

UNIVERSIDADE FEDERAL DE UBERLÂNDIA
FACULDADE DE CIÊNCIA DA COMPUTAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO



**RECONHECIMENTO DE ATIVIDADES SUSPEITAS EM
AMBIENTE EXTERNO VIA ANÁLISE DE VÍDEO
INFRAVERMELHO.**

HENRIQUE COELHO FERNANDES

Uberlândia - Minas Gerais

2011

UNIVERSIDADE FEDERAL DE UBERLÂNDIA
FACULDADE DE CIÊNCIA DA COMPUTAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO



HENRIQUE COELHO FERNANDES

**RECONHECIMENTO DE ATIVIDADES SUSPEITAS EM
AMBIENTE EXTERNO VIA ANÁLISE DE VÍDEO
INFRAVERMELHO.**

Dissertação de Mestrado apresentada à Faculdade de Ciência da Computação da Universidade Federal de Uberlândia, Minas Gerais, como parte dos requisitos exigidos para obtenção do título de Mestre em Ciência da Computação.

Área de concentração: Banco de Dados.

Orientadora:

Prof^a. Dr^a. Celia Aparecida Zorzo Barcelos

Uberlândia, Minas Gerais
2011

Dados Internacionais de Catalogação na Publicação (CIP)
Sistema de Bibliotecas da UFU

F363r Fernandes, Henrique Coelho, 1986-
Reconhecimento de atividades suspeitas em ambiente externo via
análise de vídeo infravermelho / Henrique Coelho Fernandes. - 2011.
103 f. : il.

Orientadora: Célia Aparecida Zorzo Barcelos.

Dissertação (mestrado) – Universidade Federal de Uberlândia, Programa de Pós-Graduação em Ciência da Computação.

Inclui bibliografia.

1. Computação - Teses. 2. Banco de dados - Teses. I. Barcelos, Célia Aparecida Zorzo. II. Universidade Federal de Uberlândia. Programa de Pós-Graduação em Ciência da Computação. III. Título.

CDU: 681.3

UNIVERSIDADE FEDERAL DE UBERLÂNDIA
FACULDADE DE CIÊNCIA DA COMPUTAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Os abaixo assinados, por meio deste, certificam que leram e recomendam para a Faculdade de Ciência da Computação a aceitação da dissertação intitulada “**Reconhecimento de atividades suspeitas em ambiente externo via análise de vídeo infravermelho.**” por **Henrique Coelho Fernandes** como parte dos requisitos exigidos para a obtenção do título de **Mestre em Ciência da Computação**.

Uberlândia, 26 de Outubro de 2011

Orientadora:

Prof^a. Dr^a. Celia Aparecida Zorzo Barcelos
Universidade Federal de Uberlândia

Banca Examinadora:

Prof. Dr. Bruno Augusto Nassif Travençolo
Universidade Federal de Uberlândia

Prof^a. Dr^a. Fátima Nelsizeuma Sombra de Medeiros
Universidade Federal do Ceará

Prof. Dr. Marcos Aurélio Batista
Universidade Federal de Goiás

UNIVERSIDADE FEDERAL DE UBERLÂNDIA
FACULDADE DE CIÊNCIA DA COMPUTAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Data: Outubro de 2011

Autor: **Henrique Coelho Fernandes**
Título: **Reconhecimento de atividades suspeitas em ambiente externo
via análise de vídeo infravermelho.**
Faculdade: **Faculdade de Ciência da Computação**
Grau: **Mestrado**

Fica garantido à Universidade Federal de Uberlândia o direito de circulação e impressão de cópias deste documento para propósitos exclusivamente acadêmicos, desde que o autor seja devidamente informado.

Autor

O AUTOR RESERVA PARA SI QUALQUER OUTRO DIREITO DE PUBLICAÇÃO DESTE DOCUMENTO, NÃO PODENDO O MESMO SER IMPRESSO OU REPRODUZIDO, SEJA NA TOTALIDADE OU EM PARTES, SEM A PERMISSÃO ESCRITA DO AUTOR.

Dedicatória

Aos meus pais Aparecida Coêlho Fernandes e Prof. Dr. Jonas Jäger Fernandes.

A minha querida esposa Bruna Santos Fernandes.

Agradecimentos

Agradeço...

À Deus que é o criador e sustentador da minha vida.

À minha família pelo amor incondicional. Não existe família perfeita mas não há nada melhor que a família.

À minha orientadora professora Celia Aparecida Zorzo Barcelos, por me ensinar, incentivar, direcionar e por ser um exemplo a quem sempre seguirei.

Ao professor Xavier Maldague pelos conselhos e oportunidade de trabalhar em seu laboratório.

Ao professor Marcos Aurélio Batista pelos conselhos e atenção que me auxiliaram no desenvolvimento desse trabalho.

Aos meus amigos da UFU que se mostraram companheiros e, diretamente ou indiretamente, contribuíram para a realização deste trabalho.

À CAPES pelo apoio financeiro.

*“A imaginação é mais importante que o conhecimento.”
(Albert Einstein)*

Resumo

Vigilância se tornou, nos últimos anos, algo ubíquo em nossa sociedade. Cada dia que passa é mais notória a presença de sistemas inteligentes de vigilância em nosso dia-a-dia. Isso se deve tanto aos avanços tecnológicos alcançados nas últimas décadas (aumento da capacidade de processamento e armazenamento, miniaturização de dispositivos como detectores biométricos e câmeras de vídeo) como a constante sensação de insegurança vivida em vários países. Após os dias sombrios de 11/09, segurança e vigilância se tornaram algo primordial.

Este trabalho visa o estudo de técnicas para o desenvolvimento de um sistema de vigilância para um estacionamento externo baseado em uma câmera estacionária. Tendo em vista que em um estacionamento externo é de suma importância que a vigilância seja feita tanto de dia quanto de noite, neste trabalho utilizamos uma câmera que captura imagens infravermelhas. Uma câmera infravermelha permite que enxerguemos objetos de interesse na cena até mesmo a noite. As imagens usadas nos experimentos realizados neste trabalho foram colhidas no campus da Universidade de Laval (Canadá) durante um estágio realizado no *Canada Research Chair in Multipolar Infrared Vision*.

Um sistema de vigilância baseado em câmeras de vídeo geralmente possui três partes principais: (i) detecção de movimento, (ii) monitoramento e (iii) gerenciamento de eventos. Neste trabalho, utilizamos uma dinâmica técnica de subtração de plano de fundo para realizar a detecção de movimento (segmentação de movimento). Esta técnica se adapta às mudanças bruscas de iluminação na cena tornando o método de segmentação robusto a estas mudanças. Além disso, utilizamos análise de fluxo de movimento para restringir a segmentação somente às regiões onde existem algum movimento na cena. A técnica de monitoramento de objetos em movimento usada neste trabalho é baseada em um ciclo de dois estágios: previsão e correção. Os eventos de interesse que ocorrem na área monitorada são modelados de forma explícita sendo então reconhecidos e interpretados. O foco principal deste trabalho é o reconhecimento de eventos suspeitos. Resultados experimentais obtidos mostram que tais técnicas são adequadas para um sistema de vigilância de um estacionamento externo baseado em uma câmera estacionária infravermelha.

Palavras chave: sistema automático de vigilância, reconhecimento de eventos suspeitos, vídeos infravermelhos, segmentação de plano de fundo.

Abstract

Surveillance has become, in the last years, something ubiquity in our society. Every day it is more notorious the presence of intelligent systems for surveillance in our everyday life. This is due to technological advances achieved in recent decades (storage and processing speed increasing, miniaturization of devices like biometric detectors and video cameras) as the constant feeling of insecurity experienced in several countries. Following the dark days of 9/11, security and surveillance became paramount.

This work aims the study of techniques for the development of a surveillance system of an outdoor parking lot based on a stationary camera. Considering that in an outdoor parking lot it is very important that the surveillance is made both day and night, in this work we use an infrared camera to record images. An infrared camera allows to see objects of interest in the scene even at night. The images used for the experiments in this work were recorded by the student in Laval University campus (Canada) during an internship he held in the "Canada Research Chair in Multipolar Infrared Vision".

A surveillance system based on video cameras is usually composed of three parts: (i) motion detection, (ii) tracking and (iii) event management. In this work, we use a dynamic background subtraction technique to detect motion (motion segmentation). This technique adapts to abrupt changes on the scene's illumination making the technique robust to this changes. Besides, we use flow analysis to restrict the segmentation process only to regions where we have motion in the scene. The object tracking technique used is based on a two phase cycle: prediction and correction. The events of interest which occur in the monitored area are modeled explicitly and then recognized and interpreted. The main goal of this project is to recognize suspicious events. Experimental results show that such techniques are suitable for a surveillance system for an outdoor parking lot based on a infrared stationary camera.

Keywords: automatic surveillance system, suspicious event recognition, infrared imagery, background subtraction.

Sumário

Lista de Figuras	xiii
Lista de Tabelas	xvi
Lista de Abreviaturas e Siglas	xvii
1 Introdução	18
I Fundamentação Teórica e Trabalhos Correlatos	23
2 Vídeo	24
2.1 Introdução	24
2.2 Visão infravermelha	26
2.2.1 Definição	26
2.2.2 Espectro eletromagnético	27
2.2.3 Aplicações	28
2.3 Considerações Finais	29
3 Sistemas de vídeovigilância	30
3.1 Introdução	30
3.2 Detecção de movimento	31
3.3 Monitoramento	33
3.4 Modelagem de eventos	34
3.5 Considerações Finais	36
II Método Proposto	37
4 Módulo de segmentação	38
4.1 Introdução	38
4.2 Subtração de plano de fundo	38
4.2.1 Cálculo do <i>foreground</i>	39

4.2.2	Cálculo do <i>background</i>	40
4.2.3	Arquitetura do módulo de segmentação de movimento	41
4.2.4	Cálculos do limiar dinâmico	43
4.2.5	Cálculos da taxa de aprendizagem dinâmica	44
4.3	Análise de Fluxo	45
4.3.1	Introdução	45
4.3.2	Detecção de fluxo	46
4.4	Detecção do contorno dos objetos e operações morfológicas	49
4.5	Extração de características	52
4.6	Classificação dos objetos em movimento	53
4.6.1	A classe <i>humano</i>	54
4.6.2	A classe <i>veículo</i>	55
4.6.3	A classe <i>não identificado</i>	55
4.7	Considerações finais	56
5	Módulo de monitoramento	57
5.1	Introdução	57
5.2	Algoritmo proposto	57
5.2.1	Inicialização do monitoramento	59
5.2.2	Fase de previsão	59
5.2.3	Fase de correção	60
5.2.4	Gestão de oclusões	61
5.2.5	Gestão de impressões térmicas	61
5.3	Considerações finais	63
6	Módulo de gerenciamento de eventos	64
6.1	Introdução	64
6.2	Modelagem do evento	64
6.3	Eventos de um único ator	67
6.4	Eventos de múltiplos atores	68
6.4.1	Eventos simples	68
6.4.2	Eventos suspeitos	69
6.5	Considerações finais	73
7	Resultados experimentais	74
7.1	Sistema de aquisição	74
7.2	Banco de dados	76
7.3	Módulo de segmentação e classificação de objetos	78
7.3.1	Segmentação não restrita ao fluxo	78
7.3.2	Segmentação restrita ao fluxo	82

7.4	Módulo de monitoramento	84
7.5	Módulo de gerenciamento de eventos	88
7.5.1	Eventos de um único ator	88
7.5.2	Eventos de múltiplos atores	90
7.5.3	Eventos suspeitos	91
7.6	Considerações finais	93
8	Conclusões	95
8.1	Trabalhos Futuros	96
	Referências Bibliográficas	97

Lista de Figuras

1.1	O aumento das câmeras de vídeo em sistemas de vídeovigilância gera uma quantidade de informações de vídeo que vai além da capacidade humana de processamento, retirada de [Haering et al. 2008].	19
1.2	Visão da área do estacionamento monitorada por uma câmera de vigilância. (a) cena capturada pela câmera infravermelha, (b) cena capturada por câmera convencional	20
2.1	Esquema da composição de um vídeo: um agrupamento de imagens em seqüência.	24
2.2	Diferença entre (a) imagens de visão noturna e (b) imagens de visão infravermelha.	26
2.3	Espectro eletromagnético com destaque para a banda infravermelha, retirado de [Ibarra-Castanedo 2005].	27
4.1	Efeitos de uma aparição repentina do sol no processo de segmentação. Figuras (a) a (d) foram capturadas em um intervalo de 2 segundos. O <i>foreground</i> foi bastante afetado devido a esta aparição do sol.	41
4.2	Diagrama de blocos do processo de segmentação de movimento.	42
4.3	Falha na segmentação de objeto pequenos na cena. A Figura 4.3(a) mostra a cena do estacionamento onde há dois humanos, estando um na parte inferior da cena e outro na parte superior localizado atrás dos galhos de uma árvore. A Figura 4.3(b) mostra que a segmentação do primeiro humano foi realizada com sucesso e a segmentação do segundo humano falhou. O círculo vermelho na Figura 4.3(b) mostra o local onde um humano está, porém ele não é identificado na segmentação. A Figura 4.3(b) é o resultado da segmentação não utilizando a detecção de fluxo.	46
4.4	Processo de obtenção do fluxo de movimento associado a uma seqüência de quadros de um vídeo.	47
4.5	Detecção de fluxo de movimento. (a) quadro original da seqüência de vídeo no instante k , (b) fluxo associado obtido com $m = 30$, (c) fluxo final associado ao quadro do instante k após operações morfológicas.	47

