



WAGNER MACHADO DO AMARAL

RASTREAMENTO DE PESSOAS EM VÍDEO

CAMPINAS
2013



UNIVERSIDADE ESTADUAL DE CAMPINAS
FACULDADE DE ENGENHARIA ELÉTRICA E DE COMPUTAÇÃO

WAGNER MACHADO DO AMARAL

RASTREAMENTO DE PESSOAS EM VÍDEO

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Engenharia Elétrica da Faculdade de Engenharia Elétrica e de Computação da Universidade Estadual de Campinas para obtenção do título de Mestre em Engenharia Elétrica, na área de Engenharia de Computação.

Orientador: Prof. Dr. Clésio Luis Tozzi

ESTE EXEMPLAR CORRESPONDE À VERSÃO FINAL DA TESE
DEFENDIDA PELO ALUNO WAGNER MACHADO DO AMARAL
E ORIENTADA PELO PROF. DR. CLÉSIO LUIS TOZZI

CAMPINAS
2013

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca da Área de Engenharia e Arquitetura
Rose Meire da Silva - CRB 8/5974

Am13r Amaral, Wagner Machado do, 1986-
Rastreamento de pessoas em vídeo / Wagner Machado do Amaral. –
Campinas, SP : [s.n.], 2013.

Orientador: Clésio Luis Tozzi.
Dissertação (mestrado) – Universidade Estadual de Campinas, Faculdade de
Engenharia Elétrica e de Computação.

1. Rastreamento (posição). 2. Processamento de imagens. I. Tozzi, Clésio
Luis, 1948-. II. Universidade Estadual de Campinas. Faculdade de Engenharia
Elétrica e de Computação. III. Título.

Informações para Biblioteca Digital

Título em inglês: Tracking people in video

Palavras-chave em inglês:

People's Tracking

Image processing

Área de concentração: Engenharia de Computação

Titulação: Mestre em Engenharia Elétrica

Banca examinadora:

Clésio Luis Tozzi [Orientador]

José Eduardo Cogo Castanho

Léo Pini Magalhães

Data de defesa: 24-05-2013

Programa de Pós-Graduação: Engenharia Elétrica

COMISSÃO JULGADORA - TESE DE MESTRADO

Candidato: Wagner Machado do Amaral

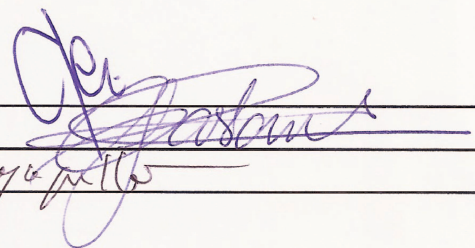
Data da Defesa: 24 de maio de 2013

Título da Tese: "Rastreamento de Pessoas em Vídeo"

Prof. Dr. Clésio Luis Tozzi (Presidente):

Prof. Dr. José Eduardo Cogo Castanho:

Prof. Dr. Léo Pini Magalhães:

Three handwritten signatures in blue ink are written over three horizontal lines. The first signature is for Prof. Dr. Clésio Luis Tozzi, the second for Prof. Dr. José Eduardo Cogo Castanho, and the third for Prof. Dr. Léo Pini Magalhães. The signatures are stylized and overlapping.

Resumo

Palavras-chave: Processamento de imagem, rastreamento de pessoas, detecção de cabeça.

O barateamento das câmeras de vídeo e o crescente poder de processamento dos computadores têm tornado a vigilância por vídeo um meio acessível e eficaz para monitorar espaços públicos e privados. Nesse contexto, o rastreamento de pessoas por vídeo tem motivado inúmeros trabalhos. O objetivo de um sistema de rastreamento de pessoas é determinar a trajetória das pessoas presentes no vídeo com base na variação de suas posições em cada quadro. Embora as soluções proposta na literatura para resolver o problema do rastreamento apresentem constante evolução, obter as trajetórias das pessoas ainda é um desafio no campo da visão computacional. O presente trabalho aborda aspectos do rastreamento de pessoas em vídeos e propõe um sistema de rastreamento baseado em abordagens descritas na literatura. O sistema proposto nesse trabalho é dividido em duas etapas principais: Detecção do *foreground* e Cálculo das Trajetórias. A etapa de detecção do *foreground* consiste na detecção das regiões na imagem correspondentes aos alvos a serem seguidos e é dividida em três partes: subtração de fundo, remoção de sombras e remoção de ruídos por morfologia matemática. A etapa do cálculo das trajetórias é dividida em duas partes: Associação e Gerenciamento das Pessoas Seguidas. A associação corresponde ao processamento realizado em cada quadro, onde as regiões do *foreground* são associadas às pessoas seguidas com base em métricas relacionadas ao centroide, *bounding box* e histograma de cor. Durante oclusões um método de detecção de cabeças é utilizado. O gerenciamento das pessoas seguidas corresponde a um procedimento mais amplo, no qual analisa-se o deslocamento de todas as pessoas presentes na cena. O sistema proposto no presente trabalho é aplicado a vídeos de bases públicas. Para cada vídeo testado, informações relacionadas ao número de pessoas no vídeo e suas posições são obtidas manualmente e comparadas aos resultados obtidos através do sistema proposto.

Abstract

Keywords: Image processing, people tracking, head detection.

The increasing of computer power has turned video surveillance into an effective and affordable means to monitor public and private areas. In this context, tracking systems is motivating researchers around the world. The main goal of a tracking system is to determine the location or direction of a target in each video frame. Although the proposed solutions for tracking problems have been improved continuously, people path detection is still a challenge in computer vision field. This work discusses the main aspects of people tracking in video and proposes an implementation based on approaches described in the literature. The system is divided into two main steps: foreground detection and trajectory calculation. The foreground detection step consists of three parts: background subtraction, shadow suppression and noise removal by morphological image processing. The trajectory calculation step is divided into two parts: matching and trajectory management. The matching is performed in each video frame to associate the foreground image regions and the tracked elements. In the implemented system matching metrics related to the centroid position, bounding box and color histogram are used. In case of occlusion a head detection method is employed. In the trajectory management procedure the foreground objects are analyzed and recorded. Public datasets were used to evaluate the system efficiency.

Dedico este trabalho a minha família.

Agradecimentos

Agradeço ao Prof. Dr. Clésio Luis Tozzi pela orientação.

Agradeço à minha irmã, Profa. Dra. Wanessa Machado do Amaral, pelas críticas e sugestões.

Agradeço aos meus pais, Jose Luis Machado do Amaral e Regina Amália Machado do Amaral, pelo apoio e incentivo.

Agradeço à minha noiva, Michelli Regina Pohl, pela compreensão e companheirismo.

Agradeço a Deus por todas as minhas conquistas.

Sumário

Lista de Figuras.....	xvii
Lista de Tabelas.....	ixx
Notação.....	xxi
Trabalhos relacionados publicados pelo autor	xxiii
1 Introdução.....	1
1.1 Motivação.....	1
1.2 Objetivo.....	3
1.3 Organização do Trabalho.....	3
2 Sistemas de Rastreamento.....	5
2.1 Revisão Bibliográfica.....	5
2.2 Sistema de Rastreamento Proposto.....	12
3 Detecção do Foreground.....	15
3.1 Subtração de Fundo.....	17
3.1.1 Subtração de Fundo Utilizada no Sistema Proposto.....	20
3.1.2 Avaliação do Método de Subtração de Fundo.....	23
3.2 Remoção de Sombras.....	36
3.2.1 Método de Remoção de Sombra Utilizado no Sistema Proposto.....	38
3.2.2 Avaliação do Método de Remoção de Sombra.....	39
3.3 Remoção de Ruídos por Morfologia Matemática.....	43
3.3.1 Operações Morfológicas.....	44
3.3.2 Operações Morfológicas Utilizadas no Sistema Proposto.....	50
3.3.3 Avaliação do Processamento Morfológico.....	54
4 Extração de Informações.....	61
4.1 Posição.....	61
4.2 Bounding Box.....	62
4.3 Histograma de Cor.....	63
4.4 Detecção de Cabeças.....	64
4.4.1 Método de Detecção de Cabeças Utilizado no Sistema Proposto.....	67
4.4.2 Avaliação do Método de Detecção de Cabeças.....	70
5 Determinação das Trajetórias.....	81
5.1 Associação.....	81
5.1.1 Previsão do Deslocamento por Filtro de Kalman.....	81
5.1.2 Métrica de Associação.....	86
5.1.3 Tratamento de Oclusões.....	88
5.2 Classificação de Estados.....	93
5.2.1 Classe Iniciante.....	95
5.2.2 Classe Madura.....	96
5.2.3 Classe Indisponível.....	96
5.2.4 Classe Desaparecida.....	97

5.2.5 Classe Fora de Cena.....	97
6 Avaliação do Sistema Proposto.....	99
6.1 Vídeo 1.....	101
6.2 Vídeo 2.....	103
6.3 Vídeo 3.....	106
6.4 Vídeo 4.....	108
7 Conclusão.....	115
Referências bibliográficas.....	117

Lista de Figuras

Figura 1: Etapas do sistema de rastreamento proposto.....	14
Figura 2: Vídeos utilizados para testar o método de detecção do foreground.....	25
Figura 3: Vídeo 1 – Revocação x Precisão para a variação do limiar k.....	28
Figura 4: Vídeo 1 – Resultado da subtração de fundo para a variação do limiar k.....	29
Figura 5: Vídeo 2 – Revocação x Precisão para a variação do limiar k.....	30
Figura 6: Vídeo 2 – Resultado da subtração de fundo para a variação do limiar k.....	31
Figura 7: Vídeo 3 – Revocação x Precisão para a variação do limiar k.....	32
Figura 8: Vídeo 3 – Resultado da subtração de fundo para a variação do limiar k.....	33
Figura 9: Vídeo 4 – Revocação x Precisão para a variação do limiar k.....	34
Figura 10: Vídeo 4 – Resultado da subtração de fundo para a variação do limiar k.....	35
Figura 11: Vídeo 3 – Variação das taxas de revocação e precisão após a remoção de sombras.....	40
Figura 12: Vídeo 3 – Comparação entre o foreground antes e após a remoção de sombra.....	41
Figura 13: Vídeo 4 – Comparação entre o foreground antes e após a remoção de sombra.....	42
Figura 14: Vídeo 4 – Variação das taxas de revocação e precisão após a remoção de sombras.....	42
Figura 15: Exemplo de ruídos no foreground.....	44
Figura 16: Reconstrução por marcadores.....	49
Figura 17: Operação de Fechamento.....	51
Figura 18: Operação de Abertura.....	52
Figura 19: Abertura por reconstrução.....	53
Figura 20: Comparação entre o foreground antes e após o processamento morfológico.....	53
Figura 21: Vídeo 3 – Revocação x Precisão para subtração de fundo, remoção de sombras e morfologia	55
Figura 22: Vídeo 3 – Foreground após subtração de fundo, remoção de sombras e morfologia.....	56
Figura 23: Vídeo 4 – Revocação x Precisão para subtração de fundo, remoção de sombras e morfologia	58
Figura 24: Vídeo 4 – Foreground após subtração de fundo, remoção de sombras e morfologia.....	59
Figura 25: Bounding box da região correspondente a uma pessoa.....	62
Figura 26: Modelo RGB.....	63
Figura 27: Histograma de cores de uma pessoa.....	64

Figura 28: Etapas do método de detecção de cabeças.....	67
Figura 29: Pessoa com os braços erguidos.....	69
Figura 30: Amostras de quadros dos vídeos utilizados para avaliar o método de detecção de cabeças. .	70
Figura 31: Resultado do método de detecção de cabeças.....	72
Figura 32: Resultado do método de detecção de cabeças durante união de blobs.....	73
Figura 33: Exemplo de falso negativo ocorrido durante a detecção de cabeças.....	74
Figura 34: Exemplo de falso positivo ocorrido durante a detecção de cabeças.....	75
Figura 35: Resultado do método de detecção de cabeças.....	77
Figura 36: Resultado do método de detecção de cabeças durante oclusão entre pessoas.....	78
Figura 37: Exemplo de falso negativo ocorrido durante a detecção de cabeças.....	79
Figura 38: Iteração do Filtro de Kalman.....	84
Figura 39: Ilustração do uso da bounding box durante a associação.....	91
Figura 40: Regras de transição entre as classes.....	94
Figura 41: Exemplos de regiões de entrada de pessoas.....	98
Figura 42: Vídeo 1 – Trajetórias determinadas com e sem o uso do método de detecção de cabeças. .	103
Figura 43: Vídeo 2 – Trajetórias determinadas com e sem o uso do método de detecção de cabeças. .	105
Figura 44: Vídeo 3 – Trajetórias determinadas com o uso do método de detecção de cabeças.....	107
Figura 45: Vídeo 4 – Trajetórias determinadas com e sem o uso do método de detecção de cabeças...	110
Figura 46: Vídeo 4 – Trajetórias determinadas pelo sistema durante uma oclusão.....	111
Figura 47: Vídeo 4 – Exemplo de negativo verdadeiro, onde o blob correspondente a uma bolsa é ignorado.....	112
Figura 48: Vídeo 4 – Exemplo de falso positivo ocasionado pelo reflexo de uma pessoa.....	113
Figura 49: Vídeo 4 – Exemplo de falso positivo ocasionado por uma porta em movimento.....	113
Figura 50: Vídeo 4 – Exemplo de falso negativo ocasionado pelo posicionamento das pessoas.....	114

Lista de Tabelas

Tabela 1: Sistemas de rastreamento.....	12
Tabela 2: Métodos de subtração de fundo.....	19
Tabela 3: Detalhes do vídeo 1.....	26
Tabela 4: Detalhes do vídeo 2.....	26
Tabela 5: Detalhes do vídeo 3.....	27
Tabela 6: Detalhes do vídeo 4.....	27
Tabela 7: Vídeo 1 – Revocação x Precisão para a variação do limiar k.....	28
Tabela 8: Vídeo 2 – Revocação x Precisão para a variação do limiar k.....	30
Tabela 9: Vídeo 3 – Revocação x Precisão para a variação do limiar k.....	32
Tabela 10: Vídeo 4 – Revocação x Precisão para a variação do limiar k.....	34
Tabela 11: Vídeo 3 – Revocação x Precisão para a subtração de fundo e remoção de sombras.....	39
Tabela 12: Vídeo 4 – Revocação x Precisão para a subtração de fundo e remoção de sombras.....	41
Tabela 13: Vídeo 3 – Revocação x Precisão para subtração de fundo, remoção de sombras e morfologia	54
Tabela 14: Vídeo 4 – Revocação x Precisão para subtração de fundo, remoção de sombras e morfologia	57
Tabela 15: Análise quantitativa da detecção de cabeças.....	71
Tabela 16: Análise quantitativa da detecção de cabeças.....	76
Tabela 17: Informações utilizadas pela métrica de associação.....	86
Tabela 18: Abordagens de resolução de oclusões.....	90
Tabela 19: Classes do sistema de gerenciamento de pessoas	94
Tabela 20: Taxas de avaliação utilizadas no presente trabalho.....	100
Tabela 21: Vídeos utilizados para avaliar o sistema proposto.....	101
Tabela 22: Vídeo 1 – Análise quantitativa dos resultados obtidos com e sem o uso do método de detecção de cabeças.....	102
Tabela 23: Vídeo 2 – Análise quantitativa dos resultados obtidos com e sem o uso do método de detecção de cabeças.....	104
Tabela 24: Vídeo 3 – Análise quantitativa dos resultados obtidos com e sem o uso do método de detecção de cabeças.....	106

Tabela 25: Vídeo 4 – Análise quantitativa dos resultados obtidos com e sem o uso do método de detecção de cabeças.....	108
--	-----

Notação

$I^t(p).H$ - Valor da camada H no modelo de cor HSV do pixel p na imagem no tempo t

$B^t(p).H$ - Valor da camada H no modelo de cor HSV do pixel p no modelo do fundo no tempo t

$A \oplus B$ - Dilatação de A por B

$A \ominus B$ - Erosão de A por B

$A \cdot B$ - Fechamento de A por B

$A \circ B$ - Abertura de A por B

$(A \oplus B)^n$ - N-dilatação

$(A \oplus_C B)$ - Dilatação de A por B condicional a C

$A \Delta_B M$ - Reconstrução de A pelo marcador M com conectividade B

$A \circ_E B$ - Abertura reconstitutiva de A por B com conectividade E

$|(x)|$ - Módulo de x

A^T - Transposta da Matriz A

A^{-1} - Inversa da Matriz A

$\hat{x}_k^-$ - Previsão do estado da pessoa pelo filtro de Kalman no tempo k

$\hat{x}_k$ - Correção da previsão do estado da pessoa pelo filtro de Kalman no tempo k

(x_C, y_C) - Coordenada do centroide do *blob*

$M_{p,k}^D$ - Métrica de associação relacionada à distância euclidiana

H_s - Histograma de cor do *blob*

$M_{p,k}^H$ - Métrica de associação relacionada ao histograma de cor

$M_{p,k}^{final}$ - Métrica de associação calculada entre a pessoa p e o centroide k

Trabalhos relacionados publicados pelo autor

Amaral, W. M. e Tozzi, C. L., “Tracking People Through Occlusions Using a Head Detection Method”, *Workshop de Visão Computacional (WVC’2013)*, Rio de Janeiro, Brasil, 2013

Amaral, W. M. e Tozzi, C. L., “Uso de Morfologia Matemática para Remoção de Ruídos da Subtração de Fundo”, *Workshop de Visão Computacional (WVC’2011)*, Curitiba, Brasil, 2011

Capítulo 1

Introdução

1.1 Motivação

O barateamento das câmeras de vídeo e dispositivos de armazenamento de dados e o poder de processamento dos computadores têm tornado a vigilância por vídeo um meio acessível e eficaz para monitorar espaços públicos e privados. Atualmente, sistemas de monitoramento por vídeo são comumente encontrados nas ruas, comércios, bancos, praças, aeroportos e demais lugares. Nesse contexto, o rastreamento de pessoas por vídeo tem motivado inúmeros trabalhos.

O objetivo de um sistema de rastreamento de pessoas é determinar a trajetória das pessoas presentes na cena com base na variação de suas posições em cada quadro do vídeo. Basicamente, um sistema de rastreamento é dividido em duas etapas:

- Detecção – Durante a etapa de detecção, usualmente são utilizadas técnicas de subtração de fundo. O resultado dessas técnicas consiste em um conjunto de imagens binárias que representam, para cada quadro do vídeo, os *pixels* correspondentes ao fundo da cena e os *pixels* correspondentes aos demais objetos presentes no vídeo.
- Determinação das trajetórias – Para determinar as trajetórias, são utilizadas métricas de associação relacionadas às informações obtidas a partir das regiões detectadas, como posição do

centroide, histograma de cor, forma e área da região.

Em situações ideais, rastrear uma pessoa é uma tarefa trivial. Entretanto, em vídeo obtidos em cenários reais alguns fatores dificultam os processos de detecção e determinação das trajetórias, e os principais são:

1. Variações de iluminação;
2. Ruídos nas imagens;
3. Objetos em movimento no fundo da cena irrelevantes no contexto do rastreamento de pessoas;
4. Surgimento e desaparecimento de pessoas;
5. Oclusão de pessoas por objetos da cena;
6. Oclusão entre pessoas.

Variações de iluminação e ruídos nas imagens podem dificultar a modelagem do fundo da cena, o que prejudica a detecção das regiões correspondentes às pessoas. Objetos em movimento no fundo da cena, como galhos de árvore ao vento, aumentam a dificuldade da detecção das pessoas.

Além das dificuldades inerentes à identificação dos *pixels* que diferem do fundo da cena, a movimentação e o posicionamento das pessoas em relação à câmera influenciam o processo de detecção. Ao longo do vídeo, pessoas podem entrar e sair do ângulo de visão da câmera, ou serem ocluídas por objetos ou por outras pessoas presentes na cena.

Quando uma pessoa é ocluída por um objeto disposto no ambiente, a região correspondente a essa pessoa não é detectada na imagem. Nessa situação, não é possível continuar o rastreamento da pessoa, a não ser que ela deixe o estado de oclusão e seja detectada novamente na imagem.

Quando uma pessoa é ocluída por outra pessoa presente na cena, as regiões correspondentes a essas pessoas são conectadas. Nessas ocasiões, abordagens de rastreamento baseadas apenas no

deslocamento das regiões detectadas durante a subtração de fundo não são capazes de identificar a posição das pessoas, pois a única região detectada corresponde às partes visíveis das pessoas oclusas. Técnicas adicionais podem ser utilizadas para tratar esses eventos, identificando as posições das pessoas envolvidas na oclusão.

Apesar dos avanços na área de rastreamento por vídeo, identificar os fatores que influenciam o rastreamento e propor técnicas capazes de tratar tais fatores em tempo real ainda é um desafio no campo da visão computacional, e demanda estudos aprofundados.

1.2 Objetivo

Basicamente, os objetivos do presente trabalho são:

1. Abordar as dificuldades inerentes ao rastreamento por vídeo, e rever bibliograficamente as principais técnicas de detecção e rastreamento presentes na literatura.
2. Adaptar e combinar técnicas existentes na literatura para implementar um sistema de rastreamento de pessoas e tratamento de oclusões por detecção de cabeças.
3. Avaliar a contribuição das técnicas empregadas em cada etapa do sistema implementado.

1.3 Organização do Trabalho

O presente trabalho é dividido em 7 capítulos. O capítulo 1 introduz o conceito de rastreamento de pessoas por vídeo, e descreve os objetivos do presente trabalho. O capítulo 2 apresenta uma revisão bibliográfica de sistemas de rastreamento e a estrutura do sistema de rastreamento de pessoas proposto no presente trabalho. O capítulo 3 aborda a etapa de detecção do *foreground*. O capítulo 4 apresenta técnicas de extração de informações a partir das regiões detectadas, como a extração do histograma de cor e a detecção da posição da cabeça das pessoas. O capítulo 5 apresenta a etapa de determinação das trajetórias. O capítulo 6 apresenta os resultados obtidos a partir do sistema proposto. Finalmente, o

capítulo 7 apresenta a conclusão e as considerações finais.

Capítulo 2

Sistemas de Rastreamento

2.1 *Revisão Bibliográfica*

Conforme Yilmaz et al (2006), o rastreamento de objetos é uma importante tarefa no campo da visão computacional, e tem ganho bastante interesse devido à proliferação de computadores com grande poder de processamento, à disponibilidade de câmeras de boa qualidade e baixo custo, e à crescente necessidade de análise automática de vídeos. De um modo geral, o uso de sistemas de rastreamento é pertinente às seguintes tarefas:

- Reconhecimento baseado em movimento: Detecção automática de objetos através da análise do movimento, como, por exemplo, a identificação de pessoas na cena através da análise da marcha;
- Segurança automatizada: Monitoramento de cenas para detectar atividades suspeitas ou eventos não desejados;
- Indexação de vídeos: Marcação automática e recuperação de trechos de interesse em vídeos de bases de dados multimídia;
- Monitoramento de tráfego: Coleta em tempo real de estatísticas para direcionar o fluxo de tráfego;
- Navegação: Planejamento de rotas baseado em vídeo, com a capacidade de evitar colisões.

Conforme (Yilmaz, 2007), o rastreamento de objetos por vídeo é definido basicamente como o processo de determinar as trajetórias de objetos em movimento computando o deslocamento dos mesmos em uma sequência de imagens, e a dificuldade relacionada a esse processo está associada a fatores como variações de luminosidade, ruídos, perda de informação causada pela projeção do mundo 3D em imagem 2D, complexidade do movimento do objeto, não rigidez ou natureza articulada do objeto e a necessidade de processamento em tempo real.

Sistemas elementares de rastreamento encontrados na literatura ignoram alguns dos fatores citados acima e são limitados a certas condições do ambiente. Entretanto, sistemas robustos utilizam técnicas específicas para lidar com tais fatores, e são aplicáveis a situações reais e apresentam resultados promissores. A seguir é apresentada uma revisão dos sistemas de rastreamento de pessoas e objetos presentes na literatura mais relevantes para o desenvolvimento do presente trabalho. Para cada sistema abordado, são descritas as técnicas utilizadas para detectar as pessoas ou objetos na imagem, determinar suas trajetórias e lidar com oclusões. A Tabela 1 mostra um resumo das técnicas apresentadas.

Seo et al (1997) propuseram um sistema para rastrear jogadores em uma partida de futebol. Nesse sistema é assumindo que o gramado corresponde a uma grande região na imagem, possui coloração uniforme e tonalidade verde. Sendo assim, um *threshold* é aplicado em cada camada RGB dos quadros do vídeo para identificar os *pixels* que possuem a coloração esperada para a grama, e o componente conexo resultante do *threshold* que possui maior área é considerado o gramado. As regiões contidas no gramado que não possuem a coloração esperada para a grama são os jogadores, e a partir dessas regiões o rastreamento é realizado. Cada jogador seguido é representado por um modelo constituído pela *bounding box* e pelo histograma das cores da região correspondente ao seu corpo na imagem. Durante o rastreamento, as trajetórias são determinadas através do uso do filtro de Kalman e do modelo das pessoas seguidas. Para determinar a posição de um jogador durante uma oclusão, os histogramas de cor do jogador e da região detectada na imagem são comparados através do algoritmo *Histogram Backprojection* (Swain e Ballard, 1991). Basicamente, através desse algoritmo o sistema consegue detectar na imagem a região que possui as cores contidas no modelo do jogador. Contudo, oclusões entre jogadores do mesmo time não são tratadas, pois as cores contidas em seus modelos são muito parecidas. Oclusões totais (quando um jogador fica totalmente ocluído por outro jogador) também não são tratadas pelo sistema.

Haritaoglu et al (2000) propuseram um sistema de segurança denominado W4, destinado ao monitoramento em tempo real das atividades de múltiplas pessoas em ambientes externos através de imagens monocromáticas capturadas por uma única câmera fixa. Nesse sistema, as regiões na imagem correspondentes às pessoas são detectadas através de uma técnica de subtração de fundo proposta pelo próprio autor, e a etapa de associação é realizada através de métricas relacionadas à posição do centroide, *bounding box* e silhueta da região detectada. Durante o rastreamento, quando duas ou mais pessoas se ocluem, W4 detecta a posição da cabeça das pessoas através do sistema denominado Hydra (Haritaoglu et al, 1999). Através da posição das cabeças, a posição de cada pessoa envolvida na oclusão é estimada e as trajetórias dessas pessoas são determinadas individualmente. Em casos de desaparecimento da pessoa seguida, causado por uma oclusão por um objeto da cena ou por uma falha no método de detecção do *foreground*, W4 mantém armazenadas as informações relacionadas à textura e forma da região correspondente à pessoa. Quando essa pessoa é detectada novamente na cena, ela é identificada através dessas informações e seu rastreamento é retomado.

Lu e Tan (2001) propuseram um sistema de rastreamento que utiliza a informação relacionada às cores das pessoas seguidas. Nesse sistema, as pessoas são detectadas através de uma técnica de subtração de fundo baseada em Wren et al (1997), onde a cor de cada *pixel* na imagem é modelada por uma distribuição gaussiana. Para evitar que as regiões correspondentes às sombras das pessoas sejam identificadas como *foreground* e prejudiquem o resultado do rastreamento, um método de detecção e remoção de sombras é aplicado ao resultado da subtração de fundo. Quando uma pessoa entra na cena, inicia-se seu rastreamento através da análise da variação do centroide, e calcula-se o histograma de cor dos *pixels* da região correspondente à pessoa para utilizações futuras em casos de oclusões. Quando duas ou mais pessoas se tocam ou se ocluem, de modo que as regiões correspondentes a seus corpos sejam unidas, as trajetórias dessas pessoas são determinadas conforme o deslocamento da região resultante da união. Quando a oclusão termina e as pessoas se separam, os histogramas de cor dessas pessoas são utilizados para identificar cada pessoa na imagem e retomar o rastreamento individualmente.

Zhao e Nevatia (2004) propuseram um sistema que realiza o rastreamento de múltiplas pessoas através do uso dos modelos de câmera, do fundo da cena e do que representa o corpo de uma pessoa. Para detectar as regiões na imagem correspondentes às pessoas em movimento, é utilizado o método de subtração de fundo proposto por Wren et al (1997), onde a cor de cada *pixel* na imagem é modelada por

uma distribuição gaussiana. Para que as sombras das pessoas não prejudiquem o processo de determinação das trajetórias, o sistema identifica e remove as regiões que correspondem às sombras das pessoas (assumindo que há uma única fonte de luz, e todas as sombras possuem a mesma direção). A detecção de pessoas na imagem é um processo recursivo: inicialmente são detectadas no *foreground* as posições de todas as cabeças; em seguida é posicionada abaixo de cada cabeça detectada uma elipse que representa o modelo de uma pessoa; finalmente, a região correspondente a cada pessoa é removida do *foreground*. Esse processo é repetido até que nenhuma cabeça seja detectada. Como as características de uma pessoa, como posição e altura, são modeladas através dos atributos de uma elipse, como centroide e os eixos da elipse, a melhor métrica de associação é aquela que minimiza a transformação de uma elipse para seu equivalente no quadro seguinte. Esse sistema consegue lidar adequadamente com oclusões parciais (desde que as cabeças das pessoas oclusas possam ser detectadas). No entanto, oclusões totais não são tratadas pelo sistema.

Lei e Xu (2006) propuseram um sistema de rastreamento de objetos e classificação de estados onde a detecção de objetos em movimento é realizada através do método de subtração de fundo MoG (Stauffer e Grimson, 1999), e as regiões correspondentes às sombras são eliminadas através de um método baseado em Horprasert et al (1999). As trajetórias são determinadas através de métricas de associação relacionadas à posição do centroide, área, cor e ao tamanho dos eixos da elipse contida na região detectada na imagem. Em casos de oclusões entre os objetos seguidos, as trajetórias dos mesmos são determinadas conforme o deslocamento da região resultante da união dos objetos. Para gerenciar o rastreamento de múltiplos objetos, é proposto um sistema de classificação de estados que descreve a situação dos objetos seguidos através do seguinte conjunto de classes: Aparecido, Maduro, Ocluso, Temporariamente Indisponível, Desaparecido, Reaparecido e Fora de Cena. Tais classes dão suporte ao tratamento de eventos como: oclusões, desaparecimento de objetos na cena e aparecimento de objetos na cena.

Freitas (2010) propôs um sistema de rastreamento de objetos baseado em Lei e Xu (2006). No sistema proposto por Freitas (2010), os objetos em movimento são detectados através do uso do método de subtração de fundo *Eigenbackground* (Oliver et al, 2000). As trajetórias são determinadas através do uso de métricas de associação relacionadas à posição do centroide, cor e área. Para gerenciar o rastreamento de múltiplos objetos, é proposto um sistema de classificação de estados que descreve a situação dos objetos seguidos através do seguinte conjunto de classes: Iniciante, Fundo, Fictício,

Temporário, Permanente, Separado, Partes, Ocluso, Indeterminado e Fora de Cena. Através desse conjunto de classes, o sistema é capaz de lidar com eventos como oclusões, desaparecimento de objetos e surgimento de objetos na cena. Contudo, a principal contribuição desse sistema está relacionada ao rastreamento de objetos que são segmentados durante a detecção. Nesse sistema, quando a região do *foreground* correspondente a um objeto seguido é segmentada, a trajetória do mesmo é determinada através do deslocamento das partes. Quando as partes se unem novamente, o rastreamento do objeto é retomado normalmente. No entanto, se as partes se separarem progressivamente, considera-se que houve uma separação de objetos que estavam oclusos, e cada parte é rastreada individualmente. Esse comportamento do sistema possibilita o rastreamento de um objeto, mesmo quando sua região no *foreground* é segmentada devido a uma oclusão parcial por objetos da cena ou devido a falhas durante a detecção.

Alam e Boufama (2011) propuseram um sistema para rastrear pessoas oclusas. Nesse sistema, o *foreground* é detectado pelo método de subtração de fundo proposto por Indupalli et al (2006), e os ruídos são removido por operações morfológicas, como a dilatação e erosão. Conforme Alam e Boufama (2011), esse método de subtração de fundo utiliza imagens no modelo de cores HSV, e trata os problemas relacionados as sombras e mudanças de iluminação. As métricas de associação utilizadas são relacionadas às seguintes características (obtidas a partir dos componentes conexos extraídos do *foreground*): centroide, velocidade, área, tamanho da *bounding box* e histograma de cor. O tratamento de oclusões ocorre da seguinte forma: Inicialmente a região correspondente a união das pessoas é segmentada em "sub-regiões"; em seguida as características de cada "sub-região" são comparadas às características das pessoas presentes no quadro anterior; o par sub-região/pessoa que apresenta o maior valor de comparação é associado, seguido do par que possui o segundo maior valor e assim por diante, até que todas as "sub-regiões" tenham sido associadas. Embora não sejam apresentados detalhes sobre o algoritmo de segmentação utilizado, os resultados apresentados em casos de oclusões são promissores.

Kim et al (2011) propuseram um sistema para rastrear grupos de objetos. Para detectar os objetos em movimento, o *foreground* é extraído através do uso de um modelo de fundo RGB baseado em Haritaoglu et al (2000), e os ruídos são eliminados através da operação morfológica de abertura. Para agrupar os objetos, o método de rotulação proposto por Chang et al (2004) é utilizado. Em cada quadro do vídeo, a posição dos grupos seguidos é prevista com base na posição, direção e velocidade

do deslocamento do centroide no quadro anterior. Os grupos seguidos são associados aos grupos detectados com base na distância euclidiana entre suas posições. Esse sistema é capaz de tratar o surgimento, desaparecimento e união entre objetos ou grupos de objetos, e devido à simplicidade das técnicas utilizadas, sua utilização em tempo real é factível. Contudo, o sistema não é capaz de manter a identidade de objetos que reaparecem na cena após oclusões.

