 1068 – 1075.
- [Takagi et al., 1997] Takagi, Y., Kato, M., Miyakita, T., Terazono, K., Jikan, S., Tamezawa, T., and Hara, T. (1997). Method and aparatus for counting underwater objects using an ultrasonic wave. United Estaes Patent.
- [Terazono et al., 2000] Terazono, K., Matsuo, I., Matsumoto, Y., and Murakami, O. (2000). Device for count fish population passing through a fish pass. European Pattent Office.

-
- [Trucco and Verri, 1998] Trucco, E. and Verri, A. (1998). *Introductory Techniques for 3-D Computer Vision*. Prentice Hall.
- [White, 1976] White, B. (1976). A transportable acoustic fish census system for lake and river studies. In *Proc. IEEE Oceans*, volume 8.