4.6	Segmentação com e sem restrição ao fluxo. As Figuras 4.6(a) e 4.6(d) mostram frames da cena original, as Figuras 4.6(b) e 4.6(e) mostram o resultado da segmentação sem a restrição das regiões de fluxo e as Figuras 4.6(c) e 4.6(f) mostram o resultado da segmentação com a restrição de fluxo.	48
4.7	Resultado da combinação do objeto com o seu contorno. Figuras 4.7(a) e 4.7(b) mostram a cena capturada pela câmera de vigilância, Figuras 4.7(c) e 4.7(d) mostram os contornos dos objetos, Figuras 4.7(e) e 4.7(f) mostram os "borrões" obtidos pela aplicação da técnica de subtração de plano de fundo e Figuras 4.7(g) e 4.7(h) mostram o resultado da combinação dos "borrões" com o contorno dos objetos.	50
4.8	Resultado das operações morfológicas aplicadas nos objetos segmentados pela técnica de subtração de plano de fundo usada neste trabalho. A Figura 4.8(a) mostra a cena do vídeo de vigilância em um instante qualquer, a Figura 4.8(b) objetos resultantes da aplicação da técnica de subtração de plano de fundo na cena da Figura 4.8(a) e a Figura 4.8(c) resultado das operações morfológicas aplicadas na Figura 4.8(b).	51
4.9	Objeto da classe humano com sua caixa delimitadora.	52
4.10	Classificação correta de dois humanos pelo algoritmo.	55
4.11	Classificação correta de um veículo pelo algoritmo.	55
5.1	Diagrama de previsão e correção.	58
5.2	Estágio de correção realizado através da sobreposição de caixas delimitadoras.	60
5.3	Impressão térmica anexada ao veículo no momento de sua saída após estar estacionado durante um tempo.	62
6.1	Classes de eventos que podem ser tratados pelo módulo de gerenciamento de eventos.	66
6.2	Exemplo de caso onde pode ocorrer indecisão no momento da entrada do humano no veículo.	69
6.3	Evento suspeito do tipo C: (a) humano se aproximando de um veículo estacionado na posição x, (b) humano ficando próximo deste veículo por um tempo, (c) humano entrando dentro do veículo, (d) humano se afastando deste veículo.	73
7.1	Câmera infravermelha utilizada neste projeto.	74
7.2	Curva de resposta espectral típica da câmera Photon 640.	75
7.3	Esquema do sistema de aquisição.	75
7.4	Sistema de aquisição usado para as gravações.	76
7.5	(a) imagem retirada do banco de dados usados em [Maadi e Maldague 2007] e (b) imagem retirada do banco de dados construído para este trabalho.	77

7.6	Segmentação sem restrição de fluxo e classificação de humanos em três diferentes cenas do banco de dados: (a) cena original e (b) resultado da segmentação da cena em (a); (c) cena original e (d) resultado da segmentação da cena em (c); (e) cena original e (f) resultado da segmentação da cena em (e).	79
7.7	Segmentação sem restrição de fluxo e classificação de veículos em três diferentes cenas do banco de dados: (a) cena original e (b) resultado da segmentação da cena em (a); (c) cena original e (d) resultado da segmentação da cena em (c); (e) cena original e (f) resultado da segmentação da cena em (e).	80
7.8	Comparação entre os três métodos de segmentação e classificação: método usado por [Maadi e Maldague 2007], método usado neste trabalho sem restrição de fluxo e método usado neste trabalho com restrição de fluxo. Podemos ver que o último método usando restrição de fluxo é o que apresentou melhor desempenho.	83
7.9	Exemplo de uma seqüência de imagens originais: os quadros (a) a (o) são consecutivos e contém diversos objetos em movimento.	85
7.10	Exemplo de resultado de monitoração: os quadros (a) a (o) são uma seqüência de quadros binários que representam o monitoramento de diversos objetos simultaneamente.	86
7.11	Exemplo de uma seqüência de imagens que ilustra o monitoramento de dois humanos durante uma oclusão parcial e temporária.	87
7.12	Exemplo de trajetórias calculadas durante a monitoração de objetos. Cada trajetória contém as informações de posição e sentido de movimento do centro de massa do objeto.	87
7.13	Exemplo da detecção do evento "Parar caminha" para a classe <i>humano</i> . . .	88
7.14	Exemplo da detecção do evento "Caminhar novamente" para a classe <i>humano</i> . . .	88
7.15	Exemplo da detecção do evento "Estacionar veículo" para a classe <i>veículo</i> . . .	89
7.16	Exemplo da detecção do evento "Iniciar" para a classe <i>veículo</i> . Cena gravada sob forte chuva.	89
7.17	Exemplo da detecção do evento "Entrar em um veículo" envolvendo um <i>humano</i> e um <i>veículo</i>	90
7.18	Exemplo da detecção do evento "Sair do interior do veículo" envolvendo um <i>humano</i> e um <i>veículo</i>	91
7.19	Evento suspeito alvo da análise deste projeto. Na seqüência os 4 sub-eventos que o compõem: (a) c1 - homem se aproximando do veículo estacionado, (b) humano ficando próximo deste veículo por um certo tempo, (c) humano entrando dentro do veículo e (d) humano se afastando do veículo. (Utilizando segmentação sem restrição ao fluxo).	92

Lista de Tabelas

2.1	Bandas de frequência do espectro infravermelho: adaptado de [Ibarra-Castanedo 2005], [Hudson 1969], [Maldague 2001] e [Piotrowski e Rogalski 2004].	28
7.1	Resultado da classificação usando a segmentação não restrita ao fluxo utilizando o banco de dados construído neste trabalho.	81
7.2	Resultado da classificação e segmentação obtidos em [Maadi e Maldague 2007]. Método e banco de dados usados naquele trabalho.	81
7.3	Resultado da classificação usando a segmentação restrita ao fluxo utilizando o banco de dados construído neste trabalho.	83
7.4	Resultado da classificação de eventos suspeitos usando a segmentação restrita ao fluxo utilizando o banco de dados construído neste trabalho. . . .	92
7.5	Resultado da classificação de eventos suspeitos usando a segmentação restrita ao fluxo utilizando o banco de dados construído neste trabalho. . . .	93

Lista de Abreviaturas e Siglas

Pixel	Picture Element
FPS	Frames Per Second
CCTV	Closed-circuit television
IR	Infrared
NIR	Near Infrared
SWIR	Short Wave Infrared
MWIR	Medium Wavelength Infrared
LWIR	Long Wavelength Infrared
VLWIR	Very Long Wavelength Infrared
NDE	NonDestructive Evaluation
BB	Black Body
MCT	Mercury Cadmium Tellurium
Si	Silício
SiO ₂	Sílica
In	Índiu
Ga	Gálio
As	Arsênico
Pb	Chumbo
S	Enxofre
Sb	Antimônio
Se	Selênio
Pt	Platina
Hg	Mercúrio
Cd	Cádmio
Te	Telúrio

Capítulo 1

Introdução

A preocupação com a segurança em nossa sociedade cresce a cada dia. Raros são os dias em que não nos deparamos com uma notícia na mídia falando sobre esta questão. Um exemplo desta preocupação é o que ocorre na cidade de Londres. Os moradores da capital inglesa são filmados pelas câmeras de vídeo que formam o sistema de segurança público da cidade até 300 vezes ao dia [Wright et al. 2010]. Outras tecnologias como identificação por rádio frequência e identificação por dados biométricos também estão sendo amplamente utilizadas. Além disso, essas tecnologias estão sendo combinadas para construir uma rede de sistemas de vigilância mais eficiente.

Tecnologias e sistemas de vigilância não são mais restritos às autoridades policiais, agências de inteligência e ao exército. Os avanços tecnológicos presentes no nosso dia a dia mostram que vigilância é um fenômeno diário. Uma das formas de segurança mais usadas hoje é o sistema de vigilância por câmeras de vídeo. Com isto é necessário a instalação de diversas câmeras de vídeo para compor um sistema de vigilância eficiente. O conjunto destas câmeras é chamado de circuito fechado de televisão, ou em inglês *Closed-circuit television* - CCTV. Um estudo mostra que o mercado global de CCTVs pode alcançar a marca de US\$28 bilhões no final do ano de 2013 [RNCOS 2009].

A enorme disponibilidade de imagens provenientes destes CCTVs gera um constante e enorme fluxo de informações. Entretanto, a quantidade de informações as vezes é tão grande que o operador(a) nem sempre consegue ver o que ele/ela realmente precisaria ver. Fazer a vigilância de uma área somente com câmeras e um operador humano é difícil, caro e ineficiente. O custo de um monitoramento usando 25 câmeras, 24 horas por dia, 7 dias por semana custa aproximadamente US\$150.000,00 por ano. Além disso, a atenção humana aos monitores de vídeo cai a um nível inaceitável depois de 20 minutos [Haering et al. 2008]. Veja a Figura 1.1.

Sistemas de vigilância baseados em câmeras de vídeo monitoram as ruas, estacionamentos e passageiros no metrô. Serviços do governo usam este tipo de tecnologia para identificar pessoas e verificar se um indivíduo que está recebendo um benefício social realmente é quem deveria estar recebendo, patrões monitoram seus empregados, ou seja,



Figura 1.1: O aumento das câmeras de vídeo em sistemas de vídeovigilância gera uma quantidade de informações de vídeo que vai além da capacidade humana de processamento, retirada de [Haering et al. 2008].

sistemas de câmeras vigilância estão por toda parte. Algumas vezes, estes sistemas são vistos pela população como sistemas que tiram a privacidade das pessoas. Mas em outras vezes, o sentimento de segurança que estes sistemas dão a seus usuários supera o sentimento de perda de privacidade.

Monitoramento de atividades por meio de câmeras de vídeo (vídeovigilância) é uma das áreas de estudo que envolve vigilância que mais tem chamado a atenção dos pesquisadores. Diversos são os trabalhos encontrados na literatura que abordam este tópico: [Benkhalil et al. 2000], [Huang et al. 2008], [Petrushin et al. 2006], dentre outros. Neste caso, uma ou mais câmeras (estacionárias ou não) são utilizadas para monitoramento de atividades na área que está sendo "vigiada". Caso uma atividade suspeita seja detectada, o sistema pode tomar algumas medidas como: disparar um alarme sonoro presente na área monitorada, avisar um operador humano ou até mesmo avisar a polícia ou um serviço de segurança particular.

Uma câmera de vídeo é capaz de capturar várias imagens por segundo, a isto damos o nome de taxa de captura que pode ser expressa em *frames per second* (fps). A cada uma dessas imagens damos o nome de quadro. Assim com vários quadros em sequência temos a sensação de que as imagens estão em movimento visualizando esta sequência de quadros (imagens). Assim podemos dizer que um vídeo é uma sequência de imagens e, portanto podemos aplicar técnicas de processamento de imagens em cada uma destas imagens, como por exemplo, eliminação de ruído e segmentação/rastreamento de objetos que se apresentem na cena.

Na maioria das vezes, os sistemas de vídeovigilância necessitam realizar um monitoramento em tempo real, vinte quatro horas por dia, sete dias por semana. Neste caso o



Figura 1.2: Visão da área do estacionamento monitorada por uma câmera de vigilância. (a) cena capturada pela câmera infravermelha, (b) cena capturada por câmera convencional

monitoramento ficará comprometido a noite ou em ambientes fechados com pouca luminosidade caso utilizemos câmeras convencionais, pois como a cena tem pouca ou quase nenhuma iluminação, as imagens capturadas não são de boa qualidade para o processamento computacional. Para contornar este problema é possível a utilização de câmeras que capturam imagens em infravermelho. A Figura 1.2 mostra uma imagem de um vídeo convencional e uma imagem de um vídeo infravermelho.

Todo corpo sólido, a uma temperatura acima do zero absoluto (-273°C), emite radiação infravermelha, que possui um comprimento de onda invisível ao olho humano. Ao converter essa radiação em luz visível, alguns equipamentos permitem, por exemplo, que um observador enxergue no escuro. O princípio de funcionamento de uma câmera infravermelha é semelhante ao das câmeras comuns. A luz entra pela lente da câmera e é concentrada sobre uma placa revestida de algum metal semicondutor. Cada ponto da placa corresponde a um *pixel* (abreviação de *picture element*, elemento mínimo que compõe uma imagem) no visor ou na tela do computador. O sinal elétrico de cada um desses pontos é processado e convertido em um sistema de cores de acordo com a intensidade do sinal recebido. A luneta infravermelha, equipamento introduzido na Segunda Guerra Mundial, permite enxergar a noite tudo o que esteja num raio de 100 metros.

As imagens infravermelhas usadas neste trabalho foram capturas no campus da universidade de Laval que fica na cidade de Quebec (Canadá). Isto só foi possível tendo em vista o estágio de seis meses que o aluno realizou no *Canada Research Chair in Multipolar Infrared Vision*. Além disso, este trabalho foi desenvolvido em parceria com o professor canadense Xavier Maldague, Ph.D. Os trabalhos desenvolvidos no laboratório do professor Maldague também foram grandes fontes de inspiração para a escolha do tema desta dissertação.

Assim, com a imagem infravermelha em mãos, um monitoramento constante pode ser feito, reconhecimento dos objetos em movimento na cena tanto em ambientes com pouca luz ou em ambientes com muita luz. Após o reconhecimento dos objetos em movimento é desejável que ações realizadas por esses objetos (chegada e saída de veículos, passagem

de uma pessoa pela cena, entre outras) bem como possíveis interações entre eles sejam interpretadas. Por exemplo, o reconhecimento automático de um assalto a um veículo ou a objetos no interior do mesmo pode ser identificado avaliando-se: movimentação de uma pessoa próxima ao veículo, intensidade de radiação infravermelha emitida por ela, entre outras. Assim, utilizando imagens infravermelhas e técnicas de processamento de imagens é possível desenvolver um sistema de vigilância automático para um estacionamento externo usando uma câmera estacionária. Um sistema como este ajuda o operador do sistema de vigilância no desempenho de sua tarefa chamando atenção dele para eventos de interesse evitando assim que eventos suspeitos "não sejam vistos" devido a uma distração do operador provocada por cansaço ou até mesmo estresse.

Este trabalho tem como objetivo desenvolver técnicas que possam ser usadas na construção de um sistema de monitoramento de um estacionamento em um ambiente externo não controlado, utilizando uma câmera estacionária que captura imagens em infravermelho. Este sistema pode ser dividido em três subsistemas: módulo de segmentação, módulo de monitoramento e módulo de gerenciamento de eventos. Nos próximos capítulos detalharemos cada um destes módulos. Utilizamos uma técnica dinâmica de subtração de plano de fundo que se adapta às mudanças bruscas de iluminação que podem ocorrer na cena inspirados no trabalho de [Maadi e Maldague 2007]. O uso de imagens infravermelhas permite que se enxergue objetos na ausência de luz e o uso de parâmetros dinâmicos no cálculo do plano de fundo dá robustez à técnica para segmentar objetos corretamente mesmo durante mudanças bruscas de iluminação na cena. Além disso, para melhorarmos a segmentação de pequenos objetos em movimento na cena, utilizamos análise de fluxo para identificar as regiões da cena onde existe movimento restringindo a segmentação a estas regiões. Para o monitoramento dos objetos segmentados, utilizamos uma técnica bastante simples de previsão e correção. Esta técnica consegue monitorar vários objetos simultaneamente de forma satisfatória e também gerencia oclusões parciais e temporárias entre os objetos. Um módulo de gerenciamento de eventos reconhece eventos simples que ocorrem na área monitorada do tipo: a entrada/saída de um humano em um veículo, e o estacionamento de um veículo. Além do reconhecimento de eventos simples, nossa técnica também é capaz de reconhecer eventos mais complexos dividindo estes eventos complexos em sub-eventos mais simples. Reconhecendo cada um destes sub-eventos separadamente é possível reconhecer o evento complexo. Para testar o reconhecimento de eventos complexos, propomos neste trabalho o reconhecimento de eventos suspeitos como "suspeita de roubo de objeto(s) do interior de um veículo estacionado". Estes eventos suspeitos são de grande interesse para o desenvolvimento de um sistema de vídeo vigilância automático e é o principal objetivo deste trabalho.