Azari et al (2011) propuseram um sistema no qual as oclusões são tratadas através do uso do modelo que representa os alvos seguidos. Nesse sistema, o fundo da cena é modelado através do método MoG (Stauffer e Grimson, 1999), e os alvos são detectado no *foreground* através de operações morfológicas e extração de contornos. Para realizar a associação é utilizado o método *Global Nearest Neighbor* (Blackman e Popoli, 1999), no qual a diferença entre a *bounding box* de uma região detectada e a *bounding box* prevista para o alvo seguido é determinada com base nos parâmetros das *bounding boxes*, como posição do centro, tamanho horizontal e vertical e área (número de *pixels* do *foreground* contidos na *bounding box*). Em casos de segmentação da região correspondente ao alvo seguido, todos os segmentos são utilizados para atualizar a trajetória. Quando uma região do *foreground* é associada a dois ou mais alvos seguidos, considera-se que houve uma oclusão, e para determinar a posição de um alvo ocluso seu modelo é comparado aos *pixels* da referida região resultante da oclusão através do método de coeficiente de correlação (Rodgers e Nicewander, 1988). Quando o alvo não é detectado por um período de tempo que excede um *threshold*, considera-se que o alvo saiu da cena e o rastreamento do mesmo é cessado.

Mirabi e Javadi (2012) propuseram um sistema para rastrear pessoas em cenários externos usando filtro de Kalman. Para detectar o *foreground*, esse sistema usa o método de subtração de fundo proposto por Stauffer e Grimson (1999) combinado com o método de remoção de sombras proposto por Li-Qun Xu et al (2005). O filtro de Kalman é utilizado para modelar a velocidade e aceleração das pessoas rastreadas e predizer a posição em cada quadro. Durante a etapa de associação é utilizada uma adaptação do algoritmo SMP - *Stable Marriage Problem*. Conforme essa abordagem, introduzida por D'Elia et al (2010), é criada uma matriz que descreve o custo de cada associação, baseado na distância euclidiana entre as posições detectadas na imagem e as posições previstas para as pessoas. O algoritmo SMP é aplicado para encontrar a associação mais estável entre as medições e as previsões. No caso de oclusão entre pessoas, as regiões correspondentes as pessoas oclusas são separadas com base nas previsões para as *bounding boxes* das mesmas. O desaparecimento de pessoas não é tratado pelo

sistema.

Autor	Deteção do Foreground	Sombras	Associação	Tratamento de Oclusão
Seo <i>et al.</i> (1997)	<i>Threshold</i> nas camadas RGB.	Não trata.	Centroide.	<i>Histogram Backprojection</i> .
Haritaoglu <i>et al.</i> (2000)	Subtração de fundo proposta pelo autor.	Não trata.	Centroide, <i>bounding box</i> e silhueta.	Deteção de cabeças por Hydra (Haritaoglu <i>et al.</i> , 1999).
Lu e Tan (2001)	Wren <i>et al.</i> (1997)	Método do próprio autor.	Centroide.	Rastreamento do centroide do grupo. Comparação de cor para identificar os alvos após oclusões.
Zhao e Nevatia (2004)	Wren <i>et al.</i> (1997)	Método do próprio autor.	Modelo em forma de elipse.	Deteção de cabeças.
Lei e Xu (2006)	Stauffer e Grimson (1999)	Horprasert <i>et al.</i> (1999)	Centroide, área, cor e modelo em forma de elipse.	Rastreamento do centroide do grupo. Comparação de cor para identificar os alvos após oclusões.
Freitas (2010)	<i>Eigenbackground</i> (Oliver <i>et al.</i> , 2000)	Não trata.	Centroide, área e cor.	Rastreamento do centroide do grupo. Comparação de cor para identificar os alvos após oclusões.
Alam e Boufama (2011)	Indupalli <i>et al.</i> (2006)	Não trata.	Centroide, área, <i>bounding box</i> e histograma de cor.	Segmentação em sub-regiões.

Kim <i>et al.</i> (2011)	Baseado em Haritaoglu et al (2000)	Não trata.	Centroide.	Rastreamento do centroide do grupo.
Azari <i>et al.</i> (2011)	Stauffer e Grimson (1999)	Não trata.	<i>Bounding box</i> , algoritmo <i>Global</i> <i>Nearest Neighbor</i> .	Coeficiente de correlação (Rodgers e Nicewander, 1988).
Mirabi e Javadi (2012)	Stauffer e Grimson (1999)	Li-Qun Xu et al (2005)	Centroide, algoritmo SMP - <i>Stable</i> <i>Marriage Problem</i> .	Segmentação através da previsão da <i>bounding</i> <i>box</i> .

Tabela 1: Sistemas de rastreamento

2.2 Sistema de Rastreamento Proposto

O presente trabalho propõe um sistema de rastreamento de pessoas constituído pela combinação de técnicas existentes na literatura. Conforme diagrama exibido na Figura 1, o sistema de rastreamento proposto é dividido em duas etapas principais: Detecção e Cálculo das Trajetórias (representadas por retângulos pontilhados). Cada etapa é constituída por um conjunto de funções (representadas por retângulos sombreados).

A etapa de detecção dos *pixels* correspondentes às pessoas na imagem é dividida em três partes: Inicialmente o *foreground* é obtido através do uso da técnica de subtração de fundo proposta por Haritaoglu et al (2000); em seguida, as regiões correspondentes às sombras são eliminadas do *foreground*, conforme abordagem de Cucchiara et al (2003); e finalmente, operações morfológicas são utilizadas para eliminar ruídos, conforme foi descrito em trabalho anterior (Amaral e Tozzi, 2011).

A segunda etapa, relacionada ao cálculo das trajetórias das pessoas detectadas na etapa anterior, é dividida em duas partes: Associação e Gerenciamento das Pessoas Seguidas. A associação corresponde ao processamento realizado em cada quadro, no qual as pessoas detectadas na imagem são associadas às pessoas seguidas nos quadros anteriores. O gerenciamento refere-se a análise do deslocamento das pessoas presentes no vídeo, e suas interações entre si e entre objetos do fundo da cena.

A associação quadro a quadro é realizada utilizando métricas de associação relacionadas à

posição do centroide, bounding box e histograma de cor. Durante oclusões entre pessoas, é utilizado um método de detecção de cabeças para determinar a posição das pessoas envolvidas na oclusão.

O gerenciamento das pessoas seguidas é realizada através da adaptação do método de classificação de estados proposto por Lei e Xu (2006).

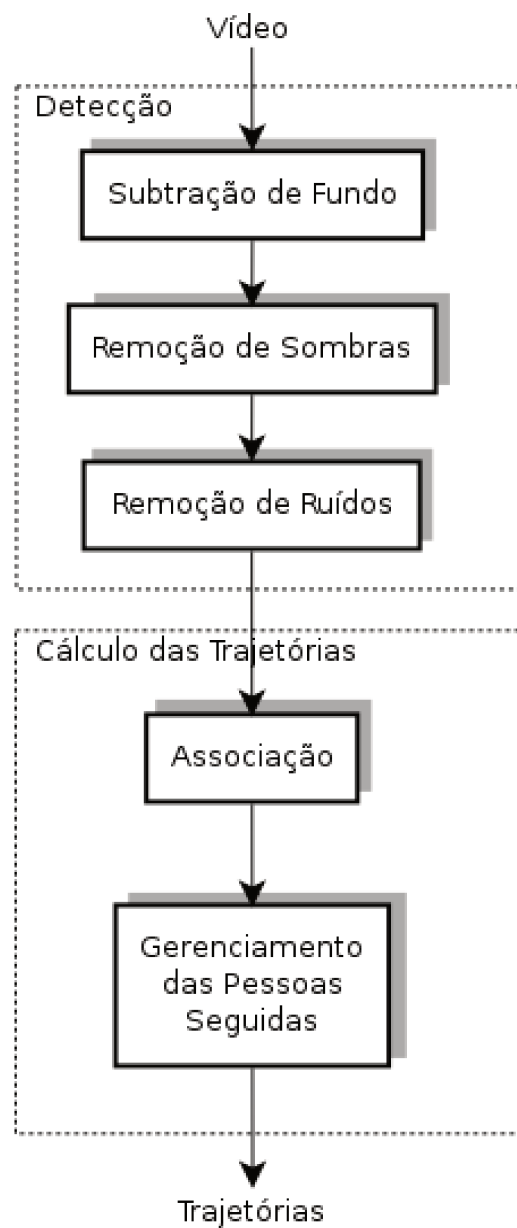


Figura 1: Etapas do sistema de rastreamento proposto

Capítulo 3

Detecção do Foreground

O rastreamento através de câmeras de vídeo é um processo recursivo que consiste em detectar as posições dos alvos seguidos em cada quadro e, com base na variação dessas posições ao longo vídeo, determinar as trajetórias. Desse modo, um método de detecção dos alvos é a etapa inicial do rastreamento.

Conforme Yilmaz et al (2006), a detecção de objetos pode ser realizada através do uso da informação temporal, que é obtida a partir da diferença entre quadros consecutivos. Uma abordagem mais complexa consiste na criação de um modelo que representa o fundo da cena, chamado de *background*. Esse modelo é comparado a cada quadro do vídeo, e qualquer variação significativa no quadro em relação ao *background* corresponde a um objeto em movimento, denominado *foreground*. Ao final desse processo, denominado subtração de fundo, usualmente é aplicado um algoritmo de detecção de componentes conexos para obter as regiões conexas correspondentes aos objetos presentes na cena.

No entanto, alguns fatores relacionados às condições do ambiente podem prejudicar a detecção do *foreground*. Tais fatores e as consequências dos mesmo (no contexto do rastreamento de pessoas) são listados abaixo:

- Partes do fundo em movimento: Objetos pertencentes ao fundo da cena em movimento repetitivo, como bandeiras ou árvores movimentando-se ao vento, são irrelevantes no contexto do rastreamento, porém, podem ser identificados como *foreground* e ocasionarem falsos positivos durante o rastreamento de pessoas.

- Variações de iluminação: Variações na iluminação da cena alteram o valor dos *pixels*, e podem fazer com que os *pixels* correspondentes ao fundo da cena difiram do modelo e sejam identificados como *foreground*. Em casos extremos, o *foreground* pode corresponder à totalidade da imagem.
- Sombras: Quando projetadas no fundo da cena, as sombras das pessoas alteram o valor dos *pixels*, e podem fazer com que os mesmos difiram do modelo e sejam identificados como *foreground*. Como usualmente as sombras estão ligadas às pessoas que as projetam, esse evento pode causar a união das regiões correspondentes às pessoas presentes na cena.
- Ruído: Os valores dos *pixels* obtidos em uma sequência de quadros do vídeo podem ser afetados por ruídos, como o ruído sal e pimenta (também conhecido como ruído de impulso). Métodos de detecção do *foreground* menos sofisticados, como os propostos por Jain e Nagel (1979) e Seo et al (1997), são prejudicados pela presença de ruídos nas imagens. Métodos mais robustos presentes na literatura, como os propostos por Stauffer e Grimson (2000) e Haritaoglu et al. (2000), obtêm melhores resultados na presença de ruídos.

Os fatores listados acima afetam a detecção do *foreground* e, conseqüentemente, a determinação das trajetórias. Desse modo, o método de detecção dos alvos a serem seguidos empregado em um sistema de rastreamento deve ser capaz de identificar a ocorrência desses fatores e lidar convenientemente com suas conseqüências, de modo a garantir que os *pixels* identificados como *foreground* correspondam aos alvos seguidos.

No presente trabalho, a detecção do *foreground* é dividida em três etapas principais: subtração de fundo, remoção de sombra e remoção de ruídos por morfologia matemática. As seções 3.1, 3.2 e 3.3, respectivamente, apresentam uma breve revisão bibliográfica e o método utilizado em cada etapa. Ao final de cada seção, são realizadas análises qualitativas e quantitativas para avaliar os resultados obtidos.

3.1 Subtração de Fundo

Métodos de detecção de objetos em movimento vêm sendo estudados desde a década de 70, como o trabalho proposto por Jain e Nagel (1979). Para encontrar as regiões correspondentes às pessoas e veículos presentes em um cenário real, Jain e Nagel aplicam um *threshold* no resultado da diferença entre quadros adjacentes do vídeo. Embora essa abordagem seja capaz de detectar as regiões correspondentes aos objetos de interesse, não são tratados fatores como variações de iluminação ou objetos em movimento repetitivo no fundo da cena.

Seo et al (1997) propuseram um método para detectar os jogadores em um campo de futebol baseado na cor dos *pixels*. Nesse método é assumido que o gramado ocupa uma grande região na imagem, possui coloração uniforme e tonalidade verde. Sendo assim, um *threshold* é aplicado em cada camada RGB dos quadros do vídeo para identificar os *pixels* que possuem a coloração esperada para a grama, e o componente conexo resultante do *threshold* que possui maior área é considerado o gramado. As regiões contidas no gramado que não possuem a coloração esperada para a grama são os jogadores. Essa abordagem é capaz de identificar as pessoas apenas em locais onde a coloração do fundo é uniforme, como, por exemplo, em um campo de futebol.

Conforme Yilmaz et al (2006), os métodos de subtração de fundo popularizaram-se com o trabalho proposto por Wren et al (1997). Para adaptar o *background* a oscilações e mudanças graduais no valor dos *pixels*, Wren et al propuseram modelar a cor de cada *pixel* $I(x, y)$ de um fundo estático através de uma gaussiana, $I(x, y) \sim N(\mu(x, y), \Sigma(x, y))$. Os parâmetros do modelo (média $\mu(x, y)$ e covariância $\Sigma(x, y)$) para cada *pixel* (x, y) são obtidos a partir da observação da cor do *pixel* em cada quadro consecutivo. Uma vez que o modelo é criado, para cada *pixel* (x, y) de um quadro, a vizinhança da cor desse *pixel* em relação à gaussiana do modelo é computada, e os *pixels* que diferem do modelo são rotulados como *foreground*. Apesar dos bons resultados apresentados, uma única gaussiana não é um bom modelo para ser usado em cenas externas (Gao et al. 2000), pois múltiplas cores podem ser observadas em uma certa localidade devido a presença de objetos no fundo da cena em movimentos repetitivos, como árvores ao vento.

Um aprimoramento substancial na modelagem do fundo através de gaussianas foi alcançado

com o uso de um modelo estatístico multicamada para descrever a cor do fundo. Stauffer e Grimson (2000) usaram um conjunto de gaussianas (normalmente de 3 a 5) para modelar a cor de cada *pixel*. Nesse método, um *pixel* em um quadro é comparado às gaussianas do modelo, e se houver correspondência entre o *pixel* e a gaussiana, a média e variância dessa gaussiana é atualizada, caso contrário uma nova gaussiana com a média igual ao valor do *pixel* e uma variância predefinida é introduzida no modelo. A vantagem dessa abordagem é que a utilização de mais de uma gaussiana permite modelar variações de iluminação, movimentações repetitivas, e alterações físicas no fundo da cena, como o abandono de objetos.

Haritaoglu et al. (2000) propuseram um método de subtração de fundo baseado em dados estatísticos de cada *pixel* adquiridos a partir de uma sequência de imagem em tons de cinza. O modelo, para cada *pixel*, é constituído por três valores: valor mínimo, valor máximo e a máxima diferença de intensidade entre quadros consecutivos observados durante o período de treinamento (de 20 à 40 segundos). A criação do modelo é dividida em duas etapas: inicialmente um filtro da mediana é aplicado para cada *pixel* para distinguir *pixels* correspondentes ao fundo de *pixels* correspondentes a objetos em movimento; em seguida, apenas os *pixels* correspondentes ao fundo são processados para criar o modelo inicial do fundo. Desse modo, o modelo pode ser obtido mesmo que pessoas e demais objetos estejam em movimento no campo de visão da câmera durante o período de treinamento. Periodicamente, o modelo de fundo é atualizado através do uso de um "mapa de suporte" para incorporar ao modelo as mudanças ocorridas no fundo da cena (como, por exemplo, variações de iluminação). A periodicidade de atualização é definida pela "taxa de atualização", que é um parâmetro inicial do sistema estipulado de acordo com a frequência de alterações observada na cena. Devido à simplicidade das inequações utilizadas durante a detecção do *foreground*, esse método possui fácil implementação.

Wei et al. (2007) propuseram um método de detecção de objetos em movimento baseado em um modelo de fundo duplo. Um modelo estatístico (SBG) é usado para descrever o intervalo de intensidade correspondente ao fundo da cena para cada *pixel*, e um outro modelo (FBG), obtido a partir da diferença entre quadros consecutivos, é utilizado para incorporar ao modelo estatístico as alterações ocorridas no fundo da cena em um curto período de tempo (como o abandono de um objeto no fundo da cena). As alterações de iluminação são identificadas a partir da diferença entre quadros consecutivos, e incorporadas ao modelo estatístico.

A Tabela 2 apresenta um resumo dos métodos de subtração de fundo abordados. Para cada método, são descritas as principais características, e é indicado se o método é capaz de lidar com alterações de iluminação (como o bloqueio do sol por nuvens), alterações físicas no ambiente (como o abandono de objetos na cena), e objetos em movimento repetitivo no fundo da cena (como galhos ao vento).

Autores	Ano	Alterações de Iluminação	Alterações Físicas no Ambiente	Objetos em Movimento Repetitivo	Principais Características
Jain e Nagel	1979	NP	NI	NP	Rapidez de processamento; Detecta movimento pela diferença entre quadros consecutivos.
Seo et al.	1997	NP	NP	NP	Possui implementação simples; Funciona apenas quando a coloração do fundo é uniforme.
Wren et al.	1997	P	P	NP	Adapta o modelo a variações graduais de iluminação ajustando os parâmetros de uma gaussiana.
Stauffer e Grimson	2000	P	P	P	Modela movimentos repetitivos através de múltiplas gaussianas.
Haritaoglu et al.	2000	P	P	P	Possui implementação simples; Adapta o modelo a alterações no ambiente através de mapas de alterações.
Wei et al.	2007	P	P	NP	Utiliza dois modelos de fundo para tratar alterações físicas no ambiente.

(P) Propõe tratamento para o evento; (NP) Não propõe tratamento para o evento; (NI) Não é influenciado pelo evento.

Tabela 2: Métodos de subtração de fundo

3.1.1 Subtração de Fundo Utilizada no Sistema Proposto

Para ser utilizado em um sistema de rastreamento eficiente, o método de subtração de fundo deve ser capaz de lidar com três eventos: alterações de iluminação na cena, alterações físicas no ambiente e objetos em movimento repetitivo no fundo da cena. Os métodos de subtração de fundo propostos por Stauffer e Grimson (2000) e Haritaoglu et al. (2000) propõem tratamento para esses eventos. Por possuir implementação simples, o método de subtração de fundo proposto por Haritaoglu et al. (2000) é utilizado no sistema de rastreamento de pessoas proposto no presente trabalho. Nessa seção esse método de subtração de fundo é descrito em detalhes.

Criação do Modelo do Fundo

A criação do modelo inicial do fundo é dividida em duas etapas principais: Na primeira etapa um filtro da mediana é aplicado para cada *pixel* durante um período do vídeo (usualmente entre 20 e 40 segundos) para distinguir “*pixels em movimento*” (correspondentes a objetos em movimento presentes na cena durante o período aquisição do modelo) de “*pixels estáticos*” (correspondentes ao fundo da cena); Na segunda etapa, apenas os “*pixels estáticos*” são processados para construir o modelo inicial do fundo. Seja V um vetor contendo N imagens consecutivas, $V^z(x)$ é o conjunto de “*pixels estáticos*” de V posicionados em x , e $\sigma(x)$ e $\lambda(x)$ são o desvio padrão e a mediana, respectivamente, do valor de intensidade do *pixel* x em todas as imagens em V . O modelo de fundo inicial para cada *pixel* x , $[m(x), n(x), d(x)]$, é obtido conforme a equação (1).

$$\begin{bmatrix} m(x) \\ n(x) \\ d(x) \end{bmatrix} = \begin{bmatrix} \min_z \{V^z(x)\} \\ \max_z \{V^z(x)\} \\ \max_z \{|V^z(x) - V^{z-1}(x)|\} \end{bmatrix} \quad (1)$$

onde $|V^z(x) - \lambda(x)| < 2\sigma(x)$

onde $m(x)$ e $n(x)$ são, respectivamente, o menor e o maior valor assumido por x , e $d(x)$ é a máxima diferença de intensidade entre quadros consecutivos.

Atualização dos Parâmetros

Não é esperado que o fundo da cena permaneça o mesmo por longos períodos de tempo. Podem ocorrer alterações de iluminação, como o sol ser bloqueado por nuvens causando alterações no brilho da imagem, ou alterações físicas, como o depósito de um objeto. Como o modelo de fundo é baseado na intensidade dos *pixels*, mudanças no valor de intensidade podem causar falsos positivos. Adicionalmente, qualquer objeto do *foreground* detectado por um longo período sem movimentação (como um carro estacionado) pode prejudicar a detecção de demais objetos ou pessoas que passem pelo local (uma pessoa saindo do carro não seria detectada). Para tratar esses possíveis eventos, o modelo de fundo é atualizado de duas maneiras distintas:

1. Um método baseado em pixel atualiza o modelo periodicamente para adaptá-lo a mudanças de iluminação no fundo da cena.
2. Um método baseado em objetos atualiza o modelo para adaptá-lo a mudanças físicas ocorridas no fundo da cena. Um objeto depositado ou removido, ou um carro estacionado deve ser adicionado ao modelo se ele não se mover por um longo período de tempo.

Para atualizar o modelo de fundo, um mapa de alterações é construído dinamicamente durante o rastreamento para determinar se o método baseado em *pixel* ou objeto será aplicado. O mapa de alterações consiste em:

- Mapa de mudanças (gS): representa o número de vezes que um *pixel* foi classificado como fundo nos últimos N quadros, sendo N o número de quadros processados desde a última atualização do modelo de fundo.

$$gS(x, t) = \begin{cases} gS(x, t-1) + 1 & \text{se } x \text{ é fundo} \\ gS(x, t-1) & \text{se } x \text{ não é fundo} \end{cases} \quad (2)$$

- Mapa de movimento (mS): representa o número de vezes que um *pixel* foi classificado como *pixel* em movimento. Um *pixel* é classificado como um *pixel* em movimento com base no resultado da subtração de três quadros consecutivos.

$$mS(x, t) = \begin{cases} mS(x, t-1) + 1 & \text{se } M(x, t) = 1 \\ mS(x, t-1) & \text{se } M(x, t) = 0 \end{cases} \quad (3)$$

onde

$$M(x, t) = \begin{cases} 1 & \text{se } (|I(x, t) - I(x, t+1)| > 2\sigma) \\ & \wedge (|I(x, t-1) - I(x, t)| > 2\sigma) \\ 0 & \text{caso contrário} \end{cases} \quad (4)$$

O mapa gS é usado para determinar as partes do modelo de fundo que serão atualizadas pelo método baseado em *pixels*, e os mapas mS e gS são usados para determinar as partes do modelo de fundo que serão atualizadas pelo método baseado em objetos. Os mapas de alterações são reiniciados após a atualização do modelo de fundo.

Durante o rastreamento, o modelo de fundo é computado separadamente para todos os *pixels* classificados como *foreground* $[m^f(x), n^f(x), d^f(x)]$ e para todos os *pixels* classificados como *background* $[m^b(x), n^b(x), d^b(x)]$. Seja $[m^c(x), n^c(x), d^c(x)]$ o modelo de fundo usado atualmente, os parâmetros do novo modelo de fundo são determinados conforme é descrito na equação (5).

$$[m(x), n(x), d(x)] = \begin{cases} [m^b(x), n^b(x), d^b(x)] & \text{se } (gS(x) > k * N) \\ [m^f(x), n^f(x), d^f(x)] & \text{se } (gS(x) < k * N \wedge mS(x) < r * N) \\ [m^c(x), n^c(x), d^c(x)] & \text{caso contrário} \end{cases} \quad (5)$$

onde k e r usualmente são 0.8 e 0.1, respectivamente.

Um outro fator importante é quão rápido o modelo de fundo se adapta a mudanças. Mudanças repentinas de iluminação, como nuvens bloqueando o sol, podem causar falhas na detecção e fazer com que grande parte da imagem seja classificada como *foreground*. Por esse motivo, quando uma grande

parte da imagem é detectada como *foreground* (acima de 80%, por exemplo), o modelo de fundo é criado novamente.

Detecção do *Foreground*

Dado o mínimo $m(x)$, máximo $n(x)$, e a mediana da maior diferença entre quadros d_μ de todos os *pixels* do modelo, o pixel x na imagem I^t é classificado como *foreground* se $B(x)=1$, conforme equação (6):

$$B(x) = \begin{cases} 0 \text{ background} & \begin{cases} (I^t(x) - m(x)) < k d_\mu \\ \vee (I^t(x) - n(x)) < k d_\mu \end{cases} \\ 1 \text{ foreground} & \text{caso contrário} \end{cases} \quad (6)$$

Em experimentos realizados por Haritaoglu et al. (2000), $k = 2$ na equação (6) apresentou altas taxas de acerto e baixas taxas de erros.

3.1.2 Avaliação do Método de Subtração de Fundo

Essa seção avalia a eficiência do método de subtração de fundo utilizado no presente trabalho. A metodologia e alguns dos vídeos utilizados nessa seção são utilizados também para avaliar o método de remoção de sombra e as operações morfológicas, etapas posteriores à subtração de fundo. Desse modo, é possível observar a evolução do resultado em cada etapa da detecção do *foreground*.

A metodologia utilizada para avaliar o resultado da detecção do *foreground* consiste em análises qualitativas e quantitativas:

- Análise qualitativa: Consistem em avaliar visualmente as imagens binárias obtidas, correspondentes ao *foreground* em cada quadro.
- Análise quantitativa: Consiste em comparar o resultado obtido ao esperado através das taxas de

precisão (*prec*) e revocação (*rev*).

Para obter as taxas de precisão e revocação, 10 quadros do vídeo onde existam pessoas são selecionados. As pessoas presentes nos quadros são segmentadas manualmente e a quantidade de *pixels* correspondentes às pessoas é calculada. Enfim, as taxas de precisão e revocação são obtidas conforme as equações (7) e (8), respectivamente.

$$prec = \frac{\text{número de pixels corretamente detectados como foreground}}{\text{número total de pixels detectados como foreground}} \quad (7)$$

$$rev = \frac{\text{número de pixels corretamente detectados como foreground}}{\text{número real de pixels do foreground}} \quad (8)$$

A taxa de precisão expressa a relação entre a quantidade de *pixels* corretamente classificados como *foreground* e o total de *pixels* classificados como *foreground*.

A taxa de revocação expressa a relação entre a quantidade de *pixels* corretamente classificados como *foreground* pelo sistema e a quantidade real de *pixels* correspondentes ao *foreground*.

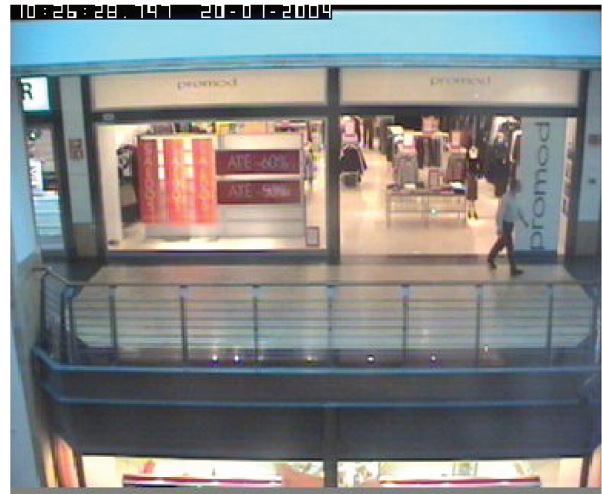
Ambos os valores de *prec* e *rev* variam entre 0 e 1, sendo que um modelo ideal de subtração de fundo deve resultar em $prec = rev = 1$.

Vídeos Utilizados

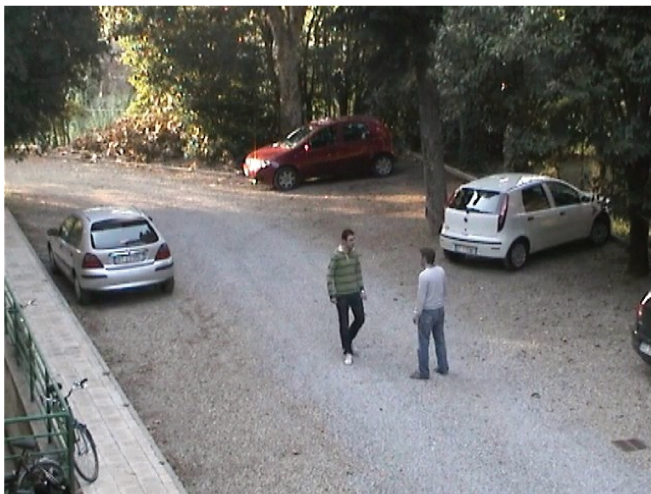
Para avaliar o método de subtração de fundo utilizado no presente trabalho foram utilizados quatro vídeos, exibidos na Figura 2. As tabelas 3, 4, 5 e 6 apresentam informações sobre os vídeos exibidos na figura.



(a) Vídeo 1



(b) Vídeo 2



(c) Vídeo 3



(d) Vídeo 4

Figura 2: Vídeos utilizados para testar o método de detecção do foreground

Vídeo 1	
Nome	CAVIAR-Meet_WalkSplit
Descrição	Duas pessoas se encontram, caminham unidas e se separam
Base	CAVIAR
Dimensões	384 x 288 pixels
Quadros por segundo	25
Formato	623 imagens comprimidas no formato MPEG2
Link para download	http://groups.inf.ed.ac.uk/vision/CAVIAR/CAVIARDATA1/Meet_WalkSplit/Meet_WalkSplit.jpg.tar.gz (acessado em 11/07/2012)

Tabela 3: Detalhes do vídeo 1

Vídeo 2	
Nome	CAVIAR-OneLeaveShop1front
Descrição	Pessoas saem de uma loja
Base	CAVIAR
Dimensões	384 x 288 pixels
Quadros por segundo	25
Formato	295 imagens comprimidas no formato MPEG2
Link para download	http://groups.inf.ed.ac.uk/vision/CAVIAR/CAVIARDATA2/OneLeaveShop1front/OneLeaveShop1front.tar.gz (acessado em 11/07/2012)

Tabela 4: Detalhes do vídeo 2

Vídeo 3	
Nome	visor_1267122885367_video_seq1
Descrição	Duas pessoas caminham em uma rua até se encontrarem. Após algum evento desconhecido, uma pessoa corre e é seguida pela segunda
Base	Visor
Dimensões	640 x 480 pixels
Quadros por segundo	6,95
Formato	avi, convertido em 831 arquivos de imagem
Link para download	http://imabelab.ing.unimore.it/visor/video_Download.asp?idVideo=340 (acessado em 11/07/2012)

Tabela 5: Detalhes do vídeo 3

Vídeo 4	
Nome	Candela-drop_bag_and_depart
Descrição	Uma pessoa abandona um objeto em um escritório
Base	Candela
Dimensões	352 x 288 pixels
Quadros por segundo	25
Formato	avi, convertido em 305 arquivos de imagem
Link para download	http://www.multitel.be/~va/videos/candela/m1.01_abandoned_object_-_dr op_bag_and_depart.avi (acessado em 11/07/2012)

Tabela 6: Detalhes do vídeo 4

Vídeo 1

O resultado foi obtido variando o valor do parâmetro k da equação 6, que estabelece o grau de *threshold* aplicado para obter o *foreground*. A Tabela (7) apresenta os valores das taxas de precisão e revocação para cada valor de k , a Figura 3 exibe o gráfico da variação das taxas, e a Figura 4 exibe o resultado da subtração de fundo para os diferentes valores do parâmetro k .

	Precisão	Revocação
k=1	0.67141	0.69507
k=2	0.86844	0.63178
k=3	0.90765	0.53406
k=6	0.91637	0.27381

Tabela 7: Vídeo 1 – Revocação $\times$ Precisão para a variação do limiar k

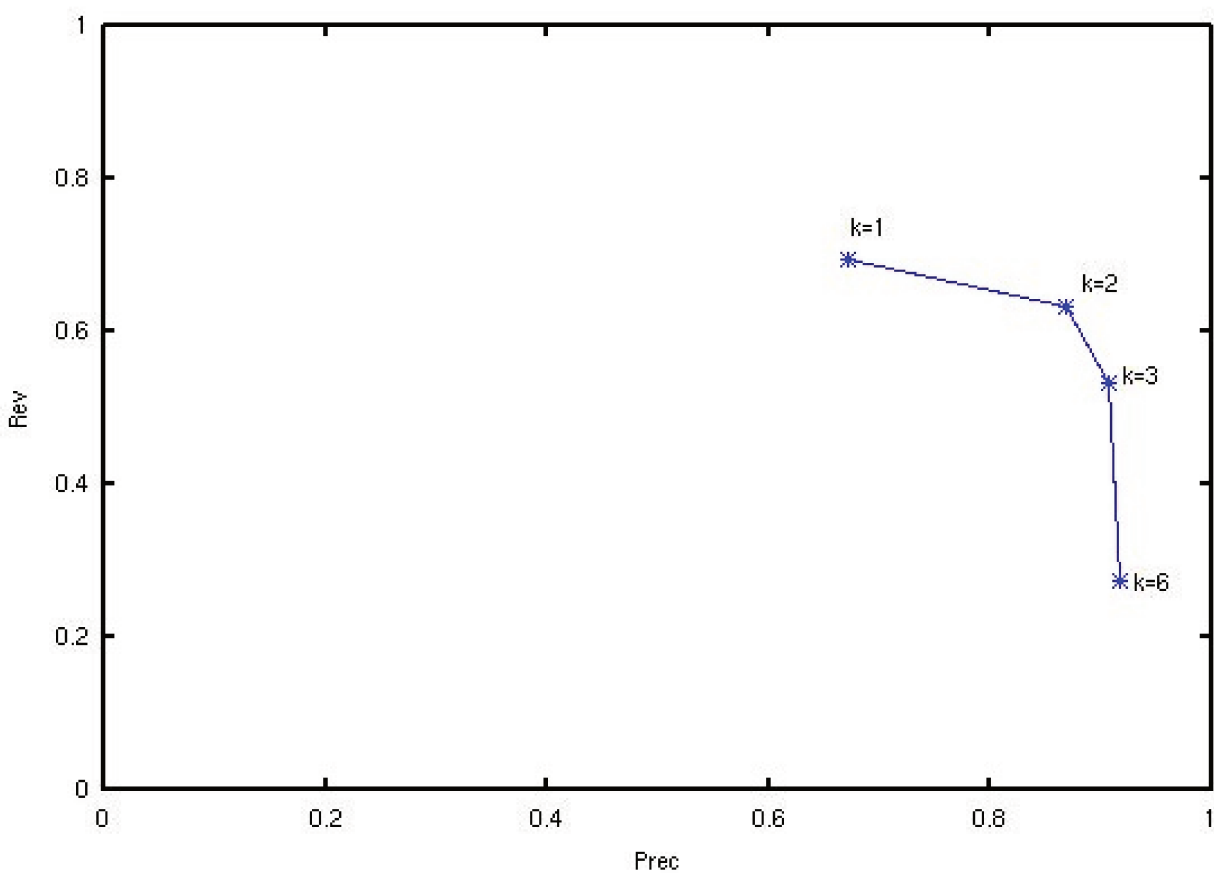


Figura 3: Vídeo 1 – Revocação $\times$ Precisão para a variação do limiar k



(a) Resultado esperado



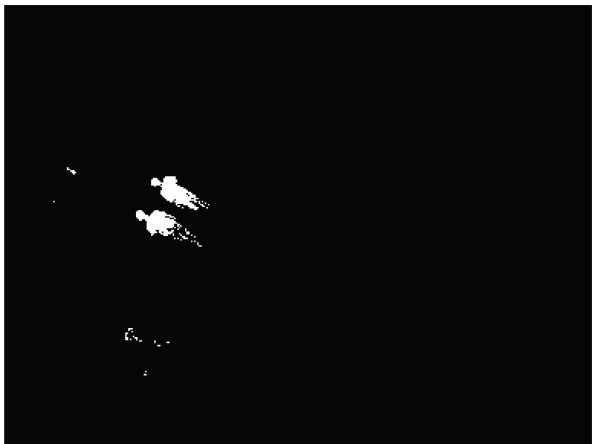
(b) $k=1$



(c) $k=2$



(d) $k=3$



(e) $k=6$

Figura 4: Vídeo 1 – Resultado da subtração de fundo para a variação do limiar k

Vídeo 2

O resultado foi obtido variando o valor do parâmetro k da equação 6, que estabelece o grau de *threshold* aplicado para obter o *foreground*. A Tabela (8) apresenta os valores das taxas de precisão e revocação para cada valor de k , a Figura 5 exibe o gráfico da variação das taxas, e a Figura 6 exibe o resultado da subtração de fundo para os diferentes valores do parâmetro k .