Esta dissertação está organizado da seguinte forma: no Capítulo 2 fazemos uma exposição de conceitos básicos sobre vídeos e algumas aplicações; no Capítulo 3 fazemos uma revisão de alguns trabalhos encontrados na literatura que tratam a respeito de segmenta-

ção, monitoramento de objetos em movimento e reconhecimento de eventos em vídeo; no Capítulo 4 descrevemos as técnicas de segmentação e classificação de objetos em movimento (humanos e veículos) utilizadas neste trabalho que foram inspiradas em [Maadi e Maldague 2007]; no Capítulo 5 descrevemos o processo utilizado para realizar o monitoramento de objetos em movimento como em [Maadi e Maldague 2007]; no Capítulo 6 detalhamos como é feita o reconhecimento de eventos neste trabalho e introduzimos nossa técnica para reconhecimento de eventos suspeitos. No Capítulo 7 mostramos alguns resultados que obtemos em cada um dos módulos que compõem este sistema (módulo de segmentação, módulo de monitoramento e módulo de gerenciamento de eventos) e no Capítulo 8 apresentamos nossas considerações finais e perspectivas futuras.

Parte I

Fundamentação Teórica e Trabalhos Correlatos

Capítulo 2

Vídeo

2.1 Introdução

Um vídeo é um agrupamento de imagens digitais que visualizadas em uma seqüência temporal nos dão uma sensação de movimento na cena. Ou seja, enquanto uma imagem digital é composta por vários *pixels*, um vídeo é composto por várias imagens digitais. Computacionalmente, uma imagem digital pode ser representada por uma matriz bidimensional de valores. Cada um desses valores representa o valor da intensidade de um *pixel* da imagem. Como o vídeo é um agrupamento de imagens, para representar computacionalmente um vídeo é necessária a utilização de mais uma dimensão. Desta forma para representarmos um vídeo no computador precisamos de uma matriz tridimensional, ou podemos chamar também de um *array* de matrizes bidimensionais.

A Figura 2.1 mostra quatro imagens consecutivas e sobrepostas que formam parte de um vídeo. Os eixos x e y são usados para representar uma única imagem. Quando temos um vídeo, o eixo temporal também está presente. Na figura, esse eixo é representado pelo eixo z. Quando variamos as imagens que compõem o vídeo neste eixo temporal visualizamos a cena em movimento.

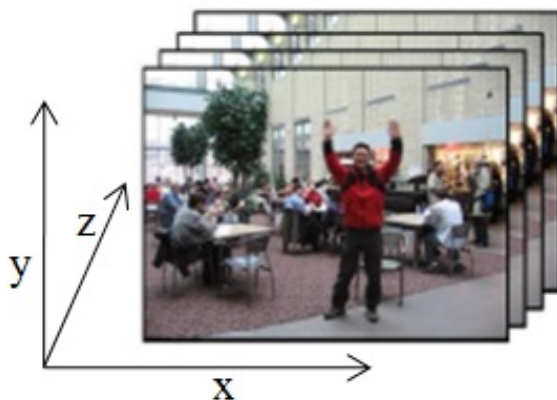


Figura 2.1: Esquema da composição de um vídeo: um agrupamento de imagens em seqüência.

Existem diversas aplicações computacionais em vídeo e diversos são os autores que tem estudado a área. Algumas delas são: vigilância ([Zhang et al. 2006], [Diamantopoulos e Spann 2005], [Huang et al. 2008]), monitoramento de tráfego ([Chen et al. 2010], [Wang 2010]), monitoramento de objetos em movimento ([Silva e Scharcanski 2010]) e detector de vagas em estacionamento ([Ichihashi et al. 2009]).

No nosso dia-a-dia estamos acostumados a ver vídeos (filmes, telejornais, vídeos do *youtube*, dentre outros). Esses vídeos na maioria das vezes são coloridos, ou seja, as imagens que os compõem utilizam o espectro de luz visível. Porém vídeos também podem ser compostos por outros tipos de imagens como é o caso das imagens infravermelhas. Neste trabalho estaremos interessados em trabalhar com vídeos infravermelhos.

Este trabalho visa o desenvolvimento de técnicas que possam ser usadas na construção de um sistema de vigilância para um estacionamento. Vários autores já propuseram a utilização de imagens do espectro visível para realizar esta tarefa. Trabalhos como [Zhang et al. 2006] e [Diamantopoulos e Spann 2005] utilizam imagens do espectro visível para implementar um sistema de vigilância. Entretanto, sistemas como esses podem ter seu funcionamento prejudicado em algumas circunstâncias como a noite e sob má condição do tempo. Nestes casos as imagens capturadas pelas câmeras convencionais serão de baixa qualidade para o processamento computacional pois são imagens que têm pouca ou nenhuma iluminação. Com isso, o sucesso na execução das tarefas do sistema automático de vigilância fica seriamente comprometido.

Uma saída para este problema é a utilização de imagens infravermelhas pois elas fornecem melhores resultados nesses casos. Isso se deve ao fato de que os objetos de interesse na cena (humanos e veículos, no caso de um sistema de vigilância de um estacionamento) são emissores naturais de raios infravermelhos mesmo na ausência da luz visível. Normalmente a temperatura do corpo humano difere da temperatura do plano de fundo da cena. Isto leva a diferentes distribuições de energia e conseqüentemente diferentes níveis de cinza entre o plano de fundo e o corpo humano em imagens infravermelhas [Xue et al. 2010]. Na próxima Seção explicaremos melhor o espectro infravermelho.

Além disso, outro fator que tem impulsionado a utilização de sensores infravermelho em sistemas de vigilância é a redução do seu custo financeiro. Com isto vários autores têm desenvolvido trabalhos que utilizam câmeras infravermelhas. Por exemplo, em [Nanda e Davis 2002] modelos probabilísticos são usados para capturar variações na forma do corpo humano para a detecção de pedestres utilizando vídeos infravermelhos e em [Xu et al. 2005] *support vector machines* e filtros de *Kalman* são adotados para realizar a segmentação de monitoramento em vídeos infravermelhos.

Assim, por prover melhores resultados em cenas que possuem pouca ou nenhuma iluminação e pelo fato dos sensores infravermelhos não possuírem um custo mais elevado, utilizaremos imagens infravermelhas para desenvolver técnicas que possam ser usadas na construção de um sistema de vídeovigilância de um estacionamento externo usando uma

câmera estacionária.

2.2 Visão infravermelha

2.2.1 Definição

Visão infravermelha pode ser definida como a capacidade de emissão de radiação infravermelha de sistemas biológicos ou artificiais. Visão térmica ou termografia também são comumente usadas neste contexto uma vez que as emissões infravermelhas de um corpo estão diretamente relacionadas com a sua temperatura: objetos quentes emitem mais energia no espectro infravermelho que objetos frios.

O termo visão noturna também pode ser encontrado na literatura com o significado de visão infravermelha. Isso se deve ao fato de que o propósito inicial do desenvolvimento de sistemas de visão infravermelha era localizar alvos inimigos a noite. O corpo humano, bem como vários outros objetos de interesse militar ou civil, é normalmente mais quente que o ambiente que o cerca. Uma vez que objetos quentes emitem mais energia no espectro infravermelho que objetos frios, detectá-los com um sensor infravermelho, tanto de dia como de noite, é uma tarefa simples. Entretanto, a visão noturna diz respeito a capacidade de enxergar a noite embora não seja necessariamente utilizando o espectro infravermelho.

Na verdade equipamentos de visão noturna podem ser construídos usando duas tecnologias: intensificadores de luz ou visão infravermelha. A primeira tecnologia utiliza um fótocatodo para converter luz em elétrons, amplificar o sinal e transformar novamente em fótons que serão usados na construção da imagem. Já a visão infravermelha usa sensores infravermelhos que captam radiação no comprimento de onda infravermelho que é invisível ao olho humano. Sendo assim visão infravermelha captura o "calor" emitido pelo objeto. A Figura 2.2 mostra uma imagem de uma operação militar usando intensificadores de luz e uma imagem de um civil utilizando visão infravermelha.



(a)



(b)

Figura 2.2: Diferença entre (a) imagens de visão noturna e (b) imagens de visão infravermelha.

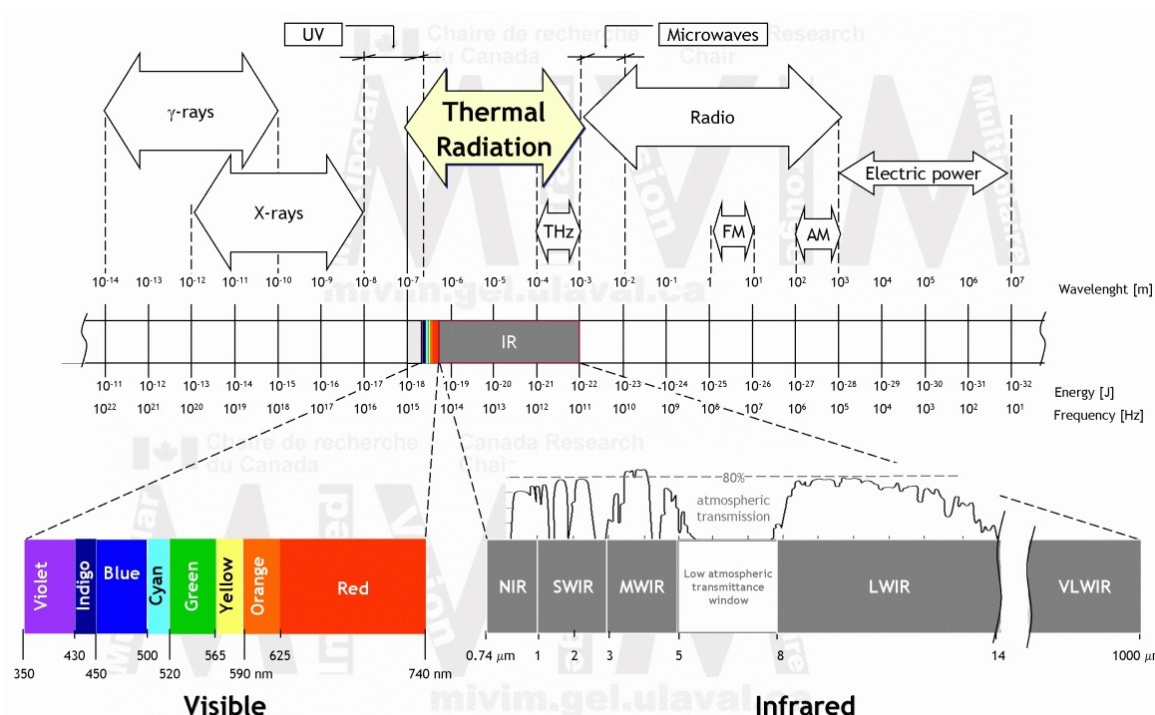


Figura 2.3: Espectro eletromagnético com destaque para a banda infravermelha, retirado de [Ibarra-Castanedo 2005].

2.2.2 Espectro eletromagnético

A Figura 2.3 (retirada de [Ibarra-Castanedo 2005]), mostra todo espectro eletromagnético destacando a parte infravermelha (IR) que fica entre a parte visível e a parte das ondas de rádio ou microondas. O espectro IR é subdividido em cinco regiões embora essa subdivisão seja arbitrária variando para cada autor. Esta subdivisão usada neste trabalho é baseada em uma combinação de transmitância atmosférica, ou seja, as regiões de comprimento de onda em que a radiação infravermelha é melhor transmitida através da atmosfera, os materiais usados para construir o sensor infravermelho para aquela subdivisão e as aplicações principais daquela subdivisão (veja a Tabela 2.1). Desta forma, a banda do infravermelho próximo (NIR) é mais usada em telecomunicações para transmissão usando fibra ótica uma vez que a sílica (SiO_2), material que compõe a fibra ótica, tem uma perda de atenuação média baixa para o comprimento de onda da banda NIR. Já a banda de infravermelho de ondas curtas (SWIR) permite telecomunicações de longa distância (sensoriamento remoto) usando uma combinação de materiais de detecção. As bandas de infravermelho de comprimento de onda médio (MWIR) e longo (LWIR) têm várias aplicações em termografia infravermelha para aplicações militares e civis, como por exemplo: identificação de alvos, vigilância, avaliação não destrutiva de materiais (NDE), entre outras. A banda de infravermelho de comprimento de onda muito longa (VLWIR) é usada em espectroscopia e astronomia.

A banda de frequência MWIR é preferida para a inspeção de objetos em alta temper-

Banda de frequência	Comprimento (μm)	Material de detecção	Aplicações
NRI	0,74 - 1	SiO_2	Telecomunicações
SWIR	1 - 3	$InGaAs$, PbS	Sensoriamento remoto
MWIR	3 - 5	$InSb$, $PbSe$, $PtSi$, $HgCdTe$	Inspeção de altas temperaturas (<i>indoor</i> , pesquisa científica)
LWIR	8 - 14	$HgCdTe$	Temperaturas ambientes (<i>outdoor</i> , pesquisa industrial)
VLWIR	14 - 1000	-	Espectroscopia, astronomia

Tabela 2.1: Bandas de frequência do espectro infravermelho: adaptado de [Ibarra-Castanedo 2005], [Hudson 1969], [Maldague 2001] e [Piotrowski e Rogalski 2004].

atura e a banda de frequência LWIR para avaliação de objetos na temperatura ambiente. Outros importantes critérios para a seleção de bandas de frequência são: a distância de operação - se a operação é *indoor-outdoor*, temperatura e emissividade dos objetos de interesse [Maldague 2001]. Em física, emissividade é a relação entre o poder emissivo de um corpo qualquer e a de um corpo negro (*black body* - BB). É conhecida como emissividade ε , e pode ter um valor máximo igual a 1, que é correspondente à de um corpo negro, e um mínimo igual a zero. BB é definido como um corpo que absorve toda a radiação que nele incide (nenhuma luz o atravessa nem é refletida).