	Precisão	Revocação
$k=1$	0.73025	0.63672
$k=2$	0.89358	0.52592
$k=3$	0.89950	0.39050
$k=6$	0.82059	0.17387

Tabela 8: Vídeo 2 – Revocação x Precisão para a variação do limiar k

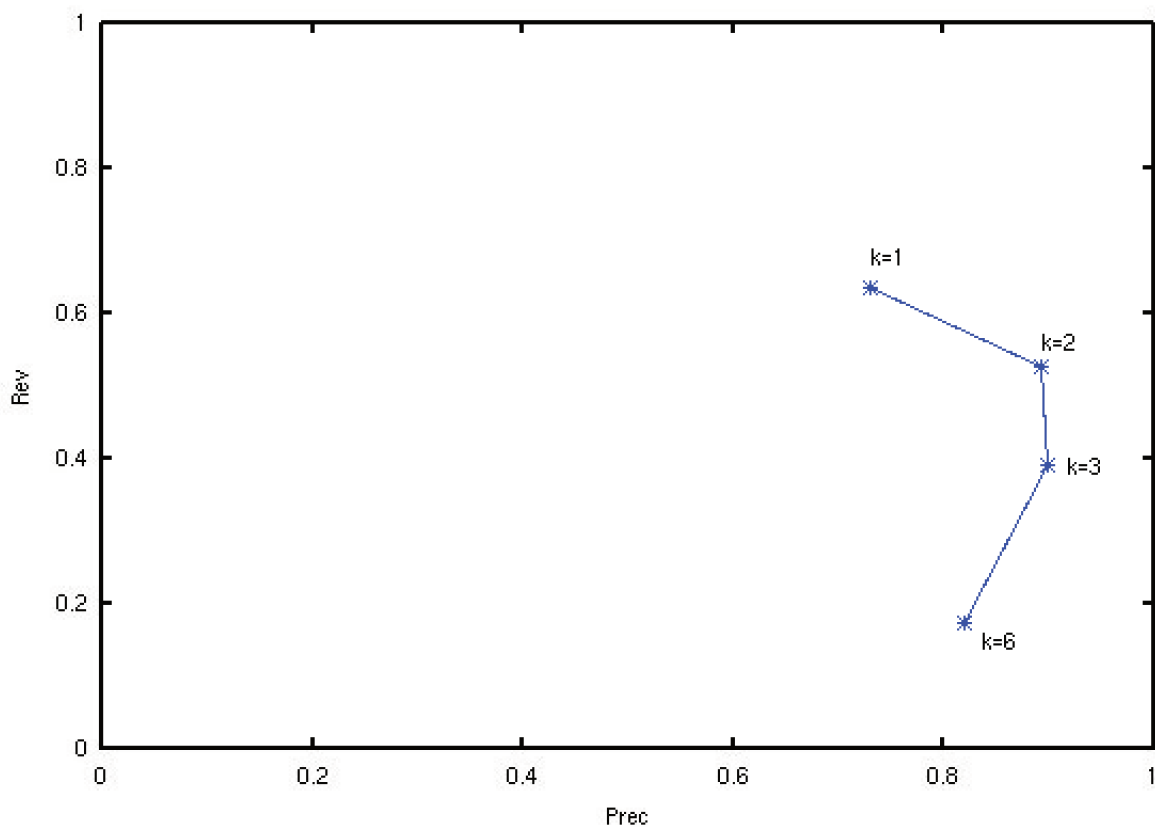
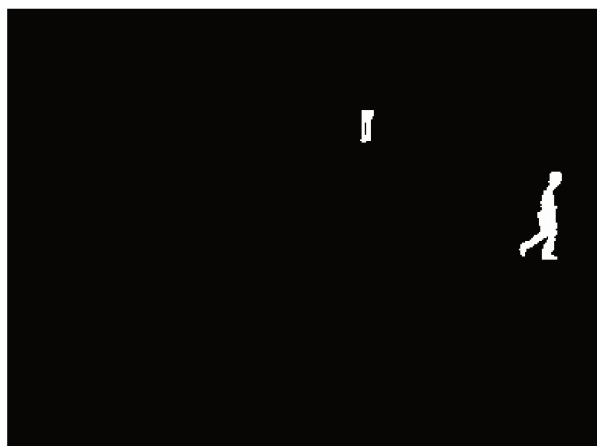


Figura 5: Vídeo 2 – Revocação x Precisão para a variação do limiar k



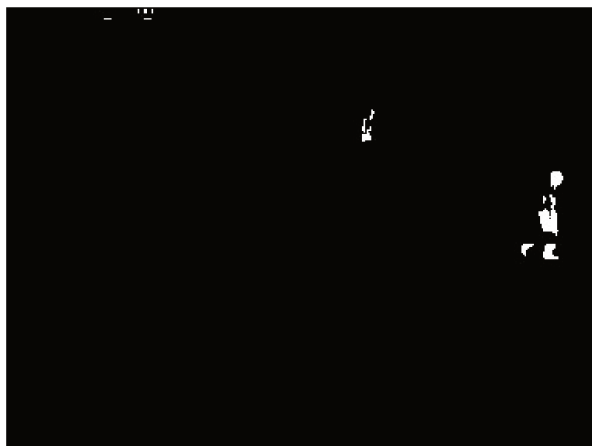
(a) Resultado esperado



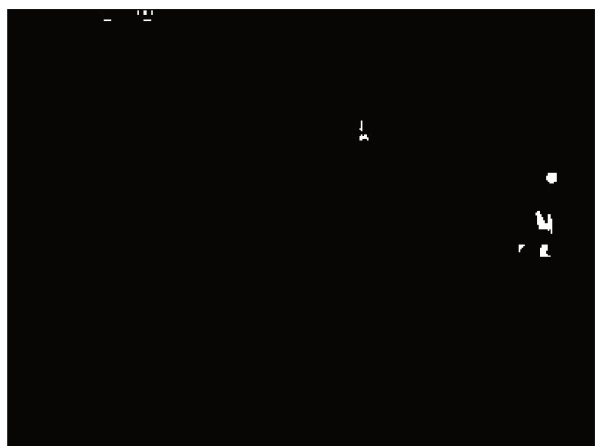
(b) $k=1$



(c) $k=2$



(d) $k=3$



(e) $k=6$

Figura 6: Vídeo 2 – Resultado da subtração de fundo para a variação do limiar k

Vídeo 3

O resultado foi obtido variando o valor do parâmetro k da equação 6, que estabelece o grau de *threshold* aplicado para obter o *foreground*. A Tabela (9) apresenta os valores das taxas de precisão e revocação para cada valor de k , a Figura 7 exibe o gráfico da variação das taxas, e a Figura 8 exibe o resultado da subtração de fundo para os diferentes valores do parâmetro k .

	Precisão	Revocação
$k=1$	0.30849	0.89292
$k=2$	0.55182	0.83889
$k=3$	0.68448	0.79276
$k=6$	0.86035	0.67367

Tabela 9: Vídeo 3 – Revocação x Precisão para a variação do limiar k

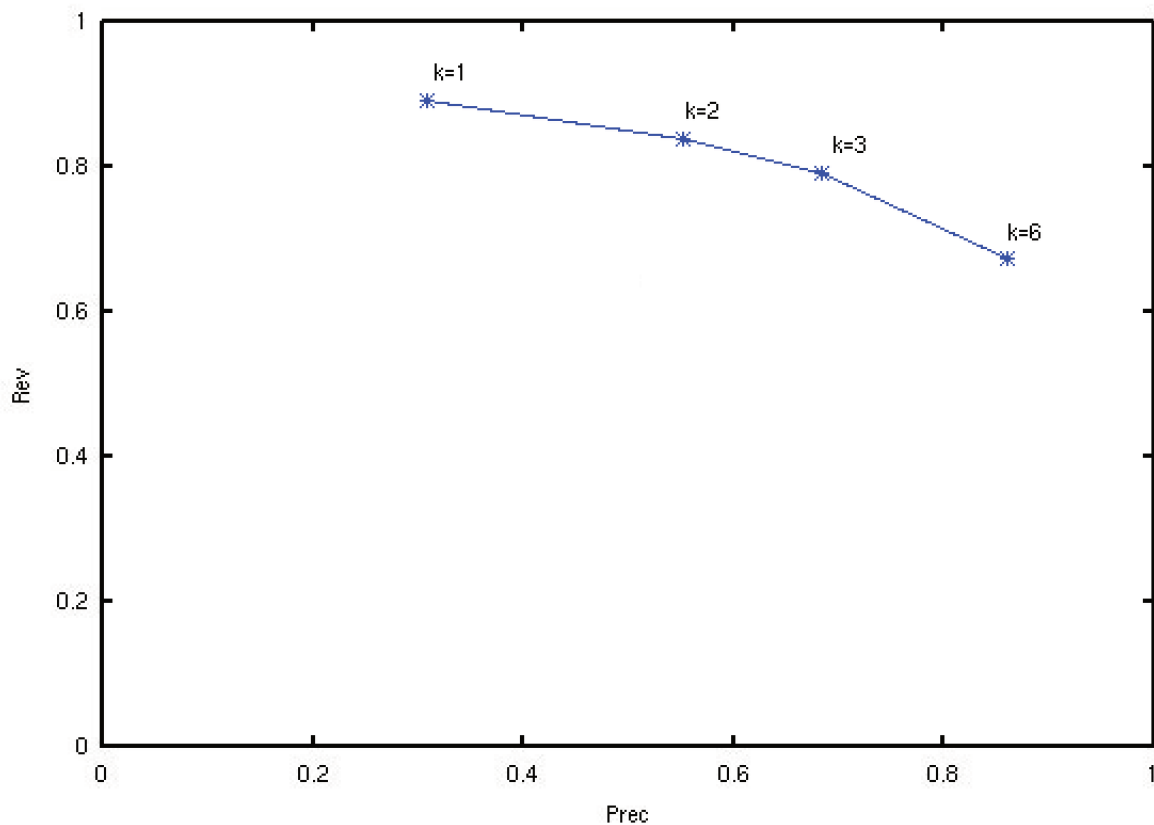
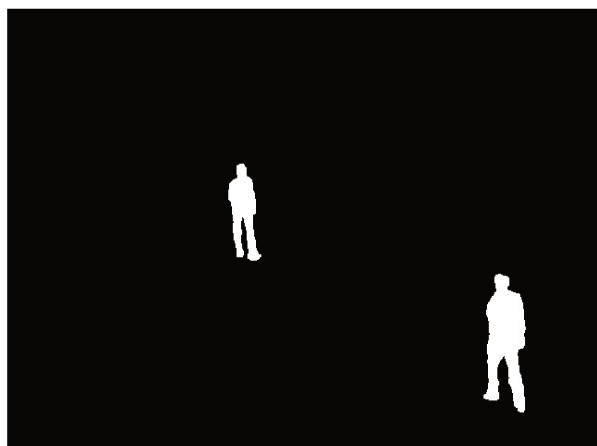


Figura 7: Vídeo 3 – Revocação x Precisão para a variação do limiar k



(a) Resultado esperado



(b) $k=1$



(c) $k=2$



(d) $k=3$



(e) $k=6$

Figura 8: Vídeo 3 – Resultado da subtração de fundo para a variação do limiar k

Vídeo 4

O resultado foi obtido variando o valor do parâmetro k da equação 6, que estabelece o grau de *threshold* aplicado para obter o *foreground*. A Tabela (10) apresenta os valores das taxas de precisão e revocação para cada valor de k , a Figura 9 exibe o gráfico da variação das taxas, e a Figura 10 exibe o resultado da subtração de fundo para os diferentes valores do parâmetro k .

	Precisão	Revocação
$k=1$	0.24253	0.89316
$k=2$	0.44671	0.83635
$k=3$	0.55774	0.78510
$k=6$	0.78336	0.67976

Tabela 10: Vídeo 4 – Revocação x Precisão para a variação do limiar k

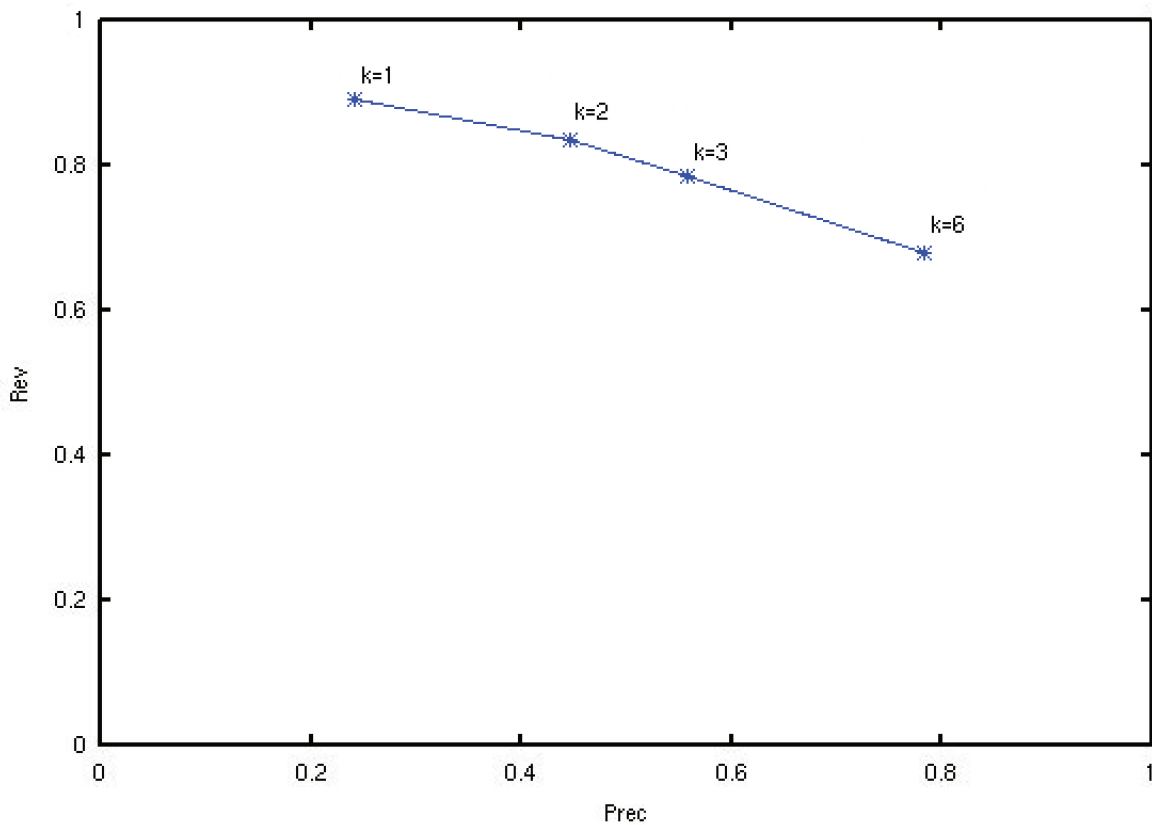


Figura 9: Vídeo 4 – Revocação x Precisão para a variação do limiar k



(a) Resultado esperado



(b) $k=1$



(c) $k=2$



(d) $k=3$



(e) $k=6$

Figura 10: Vídeo 4 – Resultado da subtração de fundo para a variação do limiar k

Para todos os vídeos analisados, os resultados esperados foram obtidos ao variar o parâmetro k da equação 6. Ou seja, a medida que o valor de k diminuiu, mais *pixels* foram classificados como *foreground*. Ainda, os melhores resultados observados foram obtidos com o valor de k entre 2 e 3, dependendo do vídeo, conforme mencionado por Haritaoglu et al (2000).

3.2 Remoção de Sombras

Quando uma sombra é projetada em uma superfície, a aparência dessa superfície é alterada. Como tipicamente o *foreground* é obtido através da diferença entre quadros consecutivos ou de técnicas de subtração de fundo, todos os *pixels* correspondentes aos objetos na cena e às sombras são detectados. Ainda, as sombras usualmente estão posicionadas adjacentes aos objetos que as projetam e, através dos métodos mais comuns de detecção do *foreground*, os objetos seguidos e as sombras são unidas em um mesmo *blob* (região conexa do *foreground*). Isso implica em dois problemas principais:

1. A forma do *blob* correspondente ao objeto é alterada, e todas as medidas geométricas são afetadas (como a posição do centroide). Outras características, como cor principal e textura, também são afetadas. Isso prejudica a classificação dos objetos.
2. A sombra de dois ou mais objetos podem criar uma falsa adjacência entre objetos, fazendo com que eles sejam unidos em um único *blob*. Isso afeta tarefas como contagem, classificação e rastreamento de objetos.

Devido aos problemas resultantes da classificação de sombras como *foreground*, um sistema de rastreamento deve tomar medidas para identificar a presença de regiões correspondentes a sombras no *foreground*, e eliminá-las. Entre as abordagens presentes na literatura para identificar sombras, uma das propriedades mais utilizadas é a diferença existente entre o valor da iluminação de um *pixel* afetado por uma sombra e o valor do mesmo *pixel* na ausência de sombra (Amato et al, 2011).

Para prevenir a classificação de sombras como *foreground*, Horprasert et al. (1999) propuseram um método que modela cada *pixel* do fundo através de uma 4-tupla $\langle E_i, s_i, a_i, b_i \rangle$, onde E_i é o valor RGB esperado para o *pixel*, s_i é a variação do valor RGB, a_i é a variação da intensidade, e

b_i é a variação da cromaticidade. Então, através da combinação de três *thresholds*, cada *pixel* é classificado como: Fundo, se intensidade e cromaticidade forem similares aos valores esperados; Fundo sombreado, se a cromaticidade for similar e a intensidade for menor que o valor esperado; Fundo iluminado, se a cromaticidade for similar e a intensidade maior que valor esperado; *Foreground*, se a cromaticidade for diferente do valor esperado. Essa abordagem parte do pressuposto que uma sombra não produz alterações significativas na direção do vetor da cor no modelo RGB (Amato et al, 2008).

No método de detecção proposto por McKenna et al. (2000), para evitar a classificação de sombras como *foreground*, são utilizadas informações relacionadas às cores e ao gradiente da imagem. Cada *pixel* é modelado usando a média e variância do gradiente, obtido através do operador de Sobel aplicado nas direções x e y. Dado um novo quadro do vídeo, calcula-se o gradiente para as camadas R, G e B. Então um *pixel* é classificado como *foreground* se, para qualquer camada, a diferença entre o gradiente e o valor do modelo for maior que um *threshold*. Conforme McKenna et al. (2000), essa abordagem ajuda a eliminar as sombras, e a eficiência do método é diretamente proporcional ao tamanho da penumbra.

As propriedades das sombras foram ainda explorada no espaço de cores HSV. Para distinguir sombras de objetos em movimento no *foreground*, Cucchiara et al. (2001) realizam uma combinação de *thresholds* nas camadas *Hue*, *Value* e *Saturation*. Testes realizados pelos autores comprovam que sombras decrescem o valor da camada *Saturation*, enquanto alteram o valor das demais camadas dentro de certos limites.

Diferentemente dos métodos tradicionais de remoção de sombra que apenas consideram propriedades relacionadas à iluminação, Hsieh et al (2004) introduzem a utilização das faixas da estrada para remover sombras que causam a junção das regiões correspondentes a veículos distintos. Inicialmente, as faixas que separam as pistas são detectadas. Em seguida, as intensidades dos *pixels* são analisadas e se, para cada *pixel* contido em uma reta paralela à faixa, a probabilidade desse *pixel* ser sombra for maior que um *threshold*, os mesmos são removidos do *foreground* (a probabilidade de um *pixel* ser sombra é definida a partir do valor médio e variância da intensidade). Dependendo da direção das faixas da estrada e da iluminação, esse método pode detectar as sombras com precisão (Hsieh et al, 2004).

Uma boa revisão e comparação entre algoritmos de remoção de sombras é apresentado por Prati et al (2003).

3.2.1 Método de Remoção de Sombra Utilizado no Sistema Proposto

No presente trabalho o método de remoção de sombra proposto por Cucchiara et al. (2001) é utilizado. Esse método baseia-se na constatação de que quando uma sombra é projetada em uma região, o valor da saturação dos *pixels* correspondentes à região é sempre decrescido, e o valor da cromaticidade nunca é alterado significativamente. Para formular esse conceito, o modelo de cor HSV é utilizado para representar a cor de um *pixel*, porque nesse modelo a cromaticidade é separada explicitamente da luminosidade, o que torna a formulação matemática mais simples do que se utilizado outro modelo de cor, como o modelo RGB por exemplo. Outra vantagem da utilização desse método é que as sombras são detectadas, independentemente da direção da iluminação.

Um pixel x no tempo t é considerado sombra se $SP^t=1$, sendo SP^t obtido a partir da comparação entre o valor das camadas H, S e V do fundo da cena (*background*) no tempo t (B_H^t , B_S^t e B_V^t , respectivamente) e o valor das camadas H, S e V da imagem correspondente ao quadro do vídeo no tempo t (I_H^t , I_S^t e I_V^t , respectivamente), conforme é descrito na equação (9).

$$SP^t[x] = \begin{cases} 1, & \text{se } \alpha \leq \frac{I_V^t[x]}{B_V^t[x]} \leq \beta \\ & \wedge |I_S^t[x] - B_S^t[x]| \leq \tau_S \\ & \wedge D_H^t[x] \leq \tau_H \\ 0, & \text{caso contrário} \end{cases} \quad (9)$$

onde $\alpha \in [0,1]$, $\beta \in [0,1]$ e D_H^t corresponde a diferença angular entre o valor da camada H da imagem e do fundo da cena, conforme é descrito na equação (10).

$$D_H^t[x] = \min(|I_H^t[x] - B_H^t[x]|, 360 - |I_H^t[x] - B_H^t[x]|) \quad (10)$$

A análise da sensibilidade dos parâmetros das equações (9) e (10) é realizada por Cucchiara *et al.* (2001).

3.2.2 Avaliação do Método de Remoção de Sombra

Para avaliar o método de remoção de sombra utilizado no presente trabalho são utilizados os vídeos 3 e 4 exibidos na Figura 2 em (c) e (d), respectivamente. O primeiro foi escolhido por corresponder a uma cena externa que contem sombras que influenciam o resultado da subtração de fundo, conforme é exibido na Figura 8. O segundo foi escolhido por corresponder a uma cena interna, com sombras projetadas no chão e nas paredes que influenciam o resultado da subtração de fundo, conforme é exibido na Figura 10.

Inicialmente o método de subtração de fundo é aplicado para diferentes valores do limiar k . Após a análise qualitativa e quantitativa do resultado da subtração de fundo para os valores de k avaliados, o método de remoção de sombras é aplicado ao melhor resultado obtido durante a subtração de fundo. Finalmente, as taxas de precisão e revocação são calculadas novamente e as imagens binárias correspondentes ao *foreground* antes e depois da aplicação do método de remoção de sombras são comparadas.

Vídeo 3

A Figura 11 exibe as taxas de precisão e revocação para a variação do parâmetro k durante a subtração de fundo e a variação das mesmas taxas após a aplicação do método de remoção de sombras para $k = 3$. Conforme é exibido na Figura 11 e na Tabela 11, após a aplicação do método de remoção de sombras a taxa de precisão aumentou, devido à remoção dos *pixels* correspondentes às sombras. No entanto, a taxa de revocação diminuiu, pois alguns *pixels* correspondentes às pessoas foram erroneamente eliminados do *foreground*.

	Precisão	Revocação
Subtração de fundo	0.68448	0.79276
Remoção de sombra	0.72044	0.78536

Tabela 11: Vídeo 3 – Revocação x Precisão para a subtração de fundo e remoção de sombras

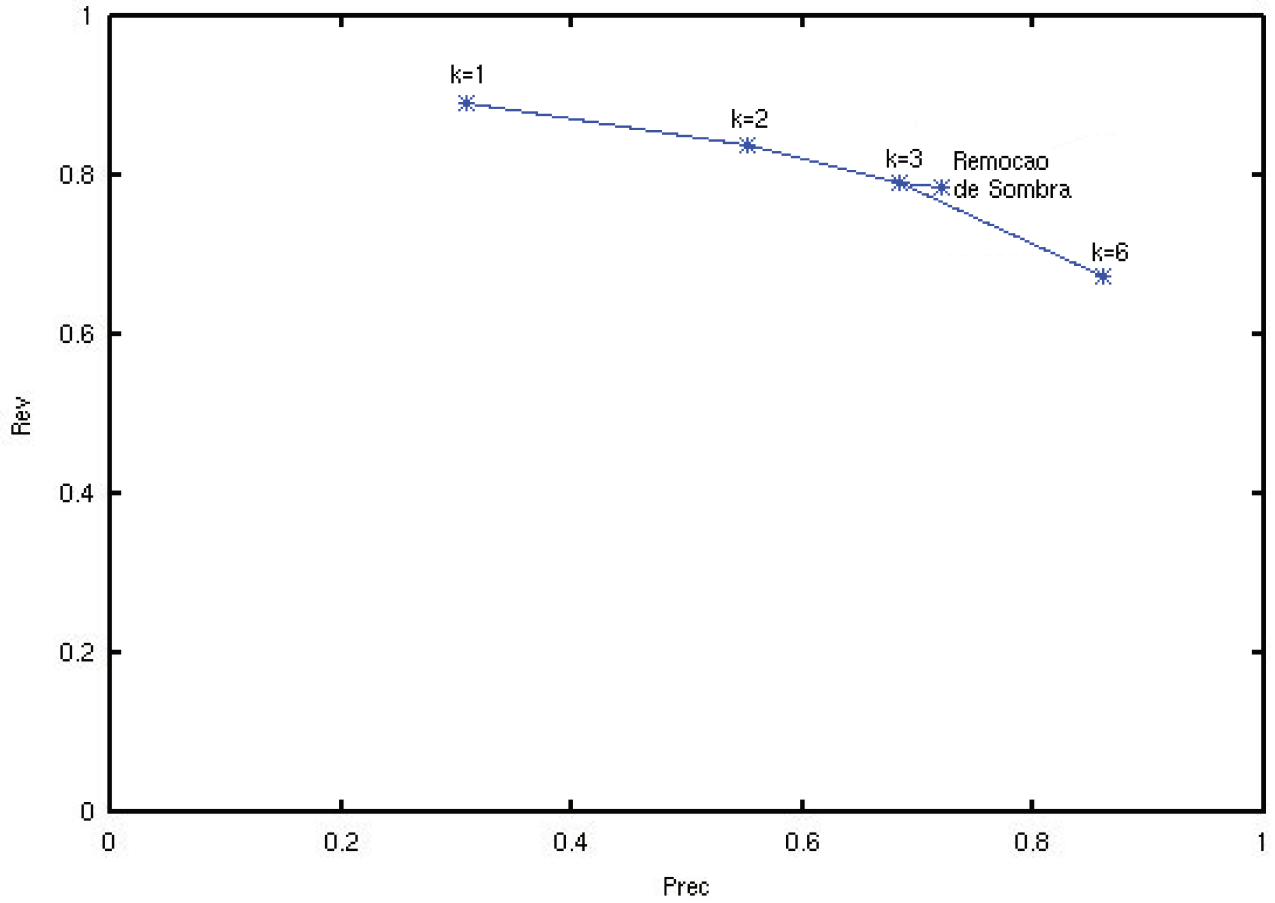


Figura 11: Vídeo 3 – Variação das taxas de revocação e precisão após a remoção de sombras

A Figura 12 exibe o *foreground* detectado antes e depois da remoção das sombras. Conforme é exibido em (b), os *pixels* correspondentes às sombras que causavam a união dos *blobs* em (a) foram removidos do *foreground*.

Para esse vídeo, os parâmetros da equação 9 assumiram os seguintes valores: $\alpha=0.8$, $\beta=1$, $\tau_H=0.05$ e $\tau_S=0.05$.

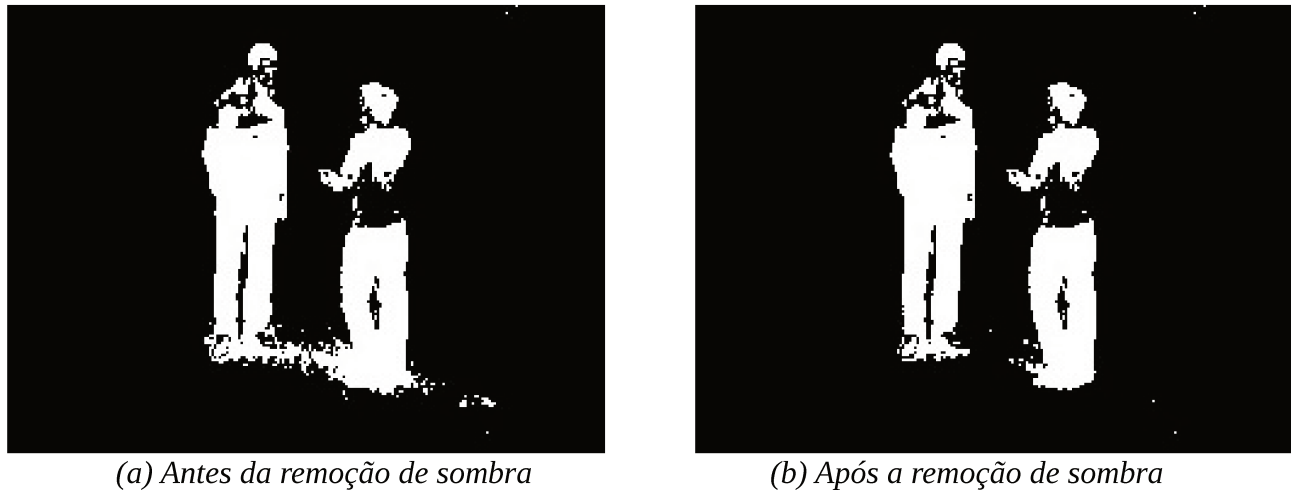


Figura 12: Vídeo 3 – Comparação entre o foreground antes e após a remoção de sombra

Vídeo 4

A Figura 14 exibe as taxas de precisão e revocação para a variação do parâmetro k durante a subtração de fundo e a variação das mesmas taxas após a aplicação do método de remoção de sombras para $k = 2$. A Tabela 12 compara as taxas de precisão e revocação antes e depois a aplicação do método de remoção de sombras.

	Precisão	Revocação
Subtração de fundo	0.44671	0.83635
Remoção de sombra	0.72865	0.78877

Tabela 12: Vídeo 4 – Revocação x Precisão para a subtração de fundo e remoção de sombras

A Figura 13 exibe o *foreground* detectado antes e depois da remoção das sombras. Como aproximadamente 53% do *foreground* detectado em (a) corresponde às sombras, a melhora no resultado exibido em (b) é expressiva.

Para esse vídeo, os parâmetros da equação 9 assumiram os seguintes valores: $\alpha=0.7$, $\beta=1$, $\tau_H=0.9$ e $\tau_S=0.1$.