Na verdade, ondas longas (LWIR) são preferidas para aplicações em ambientes externos uma vez que são menos afetadas pela radiação do Sol. Câmeras que operam na banda LWIR são tipicamente câmeras não resfriadas que usam um microbolômetro (um tipo de detector usado para fazer detecções térmicas) que é usado em aplicações industriais que usam IR, embora detectores resfriados para LWIR que usam *cádmio mercúrio telúrio* (MCT) também existam. Por outro lado, ondas médias (MWIR) necessitam de detectores resfriados. Este resfriamento pode ser feito por exemplo com nitrogênio líquido. Um resfriamento de $-196^\circ C$ permite obter excelentes resultados, porém isso pode restringir o uso a ambientes controlados.

2.2.3 Aplicações

Visão infravermelha começou sendo usada intensivamente pelo militares para visão noturna, navegação, vigilância e detecção de alvos. Durante anos, o desenvolvimento desta área foi lento devido ao alto custo dos equipamentos usados e a baixa qualidade das imagens adquiridas. Entretanto, desde o aparecimento da primeira câmera infravermelha de uso comercial na segunda metade da década de 60, a disponibilidade de melhores equipamentos de captura infravermelha tem crescido juntamente com o crescimento da capacidade de processamento dos micro-computadores e com isso cada vez mais pesquisadores

têm desenvolvido aplicações nas mais diferentes áreas utilizando imagens infravermelhas. Algumas das aplicações desenvolvidas utilizando imagens infravermelhas: detecção e monitoramento de pedestres ([Dai et al. 2007], [Li et al. 2010]), detecção de silhuetas de humanos ([Xue et al. 2010]), sensoriamento do clima ([Qihao e Weng 2009]), detecção de alvos ([Zhang et al. 2009], [Deng et al. 2010], [Yilmaz et al. 2003], [Tingjun et al. 2010]), inspeção industrial ([Wong et al. 2009], [Younus e Yang 2011]), realidade aumentada ([Amici et al. 2010]), vigilância ([Maadi e Maldague 2007], [Maadi e Maldague 2006]), infraestruturas e edificações ([Barreira e Freitas 2007], [Balaras e Argiriou 2002]), componentes aeroespaciais ([Avdelidis et al. 2003]), entre outros.

2.3 Considerações Finais

Neste Capítulo apresentamos alguns conceitos básicos sobre vídeo. Apresentamos também as definições de visão infravermelha que nortearam a construção do nosso sistema. O nosso objetivo é desenvolver técnicas que possam ser usadas na construção de um sistema de vídeovigilância automático de um estacionamento ao ar livre utilizando uma câmera infravermelha estacionária. A escolha da utilização de imagens infravermelhas deve-se ao fato de que em cenas que possuem pouca ou nenhuma iluminação elas permitem que o sistema alcance melhores resultados. O nosso sistema será dividido em três módulos: segmentação, monitoramento e gerenciamento de evento. Nos próximos capítulos faremos uma breve revisão de alguns trabalhos encontrado na literatura que tratam destes assuntos.

Capítulo 3

Sistemas de vídeovigilância

3.1 Introdução

Com o avanço da tecnologia e com a crescente onda de insegurança vivida pela sociedade, cada vez mais a tecnologia está sendo aplicada para aumentar o nível de segurança das pessoas. A área de visão computacional que envolve sistemas de vídeovigilância para cenas dinâmicas, especialmente envolvendo humanos e veículos, é atualmente uma das áreas de pesquisas mais ativas em visão computacional. Estes sistemas têm amplas possibilidades de aplicações, incluindo:

- **autorização de acesso a áreas controladas:** em algumas instalações com restrição de acesso como bases militares e importantes instalações do governo, somente pessoas autorizadas previamente têm a entrada permitida. Um banco de dados biométrico incluindo visitantes autorizados é construído antecipadamente usando técnicas biométricas (reconhecimento de digitais, íris ou face por exemplo). No momento que alguém está solicitando acesso a área o sistema pode em tempo real extrair tais características biométricas do visitante e então decidir se ele pode ter o acesso liberado. Em [Arandjelovic e Cipolla 2004], os autores usam imagens extraídas de vídeos de segurança para autorizar acesso de pessoas por meio de reconhecimento de face;
- **identificação de humanos à distância:** identificação de pessoas à distância através de sistemas inteligentes de vigilância pode ajudar a polícia a capturar suspeitos. A polícia pode construir um banco de dados com características biométricas de suspeitos e instalar um sistema de vídeovigilância onde os suspeitos costumam aparecer (estações de metrô, cassinos, etc.). O sistema automaticamente reconhece e julga se o indivíduo na cena é ou não um suspeito. Em caso positivo, o sistema soa um alarme imediatamente. Em [Hamdoun et al. 2008] os autores reconhecem humanos à distância analisando pequenas sequências de vídeos;

- **análise de congestionamento e tráfego:** usando técnicas de detecção de pessoas, sistemas de vídeovigilância podem automaticamente detectar o "fluxo" de pessoas em áreas públicas importantes como nas proximidades de estádios e grandes lojas de departamentos, e então promover uma análise de congestionamento para auxiliar o "fluxo" de pessoas. Outra aplicação que envolve análise de congestionamento e tráfego é o monitoramento de rodovias e avenidas. Um exemplo deste tipo de aplicação pode ser encontrado em [Chen et al. 2010];
- **identificação de comportamentos anormais:** em algumas circunstâncias é necessário analisar o comportamento de humanos e veículos e decidir se este comportamento é normal ou não. Por exemplo, um sistema de monitoramento em um estacionamento ou em um supermercado pode ajudar a detectar furtos ocorridos nessas áreas. Normalmente existem dois modos de "soar" o alarme. Primeiro pode-se soar um sinal sonoro no local e informar que alguma coisa anormal está ocorrendo. Segundo, pode-se informar o operador humano do sistema ou até mesmo informar a polícia automaticamente. Em [Xiang e Gong 2008] os autores propõem um sistema automático para reconhecimento de eventos anormais onde o sistema automaticamente diferencia o que é normal do que é anormal sem nenhum conhecimento prévio;
- **vigilância usando múltiplas câmeras:** para segurança social, vigilância cooperativa usando múltiplas câmeras pode ser usada para assegurar a segurança de uma comunidade inteira, por exemplo, monitorando suspeito por uma ampla área utilizando a cooperação entre múltiplas câmeras. Em monitoramento de tráfego, um sistema com múltiplas câmeras pode ajudar a polícia rodoviária a descobrir, monitorar e capturar veículos envolvidos em infrações de trânsito. Em [Javed et al. 2009] os autores apresentam um sistema de vigilância para uma área ampla que detecta, monitora e classifica objetos em movimento através de múltiplas câmeras.

Em geral, estes sistemas geralmente podem ser divididos em três módulos: um módulo de segmentação de movimento, um módulo de monitoramento de objetos e um módulo que gerencie as interações entre estes objetos. Neste Capítulo apresentaremos algumas técnicas encontradas na literatura para realizar tais tarefas.

3.2 Detecção de movimento

O primeiro passo de quase todos os sistemas de vídeovigilância é a detecção (ou segmentação) de objetos em movimento. Segmentação de movimento busca separar regiões em movimento do restante da imagem. Processos subsequentes dependem desta primeira etapa. Algumas abordagens para segmentação de movimento encontradas na literatura são descritas a seguir.

- **Diferença temporal:** Utiliza a diferença em *pixels* de dois ou três quadros consecutivos do vídeo para segmentar a região que está se movimentando na cena. Esta técnica é bastante adaptável a ambientes dinâmicos, mas geralmente não produz um bom resultado na segmentação de todas as regiões relevantes do movimento. Como um exemplo da aplicação deste método, em [Lipton et al. 1998] os autores detectam alvos em movimento utilizando diferenciação temporal. Depois que a diferença absoluta entre o quadro corrente e o quadro anterior é obtida, uma função limiar é utilizada para determinar as mudanças. Utilizando a análise de componentes ligados, os pixels em movimentos que foram segmentados são agrupados em regiões de movimento.
- **Fluxo óptico:** Segmentação de movimento utilizando fluxo óptico utiliza características dos vetores de fluxo dos objetos em movimento sobre o tempo para detectar regiões de movimento em vídeo. Por exemplo, em [Meyer e Denzler 1997] os autores calculam o campo do vetor de deslocamento para inicializar o algoritmo, chamado raios ativo, para a extração de objetos em movimento. Os resultados são usados na análise de movimento. Métodos baseados em fluxo óptico podem ser usados para detectar objetos se movimentando separadamente mesmo com a câmera em movimento. Entretanto, a maioria dos métodos de cálculo de fluxo são computacionalmente complexos e muito sensíveis a ruído, e não podem ser aplicados a transmissões de vídeos em tempo real sem a utilização de um *hardware* apropriado.
- **Subtração de plano de fundo:** É um método popular para a segmentação de movimento, principalmente nas situações onde o plano de fundo é relativamente estático. Métodos deste tipo detectam as regiões em movimento de um vídeo rastreando a diferença entre o quadro corrente e o plano de fundo de referência. O uso desta técnica pode ser encontrado por exemplo no trabalho [Grimson et al. 1998] e [Stauffer e Grimson 1999]. É um método simples de ser aplicado quando temos um ambiente bem controlado, porém muito sensível a mudanças na cena provenientes, por exemplo, de mudanças bruscas de iluminação na cena. Portanto, é muito dependente de um bom modelo de plano de fundo para reduzir o impacto destas mudanças.

Para o desenvolvimento deste trabalho optamos pela técnica de subtração de plano de fundo por sua simplicidade e resultados satisfatórios para o tipo de aplicação proposta. O método usado neste trabalho é inspirado no método proposto em [Maadi e Maldague 2007] e será detalhado no Capítulo 4.

3.3 Monitoramento

Após a segmentação do objeto em movimento é preciso monitorá-lo. Monitoramento em vídeo é o processo de localizar um objeto em movimento (ou vários objetos) ao longo do tempo através de uma câmera. Podemos classificar as técnicas de monitoramento em quatro categorias que são apresentadas a seguir.

- **Monitoramento baseado em uma região em movimento:** Este método identifica e rastreia um "borrão" ou uma caixa delimitadora para os componentes conexos no espaço 2D. O método baseia-se em propriedades destes objetos ("borrões") como tamanho, cor, forma, velocidade e centro de massa para realizar o monitoramento. A vantagem deste método é que o tempo de processamento é eficiente e ele funciona bem para um pequeno número de objetos em movimento. Um ponto fraco da técnica é que problemas de oclusão não podem ser resolvidos adequadamente em situações onde existe uma grande densidade de objetos. Regiões agrupadas irão formar uma "bolha" de objetos e causarão erros de rastreamento. Por exemplo, [Masoud e Papanikolopoulos 2001] apresentam um método para monitorar regiões em movimento. Filtros de *Kalman* ([Welch e Bishop 1995]) são usados para estimar os parâmetros utilizados no monitoramento de pedestres;
- **Monitoramento baseado em contorno ativo de objetos em movimento:** Algoritmos de monitoramento baseados em contornos ativos rastreiam objetos representando suas fronteiras através de contornos delimitadores e atualizando estes contornos em quadros sucessivos. Exemplos do uso deste método pode ser encontrado em [Baumberg 1995], [Mohan et al. 2001], [Galata et al. 2011] e [Wu e Huang 2001]. Esses algoritmos visam extrair a forma propriamente dita do objeto ao invés de trabalhar com uma caixa delimitadora para o objeto. Desta forma ele provê uma descrição mais efetiva, em termos de forma, do objeto monitorado do quando utilizamos algoritmos baseados em regiões. [Peterfreund 2002] explora um modelo de contorno ativo para monitoramento de objetos, tais como pessoas, baseado em filtros de *Kalman* ([Welch e Bishop 1995]);
- **Monitoramento baseado em um modelo de objeto em movimento:** Normalmente é baseado em um modelo 3D do objeto em movimento. Este método define uma geometria paramétrica em 3D do objeto. Ele pode resolver parcialmente o problema de oclusão, mas o tempo de processamento cresce muito se utilizarmos um modelo detalhado do objeto geométrico. Este método só consegue bons resultados se apenas uma pequena quantidade de objetos em movimento estiver envolvida na cena. Por exemplo, [Roller et al. 1993] resolveu o problema de oclusão parcial, considerando os modelos 3D. A definição de modelos de veículos parametrizados permite explorar um conhecimento *a-priori* sobre a forma de objetos típicos de cenas de trânsito;

- **Monitoramento baseado em características selecionadas de objetos em movimento:** Monitoramento baseado em características selecionadas realiza o reconhecimento e monitoramento de objetos em movimento extraindo elementos (características) destes objetos agrupando estes elementos em características de alto nível. O método realiza o casamento entre destas características dos objetos nas várias imagens do vídeo para realizar o monitoramento. Um bom exemplo de monitoramento utilizando características selecionadas pode ser encontrado no trabalho [Polana e Nelson 1994]. Neste trabalho os autores representam uma pessoa através de uma caixa delimitadora do "borrão" que representa esta pessoa. A característica utilizada para realizar o monitoramento dos objetos é o centróide desta caixa delimitadora. O monitoramento é feito corretamente mesmo quando uma oclusão parcial ocorre uma vez a velocidade instantânea dos centróides podem ser estimadas.

A escolha de qual tipo de algoritmo usar depende de restrições de tempo de processamento. Métodos que usam aspectos mais gerais e simplificados dos objetos como monitoramento baseado em uma região em movimento em monitoramento baseado em características selecionadas de objetos em movimento possuem um tempo de execução menor e são indicados para sistemas que possuem restrições quanto ao tempo de processamento. Já métodos que usam aspectos mais detalhados como monitoramento baseado em contorno ativo de objetos em movimento e monitoramento baseado em um modelo de objeto em movimento possuem um tempo de execução maior porém fornecem melhores resultados. Este último tipo de método são indicados quando se precisa de uma alta precisão no monitoramento.

Na literatura encontramos diversos outros trabalhos que se preocupam com o monitoramento de objetos em cenas de vídeos ([Dai et al. 2007], [Xu et al. 2005], entre outros). Neste trabalho usaremos um método de monitoramento baseado em um ciclo de *previsão* e *correção* que monitora objetos através de características selecionadas assim como é feito em [Maadi e Maldague 2007]. Este método foi escolhido por ser um método simples porém eficaz para realizar monitoramento em tempo real de objetos em movimento. No Capítulo 5 explicaremos esta técnica.