(a) Antes da remoção de sombra



(b) Após a remoção de sombra

Figura 13: Vídeo 4 – Comparação entre o foreground antes e após a remoção de sombra

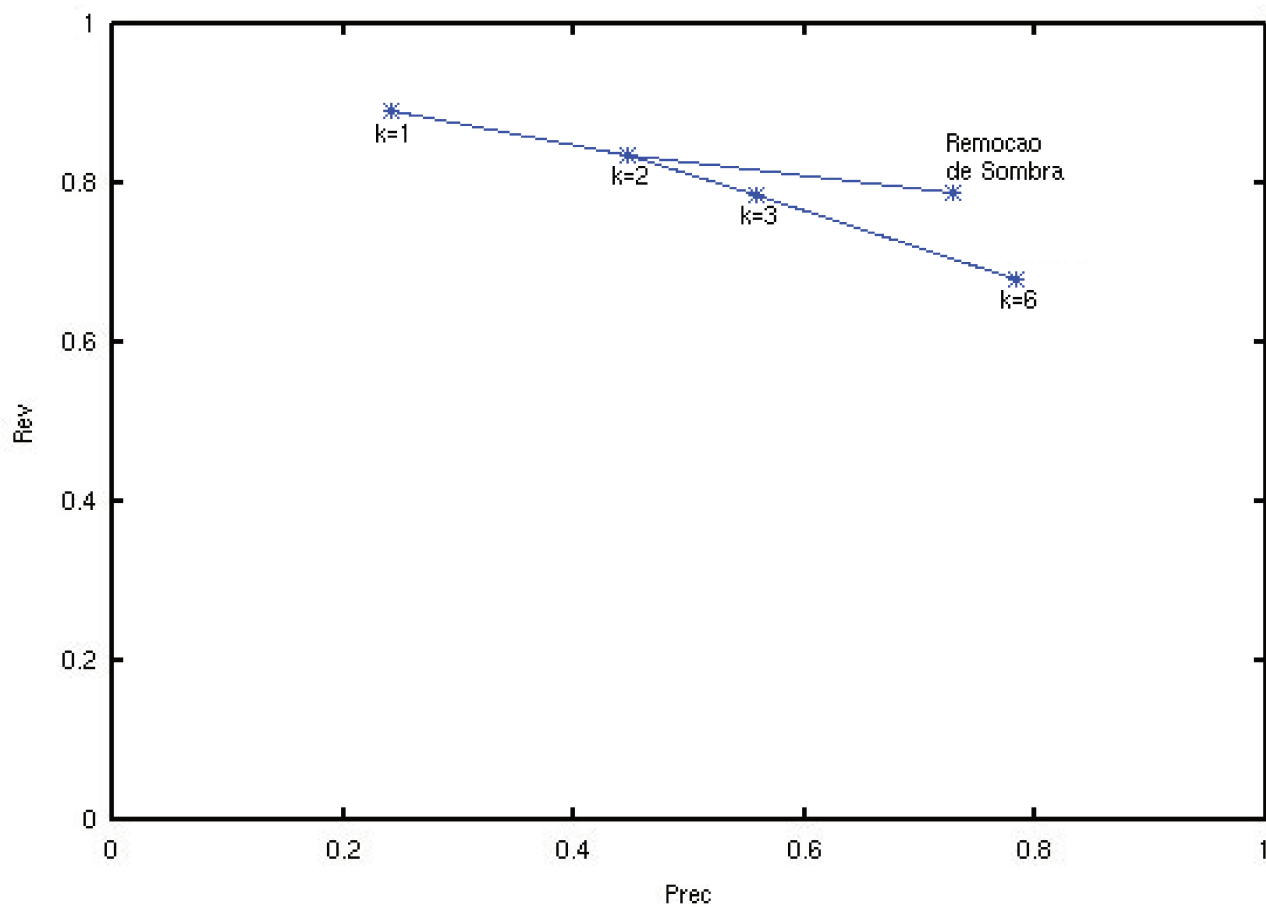


Figura 14: Vídeo 4 – Variação das taxas de revocação e precisão após a remoção de sombras

3.3 Remoção de Ruídos por Morfologia Matemática

Conforme foi apresentado nas seções 3.1 e 3.2, um método de subtração de fundo é utilizado para detectar o *foreground*, e um método de detecção e remoção de sombras é aplicado ao resultado da subtração de fundo para evitar que as sombras das pessoas prejudiquem o rastreamento. No entanto, ao final dessas duas etapas, é possível que o resultado ainda contenha erros, também conhecidos como ruídos, denominados falsos positivos ou falsos negativos:

- Um falso positivo é um *pixel* que deveria ser classificado como *background*, por corresponder ao fundo da cena, mas é erroneamente classificado como *foreground*.
- Um falso negativo é um *pixel* que deveria ser classificado como *foreground*, por corresponder a uma pessoa ou objeto em movimento na cena, mas é erroneamente classificado como *background*.

A Figura 15 exibe em (a) um quadro no qual é possível observar uma pessoa caminhando, e em (b) o resultado da subtração de fundo para esse quadro. Observe que, embora a região correspondente à pessoa tenha sido detectada, ocorreram falsos positivos e falsos negativos no resultado.

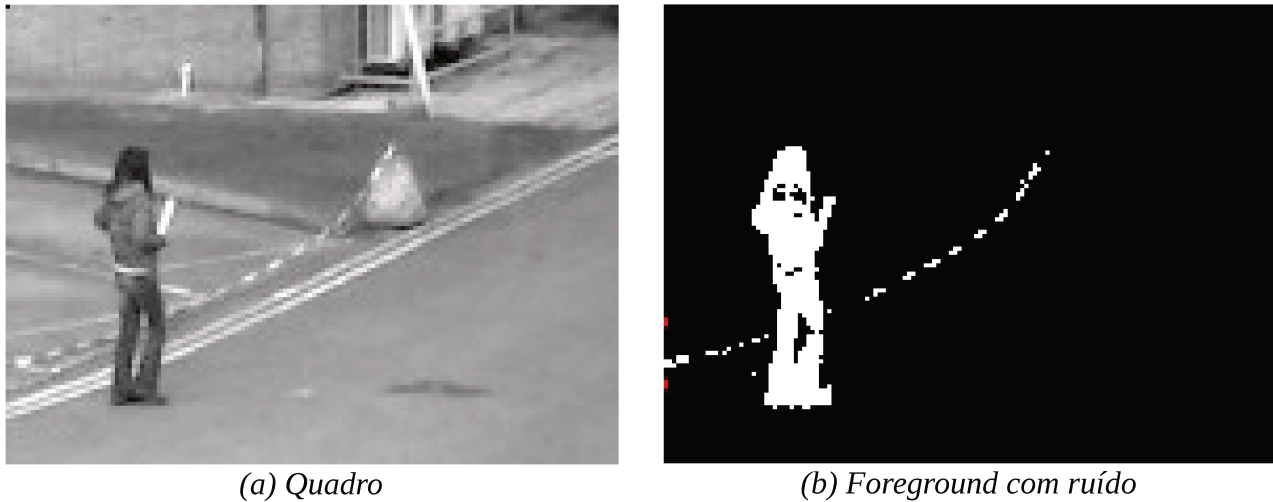


Figura 15: Exemplo de ruídos no foreground

Para evitar que erros presentes no resultado da subtração de fundo interfiram no processo de rastreamento de pessoas, técnicas de processamento de imagens por morfologia matemática são comumente utilizadas. Conforme foi demonstrado em trabalho anterior (Amaral e Tozzi, 2011), essas técnicas destacam-se por serem eficientes ao eliminar esses ruídos e aplicáveis à sistemas em tempo real.

No presente trabalho, as operações de fechamento e abertura por reconstrução são utilizadas para eliminar ruídos da subtração de fundo. Essas operações são descritas a seguir, bem como o efeito obtido quando aplicadas para esse fim. No entanto, antes de abordá-las é necessário apresentar os conceitos básicos do processamento de imagens binárias por morfologia matemática.

3.3.1 Operações Morfológicas

No contexto de morfologia matemática, uma imagem é interpretada como um conjunto de *pixels* reunidos em grupos, e o processamento morfológico consiste em extrair a geometria e topologia desse conjunto desconhecido de *pixels* através de um conjunto conhecido, denominado elemento estruturante. As operações básicas da morfologia matemática são a erosão e a dilatação. As demais operações morfológicas, inclusive o fechamento e a abertura por reconstrução (operações utilizadas no

presente trabalho), são constituídas por essas duas operações.

Erosão

Conforme Dougherty e Lotufo (2003), a erosão pode ser compreendida como o “encolhimento” das regiões da imagem binária. Uma possível forma de interpretar a erosão de uma imagem binária A por um elemento estruturante B é mover a origem do elemento estruturante por toda a imagem a ser erodida e marcar todas as posições da origem do elemento estruturante nas quais o elemento estruturante estiver totalmente contido na imagem. A união de todas as posições marcadas corresponde ao resultado da erosão, denotada por $A \ominus B$, conforme é descrito na equação (11).

$$A \ominus B = \{x : B_x \subset A\} \quad (11)$$

onde B_x é a translação do elemento estruturante B por x , conforme é descrito na equação (12).

$$B_x = \{b + x : b \in B\} \quad (12)$$

O resultado da erosão pode ser observado na equação (13), onde uma matriz binária é erodida por um elemento estruturante definido por uma matriz 3x3 (com a origem no centro da matriz).

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \ominus \begin{bmatrix} 0 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (13)$$

Dilatação

Conforme Dougherty e Lotufo (2003), a dilatação pode ser compreendida como a “expansão” das regiões da imagem binária. Uma possível forma de compreender a dilatação de uma imagem binária A

por um elemento estruturante B é mover a origem do elemento estruturante por toda a imagem a ser dilatada e marcar as posições de cada ponto do elemento estruturante quando sua origem estiver contida na imagem. A união de todos os pontos marcados corresponde ao resultado da dilatação, denotada por $A \oplus B$, conforme é descrito na equação (14).

$$A \oplus B = \bigcup_{a \in A} B_a \quad (14)$$

O resultado da dilatação pode ser observado na equação (15), onde uma matriz binária é dilatada por um elemento estruturante definido por uma matriz 3x3 (com origem no centro da matriz).

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \oplus \begin{bmatrix} 0 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 & 0 \end{bmatrix} \quad (15)$$

Fechamento

A operação de fechamento de uma imagem binária A por um elemento estruturante B , denotada por $A \cdot B$, é definida como uma dilatação seguida de uma erosão, conforme é descrito na equação (16).

$$A \cdot B = (A \oplus B) \ominus B \quad (16)$$

A aplicação prática dessa operação morfológica é “preencher buracos” com área menor que a área de um determinado elemento estruturante existentes nas regiões da imagem binária. O resultado dessa operação pode ser observado na equação (17).

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (17)$$

Abertura

A abertura de uma imagem binária A por um elemento estruturante B , denotada por $A \circ B$, é definida como uma erosão seguida de uma dilatação, conforme é descrito na equação (18).

$$A \circ B = (A \ominus B) \oplus B \quad (18)$$

A aplicação prática dessa operação morfológica é eliminar da imagem binária as regiões com área menor que a área de um determinado elemento estruturante. Na equação (19) a abertura é aplicada a uma matriz binária contendo três regiões. Observe que apenas a região que contém área maior que a área do elemento estruturante manteve-se no resultado, e as demais regiões foram eliminadas.

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \circ \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (19)$$

Dilatação Condicional

Se uma imagem binária for dilatada por um elemento estruturante que contém a própria origem, a imagem será expandida, e a maneira que a imagem expandirá dependerá unicamente do formato do elemento estruturante. Se a dilatação for repetida sucessivamente, a imagem original expandirá sem limites. Em determinadas situações é importante limitar a expansão provocada pela dilatação, e isso

pode ser realizado através da dilatação condicional (Dougherty e Lotufo, 2003).

A dilatação da imagem binária A pelo elemento estruturante B condicionada por C , denotada por $A \oplus_C B$, é descrita na equação (20).

$$A \oplus_C B = (A \oplus B) \cap C \quad (20)$$

onde C é um subconjunto de A .

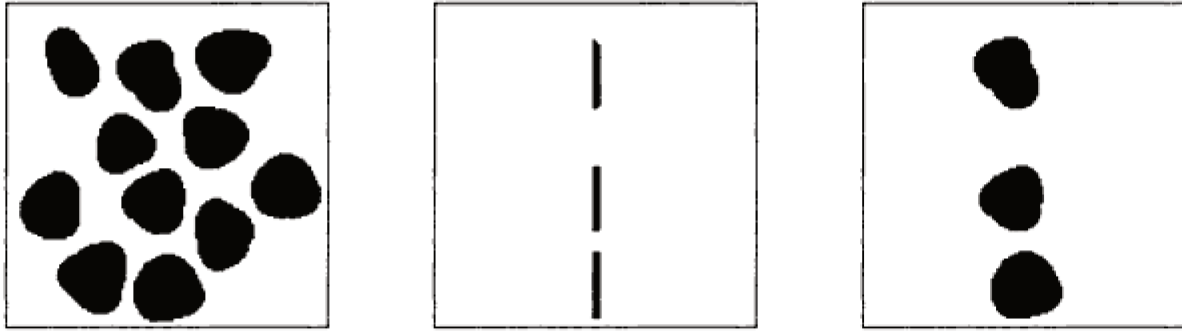
Reconstrução por Marcadores

A reconstrução por marcadores considera a imagem binária A como a união dos componentes conexos contidos em A , conforme equação (21).

$$A = \bigcup_{k=1}^n C_k \quad (21)$$

Sendo M um subconjunto de A , a reconstrução de A pelo marcador M , denotada por $A \Delta M$, é definida como a união dos componentes conexos de A que interseccionam M , conforme é descrito na equação (22). Um exemplo da reconstrução por marcadores é exibido na Figura 16.

$$A \Delta M = \bigcup \{C_k : C_k \cap M \neq \emptyset\} \quad (22)$$



(a) Imagem binária A

(b) Marcador M

(c) Reconstrução de A por M

Figura 16: Reconstrução por marcadores

Fonte: Extraída de Dougherty e Lotufo (2003)

Para relacionar a reconstrução por marcadores com as demais operações morfológicas apresentadas, na equação (23) a reconstrução de A pelo marcador M é formulada como repetidas dilatações do marcador condicionadas pela imagem binária A .

$$A \Delta_B M = (M \oplus_A B)^\infty \quad (23)$$

onde B é um elemento estruturante que define a conectividade considerada entre os *pixels*, e o símbolo $^\infty$ indica que a dilatação ocorre até o momento em que o crescimento ocasionado pelas dilatações estabilize. Comumente é considerada a conectividade de 8 *pixels*, e o elemento estruturante B é definido conforme a equação (24).

$$B = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \quad (24)$$

Abertura por Reconstrução

A operação de reconstrução requer uma imagem binária, um marcador, e um elemento estruturante. A imagem binária é a imagem de entrada na qual a operação é aplicada, o marcador informa quais componentes conexos da imagem binária serão extraídos, e o elemento estruturante

define a conectividade considerada entre os *pixels*. O marcador pode ser definido manualmente ou ser obtido através de outra operação morfológica, como a abertura. A reconstrução que utiliza o resultado da abertura como marcador denomina-se abertura por reconstrução, e é definida na equação (25).

$$A \circ_E B = A \Delta_E (A \circ B) \quad (25)$$

onde A é a imagem binária, E é o elemento estruturante que define a conectividade considerada entre os *pixels*, e B é o elemento estruturante utilizado pela operação de abertura.

3.3.2 Operações Morfológicas Utilizadas no Sistema Proposto

No sistema de rastreamento de pessoas proposto no presente trabalho, as operações morfológicas de fechamento e abertura por reconstrução (apresentadas na seção 3.3.1) são utilizadas para remover ruídos correspondentes a falsos negativos e falsos positivos, respectivamente. As justificativas para a utilização dessas operações morfológicas são apresentadas a seguir.

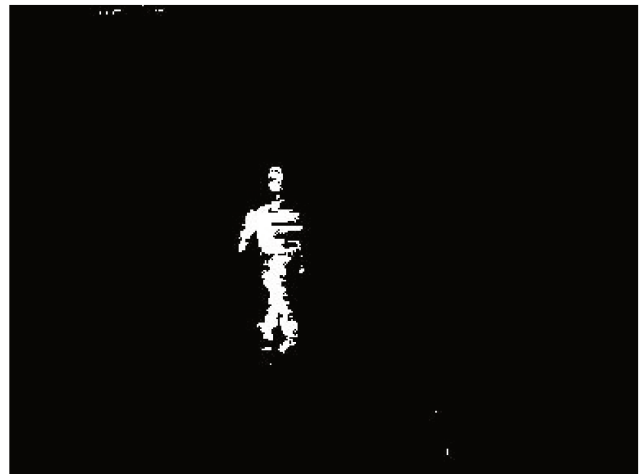
Remoção de falsos negativos

Conforme é apresentado na seção 3.3.1, a aplicação prática do fechamento é “preencher buracos” com área menor que a área de um determinado elemento estruturante existentes nos componentes conexos de uma imagem binária. No presente trabalho o fechamento é aplicado para “preencher” pequenos “buracos” existentes no resultado da subtração de fundo, correspondentes a falsos negativos.

Para exemplificar esse conceito, o resultado do fechamento aplicado ao *foreground* é exibido na Figura 17. Nessa figura é exibido em (a) um quadro de um vídeo no qual é possível observar um homem caminhando, e em (b) é exibido o *foreground* contendo falsos positivos e falsos negativos. Observe que após a aplicação do fechamento (conforme é exibido em (c)) os "buracos" na região correspondente à pessoa menores que o elemento estruturante utilizado foram "preenchidos".



(a) Quadro



(b) Subtração de fundo



(c) Operação de fechamento aplicada ao resultado da subtração de fundo

Figura 17: Operação de Fechamento.

Remoção de falsos positivos

Conforme é apresentado na seção 3.3.1, a aplicação prática da abertura é eliminar da imagem binária as regiões com área menor que a área de um determinado elemento estruturante. No contexto de remoção de ruídos da subtração de fundo, a abertura pode ser aplicada para remover pequenas regiões

do *foreground* correspondentes a falsos positivos.

Para exemplificar esse conceito, a abertura é aplicada ao resultado do fechamento (apresentado anteriormente na Figura 17). O resultado da abertura é exibido na Figura 18 (b). Observe que os ruídos correspondentes a falsos positivos foram eliminados do *foreground* após a aplicação da abertura.



(a) Resultado do fechamento

(b) Abertura aplicada ao resultado do fechamento

Figura 18: Operação de Abertura.

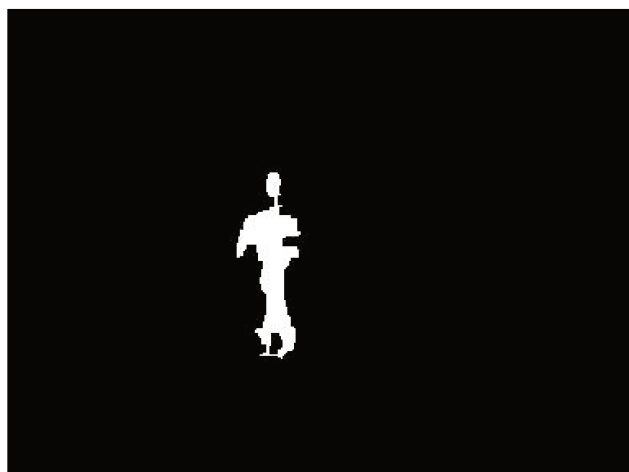
Contudo, observe na Figura 18 (b) que a forma da região correspondente a pessoa foi alterada, fazendo com que a região correspondente a cabeça fosse segmentada da região correspondente ao resto do corpo da pessoa. Para evitar que essa falha ocorra, no presente trabalho a operação de abertura por reconstrução é utilizada.

Conforme é apresentado na seção 3.3.1, a abertura por reconstrução corresponde a uma reconstrução na qual o resultado da abertura é utilizado como marcador. Na prática, todos os componentes conexos do *foreground* que intersejam o resultado da abertura são mantidos, e os demais *pixels* são eliminados.

A Figura 19 exibe em (a) o resultado do fechamento obtido conforme demonstrado na Figura 17, e em (b) o resultado da operação de abertura por reconstrução aplicada ao resultado do fechamento. Observe que os falsos positivos foram eliminados e, ao contrário do ocorrido com a operação de abertura, a forma da região correspondente à pessoa não foi alterada.



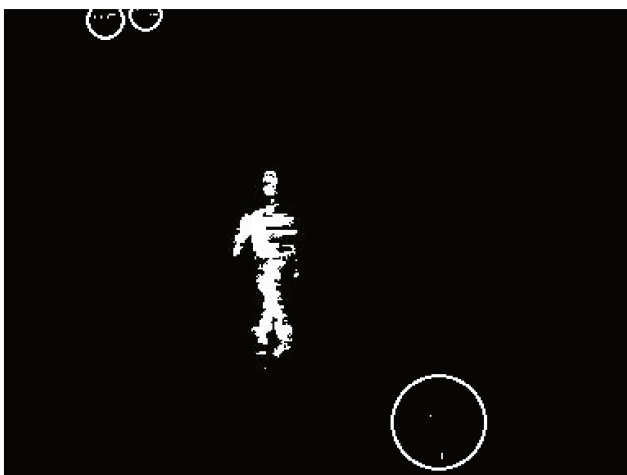
(a) Resultado do fechamento



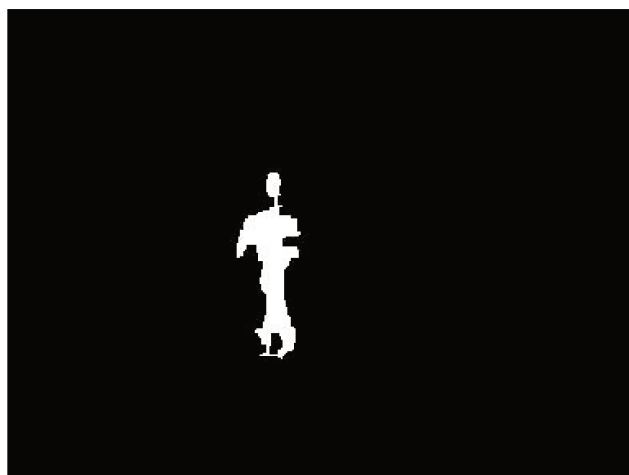
(b) Abertura por reconstrução aplicada ao resultado do fechamento

Figura 19: Abertura por reconstrução.

A Figura 20 exibe a comparação entre o resultado da subtração de fundo e o *foreground* após a aplicação dos operadores morfológicos (fechamento e abertura por reconstrução).



(a) Foreground antes do processamento morfológico



(b) Foreground após o processamento morfológico

Figura 20: Comparação entre o foreground antes e após o processamento morfológico.

Observe na Figura 20 (b) que os falsos positivos marcados em (a) foram removidos, e os “buracos” existentes na região correspondente a pessoa (falsos negativos) foram “preenchidos”.

3.3.3 Avaliação do Processamento Morfológico

Para avaliar a eficiência das operações morfológicas utilizadas no presente trabalho para remover ruídos, são utilizados os vídeos 3 e 4 exibidos na Figura 2 em (c) e (d), respectivamente. Esses dois vídeos foram utilizados para avaliar as etapas de subtração de fundo e remoção de sombras, e são utilizados nessa seção por apresentarem ruídos correspondentes a falsos positivos e negativos.

Para cada vídeo avaliado, o processamento morfológico é realizado logo após a etapa de remoção de sombras. As taxas de precisão e revocação são calculadas novamente e as imagens binárias correspondentes às três etapas do processo de detecção (subtração de fundo, remoção de sombras e remoção de ruídos por morfologia matemática) são comparadas.

Vídeo 3

Para esse vídeo foi utilizado um elemento estruturante circular com 7 *pixels* de tamanho para realizar o fechamento e a abertura por reconstrução.

A Tabela 13 e a Figura 21 exibem a variação das taxas de precisão e revocação após as etapas de subtração de fundo, remoção de sombras e operações morfológicas.

	Precisão	Revocação
Subtração de fundo	0.68448	0.79276
Remoção de sombra	0.72044	0.78536
Morfologia	0.86359	0.86110

Tabela 13: *Vídeo 3 – Revocação x Precisão para subtração de fundo, remoção de sombras e morfologia*

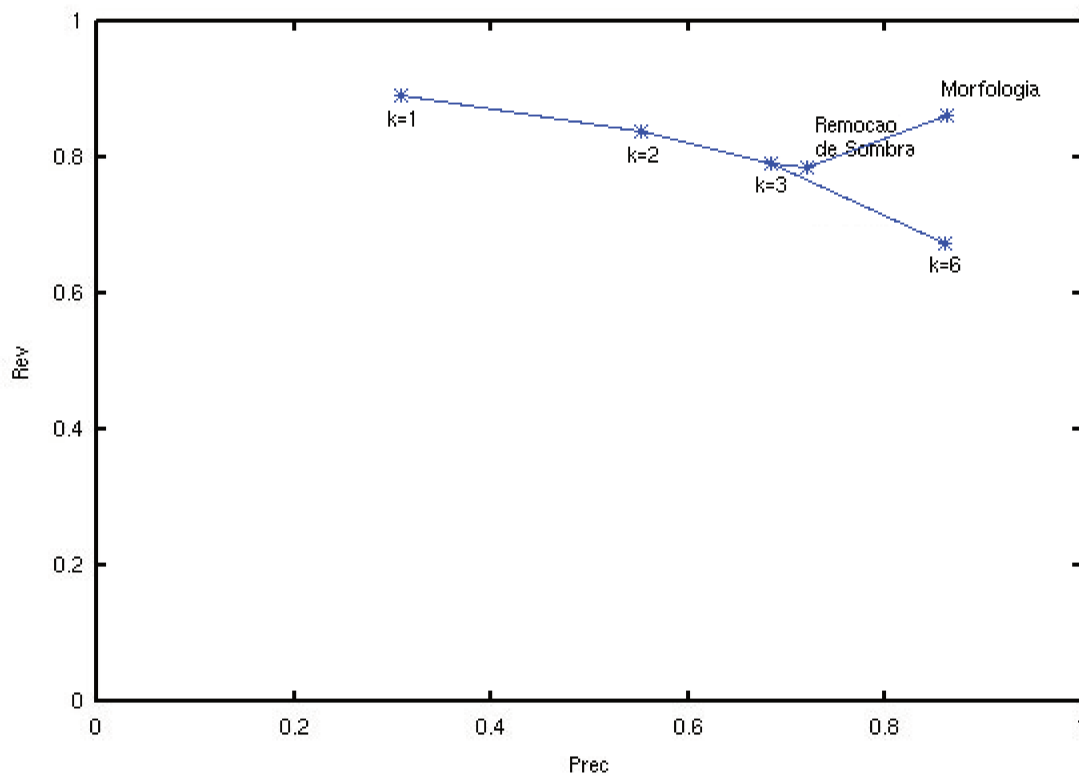
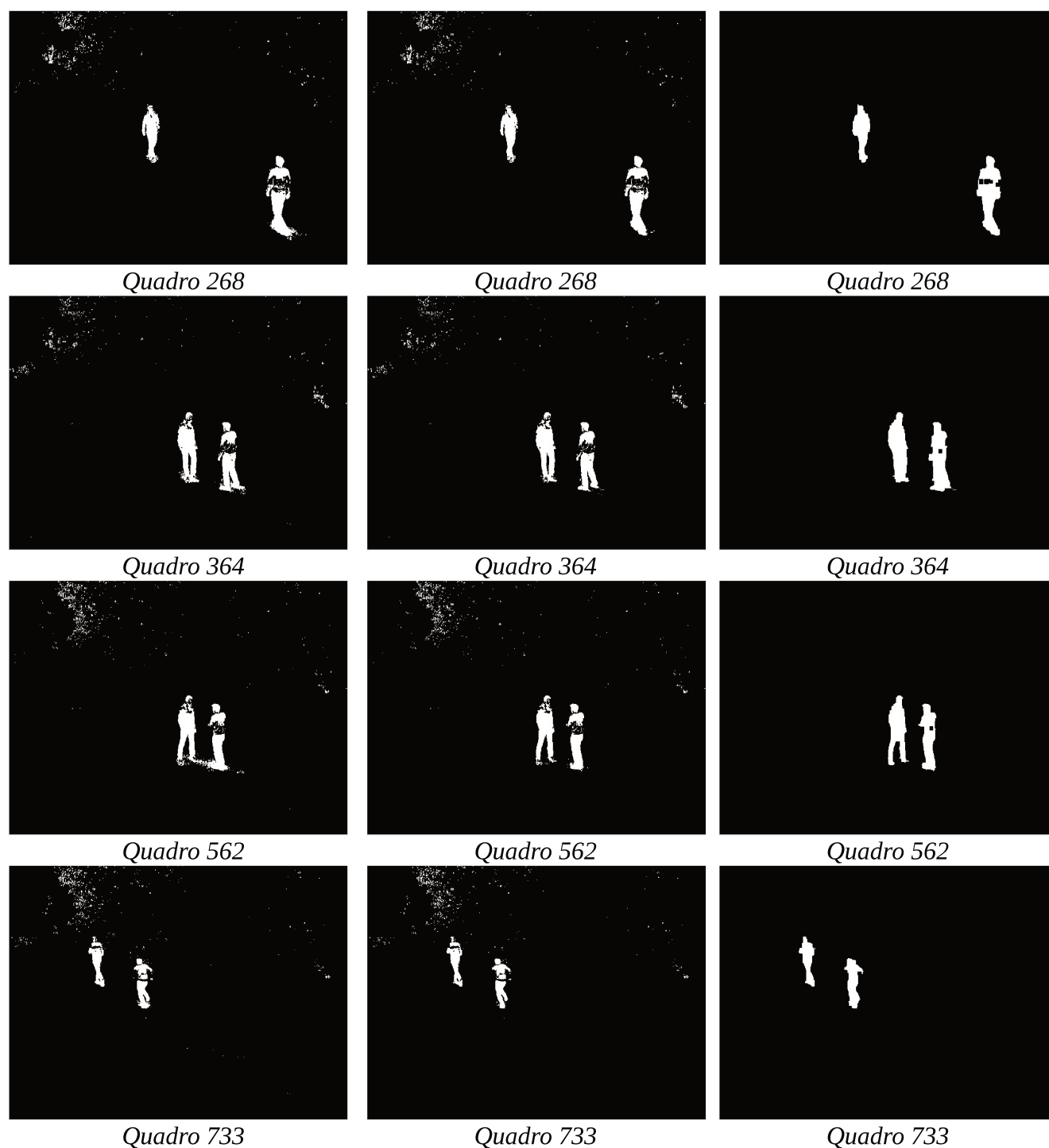


Figura 21: Vídeo 3 – Revocação $\times$ Precisão para subtração de fundo, remoção de sombras e morfologia

A Figura 22 exibe em (a) o *foreground* após a subtração de fundo, em (b) o resultado da remoção de sombra e em (c) o resultado das operações morfológicas. A figura mostra que as pequenas regiões correspondentes às sombras das pessoas, erroneamente identificadas como *foreground* pelo método de subtração de fundo, foram removidas pelo método de remoção de sombra. Como pode ser observado na Tabela 13 e no gráfico da Figura 21, a remoção de ruídos resultou em aumento da precisão. Após as operações morfológicas, os erros correspondentes a falsos positivos e negativos foram corrigidos, e ambas as taxas de precisão e revocação aumentaram.



(a) Subtração de fundo

(b) Remoção de sombra

(c) Morfologia

Figura 22: Vídeo 3 – Foreground após subtração de fundo, remoção de sombras e morfologia

Vídeo 4

Para esse vídeo foi utilizado um elemento estruturante com 3 *pixels* de tamanho para realizar o fechamento e a abertura por reconstrução.

A Tabela 14 e a Figura 23 exibem a variação das taxas de precisão e revocação após as etapas de subtração de fundo, remoção de sombras e operações morfológicas.

	Precisão	Revocação
Subtração de fundo	0.44671	0.83635
Remoção de sombra	0.72865	0.78877
Morfologia	0.73822	0.87318

Tabela 14: *Vídeo 4 – Revocação x Precisão para subtração de fundo, remoção de sombras e morfologia*

A Figura 24 exhibe em (a) o foreground após a subtração de fundo, em (b) o resultado da remoção de sombra e em (c) o resultado das operações morfológicas. Essa figura mostra que as regiões correspondentes às sombras das pessoas, erroneamente identificadas como *foreground* pelo método de subtração de fundo, foram removidas pelo método de remoção de sombra. Como pode ser observado na Tabela 14 e no gráfico da Figura 23 a remoção das sombras resultou em um aumento expressivo da precisão, devido à quantidade de *pixels* correspondentes às sombras projetadas no chão e nas paredes que foram removidos. Após as operações morfológicas, os erros correspondentes a falsos positivos e negativos foram corrigidos, e ambas as taxas de precisão e revocação aumentaram.