3.4 Modelagem de eventos

Uma das principais características que um sistema de vídeovigilância deve ter é a capacidade de identificar e interpretar interações entre objetos que ocorram na área monitorada. Essas interações são comumente chamadas de eventos. Existe uma variedade de trabalhos na literatura que focam na questão de gerenciamento de eventos ([Fernández et al. 2010], [Foresti et al. 2004], [Diamantopoulos e Spann 2005], [Michelsoni et al. 2009], [Zhang et al. 2006], [Lavee et al. 2009], [Ivanov et al. 1999], entre outros).

Existem três características comuns aos eventos presentes em quase todos os trabalhos:

- Eventos ocupam um período determinado de tempo;
- Eventos são construídos de pequenos blocos semânticos chamados sub-eventos ou eventos simples;
- Eventos são descritos usando os aspectos relevantes do vídeo de entrada.

De acordo com [Lavee et al. 2009] existem duas formas de modelagem de eventos para reconhecimento automático através de máquinas. A primeira é uma abordagem implícita e a segunda uma abordagem explícita. No primeiro tipo de abordagem, nenhum conhecimento prévio do domínio da aplicação é conhecido. O sistema automaticamente identifica padrões comuns de ação por meio dos dados observados. No segundo tipo de abordagem o sistema requer uma definição explícita da composição cada evento que estes eventos possam ser reconhecidos pelo sistema. Neste caso o sistema tenta casar os dados observados com os dados que foram fornecidos pelo operador anteriormente. Modelagem implícita torna o sistema altamente adaptável a novas e diferentes situações, mas torna o sistema pouco eficaz na detecção de eventos complexos. Por outro lado, a modelagem explícita gera melhores resultados em termos de falsos positivos e falsos negativos. Entretanto, este tipo de abordagem não é auto-adaptativa e todo conhecimento sobre os eventos deve ser passado ao sistema pelo operador. Outro problema da modelagem explícita é que ela depende do operador humano no que diz respeito a como um dado evento se caracteriza, ou seja, quais ações formam o evento. Isto é ruim pois um mesmo evento pode ser caracterizado de forma diferente, quando descrito por operadores diferentes. Logo um problema na modelagem explícita é a subjetividade na tarefa de definição dos eventos uma vez que esta tarefa depende do operador humano.

Segundo [Michelsoni et al. 2009], eventos podem ser categorizados em duas classe: simples ou compostos. Eventos simples são os eventos que representam uma única ação que é simples de ser modelada/identificada. Já eventos mais complexos são compostos por mais de um evento simples. Por esse motivo eventos simples também são chamados de sub-eventos ([Lavee et al. 2009]).

Ainda segundo [Michelsoni et al. 2009], em um ambiente urbano, um evento simples é normalmente representado por um veículo ou pedestre se movendo dentro da área monitorada. Um evento simples v é definido sobre um intervalo de tempo $[T^s \ T^f]$ e contém um conjunto de características pertencentes a um objeto O_h observado por uma sequência de m quadros consecutivos e é dado pela equação 3.1.

$$v(T^s, T^f) = \{fe_k \mid fe_k \in O_h, k = 1, 2, \dots, q\} \quad (3.1)$$

Exemplos de características são a identificação do objeto (ex. placa do veículo ou RG da pessoa), a classe do objeto detectado, sua trajetória, a velocidade de deslocamento

média, o formato do "borrão" formado pelo objeto, seu histograma de cores, entre outras. No contexto de um estacionamento, exemplos de eventos simples podem ser dados por:

- um veículo entrando na área de estacionamento;
- o veículo movendo-se com uma certa trajetória;
- o veículo parando em algum lugar específico;
- uma pessoa saindo do interior de um veículo;
- a pessoa movendo-se dentro do estacionamento;
- a pessoa saindo da área do estacionamento.

Assim, eventos simples são geralmente ações feitas em uma pequena região da área monitorada (entradas, saídas, proximidades do veículo, etc.) ou ações caracterizadas por um movimento uniforme (caminhar, levantar-se, etc.). A principal característica dos eventos simples é que eles são simples e rápidos de serem identificados. Isso faz com que eles sejam considerados como sub-eventos (unidades semânticas de construção) usados para construir eventos mais complexos na cena.

Ainda segundo [Michelsoni et al. 2009], eventos compostos são representados por um conjunto de eventos simples que estão espacialmente e/ou temporalmente relacionados. Assim, um evento composto pode ser definido sobre um intervalo de tempo por um grafo $G(V, E)$ onde o conjunto de vértices $V = \{v_1(T^s, T^f), \dots, v_r(T^s, T^f)\}$ é o conjunto de r eventos simples que compõem o evento complexo e o conjunto de arestas E é o conjunto de associações temporais e espaciais entre os eventos simples.

3.5 Considerações Finais

O desenvolvimento de sistemas automáticos de vídeo vigilância é uma das áreas que mais tem demandado esforços no meio científico e empresarial para fim de segurança pessoal, patrimonial e coletiva. Por isso, se faz necessário o estudo e desenvolvimento de técnicas que possam ser usadas na construção de tais sistemas. Estas técnicas envolvem métodos de segmentação de movimento, monitoramento de objetos e interpretação e gerenciamento de eventos. Neste Capítulo analisamos algumas dessas técnicas encontradas na literatura. Nos próximos capítulos explicaremos as técnicas que foram usadas durante o desenvolvimento deste trabalho.

Parte II

Método Proposto

Capítulo 4

Módulo de segmentação

4.1 Introdução

A detecção de objetos em movimento é um problema atual e que é largamente estudado em visão computacional ([Mandellos et al. 2011], [Fernández-Caballero et al. 2010], [V. Kastrinaki e Kalaitzakis 2003] e outros). Diversas são suas aplicações tais como sistemas de vídeovigilância e monitoramento de tráfego em rodovias. Mais recentemente uma aplicação que tem crescido muito são as aplicações multimídias desenvolvidas em realidade aumentada tais como animação de objetos computacionais e jogos de *vídeo game* orientados a movimento.

Uma etapa crucial para detecção de objetos em movimentos é a segmentação das regiões em movimento no vídeo tais como humanos e veículos. Os objetos segmentados nesta etapa serão utilizados como informação de entrada para os outros módulos do sistema: monitoramento e gerenciamento de eventos. Por isso esta etapa é crucial para o desenvolvimento do sistema proposto. Como descrito no Capítulo 3, existem várias formas de detectar regiões/objetos em movimento em um vídeo. Neste trabalho utilizamos a técnica de subtração de plano de fundo para realizar tal tarefa.

Neste Capítulo apresentamos a metodologia usada neste trabalho para a segmentação de objetos em movimento em ambientes externos utilizando imagens infravermelhas.

4.2 Subtração de plano de fundo

Dentre os vários métodos de segmentação de objetos em movimento encontrados na literatura, iremos utilizar neste trabalho o método de subtração de plano de fundo para realizar tal tarefa. A escolha do método se deu pelo fato de que ele não é dependente do tipo de cena ou ambiente de origem da cena. Além disso, a complexidade algorítmica e velocidade de execução do método de subtração de plano de fundo são baixas o que justifica o seu uso no tipo de aplicação de monitoramento proposta neste trabalho.

O método de subtração de plano de fundo consiste no cálculo de uma imagem referência que é subtraída da imagem atual do quadro corrente para obter os objetos em movimento. Essa imagem referência representa o plano de fundo da cena, ou *background*, que evolui no decorrer do tempo. Assim, para cada quadro da seqüência de vídeo, uma imagem que representa o plano de fundo é calculada. Comparando esta imagem referência com a imagem atual do quadro da seqüência de vídeo podemos extrair objetos em movimento da cena através de um limiar. A imagem atual menos o *background* nos dá o que chamamos de primeiro plano, ou *foreground*, no qual estão contidos os objetos em movimento da cena.

A maioria dos métodos de subtração de plano de fundo utilizam um limiar fixo para realizar a comparação entre a imagem atual da seqüência de vídeo e a imagem do plano de fundo estimada para obter a imagem do *foreground*. Porém este limiar fixo não consegue adequar o método de subtração de plano de fundo às mudanças bruscas ocorridas na cena presentes principalmente em ambientes externos. Por isso, neste trabalho, calculamos este limiar dinamicamente para cada imagem da seqüência de vídeo para que se tenha melhores resultados na segmentação dos objetos em movimento.

Inspirados no que é proposto em [Maadi e Maldague 2007], utilizamos neste trabalho uma técnica de subtração de plano de fundo que se adapta às mudanças de iluminação da cena por meio de um limiar dinâmico o que torna o processo de segmentação de objetos em movimento mais robusto. Introduzimos a idéia de análise de fluxo de movimento no processo de segmentação para melhorar a segmentação de objetos pequenos na cena. O uso deste limiar dinâmico e da análise de fluxo serão explicados no decorrer deste Capítulo.

4.2.1 Cálculo do *foreground*

O primeiro plano, ou *foreground*, contém os objetos em movimento na cena em um dado instante k e é calculado através da diferença entre o *background* e o quadro atual do vídeo aplicada a um limiar de binarização. Construímos $FORE_k$, o *foreground* no instante k que é representado por uma imagem binária, verificando se cada ponto (i, j) de F_k pertence ou não ao *foreground*. $FORE_k$ é dado pela equação 4.1.

$$FORE_k(i, j) = \begin{cases} 1, & \text{se } |F_k(i, j) - Bg_k(i, j)| \geq Th_k, \\ 0, & \text{caso contrário.} \end{cases} \quad (4.1)$$

onde:

- $1 \leq i \leq N$;
- $1 \leq j \leq M$;
- Th_k é o limiar de binarização aplicado que determina se um *pixel* pertence ou não ao *foreground*. Ele é calculado a cada instante k e será detalhado mais adiante;

- F_k o frame no instante k ;
- Bg_k o plano de fundo (*background*) da cena no instante k ;
- $FORE_k$ o primeiro plano (*foreground*) da cena no instante k ;
- $1 \leq k \leq q$
- $F_{-l}, F_{-l+1}, F_{-l+2}, \dots, F_{-1}, F_0, F_1, \dots, F_q$, uma seqüência de $l+q+1$ quadros de tamanho $N \times M$, sendo que usamos os l primeiros quadros desta seqüência para realizar a inicialização do *background*.

A escolha do limiar de binarização é muito importante para a qualidade da segmentação. A escolha de um valor muito elevado pode levar a perda de partes significativas do objeto em movimento detectados e a escolha de um limiar muito baixo pode introduzir ruído na segmentação fazendo com que apareçam objetos que realmente não são objetos (falso positivos) o que pode até inviabilizar a detecção dos verdadeiros objetos alvos da segmentação.

4.2.2 Cálculo do *background*

Outro problema encontrado na segmentação é o fato das cenas serem de ambientes externos os quais mudam constantemente (presença de movimento constante, mudanças gradativas de iluminação da cena, etc.), por isso a atualização do plano de fundo é essencial. Para isso, aplicamos um filtro temporal *passa-baixa* que é definido pela equação 4.2.

$$Bg_{k+1} = Bg_k + \alpha_k(F_k - Bg_k) \quad (4.2)$$

onde

- $0 \leq \alpha_k \leq 1$ é a taxa de aprendizagem dinâmica;
- $1 \leq k \leq q$
- F_k é a imagem da seqüência de vídeo no instante k ;
- Bg_{k+1} é o plano de fundo da imagem calculado no instante $k+1$;
- Bg_k é o plano de fundo da imagem calculado no instante k .

A equação 4.2 permite com que o sistema ajuste continuamente o plano de fundo da cena tratando mudanças graduais na intensidade de luminosidade e presença de objetos em movimento na cena. Quanto maior o valor de α_k , mais mudanças ocorridas na cena serão incorporadas ao plano de fundo.

Em ambientes externos, a mudança brusca na iluminação da cena pode acontecer com frequência (como a alternância entre sol-nuvem causada pelo vento). Neste caso específico,

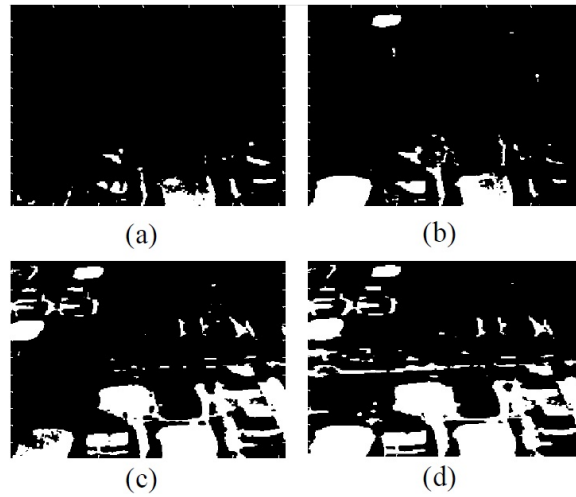


Figura 4.1: Efeitos de uma aparição repentina do sol no processo de segmentação. Figuras (a) a (d) foram capturadas em um intervalo de 2 segundos. O *foreground* foi bastante afetado devido a esta aparição do sol.

mesmo com um limiar de binarização ótimo fixado na equação 4.1, a técnica de subtração de plano de fundo pode falhar completamente porque a velocidade de ajuste do plano de fundo da cena, quando usamos um α_k constante, não é capaz de acompanhar a velocidade de mudança ocorrida na cena. Os efeitos desta mudança brusca na cena podem ser vistos na Figura 4.1 onde temos o resultado da segmentação durante a aparição repentina do sol utilizando um α_k fixo. Logo, para minimizar este problema utilizamos limiares dinâmicos para o cálculo do *foreground* e para a estimativa do *background*. Estes limiares serão calculados mais adiante neste Capítulo.

Para que possamos usar a equação 4.2 e estimar corretamente o *background* para cada instante k , precisamos primeiro inicializá-lo, ou seja, calcular o Bg_0 . Para realizar esta tarefa usamos os l primeiros quadros da sequência de vídeo e realizamos a inicialização do *background* através da equação 4.3.

$$Bg_k = \begin{cases} F_k, & \text{se } k = -l, \\ Bg_{k-1} + 0,5(F_{k-1} - Bg_{k-1}), & \text{caso contrário.} \end{cases} \quad (4.3)$$

onde

$$\bullet \quad -l \leq k \leq 0$$

O método de inicialização de plano de fundo usado neste trabalho é exatamente igual ao método usado em [Maadi e Maldague 2007].

4.2.3 Arquitetura do módulo de segmentação de movimento

Os problemas citados anteriormente se devem muitas vezes à baixa qualidade das imagens infravermelhas capturadas e à incapacidade da técnica de subtração de plano de

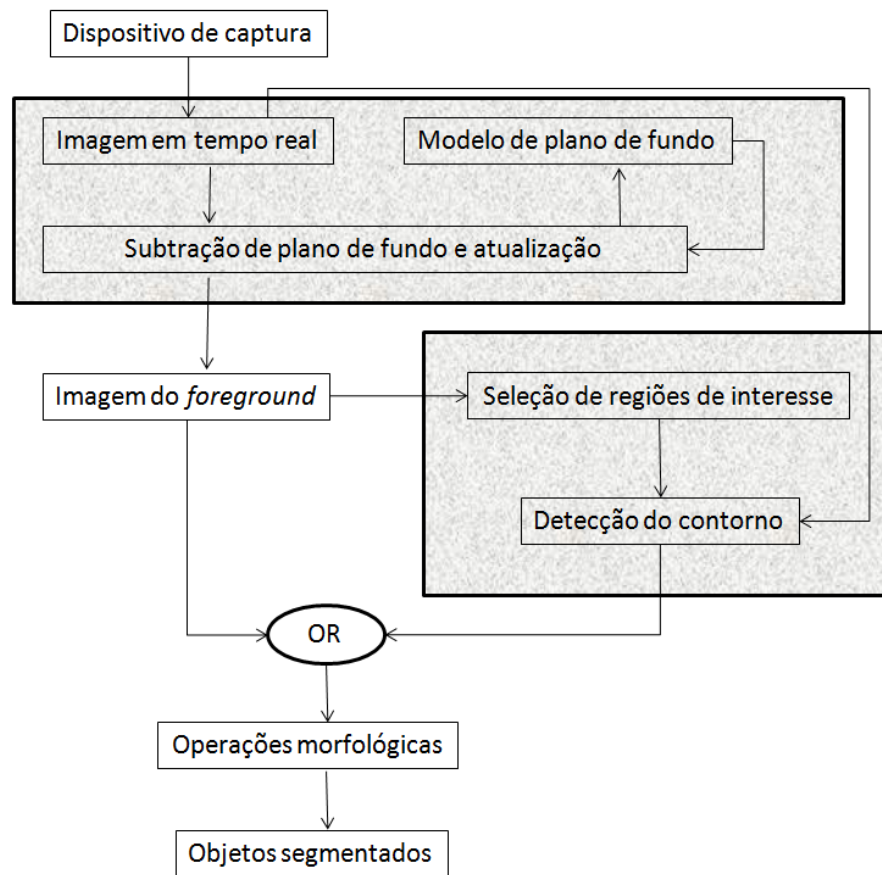


Figura 4.2: Diagrama de blocos do processo de segmentação de movimento.

fundo de se adaptar às mudanças bruscas de iluminação em ambientes externos. Uma solução é encontrar uma formulação matemática que utiliza o monitoramento temporal das mudanças na intensidade de nível de cinza dos *pixels* da imagem causados pela mudança brusca de iluminação do ambiente para ajuste dinâmico do limiar de binarização e do plano de fundo da cena. Em seguida compensar a falta de informação gerada no processo de cálculo do *foreground* usando uma ferramenta adequada. Para isso, usamos um algoritmo proposto em [Maadi e Maldague 2007] que combina subtração de plano de fundo e detecção dos contornos dos objetos segmentados. Um diagrama funcional deste modelo é apresentado na Figura 4.2.

Como é mostrado na Figura 4.2, a segmentação de movimento é feita em duas etapas principais: a primeira consiste na extração de uma máscara binária dos *pixels* pertencentes ao *foreground*, a segunda consiste em combinar o objeto resultante da primeira etapa com as bordas correspondentes da imagem original. Finalmente a segmentação é obtida através da aplicação de operações morfológicas.

A diferença do método usado neste trabalho para outros métodos de subtração de plano de fundo encontrados na literatura está no uso do limiar (equação 4.1) e da taxa de aprendizagem (ou ganho, equação 4.2) calculados dinamicamente a cada quadro do vídeo.

O ajuste dinâmico destes parâmetros permite com que o sistema se adapte rapidamente às mudanças bruscas de iluminação ocorridas na cena. Este ajuste é feito em cada quadro k do vídeo. Para esse cálculo é usado o histórico da evolução temporal da média dos níveis de cinza dos *pixels* de cada quadro da seqüência. A equação 4.4 define a média dos níveis de cinza dos *pixels* do quadro k da seqüência.

$$media_k = \frac{\sum_{i=1}^N \sum_{j=1}^M F_k(i, j)}{NM} \quad (4.4)$$

O tamanho do vetor usado para armazenar o histórico de variação de iluminação da cena, $p + 1$ (em nossa implementação $p = 4$), deve ser escolhido adequadamente a fim de comportar as variações significativas na cena observada. A escolha deste tamanho está diretamente relacionada com a taxa de quadros por segundo (*fps*) do vídeo. Quanto maior for esta taxa, maior deve ser o tamanho deste vetor. A definição deste vetor é mostrada a seguir.

$$V = (media_{k-p}, \dots, media_k), \quad k > p \quad (4.5)$$

onde $media_k$ representa a média de intensidade de níveis de cinza dos pixels do quadro do instante k .

4.2.4 Cálculos do limiar dinâmico

A qualidade de segmentação de movimento baseado em subtração de plano de fundo (*background*) depende principalmente da escolha de um limiar ótimo de binarização. Mas devido às condições não controláveis de clima e iluminação, esta escolha é em média difícil e problemática como pode ser visto em aplicações que monitoram cenas exteriores durante um longo período ininterrupto.

Para lidar com estas condições, é necessário um cálculo automático de um limiar adaptativo que se possa ajustar a toda e qualquer alteração eventual na cena. Utilizando o histórico das médias de níveis de cinza dos pixels da imagem, um limiar dinâmico pode ser calculado automaticamente para cada instante k da seqüência de vídeo pela equação 4.6.

$$Th_k = \beta(media_k + \Delta_k) \quad (4.6)$$

onde:

- Th_k : representa o valor do limiar de binarização dinâmico calculado no instante k ;
- β : representa um valor definido pelo usuário. Em nossa implementação este valor está no intervalo $[0,1 \ 0,5]$;

- $media_k$: representa a média dos níveis de cinza dos pixels do quadro no instante k do vídeo;
- Δ_k : é o módulo da variação da média dos níveis de cinza dos pixels do vídeo no intervalo $[k-p, k]$ definida pela equação 4.7.

$$\Delta_k = media_k - media_{k-p}, \quad k > p; \quad (4.7)$$

A utilização da equação 4.7 no cálculo do limiar dinâmico permite o ajuste do resultado em função das variações detectadas. Quando temos um aumento na variação dos níveis de cinza de acordo com a equação 4.6 o valor do limiar Th_k também aumenta, e quando temos uma diminuição na variação dos níveis de cinza o valor de Th_k também diminui, considerando p suficientemente pequeno.

4.2.5 Cálculos da taxa de aprendizagem dinâmica

A aplicação do limiar dinâmico permite o melhoramento da qualidade da segmentação e ajuda o sistema a se adaptar dinamicamente às variações do ambiente. Partindo deste princípio e da busca por uma melhor robustez, Maadi e Maldague propõem em [Maadi e Maldague 2007] que para cada quadro k da seqüência de vídeo seja calculado além do limiar dinâmico de binarização também fosse calculado a taxa de aprendizagem α_k usada na equação 4.2 que regula a velocidade de atualização do *background*.

Baseado no vetor do histórico de variação dos níveis de cinza dos *pixels* da imagem, α_k , a taxa de aprendizagem, pode ser estimado pela equação 4.8:

$$\alpha_k = a + b \frac{|\Delta_k|}{\max(media_k, media_{k-p})}, \quad k > p; \quad (4.8)$$

onde:

- α_k : representa a taxa de aprendizagem calculada no instante k ;
- a : é o menor valor que α_k assume, em nossa aplicação $a = 0,05$;
- b : valor parametrizado pelo usuário que caracteriza a inclinação da taxa de aprendizagem, que em nossa aplicação tem valor 0,85 obtido empiricamente;
- p : indica tamanho do histórico de variação de iluminação da cena armazenado que em nossa implementação é igual a 4.

A equação 4.8 é a expressão analítica de uma reta que passa por um ponto de origem a e tem inclinação b . A variável de entrada da equação é um valor normalizado calculado para cada instante k e é diretamente proporcional à variação da intensidade de nível de cinza nos $p + 1$ quadros usados para guardar o histórico de mudança de iluminação

ocorrido na cena. A equação tem como saída a taxa de aprendizagem para aquele instante k .

Como um aspecto prático, a aplicação da equação 4.8 com parâmetros ótimos de entrada, promove uma melhora no comportamento do processo de atualização do plano de fundo. Com isso, a velocidade de aprendizagem do sistema está sujeita a um ajuste dinâmico caracterizado por uma aceleração ou desaceleração do processo em função da variação temporal da média de intensidade dos níveis de cinza de cada imagem.

O uso deste parâmetros dinâmicos foi proposto inicialmente em [Maadi e Maldague 2007]. Utilizamos as idéias propostas lá para desenvolver o módulo de segmentação de objetos em movimento usado neste trabalho. Além do que foi proposto em [Maadi e Maldague 2007], utilizamos análise de fluxo de movimento para melhorar o módulo de segmentação. Explicaremos como isto é feito na próxima Seção.

4.3 Análise de Fluxo

4.3.1 Introdução

Através do método de subtração de plano de fundo utilizado para segmentar objetos em movimento é possível identificar humanos, veículos e outros objetos em movimento na cena. Porém em vídeos típicos de vigilância de estacionamento não é sempre que o movimento está espalhado por toda cena. Existem situações onde somente poucos e pequenos passam pela área monitorada e com isto o movimento se concentra em uma pequena região da cena. Nestes casos a área onde há movimento é muito pequena comparada a todo a cena. Quando isso acontece a identificação/segmentação destes pequenos objetos em movimento fica prejudicada se analisarmos toda a cena, porque *pixels* estacionários influenciam no cálculo do limiar dinâmico Th_k , no histórico de variação de iluminação da cena Δ_k e por conseqüência no cálculo da taxa de aprendizagem dinâmica α_k , ou seja, os *pixels* estacionários influenciam na detecção do *foreground* e do *background*. Esta influência é indesejada uma vez que o que interessa para o módulo de segmentação são somente os *pixels* em movimento. Logo, para melhorar o processo de segmentação de objetos em movimento no instante k analisamos somente as regiões que cercam os objetos em movimento para realizar a segmentação ao invés de analisar todos os *pixels* da cena no instante k .

A Figura 4.3 ilustra a situação em que temos dois humanos na cena. A Figura 4.3(a) mostra a cena capturada pela câmera de vigilância onde há um humano na parte inferior da cena e outro humano na parte superior. O humano da parte superior está longe da câmera e por isso está muito pequeno na cena. Já o humano da parte inferior está próximo da câmera e com isso está um pouco maior na cena. Podemos ver na Figura 4.3(b) que somente o humano da parte inferior foi segmentado corretamente. Isso se deve

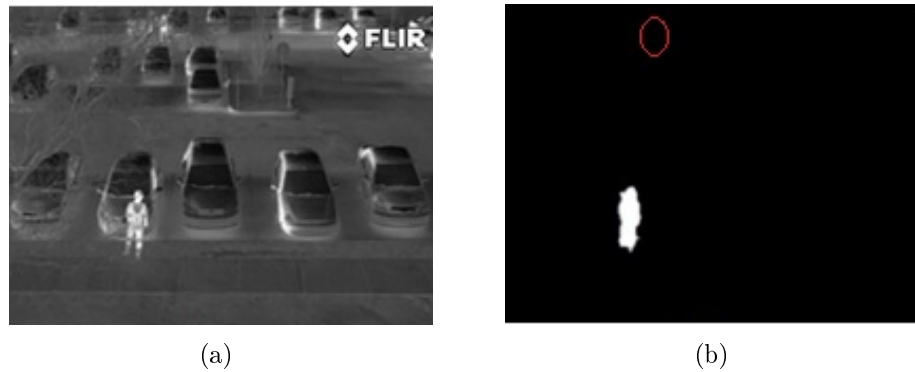


Figura 4.3: Falha na segmentação de objeto pequenos na cena. A Figura 4.3(a) mostra a cena do estacionamento onde há dois humanos, estando um na parte inferior da cena e outro na parte superior localizado atrás dos galhos de uma árvore. A Figura 4.3(b) mostra que a segmentação do primeiro humano foi realizada com sucesso e a segmentação do segundo humano falhou. O círculo vermelho na Figura 4.3(b) mostra o local onde um humano está, porém ele não é identificado na segmentação. A Figura 4.3(b) é o resultado da segmentação não utilizando a detecção de fluxo.

ao problema citado no parágrafo anterior.

O maior desafio em analisar somente as regiões onde realmente está acontecendo algum movimento é saber que um movimento está acontecendo ali. Para isto utilizamos o conceito de fluxo (que em vídeo pode ser definido como o movimento de uma região contígua de *pixels* em uma mesma direção) para encontrar as regiões da cena onde existe algum tipo de movimento. Mesmo que um objeto tenha se movido por uma pequena quantidade de quadros ou o objeto em movimento seja pequeno o fluxo de movimento associado a este objeto será detectado utilizando detecção de fluxo.

4.3.2 Detecção de fluxo

A idéia principal da detecção de fluxo é subtrair quadros consecutivos do vídeo sucessivamente para identificar as diferenças contidas entre estes quadros. Fazendo isto uma certa quantidade, o fluxo de movimento associado ao vídeo é extraído corretamente. Nos parágrafos seguintes esta técnica é explicada com mais detalhes.