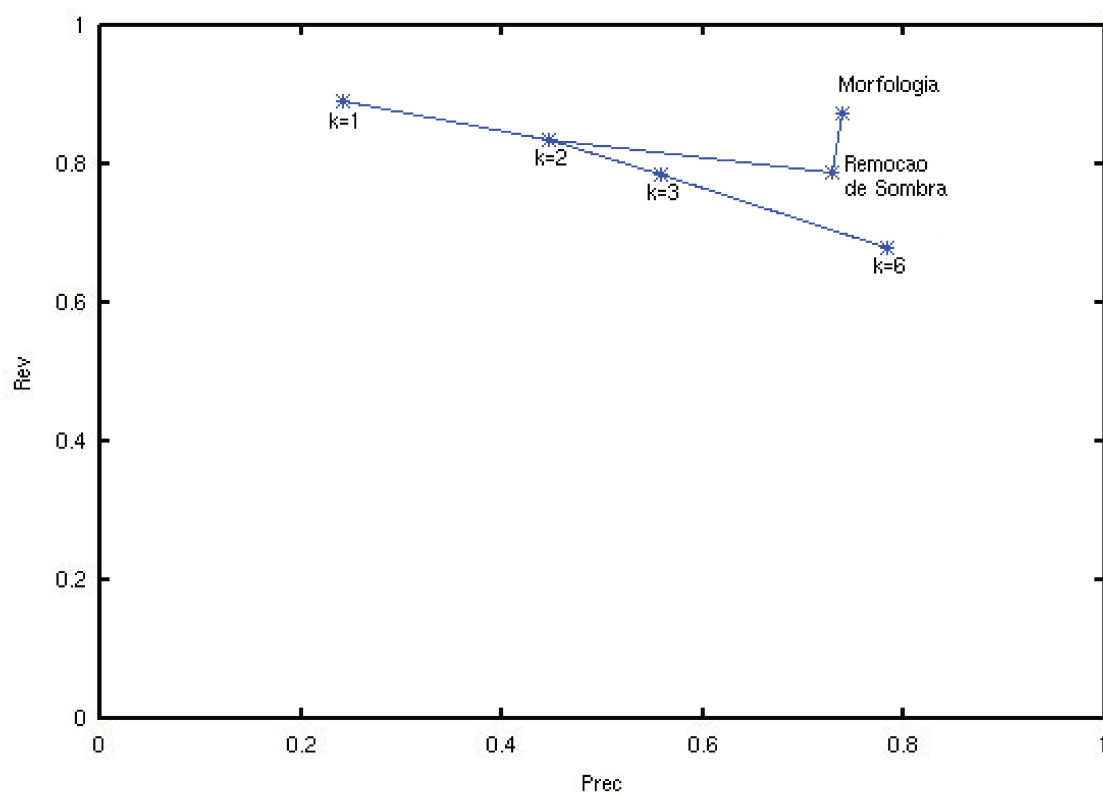
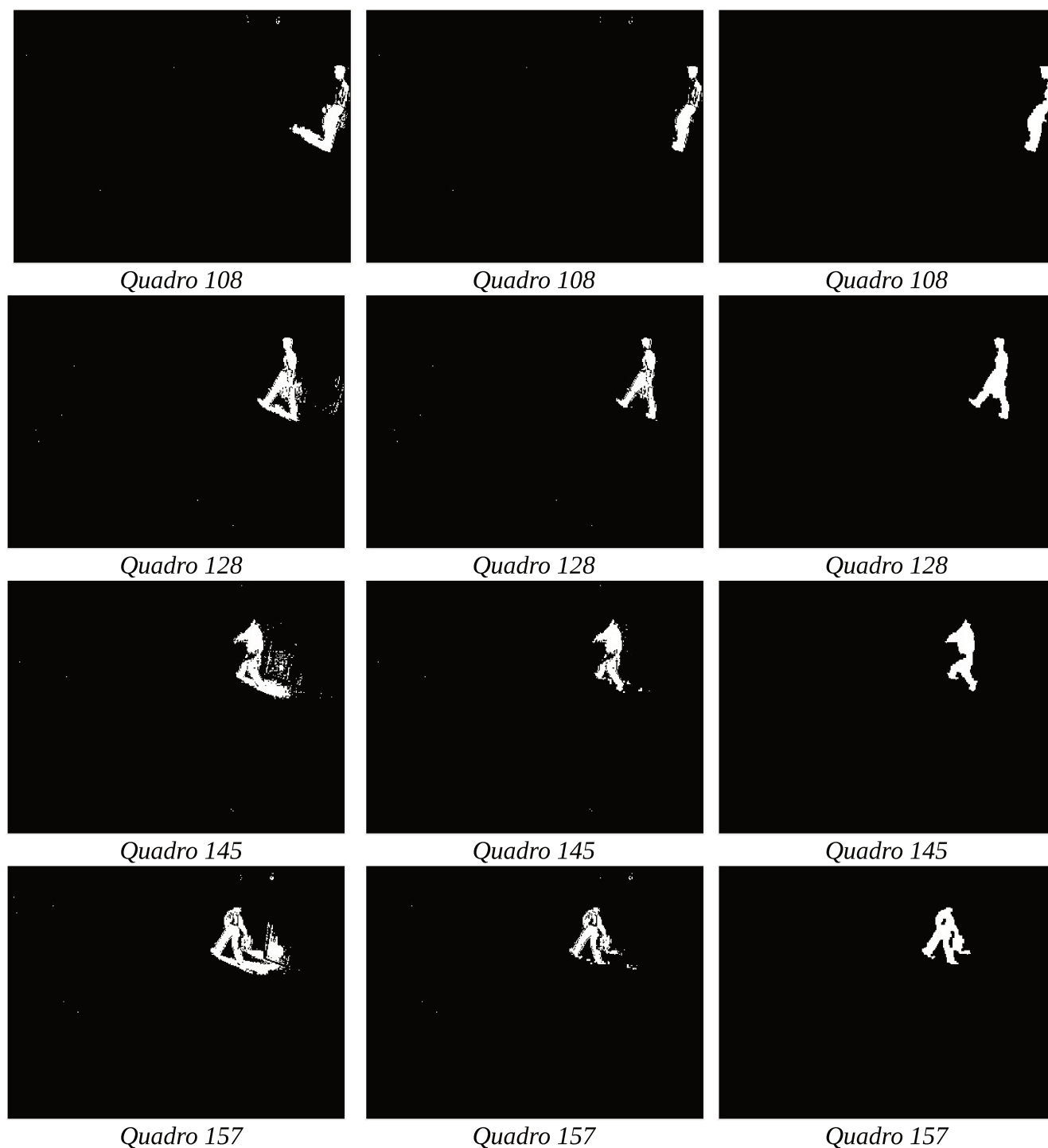


Figura 23: Vídeo 4 – Revocação $\times$ Precisão para subtração de fundo, remoção de sombras e morfologia



(a) Subtração de fundo

(b) Remoção de sombra

(c) Morfologia

Figura 24: Vídeo 4 – Foreground após subtração de fundo, remoção de sombras e morfologia

Capítulo 4

Extração de Informações

O capítulo anterior apresentou as técnicas utilizadas no presente trabalho para detectar o *foreground*, representado por uma imagem binária na qual os *pixels* com valor igual ao zero correspondem ao fundo da cena e os demais formam regiões conexas (*blobs*) que correspondem aos alvos seguidos. A partir de um conjunto de informações extraídas dos *blobs* detectados ao longo do vídeo, as trajetórias são determinadas. O presente capítulo apresenta as técnicas utilizadas para extrair informações dos *blobs* detectados.

4.1 Posição

A trajetória de uma pessoa deslocando-se na cena é determinada com base na variação de sua posição no decorrer do tempo. Logo, a posição é uma informação essencial para realizar o rastreamento. No presente trabalho, a posição de um alvo seguido é representada pelo centroide (x_c, y_c) do *blob* B . Dado um conjunto de n *pixels* $p_i = (x_i, y_i) (i=1, \dots, n)$ contidos em B , a coordenada (x_c, y_c) é definida conforme as equações (26) e (27).

$$x_c = \frac{1}{n} \sum_{i=1}^n x_i \quad (26)$$

$$y_c = \frac{1}{n} \sum_{i=1}^n y_i \quad (27)$$

4.2 Bounding Box

Em uma imagem binária a *bounding box* (caixa delimitadora) de um *blob* é o retângulo de menor área que contém todos os *pixels* do *blob* e que possui faces paralelas aos planos cartesianos. A Figura 25 exibe a região correspondente a uma pessoa detectada na imagem e a *bounding box* dessa região. Em uma imagem binária, a *bounding box* pode ser facilmente representada por duas coordenadas, por exemplo, as coordenadas do canto superior esquerdo e do canto inferior direito da *bounding box*.

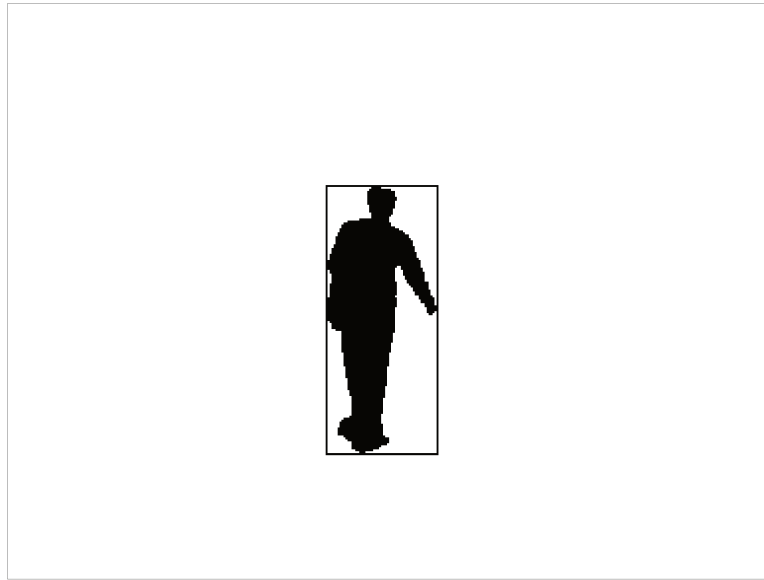


Figura 25: Bounding box da região correspondente a uma pessoa

4.3 Histograma de Cor

Em computação, as cores são descritas através de modelos matemáticos, e o mais comum entre eles é o modelo RGB, que é utilizado em dispositivos eletrônicos como em monitores de TV e computador e em câmeras digitais. Conforme Gonzalez e Woods (2002), no modelo RGB cada cor aparece nos seus componentes espectrais primários de vermelho, verde e azul. Este modelo é baseado em um sistema de coordenadas cartesianas. O subespaço de cores de interesse é o cubo exibido na Figura 26, no qual os valores RGB estão nos três cantos; ciano, magenta e amarelo estão nos outros três cantos; preto está na origem; e branco está no canto mais distante da origem. Neste modelo, a escala de cinza (pontos de iguais valores RGB) estende-se do preto até o branco ao longo da linha que junta estes dois pontos. As diferentes cores nesse modelo são pontos sobre ou dentro do cubo, e são definidas por vetores estendendo-se a partir da origem.

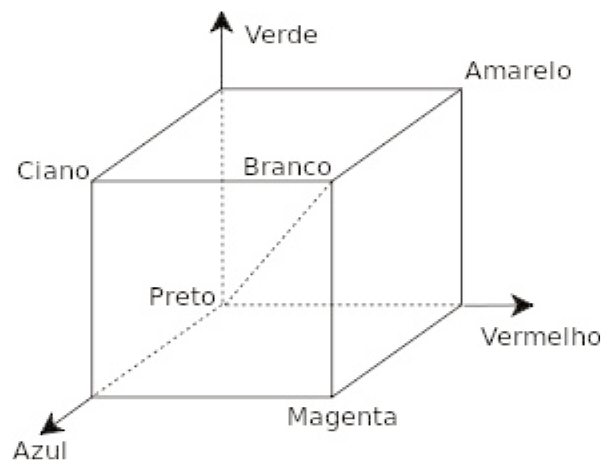


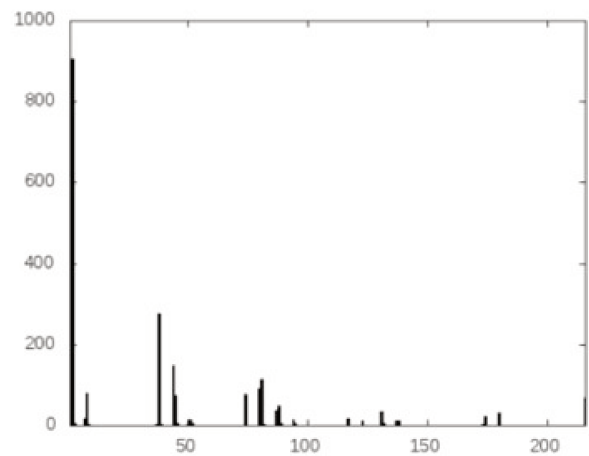
Figura 26: Modelo RGB

Considerando o espaço de cores exibido na Figura 26, os *pixels* de um *blob* associado a uma pessoa são um conjunto de pontos espalhados nesse espaço, e a posição desses pontos depende exclusivamente das cores das roupas e da pele da pessoa. Logo, essa informação relacionada às cores pode ser utilizada para associar uma pessoa a um *blob*.

No presente trabalho as cores dos *pixels* dos *blobs* no modelo RGB são extraídas e utilizadas para construir um histograma de cor. O histograma de cor é criado através da discretização das cores em faixas de cores e contagem do número de *pixels* correspondentes a cada faixa (Swain e Ballard, 1991). A Figura 27 exibe um exemplo onde um histograma de cor é gerado a partir dos *pixels* da região correspondente à pessoa presente na Figura 27 (a). Para calcular esse histograma, os canais RGB dos *pixels* foram discretizados em 6 faixas (0-42, 43-84, 85-126, 127-168, 169-210 e 211-255). Após a discretização, a quantidade de *pixels* em cada uma das cores possíveis foi contada, gerando assim o histograma exibido na Figura 27 (b).



(a) Quadro



(b) Histograma de cores da pessoa

Figura 27: Histograma de cores de uma pessoa

4.4 Detecção de Cabeças

Além de medições realizadas a partir dos *blobs* detectados, informações contextuais podem ser extraídas para serem utilizadas durante o rastreamento. No contexto do rastreamento de pessoas, a informação relacionada à posição das cabeças das pessoas presentes no vídeo é relevante, e pode ser utilizada para determinar a posição das pessoas durante oclusões.

No sistema de rastreamento proposto por Haritaoglu et al (2000), o método de detecção de

cabeças Hydra (Haritaoglu, 1999) é utilizado. Esse método assume que, devido à estrutura e formato do corpo humano, é comum que algumas partes do corpo (como mãos, pés e cabeça) apareçam nas extremidades, ou pontos máximos, na curvatura das bordas do *blob*. Assim, o método busca esses pontos para localizar possíveis posições de cabeças nas bordas do *blob*. Para tal, é utilizado um algoritmo de fecho convexo (*Graham scan*) e um algoritmo de detecção de cantos. Quando os vértices são encontrados, a vizinhança entre os vértices é analisada, e os vértices que não apresentam a curvatura esperada para uma cabeça são eliminados. Finalmente, o histograma projetado na vertical é utilizados para filtrar o resultado. Um vértice é considerado uma cabeça se, para aquele vértice, houver um número de *pixels* superior a um limite predefinido no histograma. Isso reforçar a condição de que uma cabeça deve estar acima de um torso. No entanto, esse limite predefinido é uma restrição do método, pois não é previsto que tamanho do torço de uma pessoa possa variar de acordo com a distância da pessoa em relação à câmera.

Zhao e Nevatia (2004) consideram que um ponto pode ser uma cabeça se for o ponto máximo na borda superior do *blob* dentro de um limite definido pelo tamanho médio de uma cabeça. Para verificar se há um corpo abaixo das possíveis cabeças (os picos detectados), um modelo em forma de elipse do tamanho esperado para uma pessoa é posicionado abaixo dos picos detectados. Os picos que não possuem suficientes *pixels* do *foreground* dentro do modelo são desconsiderados, e os demais são classificados como cabeças. Uma restrição desse método é a suposição com relação ao tamanho da cabeça de uma pessoa. Esse método não considera que o tamanho de uma cabeça na imagem pode variar consideravelmente de acordo com a distância entre a pessoa e a câmera.

Zhang et al (2008) propõem um algoritmo de detecção de cabeças baseado nas cores do cabelo e pele e em restrições geométricas. Conforme Zhang et al (2008), a cor natural do cabelo de um ser humano pode ser classificada em uma categoria, como preto, loiro, castanho, etc. O mesmo ocorre com a cor da pele. Desse modo, uma mistura de gaussianas é criada para modelar as categorias de cabelo e pele a partir de um conjunto de amostras. Esses modelos são utilizados para identificar os *pixels* de uma imagem que possuem a coloração esperada para o cabelo ou pele. A partir desses *pixels*, operadores morfológicos e algoritmos de rotulação são utilizados para extrair os componentes conexos. Finalmente, uma elipse é ajustada em cada componente conexo para identificar aqueles que possuem o formato esperado para uma cabeça. Uma restrição desse método é a necessidade de amostras de imagens de cabelo e pele para construir os modelos gaussianos. Falhas na detecção podem ocorrer em

casos que fujam do padrão identificado nas amostras (ex: cabelos com cores incomuns ou pessoa utilizando máscaras). Por depender diretamente da cor, esse método não é aplicável a vídeos em tons de cinza.

Merad et al (2010) propõem um método de detecção de cabeças baseado no processo de esqueletização. Inicialmente, a região conexa correspondente à pessoa é detectada através de subtração de fundo, e suavizada no domínio das frequências (Fourier). Então, o esqueleto é computado aplicando triangulação de Delaunay na silhueta da imagem binária correspondente à pessoa (ou ao grupo de pessoas). Para detectar as cabeças, os segmentos correspondentes aos nós extremos (aqueles que possuem um único vizinho) são extraídos, e os que possuem inclinação com relação ao eixo vertical dentro de um limite $[-\theta, \theta]$ são classificados como cabeças. Uma restrição desse método refere-se ao fato de que todos os segmentos verticais no esqueleto do *blob* são identificados como cabeças. Essa abordagem não se certifica que exista um torso abaixo das cabeças detectadas, como é feito por Haritaoglu et al. (2000) e Zhao e Nevatia (2004).

Um método de esqueletização “estrela” é utilizado por Fujiyoshi e Lipton (1998) para analisar a pose de uma pessoa. Esse processo consiste nos seguintes passos: 1) O *foreground* é extraído; 2) Ruídos são removidos por morfologia matemática; 3) O centroide da região é determinado; 4) As distâncias entre o centroide e os pontos da borda são calculadas; 5) O sinal obtido é suavizado através de filtros lineares ou filtragem passa baixa no domínio da frequência; 6) Máximos locais são obtidos, encontrando os pontos onde a derivada iguala-se a zero; 7) O esqueleto “estrela” é construído unindo os pontos encontrados ao centroide. Conforme Fujiyoshi e Lipton (1998), essa abordagem para obter o esqueleto demanda menos recurso computacional que as técnicas mais comuns, como a transformada de distância e o afinamento (*thinning*), e oferece informações suficientes para identificar as partes do corpo da pessoa, como os pés e a cabeça.

Win Kong et al (2012) propõem um método semelhante para detectar a posição das mãos de uma pessoa a partir da análise das bordas da região detectada. Esse processo consiste nos seguintes passos: 1) A silhueta da pessoa é detectada; 2) Ruídos são removidos por dilatações e erosões; 3) O ponto inicial do escaneamento é definido, e as bordas são detectadas através do detector de bordas de Canny; 4) É criado um gráfico a partir da distância euclidiana entre o centroide e os pontos da borda; 5) Os picos no gráfico são encontrados. Ao final desse processo o primeiro e o terceiro pico significativo no gráfico representam as mãos, e os demais a cabeça e as pernas. Adaptações nesse método podem ser

feitas para detectar a posição da cabeça no *blob*.

4.4.1 Método de Detecção de Cabeças Utilizado no Sistema Proposto

O método de detecção de cabeças utilizado no presente trabalho é baseado em Win Kong et al. (2012) e Haritaoglu et al. (2000). No presente trabalho, considera-se que um *pixel* corresponde à posição de uma cabeça se duas condições forem satisfeitas:

1. O *pixel* deve corresponder a um pico na extremidade superior do *blob*.
2. Devem existir suficientes *pixels* abaixo da cabeça, correspondentes ao corpo da pessoa.

Para verificar essas duas condições, o método proposto é dividido em 6 passos: 1) detecção do centroide; 2) detecção das bordas; 3) cálculo das distâncias entre os *pixels* da borda e o centroide; 4) detecção das proeminências e concavidades; 5) verificação de vizinhança; 6) verificação de altura. Esses passos são descritos a seguir, e a Figura 28 exemplifica a aplicação do método.

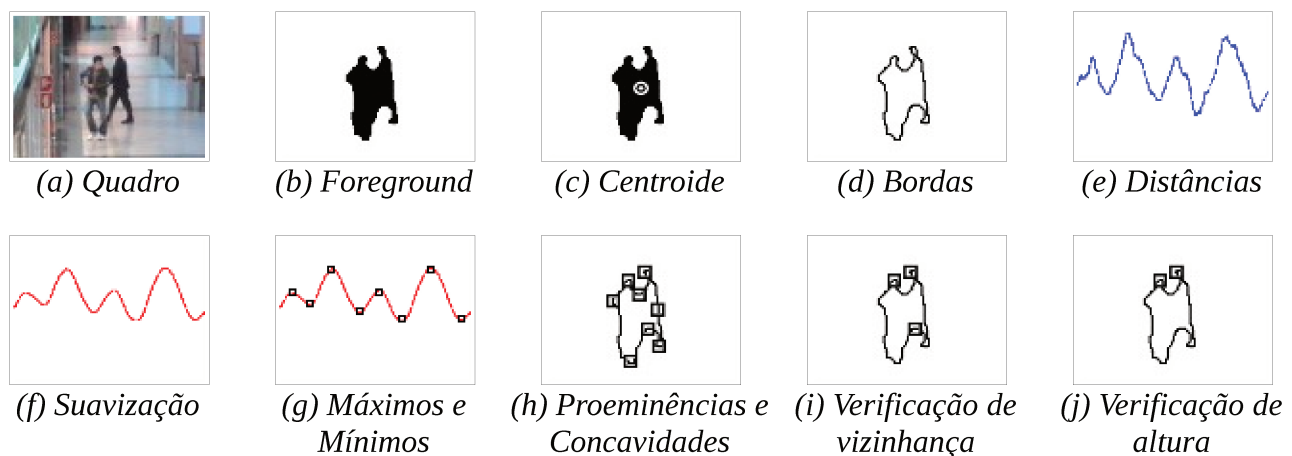


Figura 28: Etapas do método de detecção de cabeças

Passo 1: Detecção do centroide do blob

Inicialmente determina-se o centroide do *blob*, conforme é exibido na Figura 28 de (a) à (c).

Passo 2: Detecção das bordas do blob

Em seguida, as bordas do *blob* são detectadas através do algoritmo de traçado de contorno Square-Tracing (Cheriet et al, 2007), conforme é exibido na Figura 28 (d).

Passo 3: Cálculo das distâncias entre os pixels da borda e o centroide

Um gráfico é criado, onde o eixo x corresponde aos *pixels* da borda e o eixo y corresponde à distância euclidiana entre o *pixel* e o centroide, conforme é exibido na Figura 28 (e). O gráfico obtido é suavizado utilizando a média móvel simples, conforme é exibido na Figura 28 (f).

Passo 4: Detecção das proeminências e concavidades

Os pontos do gráfico exibido na Figura 28 (f) onde ocorre variação no sinal da derivada são identificados. Esses pontos correspondem aos máximos e mínimos locais do gráfico, conforme é exibido na Figura 28 (g). Os *pixels* referentes aos pontos máximos e mínimos são marcados na imagem. Esses *pixels* correspondem às proeminências e concavidades mais significativas do *blob*, conforme é exibido na Figura 28 (h).

Passo 5: Verificação de vizinhança

Cada *pixel* marcado na Figura 28 (h) é comparado aos seus dois vizinhos (o *pixel* anterior e o próximo, considerando a sequencia utilizada pelo algoritmo *Square-Tracing*). Os *pixels* que estiverem abaixo (considerando o eixo y) de um de seus vizinhos são eliminados. O resultado dessa verificação pode ser observado na Figura 28 (i).

Passo 6: Verificação de altura

Para cada *pixel* h candidato a representar uma cabeça (*pixels* restantes da verificação de vizinhança), calcula-se a sua altura a_h , definida como a distância entre h e o último *pixel* contido no *blob* e abaixo do ponto h . Em seguida, são desconsiderados todos os *pixels* que não satisfazem a inequação (28).

$$m - a_h < d.m \quad (28)$$

onde m é a maior altura calculada, e d é um valor entre 0 e 1 que determina um nível de *threshold*. Os *pixels* que satisfazem a inequação (28) são classificados como posições de cabeças (o resultado dessa verificação pode ser observado na Figura 28 (j)).

A inequação (28) reforça a restrição de que uma cabeça deve estar acima de um torso, ou seja, h é uma cabeça somente se a_h é superior a um limite. Essa restrição evita falsas detecções, como no exemplo exibido na Figura 28, onde os picos na extremidade superior do *blob* correspondentes às mãos da pessoa não são identificados como cabeças por não satisfazerem a inequação (28).



Figura 29: Pessoa com os braços erguidos

Fonte: Adaptada de Haritaoglu et al. (2000)

Dada a posição h correspondente a uma cabeça, a posição aproximada dessa pessoa pode ser

definida como o ponto médio entre h e o último *pixel* do *blob* abaixo do ponto h .

Com relação aos métodos de detecção de cabeças revisados no presente trabalho, o método proposto possui as seguintes vantagens:

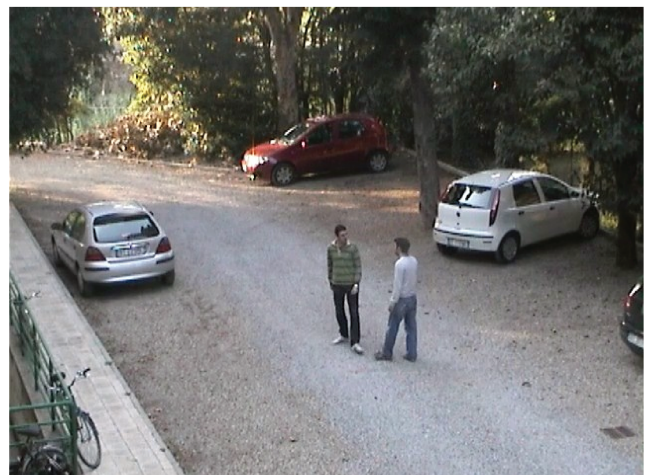
- É aplicável a vídeos em tons de cinza.
- Não necessita de amostra para construir modelos.
- Não armazena modelos em memória.
- Certifica-se que exista um torso abaixo das cabeças detectadas.
- Não faz suposições com relação ao tamanho esperado para uma cabeça.
- Não faz suposições com relação à altura esperado para uma pessoa.

4.4.2 Avaliação do Método de Detecção de Cabeças

Para avaliar o método de detecção de cabeças proposto, análises quantitativas e qualitativas são aplicadas a dois vídeos pertencentes às bases CAVIAR e Visor (exibidos na Figura 30). Na análise quantitativa o método é aplicado a 50 quadros que contém pessoas, e é calculado o total de cabeças presentes nos quadros, o total de cabeças detectas pelo sistema e a quantidade de falsos positivos e negativos. Na análise qualitativa, imagens correspondentes às principais etapas do processo de detecção são exibidas para determinados quadros.



(a) Vídeo 1: CAVIAR-OneStopEnter2cor



(b) Vídeo 2: visor_1267122885367_video_seq1

Figura 30: Amostras de quadros dos vídeos utilizados para avaliar o método de detecção de cabeças

Vídeo 1

A Tabela 15 corresponde aos dados obtidos a partir do processamento de 50 quadros. Nessa tabela são exibidos o total de cabeças presentes nos quadros, cabeças corretamente detectadas pelo sistema, falsos positivos e falsos negativos.

Total de cabeças nos quadros	94
Cabeças corretamente detectadas	86
Falsos positivos	16
Falsos negativos	8

Tabela 15: Análise quantitativa da detecção de cabeças

Conforme é exibido na Tabela 15, 86 cabeças foram detectadas corretamente. Esse valor corresponde a 91,5% do total de cabeças presentes nos quadros. As cabeças foram detectadas em situações nas quais as pessoas estavam isoladas, conforme é exibido na Figura 31, e em casos de junção de *blobs*, conforme é exibido na Figura 32.

A Figura 31 mostra ainda que o método detecta as cabeças em situações nas quais a distância entre a pessoa e a câmera é diferente para as pessoas presentes no quadro, ou seja, quando o tamanho do *blob* é diferente para cada pessoa.

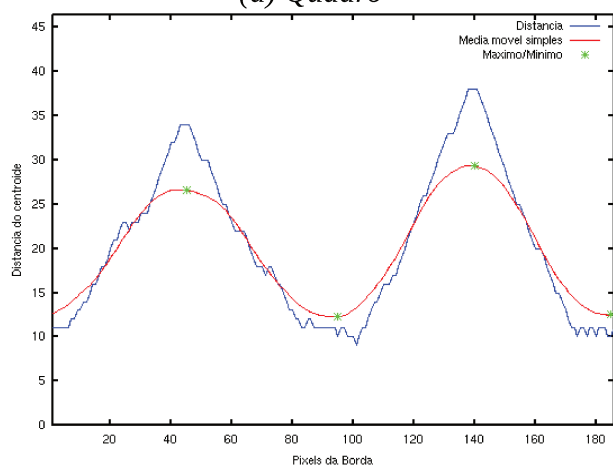
Ocorreram falsos negativos em situações nas quais a posição da cabeça da pessoa não correspondeu a um pico na extremidade superior do *blob*, conforme exemplo exibido na Figura 33. Falhas ocorridas durante a obtenção do *foreground* ocasionaram falsos positivos, conforme exemplo exibido na Figura 34.



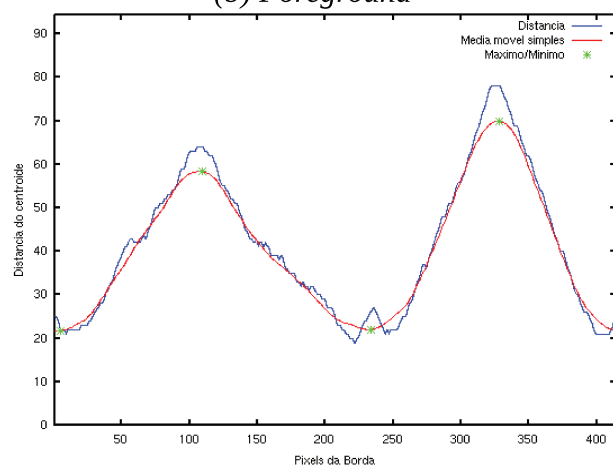
(a) Quadro



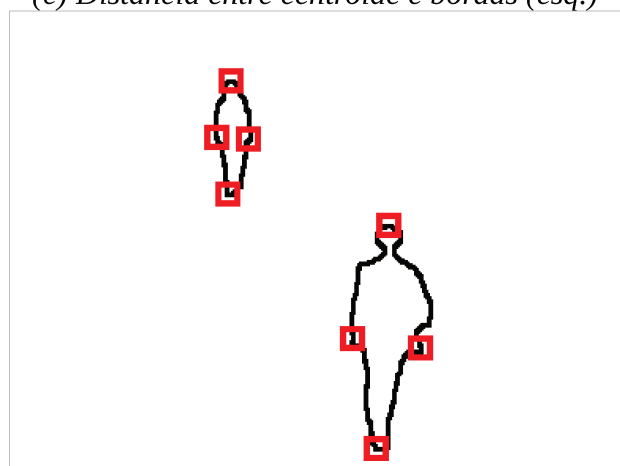
(b) Foreground



(c) Distância entre centroide e bordas (esq.)



(d) Distância entre centroide e bordas (dir.)