Seja $F_1^0, F_2^0, \dots, F_{m+1}^0$ uma seqüência de $m+1$ quadros que faz parte de uma seqüência de vídeo de u quadros no total. O fluxo de movimento associado a esta seqüência pode ser dado por um fator F construído da seguinte forma:

$$F_i^j = F_{i+1}^{j-1} - F_i^{j-1} \quad (4.9)$$

onde j representa o nível da detecção do fluxo que varia de 1 a m , ou seja, quantas vezes sucessivas a subtração entre os quadros já foi feita, e i indica o número de quadros usados para gerar o fator F e varia de 1 até $m - j + 1$. A cada nível de detecção de fluxo o número de quadros usados para gerar o fluxo daquele nível é diminuído de um quadro

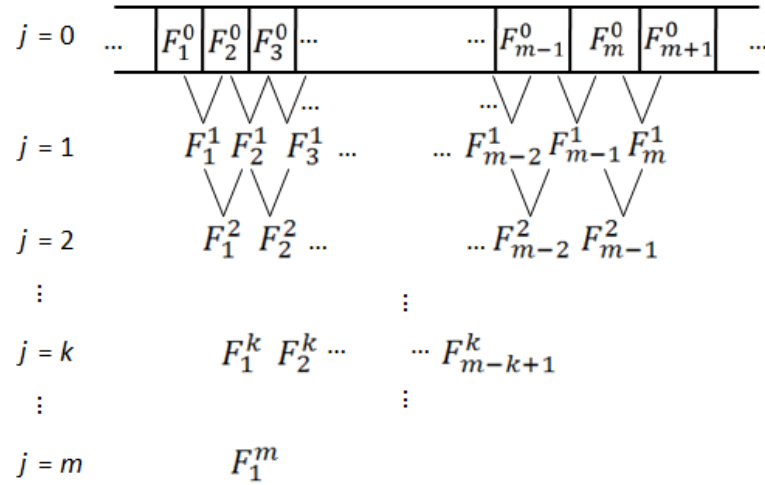


Figura 4.4: Processo de obtenção do fluxo de movimento associado a uma seqüência de quadros de um vídeo.

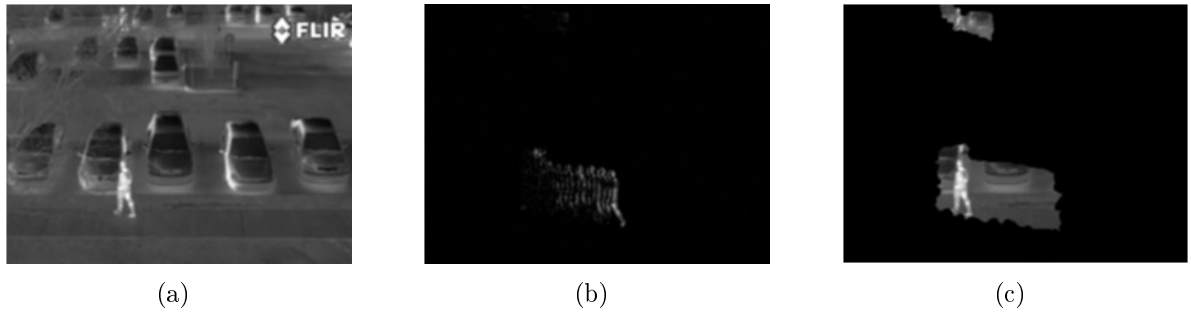


Figura 4.5: Detecção de fluxo de movimento. (a) quadro original da seqüência de vídeo no instante k , (b) fluxo associado obtido com $m = 30$, (c) fluxo final associado ao quadro do instante k após operações morfológicas.

em relação ao nível anterior. E os quadros usados para gerar o fluxo no nível j são os quadros que compõe o nível $j - 1$. Um desenho esquemático do processo de geração do fluxo associado ao nível j é mostrado na Figura 4.4. Por exemplo, F_2^2 é obtido através do cálculo $F_3^1 - F_2^1$.

O fluxo detectado por esse processo indica em que regiões do quadro estão ocorrendo movimentos. O número m de vezes que a equação 4.9 é aplicada a uma seqüência de vídeo foi escolhido experimentalmente de tal forma que o fluxo associado a todos os objetos em movimento na cena esteja presente em F_1^m (fator que representa o fluxo associado à seqüência de quadros do vídeo no nível m). A Figura 4.5 mostra este processo de identificação de fluxo. Na Figura 4.5(a) temos o quadro original do instante k , na Figura 4.5(b) temos o fluxo associado a este quadro obtido no nível 30. A Figura 4.5(c) mostra o fluxo final associado ao quadro. Este fluxo foi obtido aplicando-se operações morfológicas (erosões e dilatações) nos objetos da Figura 4.5(b). Assim na Figura 4.5(c) temos regiões consistentes onde podemos ver onde houve algum tipo de movimento na cena naquele

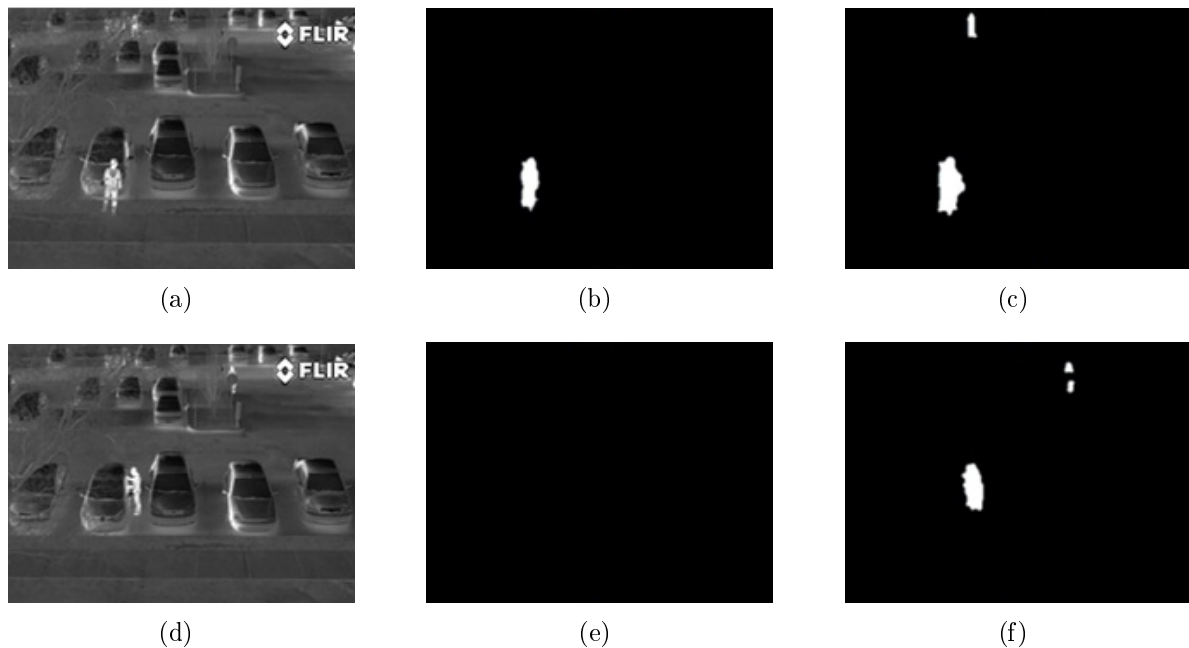


Figura 4.6: Segmentação com e sem restrição ao fluxo. As Figuras 4.6(a) e 4.6(d) mostram frames da cena original, as Figuras 4.6(b) e 4.6(e) mostram o resultado da segmentação sem a restrição das regiões de fluxo e as Figuras 4.6(c) e 4.6(f) mostram o resultado da segmentação com a restrição de fluxo.

instante. Podemos ver nas Figuras 4.5(b) e 4.5(c) que mesmo o fluxo associado com o movimento da pessoa no topo da cena que está distante da câmera e conseqüentemente é bem pequeno, é detectado pelo nosso processo de detecção de fluxo.

Como essas regiões de fluxo representam as regiões da cena onde existe movimento e este módulo de segmentação só está interessado em detectar objetos em movimento, propusemos analisar a imagem somente nas regiões onde há fluxo de movimento. Desta forma, informações desnecessárias de *pixels* estacionários não influenciaram nos cálculos do processo de subtração de plano de fundo. A Figura 4.6 mostra o resultado da segmentação de objetos em movimento com e sem a detecção de fluxo proposta. As Figuras 4.6(a) e 4.6(d) mostram a cena original, as Figuras 4.6(b) e 4.6(e) mostram o resultado da segmentação sem a restrição das regiões de fluxo e as Figuras 4.6(c) e 4.6(f) mostram o resultado da segmentação com a restrição de fluxo. Podemos ver que quando utilizamos a restrição de fluxo conseguimos detectar objetos em movimento que não são detectados quando não usamos esta restrição.

A escolha do valor de m está relacionada ao tipo de aplicação para a qual a técnica de análise de fluxo de movimento será usada bem como ao valor da taxa *fps* do vídeo. O valor de m deve ser tal que movimentos significantes possam ser capturados. Não pode ser um valor muito alto de tal forma que tenhamos um único fluxo de movimento associado ao vídeo que cubra espacialmente quase toda a cena (devido à passagem de vários objetos pela cena) o que tornaria a aplicação da técnica desnecessária, e não pode ser um valor muito baixo de tal forma que alguns movimentos não sejam detectados. Para a nossa

aplicação de um estacionamento externo onde a velocidade dos objetos é relativamente baixa e a taxa *fps* dos vídeos é igual a 5, realizamos vários testes com vários valores para m e obtivemos os melhores resultados quando assumimos $m = 30$. Caso usemos a técnica de detecção de fluxo de movimento para outro ambiente de aplicação e/ou outra taxa de *fps*, talvez outro valor para m deva ser encontrado para se obter resultados semelhantes aos obtidos neste trabalho.

4.4 Detecção do contorno dos objetos e operações morfológicas

Aplicando esta técnica de subtração de plano de fundo combinada com o limiar, com a taxa de aprendizagem dinâmica e com a detecção de fluxo o sistema é capaz de segmentar objetos em movimento na cena. Mas mesmo que esta técnica possua uma robustez em face às condições externas, como mudança na iluminação da cena, ela não compensa a falta de informação devido à baixa qualidade das imagens adquiridas. Por isso é necessário combinar a técnica de subtração de plano de fundo com outra técnica que permite a recuperação da maior quantidade de informação possível dos objetos da cena. Nesta Seção explicamos como isto é feito neste trabalho.

Detecção dos contornos, ou detecção de borda, dos objetos é uma técnica de processamento de imagem e visão computacional para determinar pontos de uma imagem digital em que a intensidade luminosa muda repentinamente. Mudanças repentinas em imagens geralmente refletem eventos importantes no cenário, como a descontinuação da profundidade (transição entre o objeto e o fundo), descontinuação da orientação da superfície, mudança das propriedades do material ou variações na iluminação da cena.

A aplicação da equação 4.1 permite ao sistema a segmentação temporal dos objetos em movimento na cena. Porém, muitas vezes devido à baixa qualidade das imagens adquiridas, estes objetos estão incompletos ou com uma segmentação de baixa qualidade. Para conseguir uma melhora na segmentação dos objetos em movimento, [Maadi e Maldague 2007] combinaram as regiões segmentadas pela equação 4.1 com as bordas, ou contornos, destes objetos. Uma diagrama de blocos que sintetiza a arquitetura da solução pode ser visto na Figura 4.2.

A detecção dos contornos não é feita em toda imagem. Ela é feita somente no resultado obtido pela aplicação da equação 4.1. As regiões detectadas pela equação 4.1 são chamadas regiões de interesse. Estas regiões de interesse são definidas como cada conjunto de *pixels* contíguos (4-conexos) em uma mesma região que formam "borrões" pela imagem. A cada um destes "borrões" detectados após a aplicação da equação 4.1 é adicionado 5% de seu tamanho para que este "borrão" se torne realmente a região de interesse da qual o contorno será extraído.

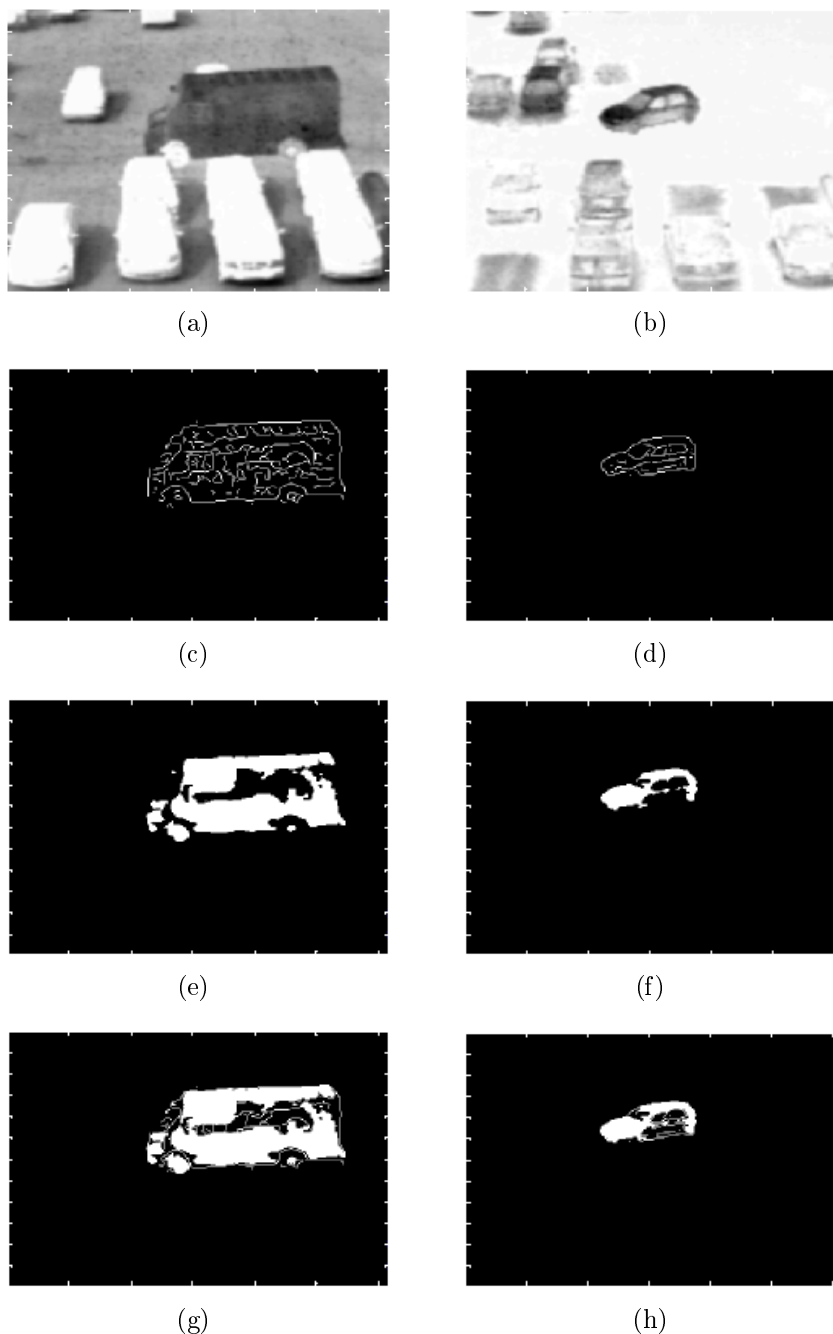


Figura 4.7: Resultado da combinação do objeto com o seu contorno. Figuras 4.7(a) e 4.7(b) mostram a cena capturada pela câmera de vigilância, Figuras 4.7(c) e 4.7(d) mostram os contornos dos objetos, Figuras 4.7(e) e 4.7(f) mostram os "borrões" obtidos pela aplicação da técnica de subtração de plano de fundo e Figuras 4.7(g) e 4.7(h) mostram o resultado da combinação dos "borrões" com o contorno dos objetos.