(e) Cantos significativos

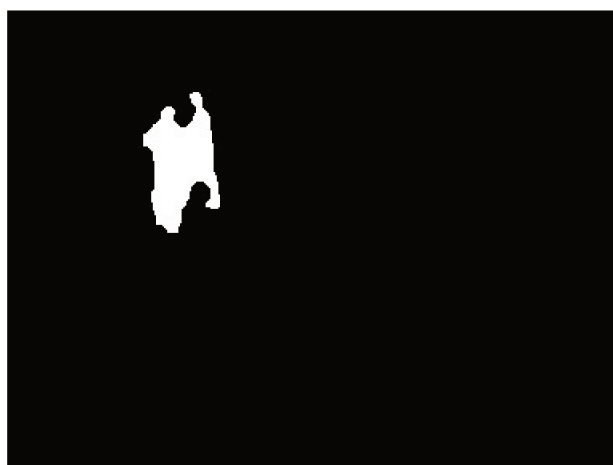


(f) Cabeças detectadas

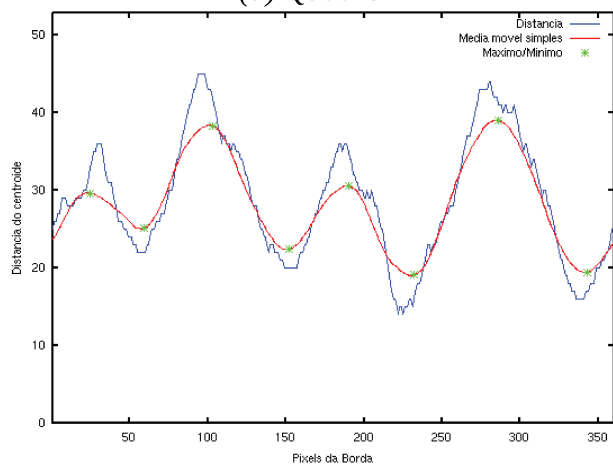
Figura 31: Resultado do método de detecção de cabeças



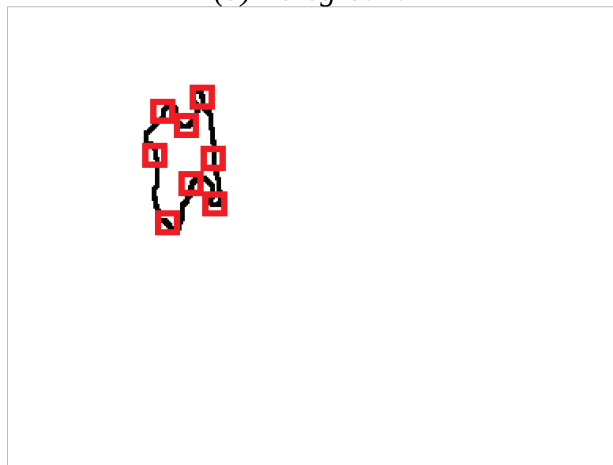
(a) Quadro



(b) Foreground



(c) Distância entre centroide e bordas



(d) Cantos significativos

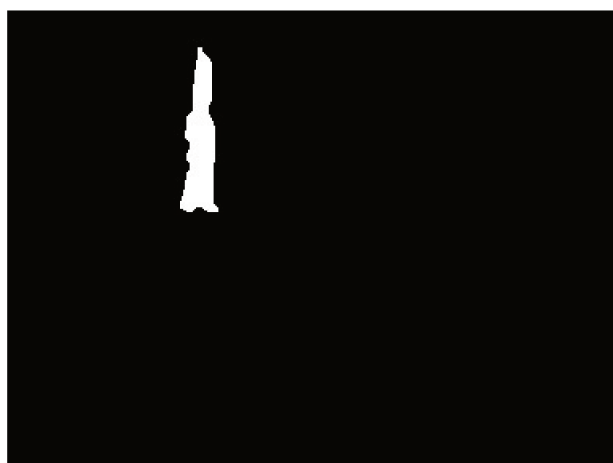


(e) Cabeças detectadas

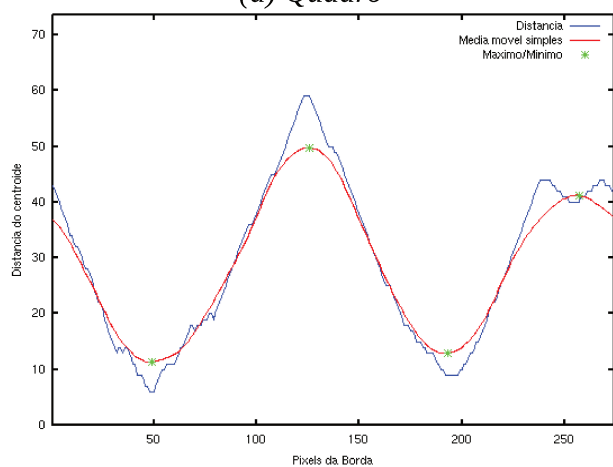
Figura 32: Resultado do método de detecção de cabeças durante união de blobs



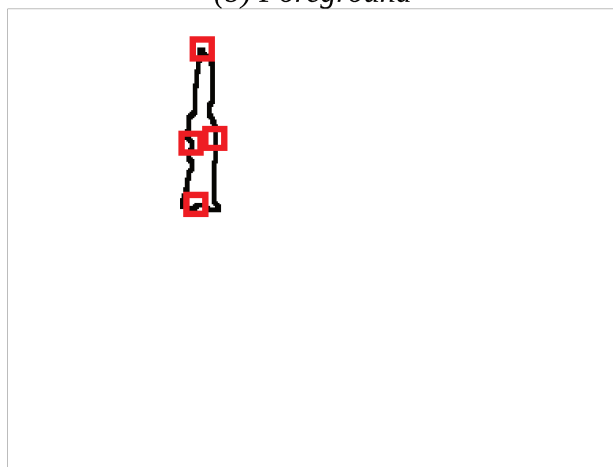
(a) Quadro



(b) Foreground



(c) Distância entre centroide e bordas



(d) Cantos significativos



(e) Cabeças detectadas

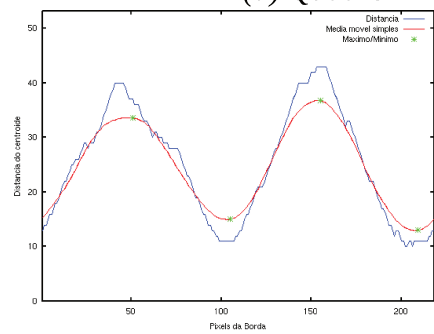
Figura 33: Exemplo de falso negativo ocorrido durante a detecção de cabeças



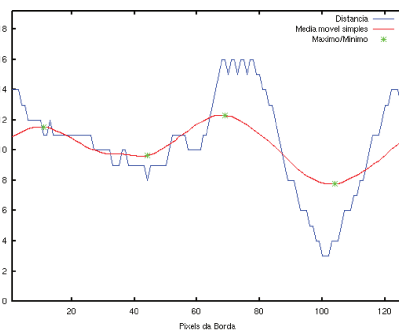
(a) Quadro



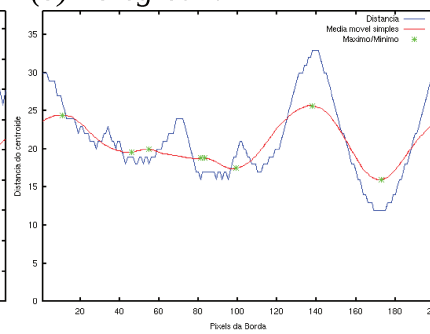
(b) Foreground



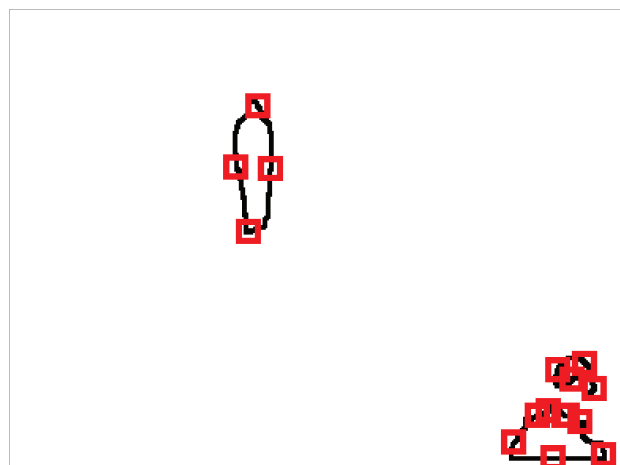
(c) Gráfico (esquerda)



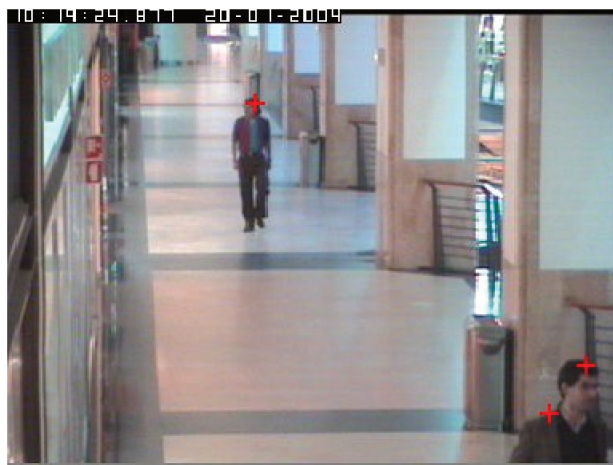
(d) Gráfico (direita, cima)



(e) Gráfico (direita, baixo)



(f) Cantos significativos



(g) Cabeças detectadas

Figura 34: Exemplo de falso positivo ocorrido durante a detecção de cabeças

Vídeo 2

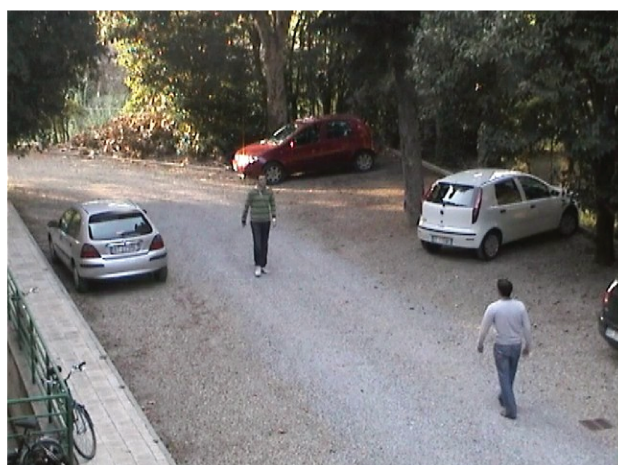
A Tabela 16 corresponde aos dados obtidos a partir do processamento de 50 quadros. Nessa tabela são exibidos o total de cabeças presentes nos quadros, cabeças corretamente detectadas pelo sistema, falsos positivos e falsos negativos.

Total de cabeças nos quadros	85
Cabeças corretamente detectadas	72
Falsos positivos	7
Falsos negativos	13

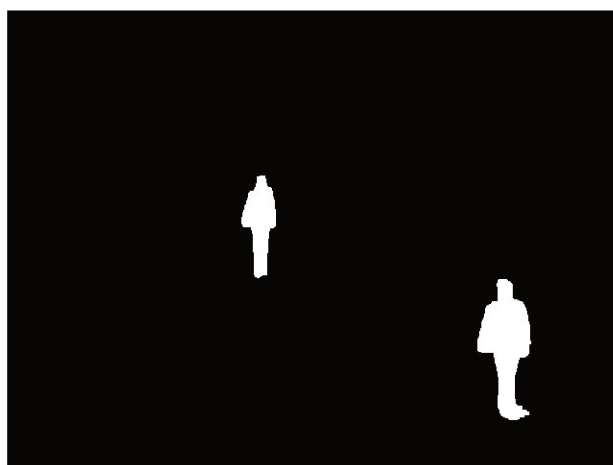
Tabela 16: Análise quantitativa da detecção de cabeças

Conforme é exibido na Tabela 16, 72 cabeças foram detectadas corretamente. Esse valor corresponde a 84.7% do total de cabeças presentes nos quadros. As cabeças foram detectadas em situações nas quais as pessoas estavam isoladas, conforme é exibido na Figura 35, e em casos de oclusões entre pessoas, conforme é exibido na Figura 36.

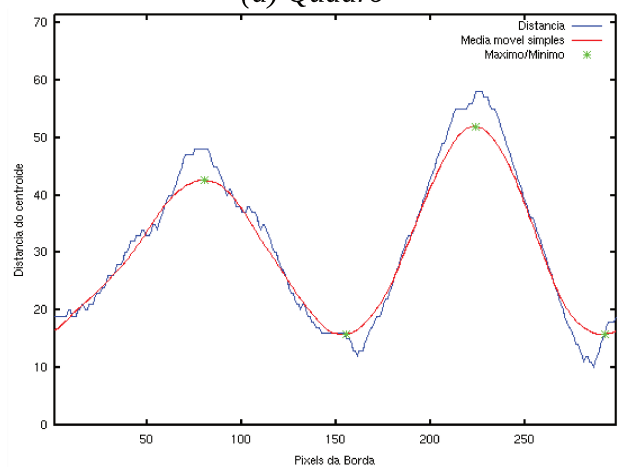
Ocorreram falsos negativos em situações nas quais a posição da cabeça da pessoa não correspondeu a um pico na extremidade superior do *blob*, conforme exemplo exibido na Figura 37.



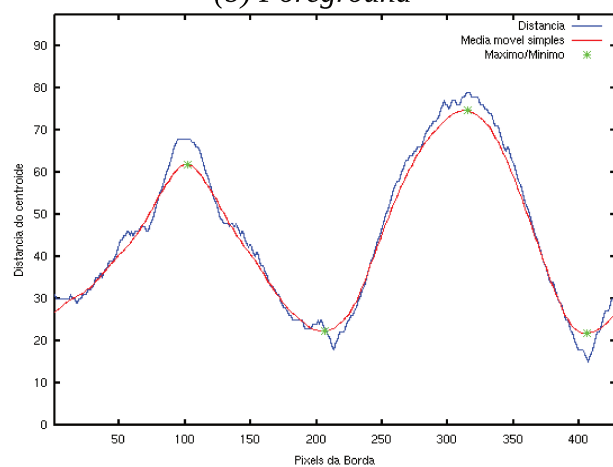
(a) Quadro



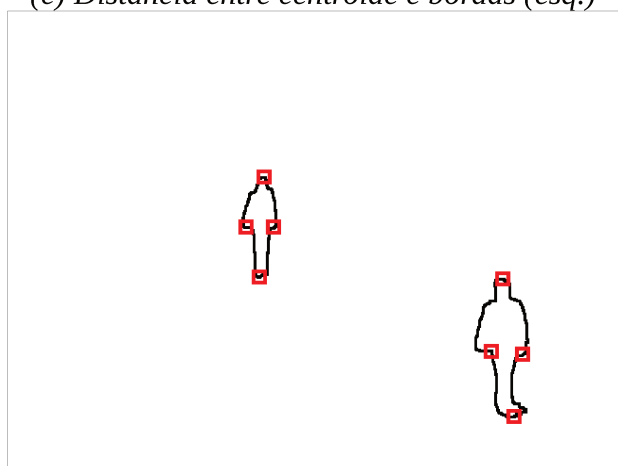
(b) Foreground



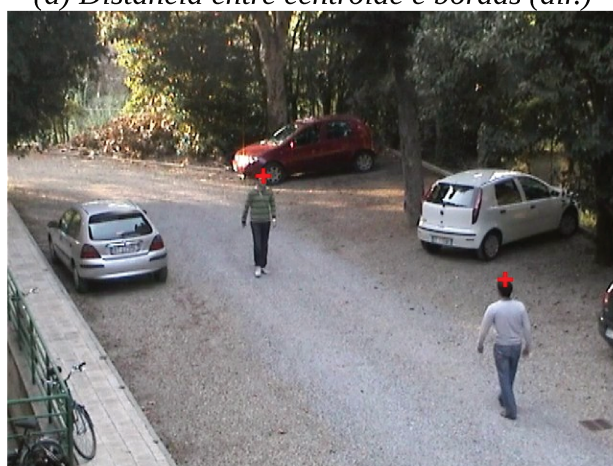
(c) Distância entre centroide e bordas (esq.)



(d) Distância entre centroide e bordas (dir.)

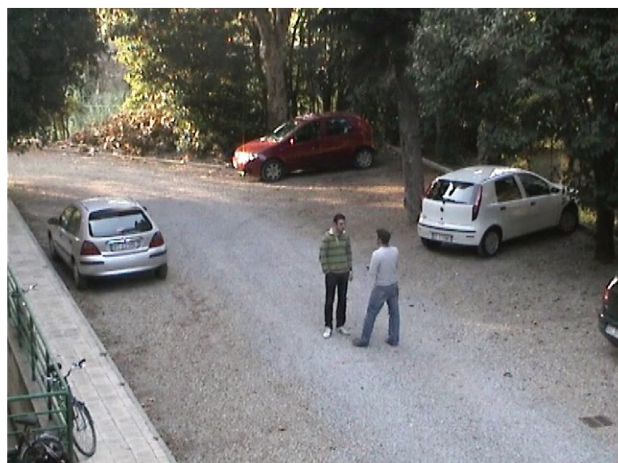


(e) Cantos significativos



(f) Cabeças detectadas

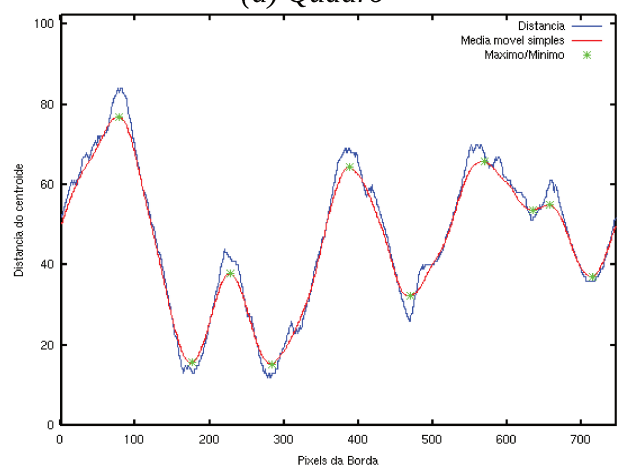
Figura 35: Resultado do método de detecção de cabeças



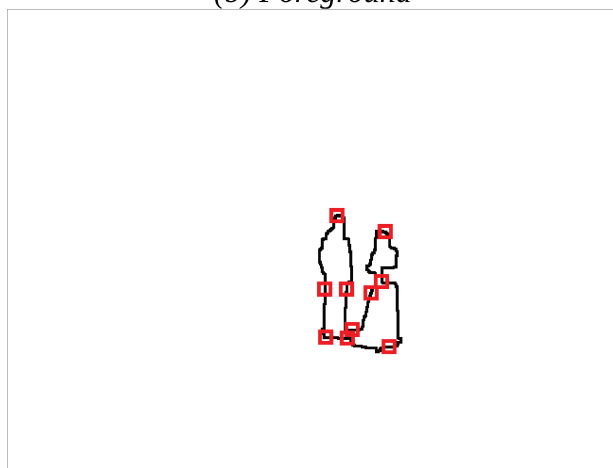
(a) Quadro



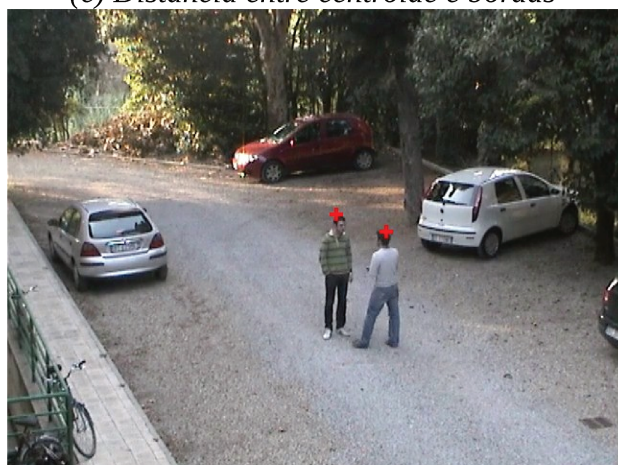
(b) Foreground



(c) Distância entre centroide e bordas



(d) Cantos significativos



(e) Cabeças detectadas

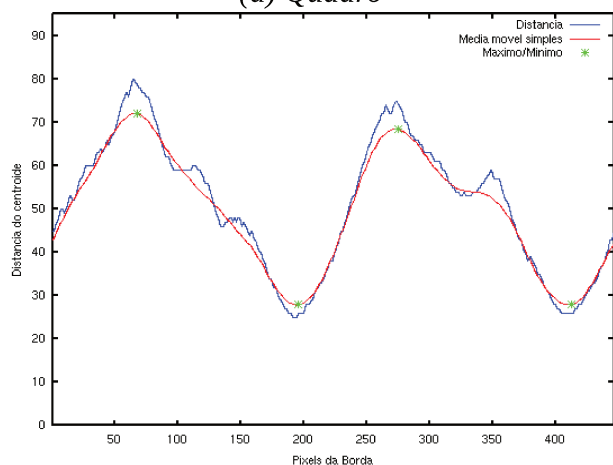
Figura 36: Resultado do método de detecção de cabeças durante oclusão entre pessoas



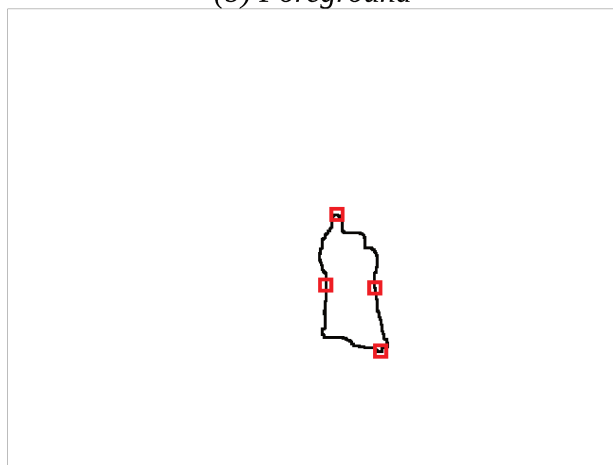
(a) Quadro



(b) Foreground



(c) Distância entre centroide e bordas



(d) Cantos significativos



(e) Cabeças detectadas

Figura 37: Exemplo de falso negativo ocorrido durante a detecção de cabeças

Capítulo 5

Determinação das Trajetórias

O capítulo anterior demonstrou as técnicas utilizadas para extrair informações dos *blobs* correspondentes às pessoas presentes em cada quadro do vídeo. O presente capítulo descreve como as informações extraídas dos *blobs* são utilizadas para associar os *blobs* detectados às pessoas seguidas, e como é gerenciado o rastreamento de múltiplas pessoas deslocando-se simultaneamente pela cena e interagindo entre si e com o ambiente.

5.1 Associação

5.1.1 Previsão do Deslocamento por Filtro de Kalman

A partir da análise de uma sequência de quadros consecutivos, é possível determinar a direção e velocidade empregada por uma pessoa em determinado instante do vídeo. Com base na direção e velocidade, facilmente estima-se a posição dessa pessoa em quadros posteriores considerando um movimento retilíneo uniformemente variado. Usualmente, sistemas de rastreamento utilizam a previsão do deslocamento para identificar o *blob* mais propenso a corresponder ao alvo seguido.

Proposto por Kalman (1960), o filtro de Kalman têm sido vastamente utilizado em sistemas de rastreamento durante a etapa de associação para estimar a posição das pessoas seguidas. O filtro trata-se de um conjunto de equações matemáticas que provêm um modo eficiente de estimar o estado de um processo a partir de uma série de medições ruidosas de modo a minimizar o erro quadrático

médio entre as informações reais e estimadas (Welch e Bishop, 1995). Esse filtro destina-se ao problema de estimar o estado $x \in R^n$ de um processo discreto no tempo controlado por uma equação diferencial estocástica, conforme equação (29)

$$x_k = Ax_{k-1} + Bu_{k-1} + w_{k-1} \quad (29)$$

com uma medição $z \in R^m$ dada pela equação (30).

$$z_k = Hx_k + v_k \quad (30)$$

As variáveis aleatórias w_k e v_k representam o ruído do processo e das medições, respectivamente. Assume-se que estes ruídos são independentes, brancos e com distribuição normal, conforme as equações (31) e (32).

$$p(w) \sim N(0, Q) \quad (31)$$

$$p(v) \sim N(0, R) \quad (32)$$

Na prática, as matrizes de covariância do ruído do processo (Q) e de covariância do ruído da medida (R) podem mudar a cada iteração ou medição, mas assumimos que são constantes.

A matriz A (de tamanho $n \times n$) na equação diferencial (29) relaciona o estado do processo no tempo anterior $k-1$ ao estado atual no tempo k . Na prática, A pode mudar a cada iteração, mas assumimos que é constante. A matriz B (de tamanho $n \times l$) relaciona a entrada de controle $u \in R^l$ ao estado x . A matriz H (de tamanho $m \times n$) na equação de medida (30) relaciona o estado à medida z_k . Na prática, H pode mudar a cada iteração ou medida, mas assumimos que é constante.

Algoritmo do Filtro de Kalman Discreto

O algoritmo do filtro de Kalman discreto é constituído por duas fases: previsão e correção. As

equações relacionadas à fase de previsão são responsáveis por projetar adiante no tempo o estado atual do processo e a covariância do erro para obter a estimativa *a priori* (antes da medição). As equações relacionadas à fase de correção são responsáveis por incorporar a medida à estimativa *a priori* para obter a estimativa *a posteriori* (após a medição).

Na fase da previsão, o estado do processo $\hat{x}_k^-$ e a covariância do erro P_k^- são obtidos pelas equações (33) e (34), respectivamente, obtendo assim a estimativa *a priori*. As matrizes A e B são da equação (29), e Q é da equação (31). O sinal menos em $\hat{x}_k^-$ e P_k^- denota que o estado e a covariância do erro correspondem a estimativa *a priori*.

$$\hat{x}_k^- = A \hat{x}_{k-1} + B u_{k-1} \quad (33)$$

$$P_k^- = A P_{k-1} A^T + Q \quad (34)$$

A fase da correção, correspondente às equações (35), (36) e (37), destina-se ao cálculo da estimativa *a posteriori* $\hat{x}_k$ como uma combinação linear entre a estimativa *a priori* $\hat{x}_k^-$ e a diferença ponderada entre a medição atual z_k e a previsão da medida $H \hat{x}_k^-$, conforme (35). A diferença entre a medição e a previsão de medida $(z_k - H \hat{x}_k^-)$ em (35) é chamada de inovação de medida, ou resíduo.

$$\hat{x}_k = \hat{x}_k^- + K_k (z_k - H \hat{x}_k^-) \quad (35)$$

A matriz K (equação (36)), de tamanho $n \times m$, é chamada de ganho de Kalman, e minimiza a covariância do erro *a posteriori* P_k .

$$\begin{aligned} K_k &= P_k^- H^T (H P_k^- H^T + R)^{-1} \\ &= \frac{P_k^- H^T}{H P_k^- H^T + R} \end{aligned} \quad (36)$$

O valor de K_k é inversamente proporcional ao valor da covariância do ruído de medida R_k . Desse modo, quanto menor o valor de R_k , maior é o valor de K_k e do peso atribuído ao resíduo.

na equação (35). Por outro lado, o valor de K_k é diretamente proporcional ao valor da covariância do erro *a priori* P_k^- . Desse modo, quanto menor o valor de P_k^- , menor é o valor de K_k e do peso atribuído ao resíduo.

Finalmente, obtém-se a estimação *a posteriori* da matriz de covariância do erro, conforme equação (37).

$$P_k = (I - K_k H) P_k^- \quad (37)$$

A Figura 39, adaptada de Welch e Bishop (1995), exibe o fluxo do processamento de uma iteração do filtro de Kalman. A cada iteração, o processo é repetido com a última estimativa *a posteriori* utilizada para prever a estimativa *a priori* da iteração seguinte. Essa natureza recursiva do filtro torna factível sua aplicação ao rastreamento.

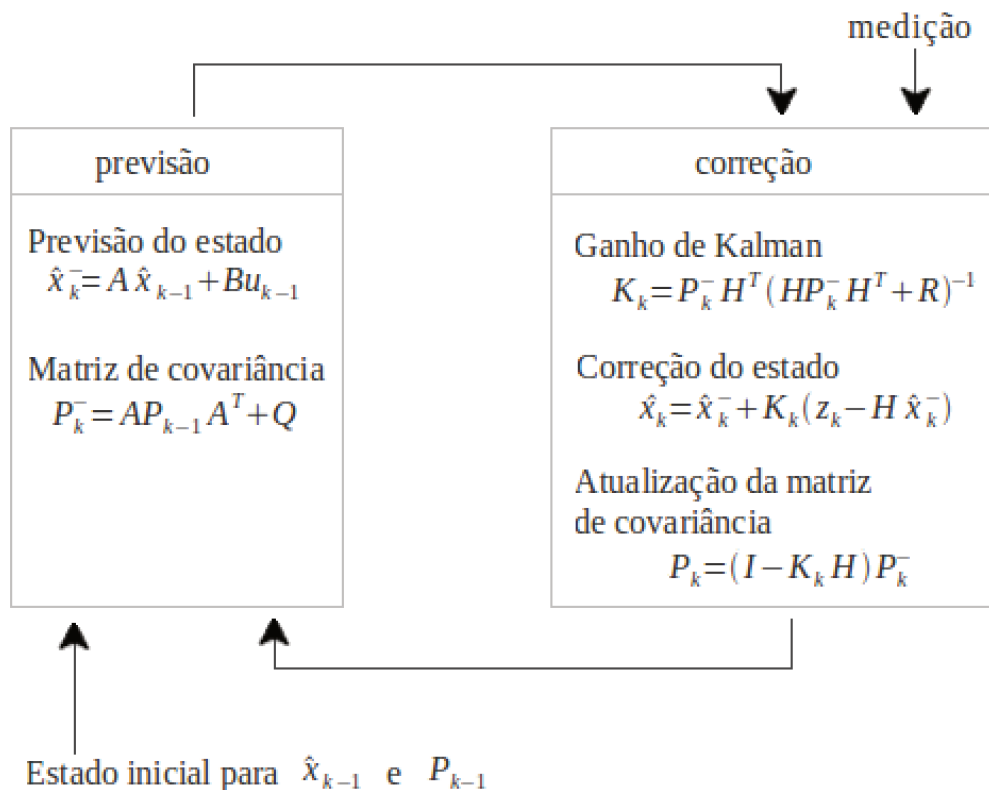


Figura 38: Iteração do Filtro de Kalman
 Fonte: Adaptada de Welch e Bishop (1995)

Implementação do Filtro de Kalman

No sistema de rastreamento de pessoas proposto no presente trabalho, a posição de uma pessoa é representada pela coordenada do centroide, e o filtro de Kalman é utilizado para estimar a posição dessa pessoa nos quadros seguintes. Desta forma, a cada novo quadro do vídeo, são fornecidas ao filtro a posição e velocidade estimada da pessoa no instante anterior. Considera-se ainda que a posição da pessoa deve variar de acordo com o sistema descrito na equação (38).

$$\begin{cases} x_C = x_C^- + \Delta_x \\ y_C = y_C^- + \Delta_y \end{cases} \quad (38)$$

sendo x_C e y_C respectivamente a posição x e y do centroide, e Δ_x e Δ_y a velocidade da pessoa na direção x e y , respectivamente, calculadas conforme equação (39).

$$\begin{cases} \Delta_x = x_C - x_C^- \\ \Delta_y = y_C - y_C^- \end{cases} \quad (39)$$

Desse modo, o vetor que representa o estado de uma pessoa é definido conforme é descrito na equação (40)

$$x = \begin{bmatrix} x_C \\ y_C \\ \Delta_x \\ \Delta_y \end{bmatrix} \quad (40)$$

e a matriz A que descreve a transição de estado é definida conforme é descrito na equação (41).

$$A = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (41)$$

A matriz H que realiza o mapeamento entre a posição medida e a prevista é descrita na equação (42)

$$H = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \quad (42)$$

A Figura 38 exibe as equações do filtro de Kalman utilizadas no sistema de rastreamento proposto.

5.1.2 Métrica de Associação

Para realizar a associação de uma pessoa seguida a um *blob* detectado, inicialmente estima-se a sua posição atual através do filtro de Kalman. Em seguida, a posição dos centroides dos *blobs* detectados são analisadas. Se a região prevista para a posição da pessoa contiver o centroide de um dos *blobs* detectados, esse *blob* é associado a pessoa. Entretanto, se dois ou mais centroides estiverem contidos na região prevista para a pessoa, uma métrica definida a partir das informações listadas na Tabela 17 é utilizada.

(x_p, y_p)	Posição prevista para a pessoa seguida
(x_k, y_k)	Posição do centroide do <i>blob</i> k
H_p	Histograma de cor da pessoa seguida
H_k	Histograma de cor do <i>blob</i> k

Tabela 17: Informações utilizadas pela métrica de associação

Dado um conjunto de *blobs* cujos centroides estão contidos na região prevista pelo filtro de Kalman para a pessoa p , a pessoa é associada ao centroide k que maximiza o valor da métrica de associação $M_{p,k}^{final}$, definida a partir da combinação das métricas relacionadas à distância $M_{p,k}^D$ e ao histograma de cor $M_{p,k}^H$.

A distância euclidiana entre a posição (x_k, y_k) do centroide k e a posição (x_p, y_p) prevista para a pessoa é definida na equação (43).

$$D_{p,k} = \sqrt{(x_p - x_k)^2 + (y_p - y_k)^2} \quad (43)$$

A métrica $M_{p,k}^D$ relacionada à distancia $D_{p,k}$ é definida na equação (44).

$$M_{p,k}^D = \frac{D_{p,m} - D_{p,k}}{D_{p,m}} \quad (44)$$

onde $D_{p,m}$ é a maior distância para todos os centroides contidos na região prevista. O índice m refere-se ao centroide mais distante de p , conforme equação (45).

$$m = \operatorname{argmax}_j (D_{p,j}) \quad (45)$$

A métrica $M_{p,k}^H$ é definida a partir da intersecção dos histogramas da pessoa H_p e do *blob* H_k , conforme a equação (46).

$$M_{p,k}^H = \frac{\sum_{i=1}^N \min(H_p(i), H_k(i))}{\min(\mathbf{h}_p, \mathbf{h}_k)} \quad (46)$$

onde N é o número de faixas do histograma após a discretização das cores, e $\mathbf{h}_n$ corresponde à somatória dos valores do histograma $\mathbf{n}$, conforme a equação (47).

$$\mathbf{h}_n = \sum_{i=1}^N H_n(i) \quad (47)$$

A métrica de associação final é definida conforme a equação (48)

$$M_{p,k}^{final} = \omega M_{p,k}^D + (1 - \omega) M_{p,k}^H \quad (48)$$

onde ω é um parâmetro que pondera a combinação entre as duas métricas, $\omega = [0...1]$. A pessoa é associada ao centroide com maior valor de $M_{p,k}^{final}$.

5.1.3 Tratamento de Oclusões

Conforme foi apresentado anteriormente, a trajetória de uma pessoa na imagem é determinada pela variação da posição do *blob* referente a essa pessoa. Entretanto, em algumas situações um *blob* detectado corresponde a mais de uma pessoa. Isso ocorre porque *blobs* correspondentes a pessoas distintas se unem quando um dos seguintes eventos ocorre:

- Uma pessoa sobrepõe a outra em relação à câmera.
- Duas ou mais pessoas se tocam.

As abordagens utilizadas para realizar o rastreamento durante uniões de *blobs* podem ser classificadas como *merge-split* (MS) ou *straight-through* (ST). Na abordagem MS, as pessoas sobrepostas são encapsuladas em um novo grupo. Quando uma das pessoas se separa do grupo, sua identidade é restabelecida com base em características extraídas das regiões detectadas, como a cor em Seo et al (1997), Lu e Tan (2001) e Freitas (2010). Essa abordagem possui duas desvantagens:

1. A posição central do grupo usualmente é diferente da posição de uma pessoa. Quanto maior for a diferença entre o centroide desse *blob* e a posição real da pessoa, maior será o erro na trajetória determinada conforme essa abordagem.
2. Se duas pessoas entrarem na cena unidas, a oclusão (definida, conforme a abordagem MS, como a união de dois *blobs* rastreados) não será identificada e uma única trajetória será determinada.

A abordagem *straight-through* (ST) propõe que os alvos sejam rastreados individualmente durante oclusões. Para tal, é necessário determinar a posição das pessoas associadas ao *blob*.

Para tratar oclusões entre as pessoas seguidas, Mirabi e Javadi (2012) propõem a segmentação do *blob* resultante da oclusão com base na *bounding box* e na previsão da posição da pessoa. Para cada pessoa, a posição é prevista através do filtro de Kalman, e uma *bounding box* é definida com centro na posição prevista. O tamanho e forma dessa *bounding box* são obtidos a partir da *bounding box* da pessoa no quadro anterior à oclusão. Então, a posição da pessoa é atualizada com base no centroide do *foreground* contido na nova *bounding box*. Desse modo, a atualização das trajetórias é realizada individualmente para cada pessoa durante a oclusão.

Alam e Boufama (2011) propõem uma abordagem baseada na divisão do *blob* em *sub-blobs*, ou seja, separação do *blob* em partes. A associação é feita com base na comparação entre modelos formados a partir de informações das pessoas seguidas nos quadros anteriores e dos *sub-blobs*. Os modelos são constituídos pelas seguintes informações: centroide, velocidade, tamanho e histograma de cor. Embora os autores não apresentem detalhes da segmentação do *blob*, os resultados são promissores.

Azari et al (2011) utilizam o método de coeficiente de correlação (Rodgers e Nicewander, 1988) para determinar a posição das pessoas oclusas através da correspondência de modelos (*Template matching*). Conforme essa abordagem, o modelo de uma pessoa é transladado pelo *foreground*, e a correlação entre os *pixels* do modelo e da imagem é calculada. Então, a região da imagem com maior valor de correlação é escolhida para atualizar o deslocamento da pessoa.

Os sistemas de rastreamento propostos por Haritaoglu et al (2000) e Zhao e Nevatia (2004) determinam a posição das pessoas através de um método de detecção de cabeças. Ambos os sistemas utilizam a posição da cabeça das pessoas oclusas para inferir a posição das mesmas, e atualizar as trajetórias individualmente durante oclusões.

A Tabela 18 apresenta uma listagem das abordagens de resolução de oclusões entre pessoas citadas acima.

Autor	Abordagem	Comentários
Mirabi e Javadi (2012)	<i>Straight-Through</i>	Determina a posição individual das pessoas durante oclusões parciais com base na previsão de <i>bounding boxes</i> . Não trata oclusão de pessoas por objetos.
Alam e Boufama (2011)	<i>Straight-Through</i>	Determina a posição individual das pessoas durante oclusões com base na posição e cor. Não apresenta detalhes do processo de divisão do <i>blob</i> em <i>sub-blobs</i> .
Azari et al (2011)	<i>Straight-Through</i>	Determina a posição individual das pessoas durante oclusões parciais utilizando <i>Template matching</i> . Não trata oclusão de pessoas por objetos.
Freitas (2010)	<i>Merge-Split.</i>	Trata oclusão de pessoas por objetos. Durante oclusões, as pessoas são associadas a posição central do grupo correspondente às pessoas oclusas.
Zhao e Nevatia (2004)	<i>Straight-Through</i>	Determina a posição individual das pessoas durante oclusões parciais com base na detecção de cabeças.
Lu e Tan (2001)	<i>Merge-Split.</i>	Durante oclusões, as pessoas são associadas a posição central do grupo correspondente às pessoas oclusas.
Haritaoglu et al (2000)	<i>Straight-Through</i>	Determina a posição individual das pessoas durante oclusões parciais com base na detecção de cabeças.
Seo et al (1997)	<i>Merge-Split.</i>	Durante oclusões, as pessoas são associadas a posição central do grupo correspondente às pessoas oclusas.

Tabela 18: Abordagens de resolução de oclusões

Tratamento de oclusão entre pessoas utilizado no presente trabalho

Quando ocorre uma oclusão entre duas ou mais pessoas, dois problemas devem ser resolvidos para manter o rastreamento:

1. O sistema deve identificar o *blob* que corresponde à pessoa oclusa.
2. O sistema deve identificar a posição da pessoa oclusa.

Quando não há oclusão entre pessoas, o *blob* correspondente à pessoa pode ser facilmente

identificado através da comparação entre a posição prevista para a pessoa e a posição dos centroides dos *blobs* detectados na imagem. No entanto, quando duas pessoas se ocluem, a posição do centroide do *blob* correspondente às pessoas tende a variar com relação à posição prevista para as pessoas. Em alguns casos, o centroide da região detectada pode inclusive estar em uma região diferente da região prevista para a posição das pessoas seguidas, resultando em falha na associação. Essa situação é demonstrada na Figura 39.

A Figura 39 ilustra uma situação onde uma pessoa caminhando na cena é ocluída por um grupo de pessoas. Nessa figura é possível observar que o centroide do *blob* detectado está posicionado em um local diferente da região prevista para a posição da pessoa seguida. Nesse exemplo, a *blob* correspondente a pessoa não seria encontrado, pois nenhum centroide está contido na região prevista para a pessoa seguida, e a consequência disso seria a descontinuação do rastreamento da pessoa.

Para evitar a ocorrência de erros de associação em casos de oclusões, no sistema proposto a posição prevista para a pessoa seguida é também comparada à *bounding box* dos *blobs* detectados. Se a posição prevista para a pessoa estiver contida na *bounding box* de um *blob*, a pessoa é associada a esse *blob*. Conforme é exibido na Figura 39, essa condição é verdadeira para esse exemplo.

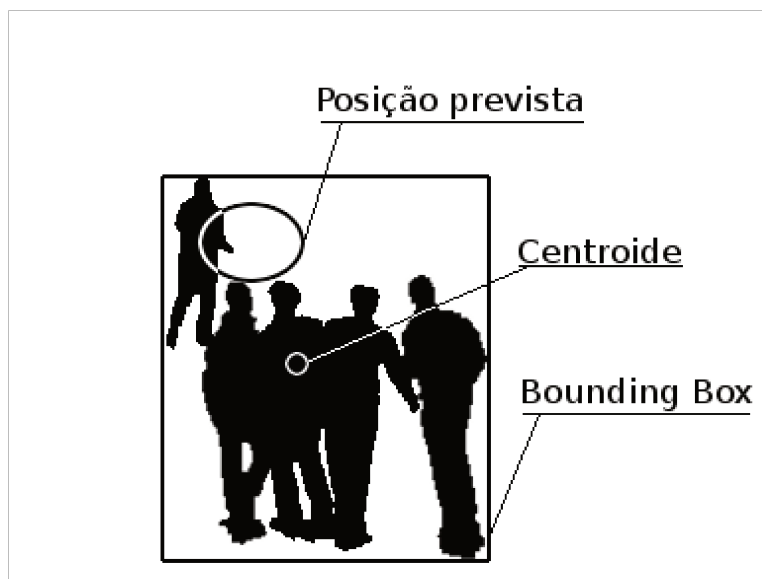


Figura 39: Ilustração do uso da *bounding box* durante a associação

Para identificar se a posição (x, y) prevista para a pessoa seguida está contida na *bounding*

box, a posição é comparada as coordenadas do canto superior esquerdo e do canto inferior direito da *bounding box*, conforme é descrito na equação (49).

$$\left\{ \begin{array}{ll} (x, y) \text{ está contida na bounding box,} & \text{se } x_{SE} < x < x_{ID} \\ & \wedge y_{SE} < y < y_{ID} \\ (x, y) \text{ não está contida na bounding box,} & \text{caso contrário} \end{array} \right. \quad (49)$$

sendo (x_{SE}, y_{SE}) o canto superior esquerdo e (x_{ID}, y_{ID}) o canto inferior direito da *bounding box*.

Após associar a pessoa a um *blob*, o sistema de rastreamento determina a posição da pessoa. Quando não há oclusão, a posição do centroide do *blob* é atribuída à posição da pessoa. Entretanto, em casos de oclusão entre pessoas, o centroide não corresponde a posição de uma pessoa, mas de um grupo de pessoas. No presente trabalho, as posições das pessoas envolvidas na oclusão são estimadas a partir do método de detecção de cabeças.

A partir do método de detecção de cabeças, as posições das pessoas são estimadas. Então, para cada pessoa envolvida na oclusão, é calculado o valor da métrica $M_{p,k}^D$ (relacionada à distância) para o centroide do *blob* e também para as posições estimadas a partir das posições das cabeças. O ponto que resulta em maior valor para $M_{p,k}^D$ é atribuído à pessoa oclusa.

Um caso particular de oclusão é aquele que ocorre antes das pessoas entrarem na cena, e continua durante e após a permanência das mesmas no campo de visão da câmera. Nesse caso específico, não ocorre a junção de dois *blobs* seguidos, e a oclusão não pode ser identificada como a associação de duas pessoas ao mesmo *blob*, pois essas pessoas não estavam sendo seguidas nos quadros anteriores. Nesse caso todas as abordagens *Merge-Split* identificam o surgimento de um *blob*, e determinam a sua trajetória como se fosse uma única pessoa. Outras abordagens *Straight-Through*, como Mirabi e Javadi (2012) e Haritaoglu et al (2000), obtêm o mesmo resultado, pois não preveem que duas pessoas possam entrar na cena oclusas e permanecer nesse estado durante todo o tempo em que estejam presentes no vídeo.

No presente trabalho, o rastreamento de uma pessoa só é efetivamente iniciado quando a mesma transita para a classe Madura. Para que a entrada na cena de pessoas em estado de oclusão seja identificada pelo sistema de rastreamento, o método de detecção de cabeças é aplicado ao *blob* seguido

enquanto ele está associado à classe Iniciante. Se duas ou mais cabeças forem detectadas, considera-se que há uma oclusão entre pessoas, e o rastreamento é realizado para cada pessoa oclusa individualmente. Caso contrário, prossegue-se o rastreamento de uma única pessoa. Desse modo, o sistema proposto é capaz de identificar a entrada de pessoas na cena em estado de oclusão.

5.2 Classificação de Estados

Em cenas onde há fluxo de pessoas, é comum a ocorrência de eventos como entrada, saída, desaparecimento e reaparecimento de pessoas. Ainda, múltiplas pessoas podem se agrupar e posteriormente se separar. Isso implica que uma pessoa pode estar em diferentes estados ao longo do vídeo. Ao modelar explicitamente esses estados, tratamentos específicos podem ser empregados em cada situação citada acima.

Lei e Xu (2006) propõem um sistema de classificação de estados, posteriormente modificado por Freitas (2010), onde as situações previstas para as pessoas seguidas são representadas por um conjunto definido de classes. Cada pessoa seguida é associada a uma das classes de acordo com a situação em que se encontra em determinado instante de vídeo, podendo variar de classe enquanto transita pela cena. Com base nas classes associadas às pessoas presentes no vídeo, o sistema de rastreamento pode utilizar diferentes abordagens para determinar as trajetórias de pessoas isoladas na cena, pessoas em grupos, e pessoas oclusas por partes do fundo da cena.

No presente trabalho o sistema de classificação de estado proposto por Lei e Xu (2006) é adaptado e utilizado para representar a situação das pessoas rastreadas. A Tabela 19 apresenta o conjunto de classes utilizadas e o diagrama exibido na Figura 40 descreve as regras de transição entre as classes.

Classe	Descrição
Iniciante	Classe associada à pessoa no momento em que ela entra na cena.
Madura	Após satisfeitas as métricas de confiabilidade relacionadas ao tempo de permanência e fator de movimento regular, o rastreamento da pessoa é efetivamente iniciado.
Indisponível	As pessoas seguidas que não são associadas a nenhum <i>blob</i> transitam para a classe Indisponível.
Desaparecida	Pessoas que permanecem associadas à classe Indisponível por um período de tempo superior a um parâmetro predefinido transitam para a classe Desaparecida, e o rastreamento é interrompido.
Fora de Cena	Pessoas que permanecem associadas à classe Desaparecida por um período de tempo superior a um parâmetro predefinido transitam para a classe Fora de Cena, e o rastreamento é cessado permanentemente.

Tabela 19: Classes do sistema de gerenciamento de pessoas

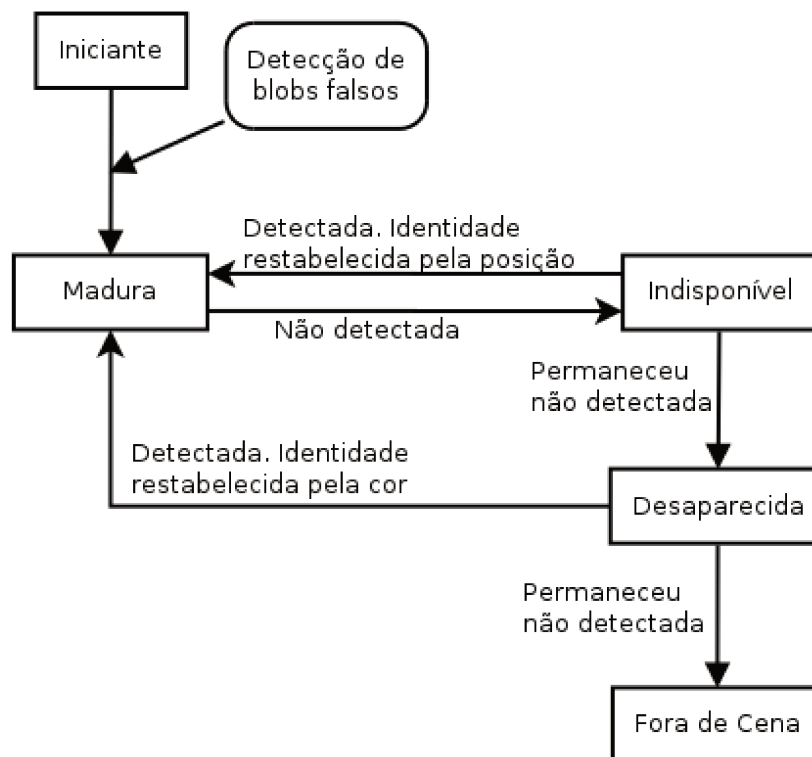


Figura 40: Regras de transição entre as classes

5.2.1 Classe Iniciante

No momento em que uma pessoa surge na cena pela primeira vez, um *blob* é detectado. Entretanto, esse *blob* não é associado a nenhuma pessoa seguida justamente porque o rastreamento dessa pessoa ainda não foi iniciado. Sendo assim, toda região detectada na imagem que não é associada a nenhuma pessoa seguida pode corresponder a uma nova pessoa entrando na cena. Contudo, antes de iniciar o rastreamento de uma pessoa, é preciso verificar se a região detectada não corresponde a um ruído. Para tal, o *blob* é associado à classe Iniciante, seu rastreamento é iniciado e em seguida é realizada a detecção de *blobs* falsos (conforme diagrama exibido na Figura 40), onde são analisadas duas métricas de confiabilidade: tempo mínimo de permanência e fator de movimento regular. O tempo de permanência é o número de quadros consecutivos que o *blob* foi detectado. O fator de movimento expressa a aleatoriedade do movimento realizado, conforme é descrito na equação (50).

$$\zeta_m = \left(\frac{\sigma_{cx}^2}{\sigma_{vx}^2 + \tau} + \frac{\sigma_{cy}^2}{\sigma_{vy}^2 + \tau} \right) \frac{1}{2} \quad (50)$$

onde σ_{cx}^2 e σ_{cy}^2 são a variância do centroide na direção x e y respectivamente, σ_{vx}^2 e σ_{vy}^2 são a variância da velocidade na direção x e y , respectivamente, e τ é uma constante que evita divergência quando a variação da posição e velocidade do centroide são iguais a zero (Freitas, 2010). O fator de movimento é considerado regular se $\zeta_m > T_\zeta$, conforme é descrito na equação (51), onde T_ζ é um *threshold*.

$$\begin{cases} \zeta_m > T_\zeta & \text{movimento regular} \\ \text{caso contrário,} & \text{movimento irregular} \end{cases} \quad (51)$$

Caso uma das duas métricas de confiabilidade não seja satisfeita, o *blob* é desconsiderado do sistema de rastreamento. Caso ambas as métricas de confiabilidade sejam satisfeitas, o *blob* é considerado uma pessoa na cena, e transita para a classe Madura.

5.2.2 Classe Madura

O rastreamento de uma pessoa é efetivamente iniciado quando as métricas de confiabilidade relacionadas ao tempo de permanência e fator de movimento regular são satisfeitas, e a pessoa transita para a classe Madura. A associação quadro a quadro é realizada utilizando métricas relacionadas às informações extraídas dos *blobs*, como centroide, *bounding box* e histograma de cor. Em casos de oclusão entre pessoas, o método de detecção de cabeças proposto no presente trabalho é aplicado para identificar a posição das pessoas envolvidas na oclusão, e atualizar a trajetória de cada pessoa individualmente (mais detalhes sobre as métricas de associação e o método de detecção de cabeças são apresentados na seção 5.1).

Uma alteração no sistema de classificação de estados utilizado no presente trabalho com relação ao proposto por Lei e Xu (2006) é a remoção da classe *Occluded*. Ao contrário do sistema de rastreamento proposto por Lei e Xu (2006) que trata oclusões conforme a abordagem *Merge-Split*, o presente trabalho utiliza a abordagem *Straight-Through*, e por isso o sistema de classificação de estados foi adaptado. Nesse caso, as pessoas ocluídas por outras pessoas permanecem associadas à classe Madura, e suas trajetórias são determinadas individualmente.

5.2.3 Classe Indisponível

Enquanto desloca-se pela cena, uma pessoa pode ser ocluída por um objeto estático ou por alguma estrutura no fundo da cena e não ser detectada na imagem. Como a trajetória é determinada a partir da variação da posição da pessoa, o rastreamento é prejudicado quando o *blob* não pode ser detectado. No entanto, é comum que a pessoa seja detectada novamente em quadros posteriores quando a oclusão termina.

Conforme é exibido no diagrama da Figura 40, quando uma pessoa seguida não é associada a nenhum *blob*, ela transita para a classe Indisponível, e sua trajetória continua sendo atualizada pela previsão do filtro de Kalman. Se em quadros posteriores essa pessoa for associada a um *blob* novamente, ela retorna à classe Madura e seu rastreamento é continuado normalmente.

5.2.4 Classe Desaparecida

Conforme é descrito acima, quando uma pessoa não é detectada na imagem, ela transita para a classe Indisponível, e sua trajetória é atualizada com base na previsão do filtro de Kalman. Se a pessoa for detectada em quadros posteriores, ela retorna à classe Madura. Contudo, quanto mais tempo a pessoa permanece associada à classe Indisponível, menos precisa se torna a previsão de seu deslocamento. Portanto, se uma pessoa permanecer associada à classe Indisponível por um período de tempo superior a um parâmetro predefinido, a pessoa transita para a classe Desaparecida, e seu rastreamento é cessado.

Entretanto, uma pessoa associada à classe Desaparecida pode ser detectada novamente após a oclusão. Para identificar o reaparecimento de uma pessoa oclusa, inicialmente é analisada a região do *blob* detectado. Os *blobs* que surgem nas áreas de entrada (imagens binárias onde os *pixels* com valor igual a 1 indicam as regiões de entrada de pessoas, conforme os exemplos exibidos na Figura 41) são considerados novas pessoas entrando na cena. Os *blobs* que surgem nas demais regiões são considerados potenciais pessoas desaparecidas. Para identificar se um *blob* detectado fora da área de entrada é uma das pessoas desaparecidas, determina-se a intersecção entre o histograma do *blob* e das pessoas desaparecidas. Se o maior valor de intersecção calculado for superior a um *threshold*, a pessoa transita para a classe Madura, e seu rastreamento é retomado.

Uma alteração no sistema de classificação de estados utilizado no presente trabalho com relação ao proposto por Lei e Xu (2006) é a remoção da classe *Reappeared*. No presente trabalho um *blob* correspondente a uma pessoa reaparecida inicialmente é associado à classe Iniciante. Caso satisfaça as métricas de confiabilidade, a identidade da pessoa é restabelecida através da comparação entre histogramas, e ela retorna à classe Madura. Logo, entende-se que a classe *Reappeared* não é necessária e por isso ela não é utilizada no presente trabalho.

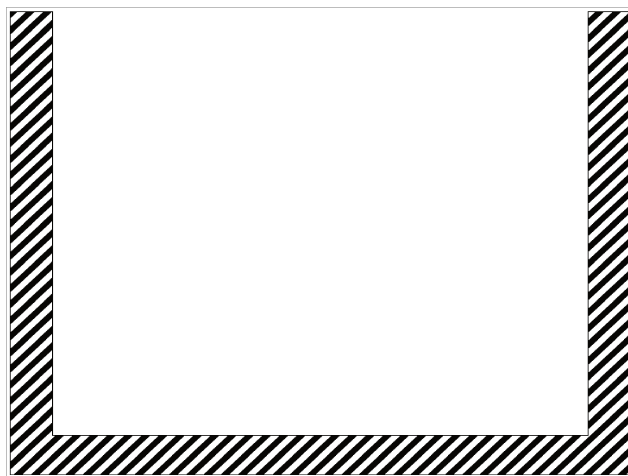
5.2.5 Classe Fora de Cena

Uma pessoa é associada à classe Fora de Cena quando permanece associada à classe

Desaparecida por um período de tempo superior a um parâmetro predefinido, e assume-se que é seguro afirmar que a pessoa abandonou a cena. Nesse caso, seu rastreamento é cessado permanentemente, e as informações armazenadas dessa pessoa podem ser removidas da memória.



(a) Quadro 1



(b) Regiões de entrada do quadro 1



(c) Quadro 2



(d) Regiões de entrada do quadro 2

Figura 41: Exemplos de regiões de entrada de pessoas

Capítulo 6

Avaliação do Sistema Proposto

O sistema de rastreamento proposto no presente trabalho foi implementado utilizando a linguagem GNU Octave. GNU Octave é uma linguagem de alto nível, de código aberto e compatível com Matlab que provê uma interface para resolução de problemas computacionais.

A metodologia utilizada para avaliar o rastreamento consiste em análises qualitativas e quantitativas. A análise qualitativa consiste em exibir as trajetórias determinadas, o que possibilita a avaliação visual do resultado obtido. A análise quantitativa consiste em comparar numericamente o resultado esperado ao obtido e calcular taxas de avaliação, conforme Ellis (2002).

Para realizar a análise quantitativa, cada quadro do vídeo é analisado e o número de pessoas corretamente detectadas é determinado. Nesse contexto, uma detecção correta pode ocorrer se uma das condições abaixo for verdadeira.

1. A pessoa seguida for associada ao centroide do *blob* correspondente à pessoa na imagem.
2. Em caso de junção de *blobs* correspondentes a pessoas distintas, a pessoa for associada à posição estimada a partir da posição de sua cabeça. Nesse trabalho, não é considerada uma detecção correta quando a pessoa é associada à posição do centroide do *blob* resultante da oclusão entre duas ou mais pessoas.

Conforme essa abordagem, para um vídeo composto por 100 quadros e que contenha 2 pessoa transitando pela cena, é esperado que ocorram 2 detecções corretas por quadro, totalizando 200 detecções corretas em todo o vídeo.

O número total de detecções corretas esperado para todo o vídeo é denominado positivos esperados, e o número total de detecções realizadas pelo sistema é denominado positivos obtidos.

Entretanto, uma detecção realizada pelo sistema não significa necessariamente uma detecção correta. Desse modo, as detecções realizadas pelo sistema dividem-se em positivos verdadeiros (PV) – detecções corretamente realizadas – e falsos positivos (FP) – trajetórias calculadas indevidamente. Casos onde a pessoa não é rastreada pelo sistema são denominados falsos negativos (FN).

Além do número de pessoas presentes em cada quadro do vídeo, é contado também o número de *blobs* detectados pelo método de subtração de fundo que não correspondem às pessoas presentes na cena. Esses *blobs*, usualmente resultantes de alterações físicas na cena como a remoção ou inclusão de objetos no fundo, devem ser desconsiderados durante o rastreamento. A soma de todos os *blobs* a serem desconsiderados durante o rastreamento denomina-se negativos esperados. Os *blobs* identificados como negativos esperados e que são corretamente desconsiderados pelo sistema são denominados negativos verdadeiros (NV), e os *blobs* que são erroneamente rastreados representam os falsos positivos (FP).

A partir da contagem de PV, FP, NV e FN as taxas exibidas na Tabela 20 são determinadas. Essas taxas expressam em números a eficiência do sistema, e podem ser utilizadas como base de comparação com outros sistemas de rastreamento.

Taxa de Sensibilidade	$PV / (PV + FN)$
Taxa de falso negativo	$FN / (FN + PV)$
Taxa de falso positivo	$FP / (FP + NV)$

Tabela 20: Taxas de avaliação utilizadas no presente trabalho

Os vídeos utilizados para avaliar o sistema proposto são listados na Tabela 21.

	Quadros	Dimensões	Link para download do vídeo (acessados em 11/07/2012)
Vídeo 1	831	640x480	http://imagelab.ing.unimore.it/visor/video_Download.asp?idVideo=340
Vídeo 2	383	384x288	http://groups.inf.ed.ac.uk/vision/CAVIAR/CAVIARDATA2/EnterExitCrossingPaths1front/EnterExitCrossingPaths1front.tar.gz
Vídeo 3	606	352x288	http://imagelab.ing.unimore.it/visor/video_Download.asp?idVideo=199
Vídeo 4	3051	720x676	http://ftp.cs.rdg.ac.uk/PETS2006/S4-T5-A.zip

Tabela 21: Vídeos utilizados para avaliar o sistema proposto

6.1 Vídeo 1

No vídeo 1 duas pessoas caminhando em sentidos opostos se encontram e permanecem oclusas. Após um período de tempo, elas se separam e começam a correr no mesmo sentido.

O rastreamento é realizado com e sem a utilização do método de detecção de cabeças para avaliar a contribuição da utilização do método. Os resultados da análise quantitativa são apresentados na Tabela 22.

Valores esperados		
Positivos esperados	1355	
Negativos esperados	0	
Valores obtidos		
	Sem detecção de cabeças	Com detecção de cabeças
Positivos verdadeiros	1008	1190
Falsos negativos	347	165
Negativos verdadeiros	0	0
Falsos positivos	0	0
Taxas de avaliação		
	Sem detecção de cabeças	Com detecção de cabeças
Taxa de Sensibilidade	0,74	0,87
Taxa de falso negativo	0,25	0,12
Taxa de falso positivo	0,00	0,00

Tabela 22: Vídeo 1 – Análise quantitativa dos resultados obtidos com e sem o uso do método de detecção de cabeças

Conforme é exibido na Tabela 22, para esse vídeo eram esperados 1355 positivos verdadeiros e 0 negativos verdadeiros. Com a utilização do método de detecção de cabeças, o total de positivos verdadeiros aumentou de 1008 para 1190 e o total de falsos negativos, que é inversamente proporcional ao total de positivos verdadeiros, diminuiu de 347 para 165. As taxas de avaliação indicam que o resultado obtido com o uso do método de detecção de cabeças foi superior, o que é expressado pelo aumento da taxa de sensibilidade e diminuição da taxa de falso negativo.

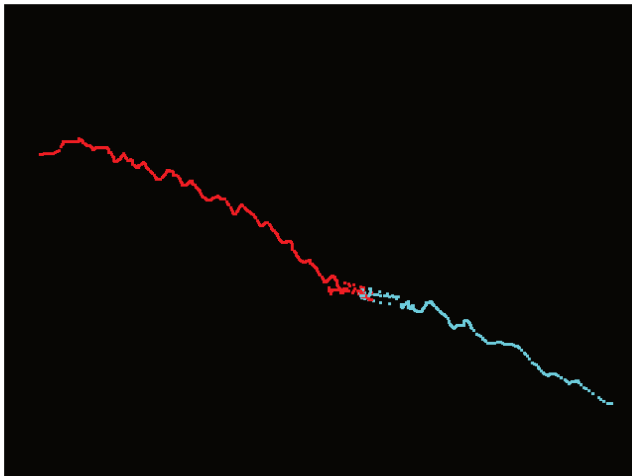
A Figura 42 exibe a comparação entre as trajetórias determinadas com e sem o uso do método de detecção de cabeças. Em (a) é exibido um quadro do vídeo durante a oclusão, e em (b) é exibido um quadro após a oclusão. A Figura 42 (c) mostra que sem o uso do método de detecção de cabeças as trajetórias se juntaram durante a oclusão, pois as duas pessoas foram associadas a posição do centroide do *blob* detectado. A Figura 42 (d) mostra que com o uso do método de detecção de cabeças as trajetórias não se juntaram durante a oclusão.



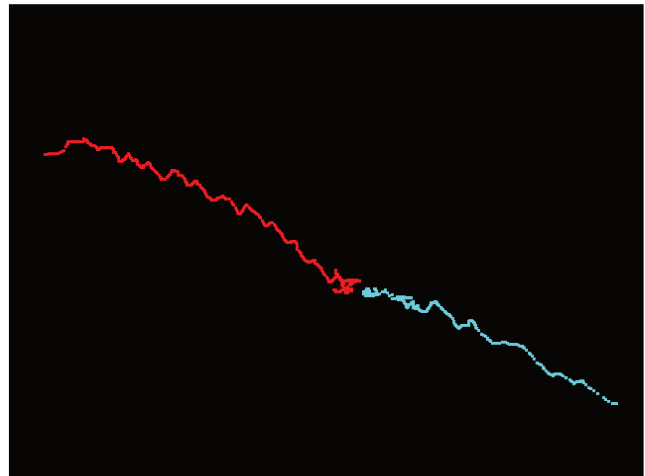
(a) Quadro 529 durante a junção dos blobs



(b) Quadro 657 após a separação dos blobs



(c) Quadros 1-657 sem detecção de cabeças



(d) Quadros 1-657 com detecção de cabeças

Figura 42: Vídeo 1 – Trajetórias determinadas com e sem o uso do método de detecção de cabeças

6.2 Vídeo 2

O vídeo 2 corresponde a uma cena onde ocorre uma oclusão entre duas pessoas. A oclusão ocorre na porta de uma loja, entre uma pessoa que está entrando e outra que está saindo da loja. No momento em que a oclusão termina, ambas as pessoas alteram o sentido do deslocamento, e caminham separadas.

O rastreamento é realizado com e sem a utilização do método de detecção de cabeças para

avaliar a contribuição da utilização do método. Os resultados da análise quantitativa são apresentados na Tabela 23.

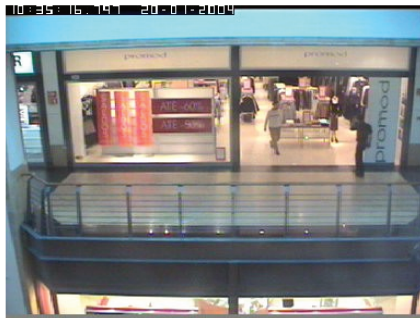
Valores esperados		
Positivos esperados	416	
Negativos esperados	0	
Valores obtidos		
	Sem detecção de cabeças	Com detecção de cabeças
Positivos verdadeiros	378	414
Falsos negativos	38	2
Negativos verdadeiros	0	0
Falsos positivos	0	0
Taxas de avaliação		
	Sem detecção de cabeças	Com detecção de cabeças
Taxa de Sensibilidade	0,908	0,995
Taxa de falso negativo	0,091	0,005
Taxa de falso positivo	0,000	0,000

Tabela 23: Vídeo 2 – Análise quantitativa dos resultados obtidos com e sem o uso do método de detecção de cabeças

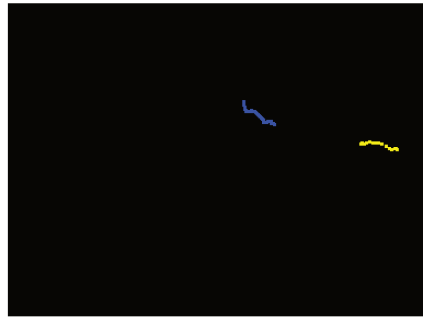
Conforme é exibido na Tabela 23, para esse vídeo eram esperados 416 positivos verdadeiros e 0 negativos verdadeiros. Com a utilização do método de detecção de cabeças, o total de positivos verdadeiros aumentou de 378 para 414 e o total de falsos negativos, que é inversamente proporcional ao total de positivos verdadeiros, diminuiu de 38 para 2. As taxas de avaliação indicam que o resultado obtido com o uso do método de detecção de cabeças foi superior, o que é expressado pelo aumento da taxa de sensibilidade e diminuição da taxa de falso negativo.

A Figura 43 exibe as trajetórias das duas pessoas. Em (a) é exibido um quadro do vídeo antes da oclusão. Em (b) e (c) são exibidas as trajetórias determinadas até o quadro exibido em (a) sem e com o uso do método de detecção de cabeças, respectivamente. Em (d) é exibido um quadro no momento da oclusão entre as pessoas. A Figura 43 (e) mostra que, sem a utilização do método de detecção de

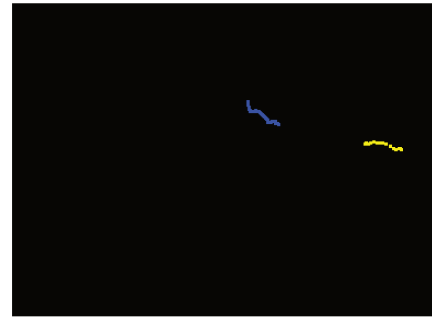
cabeças, as trajetórias se juntaram durante a oclusão. A Figura 43 (f) mostra que, com a utilização do método de detecção de cabeças, as trajetórias não se juntaram, e as pessoas continuaram sendo rastreadas isoladamente.



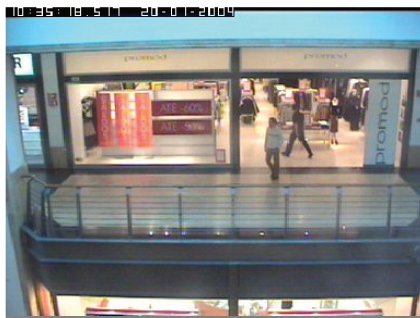
(a) Quadro 214



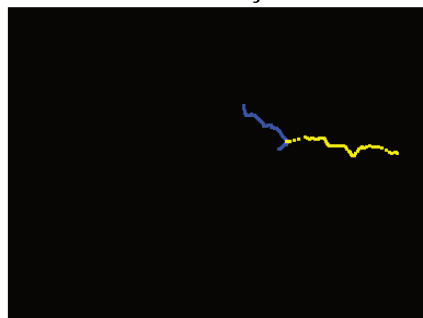
(b) Quadros 1-214 sem detecção de cabeça



(c) Quadros 1-124 com detecção de cabeça



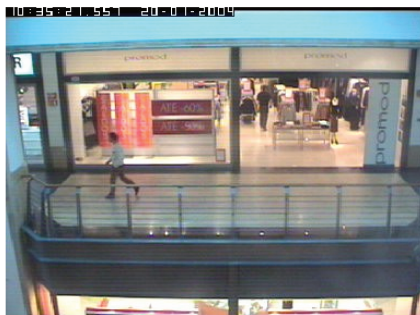
(d) Quadro 257



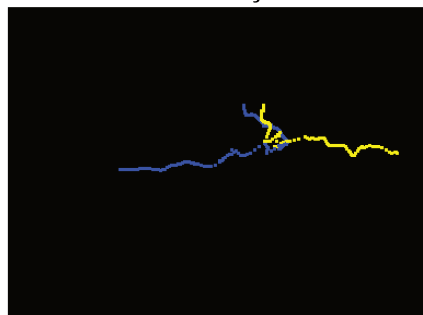
(e) Quadro 1-257 sem detecção de cabeça



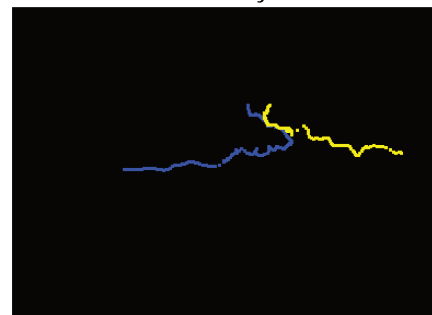
(f) Quadros 1-257 com detecção de cabeça



(g) Quadro 333



(h) Quadros 1-333 sem detecção de cabeça



(i) Quadros 1-333 com detecção de cabeça

Figura 43: Vídeo 2 – Trajetórias determinadas com e sem o uso do método de detecção de cabeças

6.3 Vídeo 3

O vídeo 3 corresponde a uma cena onde duas pessoas movimentam-se em uma sala. Diversas oclusões ocorrem entre as pessoas e entre uma placa.

O rastreamento é realizado com e sem a utilização do método de detecção de cabeças para avaliar a contribuição da utilização do método. Os resultados da análise quantitativa são apresentados na Tabela 24.

Valores esperados		
Positivos esperados	703	
Negativos esperados	0	
Valores obtidos		
	Sem detecção de cabeças	Com detecção de cabeças
Positivos verdadeiros	637	689
Falsos negativos	66	14
Negativos verdadeiros	0	0
Falsos positivos	0	0
Taxas de avaliação		
	Sem detecção de cabeças	Com detecção de cabeças
Taxa de Sensibilidade	0,91	0,98
Taxa de falso negativo	0,09	0,02
Taxa de falso positivo	0,00	0,00

Tabela 24: Vídeo 3 – Análise quantitativa dos resultados obtidos com e sem o uso do método de detecção de cabeças

Conforme é exibido na Tabela 24, para esse vídeo eram esperados 703 positivos verdadeiros e 0 negativos verdadeiros. Com a utilização do método de detecção de cabeças, o total de positivos verdadeiros aumentou de 637 para 689 e o total de falsos negativos, que é inversamente proporcional ao total de positivos verdadeiros, diminuiu de 66 para 14. As taxas de avaliação indicam que o resultado obtido com o uso do método de detecção de cabeças foi superior, o que é expressado pelo

aumento da taxa de sensibilidade e diminuição da taxa de falso negativo.

A Figura 44 exibe os principais eventos ocorridos no vídeo, e as trajetórias determinadas pelo sistema proposto. A Figura 44 (a) exibe a trajetória da pessoa determinada até o momento em que ela é parcialmente ocluída por uma placa. A Figura 44 (b) exibe o instante em que a pessoa sai de trás da placa e ajoelha-se no chão. A Figura 44 (c) mostra o momento em que ocorre a primeira oclusão entre as pessoas. A Figura 44 (d) exibe o momento em que ocorre a segunda oclusão entre as pessoas. Em todos esses eventos, as pessoas foram rastreadas corretamente.



(a) Quadros 1-121



(b) Quadros 1-219



(c) Quadros 1-313



(d) Quadros 520-529

Figura 44: Vídeo 3 – Trajetórias determinadas com o uso do método de detecção de cabeças

6.4 Vídeo 4

O vídeo 4 contém inúmeras pessoas deslocando-se na cena e a situação de maior interesse nessa avaliação corresponde a uma oclusão ocorrida entre duas pessoas. Inicialmente a primeira pessoa entra na cena e coloca uma bolsa no chão. Em seguida uma segunda pessoa para ao lado da primeira e ambas permanecem oclusas por um período de tempo. Finalmente, a primeira pessoa abandona a cena sem a bolsa, enquanto a outra permanece no local. Além de rastrear as duas pessoas, é esperado que o sistema ignore o *blob* correspondente à bolsa abandonada no chão. O total de *blobs* a serem ignorados são denominados negativos esperados.

O rastreamento é realizado com e sem a utilização do método de detecção de cabeças para avaliar a contribuição da utilização do método. Os resultados da análise quantitativa são apresentados na Tabela 25.

Valores esperados		
Positivos esperados	5128	
Negativos esperados	1058	
Valores obtidos		
	Sem detecção de cabeças	Com detecção de cabeças
Positivos verdadeiros	2298	4751
Falsos negativos	2830	377
Negativos verdadeiros	1058	1058
Falsos positivos	1872	530
Taxas de avaliação		
	Sem detecção de cabeças	Com detecção de cabeças
Taxa de Sensibilidade	0,45	0,93
Taxa de falso negativo	0,55	0,07
Taxa de falso positivo	0,64	0,33

Tabela 25: Vídeo 4 – Análise quantitativa dos resultados obtidos com e sem o uso do método de detecção de cabeças

Conforme é exibido na Tabela 25, para esse vídeo eram esperados 5128 positivos verdadeiros e 1058 negativos verdadeiros. Com a utilização do método de detecção de cabeças, o total de positivos verdadeiros aumentou de 2298 para 4751 e o total de falsos negativos, que é inversamente proporcional ao total de positivos verdadeiros, diminuiu de 2830 para 377. O total de negativos verdadeiros foi igual ao valor esperado para ambos os resultados e o total de falsos positivos diminuiu de 1872 para 530 com a utilização do método de detecção de cabeças. As taxas de avaliação indicam que o resultado obtido com o uso do método de detecção de cabeças foi superior, o que é expressado pelo aumento da taxa de sensibilidade e diminuição das taxas de falso negativo e falso positivo.

A Figura 45 exibe as trajetórias determinadas durante a oclusão principal que ocorre no vídeo. Nessa figura, é exibido em (a), (b) e (c) o *foreground* detectado antes, durante e após a oclusão. Em (d), (e) e (f) são exibidas as trajetórias determinadas sem o uso do método de detecção de cabeças e em (g), (h) e (i) são exibidas as trajetórias determinadas com o uso do método.

Durante a oclusão sem o uso do método de detecção de cabeças ambas as pessoas foram associadas à posição do centroide do único *blob* detectado, conforme é exibido na Figura 45 (e). No momento em que uma das pessoas começou a caminhar, iniciou-se a variação da posição do centroide do *blob* na direção do deslocamento da pessoa. Quando os *blobs* se separaram, as duas pessoas foram associadas ao *blob* correspondente à pessoa que estava caminhando, pois o seu centroide estava mais próximo da posição prevista para ambas as pessoas. Nos quadros seguintes, o *blob* correspondente a pessoa que permaneceu no centro da cena foi erroneamente identificado como uma nova pessoa na cena, conforme é exibido na Figura 45 (f).

Durante a oclusão com o uso do método de detecção de cabeças as posições das pessoas foram determinadas corretamente, conforme é exibido na Figura 45 (h). Quando os *blobs* se separaram, ambas as pessoas foram associadas aos respectivos *blobs*, e o rastreamento foi continuado corretamente, conforme é exibido na Figura 45 (i).

Ao longo do vídeo, outras pessoas surgem na cena. Em alguns casos ocorrem oclusões, como o exemplo mostrado na Figura 46, onde duas pessoas caminhando em direções opostas se ocluem. A Figura 46 exibe em (a), (b) e (c) o *foreground* detectado, e em (d), (e) e (f) as trajetórias das duas pessoas em questão.

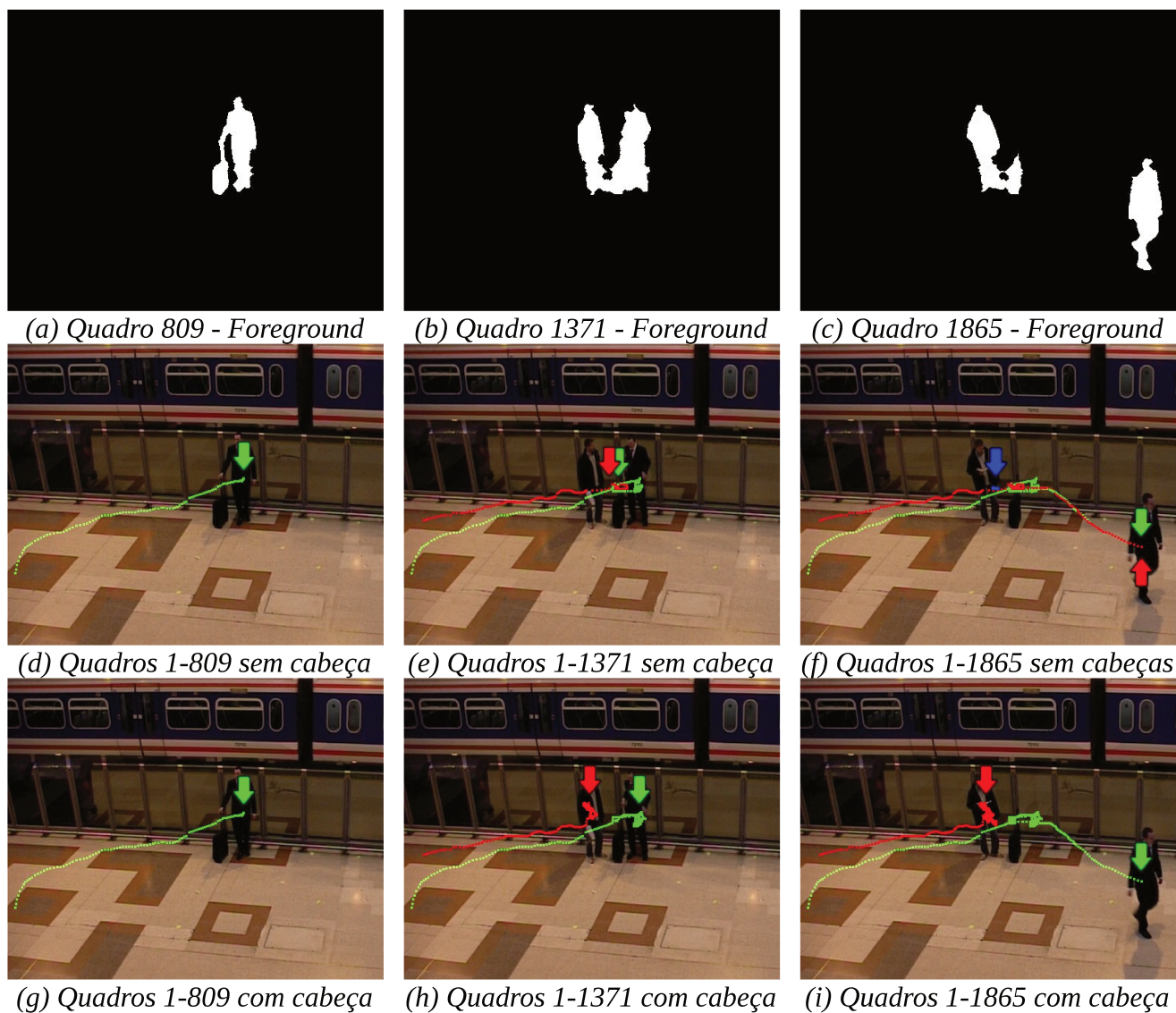


Figura 45: Vídeo 4 – Trajetórias determinadas com e sem o uso do método de detecção de cabeças

Nesse vídeo uma bolsa é abandonada na cena por um período de tempo correspondente a 1058 quadros. Por diferir do fundo da cena, os *pixels* correspondentes à bolsa são identificados como *foreground*, e um *blob* é detectado, conforme é exibido na Figura 47 (a). Esse *blob* pode ocasionar falhas no rastreamento se for erroneamente associado a uma pessoa.

Métodos de subtração de fundo presentes na literatura propõem maneiras de atualizar o modelo do fundo, afim de incorporar ao modelo as alterações de iluminação e objetos abandonados no fundo na cena. Entretanto, a frequência com que o modelo é atualizado é usualmente determinada por meio da

configuração de parâmetros. Quanto maior for a frequência de atualização do modelo de fundo, mais rápido serão incorporados ao modelo os objetos abandonados na cena. Contudo, pessoas que permanecerem imóveis na cena serão igualmente incorporadas ao modelo de fundo, ocasionando falhas de detecção das mesmas. Desse modo, a frequência de atualização pode ser configurada de acordo com a situação, afim de minimizar a detecção de objetos abandonados no fundo da cena, sem prejudicar a detecção das pessoas.

Como a frequência de atualização do modelo de fundo pode ser alterada mediante a configuração de parâmetros, foi considerado nos testes realizados que todos os 1058 *blobs* correspondentes à bolsa detectados ao longo do vídeo correspondem a negativos verdadeiros, ou seja, são *blobs* detectados que não correspondem a uma pessoa na cena, e devem ser desconsiderados durante o rastreamento. Graças à métrica de confiabilidade relacionada ao fator de movimento regular do *blob*, descrita nas equações (50) e (51), a bolsa é desconsiderada durante o rastreamento, conforme é exibido na Figura 47 (b).

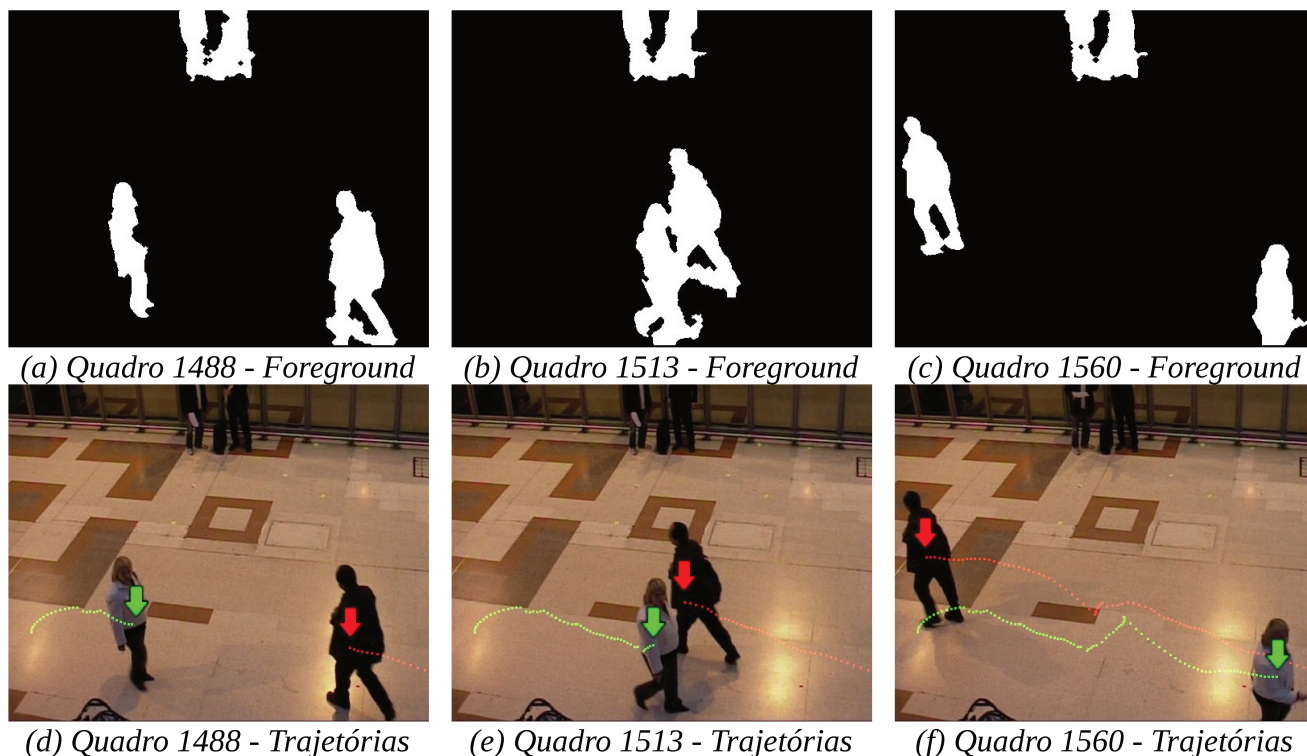


Figura 46: Vídeo 4 – Trajetórias determinadas pelo sistema durante uma oclusão

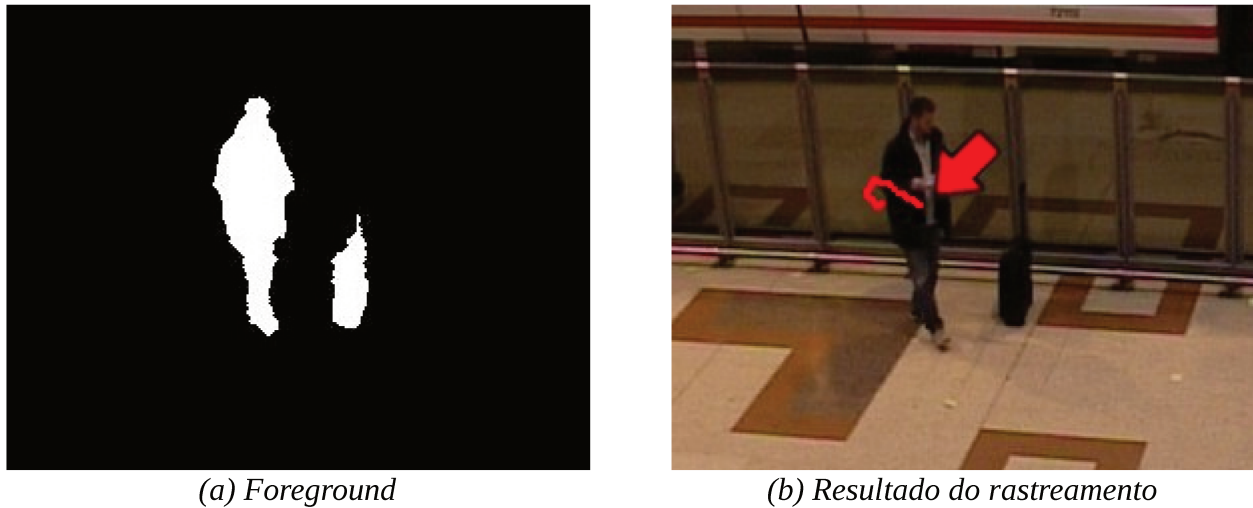


Figura 47: Vídeo 4 – Exemplo de negativo verdadeiro, onde o *blob* correspondente a uma bolsa é ignorado

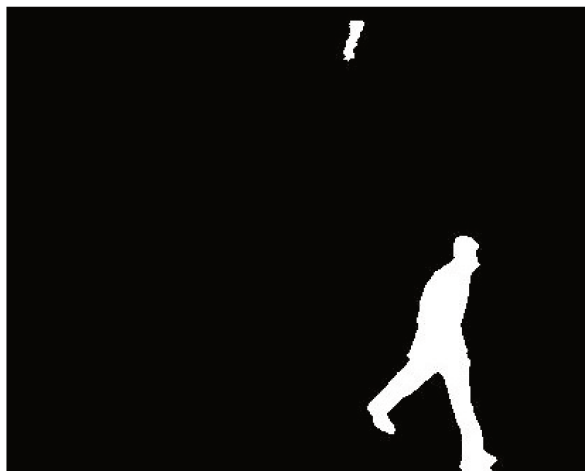
Os resultados obtidos pelo sistema proposto apresentaram falsos positivos em determinadas situações onde ocorreram falhas durante a obtenção do *foreground*. Nesse trabalho todos os *blobs* ocasionados por falhas de detecção do *foreground* que foram associados à classe Madura e tiveram a trajetória determinada pelo sistema foram considerados falsos positivos.

A Figura 48 exibe a ocorrência de uma falha na detecção do *foreground* ocasionada pelo reflexo de uma pessoa. A Figura 48 (a) exibe o *blob* correspondente a pessoa na cena e o *blob* correspondente ao reflexo da pessoa. O *blob* correspondente ao reflexo satisfaz os critérios relacionados ao tempo de permanência na cena e fator de movimento regular, foi associado à classe madura e teve a trajetória determinada, conforme é exibido na Figura 48 (b), ocasionando o falso positivo.

A Figura 49 exibe uma situação de falso positivo onde uma porta é aberta e é detectada pelo método de subtração de fundo, conforme é exibido na Figura 49 (a). Conforme a porta se desloca na cena o *blob* detectado satisfaz o critério relacionado ao tempo de permanência na cena e a variação da posição do centroide do *blob* satisfaz o critério relacionado ao fator de movimento regular. Desse modo, o *blob* é associado a classe Madura e têm a trajetória determinada pelo sistema, conforme é exibido na Figura 49 (b), ocasionando o falso positivo.

Falsos negativos ocorreram durante oclusões em situações nas quais houveram falhas durante a detecção de cabeças. A Figura 50 exibe uma situação onde duas pessoas surgem na cena em estado de

occlusão. Entretanto, a cabeça de uma das pessoas não corresponde a um pico na extremidade superior do *blob*, conforme é exibido na Figura 50 (a). Como o método de detecção de cabeças utilizado no presente trabalho identifica as cabeças a partir dos picos na extremidade do *blob*, uma única pessoa foi identificada na cena e uma única trajetória foi determinada, conforme é exibido na Figura 50 (b).



(a) Foreground



(b) Resultado do rastreamento

Figura 48: Vídeo 4 – Exemplo de falso positivo ocasionado pelo reflexo de uma pessoa



(a) Foreground



(b) Resultado do rastreamento

Figura 49: Vídeo 4 – Exemplo de falso positivo ocasionado por uma porta em movimento



(a) Foreground



(b) Resultado do rastreamento

Figura 50: Vídeo 4 – Exemplo de falso negativo ocasionado pelo posicionamento das pessoas

Capítulo 7

Conclusão

O presente trabalho abordou aspectos do rastreamento de pessoas em vídeo e propôs um sistema de rastreamento de pessoas baseado em abordagens descritas na literatura. Esse sistema é dividido em duas etapas principais: detecção do *foreground* e determinação das trajetórias.

O *foreground* é detectado através do método de subtração de fundo proposto por Haritaoglu *et al.* (2000). Esse método, além de possuir implementação simples, é capaz de lidar com eventos como alterações de iluminação, alterações físicas no ambiente e objetos em movimento repetitivo no fundo da cena. Para melhorar o resultado da subtração de fundo, é utilizado um pós-processamento baseado em remoção de sombras (Cucchiara *et al.*, 2001) e processamento morfológico. Testes realizados neste trabalho demonstraram que, para os vídeos avaliados, o pós-processamento utilizado melhorou o resultado da detecção do *foreground* e, conseqüentemente, do rastreamento.

Para determinar as trajetórias são utilizadas métricas de associação relacionadas à posição do centroide, *bounding box* e histograma de cor. Durante oclusões, é utilizado um método de detecção de cabeças baseado na análise das bordas do *blob*. Com taxas de precisão entre 87% e 99%, o resultado do rastreamento foi considerado satisfatório para os vídeos avaliados.

O método de classificação de estados proposto por Lei e Xu (2006) e modificado por Freitas (2010) foi adaptado e utilizado para classificar as pessoas seguidas. Com base nos testes realizados é possível afirmar que um sistema de classificação de estado oferece uma maneira organizada e simples para gerenciar o rastreamento das pessoas presentes no vídeo.

Todos os objetivos do presente trabalho foram plenamente alcançados e a seguir são listadas possíveis continuações:

- Propor melhorias na etapa de detecção do *foreground*, através do estudo de outras técnicas descritas na literatura ou do desenvolvimento de novas abordagens.
- Propor melhorias na etapa de detecção de cabeças, através do estudo de outras técnicas descritas na literatura ou do desenvolvimento de novas abordagens.
- Incluir um método de classificação na etapa de detecção para distinguir pessoas de objetos, veículos e animais.
- Adaptar o sistema proposto no presente trabalho para utilizar imagens capturadas simultaneamente por duas ou mais câmeras e avaliar a contribuição de uma abordagem fundamentada em multi-visão aplicada ao tratamento de oclusões.

Referências bibliográficas

Alam, M. O. e Boufama, B., “Occlusion Handling Based on Sub-blobbing in Automated Video Surveillance System”, In *Proceedings of The Fourth International Conference on Computer Science and Software Engineering (C3S2E '11)*. ACM, New York, NY, USA, 139-143. 2011

Amaral, W. M. e Tozzi, C. L., “Uso de Morfologia Matemática para Remoção de Ruídos da Subtração de Fundo”, *Workshop de Visão Computacional (WVC'2011)*, Curitiba, Brasil, 2011

Amato, A.; Mozerov, M.G.; Bagdanov, A.D.; Gonzalez, J., "Accurate Moving Cast Shadow Suppression Based on Local Color Constancy Detection,"*IEEE Transactions on Image Processing*, vol.20, no.10, pp.2954-2966, Oct. 2011

Azari, M., Seyfi, A., Rezaie, A.H., , "Real Time Multiple Object Tracking and Occlusion Reasoning Using Adaptive Kalman Filters," *7th Iranian Machine Vision and Image Processing (MVIP)*, pp.1-5, 16-17 Nov. 2011

Blackman, S. e Popoli, R., "Design and Analysis of Modern Tracking Systems", Norwood, MA: Artech House, 1999

Bovik, A., *Handbook of Image and Video Processing*. New York: Academic. 2000

Castelli, V. e Bergman, L. D., “Digital Imagery: Fundamentals, in *Image Databases: Search and Retrieval of Digital Imagery*”, (eds V. Castelli and L. D. Bergman), John Wiley & Sons, Inc., New York, USA. 2002

Chang, F., Chen, C. J. e Lu, C. J., “A Linear-time Component Labeling Algorithm Using Contour Tracing Technique,” *Computer Vision and Image Understanding*, Vol. 93, No. 2, pp. 206-220, 2004.

Cheriet, M.; Kharma, N.; Cheng-Lin Liu, “Character Recognition Systems: A Guide for Students and

Practitioners”, Wiley Publications, 2007.

Cucchiara, R., Grana, C., Piccardi, M., Prati, A., Sirotti, S., “Improving shadow suppression in moving object detection with HSV color Information”. *Proceedings of IEEE Conference on Intelligent Transportation Systems*. Oakland, USA, 2001:334-339, 2001

Cucchiara, R., Grana, C., Piccardi, M., Prati, ^a, "Detecting moving objects, ghosts, and shadows in video streams," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.25, no.10, pp. 1337-1342. 2003

Del Bimbo, A., “Visual Information Retrieval”. *Morgan Kaufmann Publishers Inc.*, San Francisco, CA, USA. 1999

D’Elia, Edoardo; Taviani, Carlo e Zugno, Alberto, "Simulation and Multi-Target Tracking". University of Padua. 15 Feb 2010

Dougherty, E. e Lotufo, R.A., “Hands-on morphological image processing”. Bellingham: SPIE. 272 p., il. (Tutorial texts in optical engineering; v. v. TT59). ISBN 081944720X, 2003

Ellis, T. J., “Performance metrics and methods for tracking in surveillance”. In 3rd IEEE Workshop on Performance Evaluation of Tracking and Surveillance, Copenhagen, Denmark, 2002

Freitas, G. M., “Rastreamento de objetos em vídeos e separação em classes”. Dissertação de Mestrado. Universidade Estadual de Campinas, Faculdade de Engenharia Elétrica e de Computação, 2010. <http://cutter.unicamp.br/document/?code=000771490>.

Fujiyoshi, H.; Lipton, A.J., "Real-time human motion analysis by image skeletonization,". *Proceedings of Fourth IEEE Workshop on Applications of Computer Vision (WACV '98)*, pp.15-21, 19-21 Oct 1998

Girgensohn, A., Kimber, D., Vaughan, J., Yang, T., Shipman, F., Turner, T., Rieffel, E., Wilcox, L., Chen, F. e Dunnigan, T., “DOTS: support for effective video surveillance”, *Proceedings of the 15th*

international conference on Multimedia, 2007

Gonzalez, R. C. e Woods, R. E., “Digital image processing”. 2nd ed. Upper Saddle River, N.J.: Prentice-Hall, 2002

Haritaoglu, I., Harwood, D. e Davis, L. S., “Hydra: Multiple people detection and tracking using silhouettes”. *IEEE Workshop on Visual Surveillance*, 1999

Haritaoglu, I., Harwood, D., e Davis, L. S., “W4: real-time surveillance of people and their activities”. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):809-830, 2000

Horprasert, T., Harwood, D., Davis, L.S., “A statistical approach for real-time robust background subtraction and shadow detection”. In: *Proc. of IEEE ICCV’99 Frame-Rate Workshop*, Corfu, Greece, Sept 1999

Hsieh, J. W. ; Yu, S. H.; Chen, Y. S.; Hu, W. F.; BA, shadow elimination method for vehicle analysis”, in *Proc. IEEE Int. Conf. Pattern Recognition*, Cambridge, U.K., pp. 372–375, 2004

Hu, W., Tan, T., Wang, L. and Maybank, S., “A survey on visual surveillance of object motion and behaviors”. *IEEE Transactions on Systems, Man, and Cybernetics*, Part C: Applications and Reviews. pp. 334-352, 2004

Indupalli, S., Ali, M.A., Boufama, B., "A Novel Clustering-Based Method for Adaptive Background Segmentation," 3rd Canadian Conference on Computer and Robot Vision, pp. 37, 07-09 Jun 2006

Jain, R. e Nagel, H. H., “On the analysis of accumulative difference pictures from image sequences of real world scenes”. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1(2):206–214, 1979

Kalman, R. E., “A new approach to linear filtering and prediction problems”. *Transactions of the ASME, Journal of Basic Engineering*, (82 (Series D)):35–45, 1960

Kim, J. S., Yeom, D. H., Joo, Y. J., "Fast and robust algorithm of tracking multiple moving objects for intelligent video surveillance systems," *IEEE Transactions on Consumer Electronics*, vol.57, no.3, pp.1165-1170, Aug 2011

Ko, T., "A survey on behavior analysis in video surveillance for homeland security applications". *37th IEEE on Applied Imagery Pattern Recognition Workshop (AIPR)*, 2008

Lei, B. e Xu, L.-Q., "Real-time outdoor video surveillance with robust foreground extraction and object tracking via multi-state transition management". *Pattern Recognition Letters*, 27(15):1816–1825. Vision for Crime Detection and Prevention. 2006

Li-Qun Xu; Landabaso, J.L.; Pardas, M., "Shadow Removal with Blob-Based Morphological Reconstruction for Error Correction,". In *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '05)*, vol.2, pp. 729- 732, 18-23 Mar 2005

Lu, W. e Tan, Y. P., "A color histogram based people tracking system". *IEEE International Symposium on Circuits and Systems*. Vol. 2, pp. 137-140, 2001

McKenna, S. J.; Jabri, S.; Duric, Z.; Rosenfeld, ^a, "*Tracking groups of people*" *Comput. Vision Image Understanding*, no. 80, pp. 42–56, 2000

Merad, D.; Aziz, K.; Thome, N., "Fast People Counting Using Head Detection from Skeleton Graph," *7th IEEE International Conference on Advanced Video and Signal Based Surveillance*, pp. 233-240, 2010

Mirabi, M. e Javadi, S., "People Tracking in Outdoor Environment Using Kalman Filter", *Third International Conference on Intelligent Systems, Modelling and Simulation*, pp. 303-307, 2012.

Oliver, N.M., Rosario, B. e Pentland, A.P., "A bayesian computer vision system for modeling human interactions". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):831-843. 2000

Pinho, R. R., Tavares, J. M. R. S., Correia, M. V., “Human Movement Tracking and Analysis with Kalman Filtering and Global Optimization Techniques” . *ICCB2005 - II International Conference on Computational Bioengineering*, pp. 915-926, Instituto Superior Técnico, Lisboa, Portugal, 14-16 Sept. 2005

Pinho, R. R., Tavares, J. M. R. S., “Tracking Features in Image Sequences with Kalman Filtering, Global Optimization, Mahalanobis Distance and a Management Model”. *Computer Modeling in Engineering & Sciences*, Tech Science Press, ISSN: 1526-1492 (print) - 1526-1506 (online), vol. 46, no. 1, pp. 51-75. 2009

Prati, A.; Mikic, I. ; Trivedi, M.M.; Cucchiara, R., “Detecting Moving Shadows: Algorithms and Evaluation”, in *IEEE Transaction on PAMI*, vol. 25, n.7, pp. 918-923, July, 2003

Rodgers, J. L. e Nicewander, W. ^a, “Thirteen ways to look at the correlation coefficient,” *American Statistician* 42: 59–66.1. 1988

Seo, Y., Choi, S., Kim, H. e Hong, K.,, “Where Are the Ball and Players? Soccer Game Analysis with Color Based Tracking and Image Mosaick”. In *Proceedings of the 9th International Conference on Image Analysis and Processing-Volume II (ICIAP '97)*, Alberto Del Bimbo (Ed.). Springer-Verlag, London, UK, 196-203. 1997

Stauffer, C. e Grimson, W.E.L., “Adaptive background mixture models for real-time tracking”. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 2, pp. 252, 1999

Swain, M. J. e Ballard D. H., “Color Indexing”. *International Journal of Computer Vision*. 1991

Zhang, Z.; Gunes, H.; Piccardi, M., "An accurate algorithm for head detection based on XYZ and HSV hair and skin color models", *15th IEEE International Conference on Image Processing ICIP*, pp.1644-1647, 12-15 Oct. 2008

Zhang, Z e Piccardi, M., “A review of tracking methods under occlusion”. *IAPR Conference on Machine Vision Applications*. pp. 146-149. 2007

Zhao, T. e Nevatia, R., “Tracking multiple humans in complex situations”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.26, no.9, pp.1208-1221, 2004

Zui Zhang; Gunes, H.; Piccardi, M., "An accurate algorithm for head detection based on XYZ and HSV hair and skin color models," 15th IEEE International Conference on Image Processing ICIP, pp.1644-1647, 12-15 Oct. 2008

Wei Li, Xiaojuan Wu, Matsumoto, K., Hua-An Zhao, "Foreground detection based on optical flow and background subtract," *International Conference on Communications, Circuits and Systems (ICCCAS)*, pp.359-362, 28-30 Jul 2010

Wei P., Kun W., Zhi C., Zhi-Sheng Y., "A Background Reconstruction Method Based on Double-background," Fourth International Conference on Image and Graphics (ICIG), pp.502-507, 22-24 Aug. 2007

Win Kong; Hussain, A.; Saad, M.H.M.; Tahir, N.M., "Hand detection from silhouette for video surveillance application," IEEE 8th International Colloquium on Signal Processing and its Applications (CSPA), pp.514-518, 23-25 Mar 2012

Wren, C.R., Azarbayejani, A., Darrell, T. e Pentland, A.P., “Pfinder: real-time tracking of the human body”. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 7., pp. 780-785. Jul 1997

Yilmaz, A., Javed, O., SHAH, M., “Object Tracking: A survey”. *ACM Comput. Surv.* 38. p. 45. Article 13. 2006

Yilmaz, ^a, “Object Tracking by Asymmetric Kernel Mean Shift with Automatic Scale and Orientation Selection”. Ohio State University. IEE 1-4244-1180-7/07. 2007