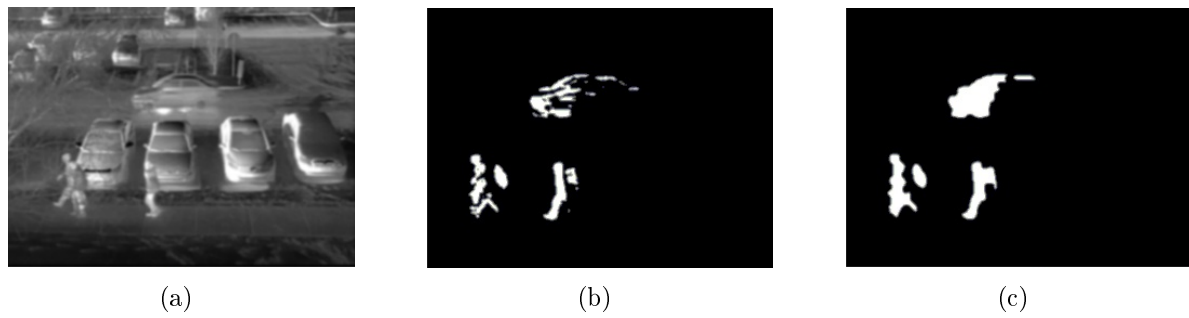


Figura 4.8: Resultado das operações morfológicas aplicadas nos objetos segmentados pela técnica de subtração de plano de fundo usada neste trabalho. A Figura 4.8(a) mostra a cena do vídeo de vigilância em um instante qualquer, a Figura 4.8(b) objetos resultantes da aplicação da técnica de subtração de plano de fundo na cena da Figura 4.8(a) e a Figura 4.8(c) resultado das operações morfológicas aplicadas na Figura 4.8(b).

Neste trabalho, assim como em [Maadi e Maldague 2007], utilizamos um operador bastante conhecido: o operador de *Canny*. No trabalho de [Maadi e Maldague 2007] os autores realizaram testes com diversos operadores e o operador de *Canny* mostrou resultados satisfatórios. Em [Ziou e Tabbone 1998] é possível realizar uma leitura mais completa sobre detectores de borda.

A Figura 4.7 mostra o resultado da combinação da região de interesse do objeto detectada com o contorno deste objetos. As Figuras 4.7(a) e 4.7(b) mostram a cena capturada pela câmera de vigilância, as Figuras 4.7(c) e 4.7(d) mostram os contornos dos objetos, as Figuras 4.7(e) e 4.7(f) mostram os "borrões" obtidos pela aplicação da técnica de subtração de plano de fundo e por fim as Figuras 4.7(g) e 4.7(h) mostram o resultado da combinação dos "borrões" com o contorno dos objetos. É possível notar que a combinação dos "borrões" com o contorno dos objetos provê objetos melhor definidos, quando comparados aos "borrões" obtidos somente pela subtração de plano de fundo.

A combinação entre os objetos detectados pela equação 4.1 e o contorno destes objetos (detectado na imagem original) é feita através de um operador "ou" entre as duas imagens como pode ser visto na Figura 4.2. Após este "ou" algumas operações morfológicas matemáticas são aplicadas para eliminar ruído e falsas detecções. Duas dilatações e duas erosões são aplicadas consecutivamente nos objetos resultantes do "ou" entre as duas imagens. O tamanho e a forma dos operadores usados nessas operações morfológicas foram escolhidos experimentalmente de forma a otimizar a segmentação dos objetos em questão. Por fim é necessário realizar uma análise de componentes conexos. A técnica consiste em reagrupar os *pixels* de uma vizinhança chamados "componentes conexos". Cada um dos agrupamentos de *pixels* são disjuntos entre si e cada componente conexa pode ser 4-conexa ou 8-conexa dependendo do critério de vizinhança utilizado. Neste trabalho utilizamos componentes 4-conexas pois foi usado o critério de vizinhança de 4 vizinhos. É possível encontrar maiores informações sobre análise de componentes conexas e operações morfológicas em [Rafael C. González 2008].

A Figura 4.8 mostra o resultado das operações morfológicas aplicadas nos objetos segmentados pela técnica de segmentação de plano de fundo utilizada neste trabalho. A Figura 4.8(a) mostra a cena presente no vídeo em um determinado instante. A Figura 4.8(b) mostra os objetos segmentados no processo de segmentação de plano de fundo e a Figura 4.8(c) mostra o resultado final da segmentação após as operações morfológicas.

Após a aplicação destas operações morfológicas, temos em mãos "borrões" espalhados pela imagem. Esses "borrões" são os objetos em movimento na cena. Para representar estes objetos traçamos uma caixa delimitadora em torno de cada um desses "borrões". A Figura 4.9 mostra um objeto da classe humano e sua caixa delimitadora.



Figura 4.9: Objeto da classe humano com sua caixa delimitadora.

4.5 Extração de características

Uma vez que a segmentação dos objetos em movimento é concluída, a próxima etapa é a seleção de características que possam sintetizar da melhor maneira possível as características da imagem. Isto pode ser feito através da *extração de características* destes objetos. Estas características representam as informações elementares que são usadas: para identificar qual a *classe* de cada objeto, para monitorar o objeto e para reconhecer eventos de interesse.

Em função do objetivo fixado e da metodologia de segmentação escolhida, as características extraídas podem ser estruturais, estatísticas, topológicas ou métricas. Sendo assim, a fim de simplificar a complexidade algorítmica e satisfazer os objetivos do sistema, um vetor de características contendo os seguintes dados foi criado:

- coordenadas do centro de massa da caixa delimitadora do objeto;
- área total da caixa delimitadora do objeto;
- posição da caixa delimitadora do objeto (representada por quatro valores que são as posições espaciais extremas do objeto na imagem, dois na horizontal e dois na vertical);
- velocidade de deslocamento do centro de massa (calculada pela equação 5.2);
- primeiro e último quadros em que o objeto apareceu (preenchido pelo módulo de monitoramento);

- classe do objeto (preenchido após o objeto ser classificado em uma das três classes: humano, veículo ou não-identificado).

Em imagens infravermelhas, uma possível característica é a temperatura média do objeto. A temperatura média do objeto está relacionado com o nível de cinza de cada *pixels* que compõem o objeto. Quando a temperatura do ambiente que cerca o objeto é constante, podemos assumir que a temperatura média de um objeto é maior quanto maior forem os níveis de cinza de seus *pixels*. Como os vídeos usados neste trabalho foram gravados em três dias consecutivos consideramos a temperatura do ambiente constante em todos os vídeos. Por isso em nosso trabalho a temperatura média dos objetos está diretamente relacionada com os níveis de cinza de seus *pixels*. Para analisar vídeos onde a temperatura do ambiente são diferentes é necessário aplicar uma equação que converta o nível de cinza para temperatura real. Para isto podemos usar a equação proposta em [Marinetti et al. 1997] para descobrir a real temperatura média do objeto. Em nosso trabalho não usamos a característica temperatura média pelo fato de que a temperatura do ambiente foi considerada constante e assim podemos olhar diretamente para os níveis de cinza dos *pixels* para realizar comparações entre os objetos.

4.6 Classificação dos objetos em movimento

Para um humano é fácil a detecção de objetos em movimento bem como a identificação das interações entre estes objetos. Para que uma máquina atinja este nível de inteligência, uma tarefa essencial é a classificação dos objetos em movimento em diferentes classes. A classificação destes objetos é feita com base nas características extraídas anteriormente. Nesta Seção detalharemos o algoritmo de classificação dos objetos segmentados.

Para cada região segmentada da imagem onde existe possibilidade da presença de um objeto, o algoritmo de classificação atribuirá a classe mais adequada para este objeto. Propomos neste Capítulo um algoritmo que permite distinguir humanos e veículos. Esta restrição em duas classes pode ser justificada pelo fato de que a ocorrência de objetos destas classes é muito freqüente em um ambiente de um estacionamento. Buscamos os seguintes objetivos na construção do algoritmo de classificação:

- Encontrar uma melhor combinação de critérios utilizando diferentes características primitivas do objeto;
- A classificação é feita sem exemplos de aprendizagem dos objetos;

Dentre as várias abordagens de classificação existentes, inspirados em [Maadi e Maldague 2006], focamos nossos esforços em três condições básicas para o funcionamento em tempo real do sistema: simplicidade, rapidez (baixo tempo de máquina) e eficiência (alto índice

de acerto na classificação). Para isso, optamos, após vários testes, pela combinação de três características diferentes:

1. A relação entre a altura e a largura do retângulo da caixa delimitadora do objeto candidato, usado para dar ao algoritmo uma invariância em alterações de escala da imagem;
2. A área ocupada pelo retângulo da caixa delimitadora do objeto;
3. O módulo da velocidade de deslocamento do objeto na cena, representada pela velocidade do centro do retângulo da caixa delimitadora do objeto que pode ser calculado pela equação 5.2.

Os elementos de entrada para o nosso método de classificação são as siluetas 2D com qualquer forma e em qualquer posição conhecida na imagem. Como saída, as etiquetas "humano" ou "veículo" são atribuídos a cada um dos objetos de acordo com a classificação mais adequada para este objeto. Caso o objeto não obedeça nenhum dos critérios das duas classes, uma terceira classe de objetos não identificados agrupa estes objetos. Dentre os objetos desta classe estão cães, pássaros, ciclistas e qualquer outro objeto que possa passar pela cena (que não seja um humano ou um veículo). A aparição de ruído na cena devido à baixa qualidade de imagens capturadas ou qualquer outro fator externo ao sistema também pode gerar objetos não identificados pelo sistema (borrões disformes). Estes objetos também serão colocados nesta terceira classe. Os objetos desta terceira classe passarão pelo algoritmo de classificação a cada quadro até que este objeto saia do campo de visão da câmera e não esteja mais presente na cena. Isto para tentar garantir que um objeto (por exemplo um humano) que foi classificado como não identificado nos primeiros quadros seja corretamente classificado nos quadros seguintes.

Assim, inspirados em [Maadi e Maldague 2006], propomos algumas mudanças nos critérios de classificação propostos por eles para criar os nossos critérios de classificação a fim de obter melhores resultados na classificação de objetos.

4.6.1 A classe *humano*

O vetor de características criado para cada objeto contém todos os dados para o monitoramento e para a classificação: área, posição do centro de massa, coordenadas da caixa delimitadora do objeto, velocidade de deslocamento, etc. Um objeto é classificado como humano se ele obedece o seguintes critério: a velocidade de deslocamento é diferente de zero, em outras palavras, o objeto não está parado; a área do objeto é menor que o maior tamanho possível para objetos da classe humano; e a relação altura/largura é maior que 1,5. O valor para o maior tamanho possível para objetos da classe humano foi escolhido experimentalmente e em nossa implementação esse valor é de 1500 *pixels*. A Figura 4.10 mostra a classificação correta de dois humanos.



Figura 4.10: Classificação correta de dois humanos pelo algoritmo.

4.6.2 A classe *veículo*

Para que um grupo de pessoas não seja confundido com um veículo, o parâmetro velocidade tem papel fundamental para classificação de um objeto como sendo veículo. A unidade de medida utilizada é *pixel*/quadro e é calculada entre duas imagens consecutivas a partir das coordenadas do centro de massa do objeto. Um objeto é classificado como veículo, quando um dos seguintes critérios é verificado: a área do objeto é maior que um limiar pré-definido (em nossa implementação: 500 *pixels*) e a velocidade de deslocamento do centro de massa é maior que outro limiar (em nossa implementação: 4); ou a área do objeto é maior que a maior área dos objetos da classe *humano* (em nossa implementação: 1500 *pixels*); ou a velocidade de deslocamento do centro de massa do objeto é maior que a maior velocidade de deslocamento dos objetos da classe *humano* (em nossa implementação: 20). Novamente todos os limiares utilizados para a criação destes critérios foram escolhidos experimentalmente e estes valores foram os que mostraram melhores resultados. A Figura 4.11 mostra a classificação correta de um veículo.

4.6.3 A classe *não identificado*

Neste trabalho esta terceira classe agrupa todos os demais objetos que não passaram pelos testes das outras duas classes. Os objetos desta classe só são utilizados pelo módulo de monitoramento e não serão utilizados pelo módulo de gerenciamento de eventos. Nesta classe ocorrem vários tipos de objetos como animais, grupos de pessoas, ciclistas, etc. Os objetos desta terceira classe passaram pelo algoritmo de classificação a cada quadro até que este objeto saia do campo de visão da câmera e não esteja mais presente na cena.



Figura 4.11: Classificação correta de um veículo pelo algoritmo.

Isto para tentar garantir que um objeto (por exemplo um humano) que foi classificado como não identificado nos primeiros quadros seja corretamente classificado nos quadros seguintes.

4.7 Considerações finais

Neste trabalho propomos técnicas para um sistema para vigilância de áreas externas utilizando uma câmera infravermelha estacionária. A primeira e mais importante parte que compõem este sistema é o módulo de segmentação de objetos em movimento. Inspirados em [Maadi e Maldague 2007] utilizamos uma técnica de subtração de plano de fundo que se adapta dinamicamente às mudanças repentinas de iluminação da cena. Para isso calculamos para cada instante k da seqüência de vídeo um limiar dinâmico para o cálculo do *foreground* e uma taxa de aprendizagem dinâmica para a estimativa do *background*. Também usamos a idéia de fluxo para podermos analisar o vídeo somente nas regiões onde realmente existe movimento.

Também apresentamos os critérios usados para a classificação dos objetos segmentados em duas classes: *humano* e *veículo*. As características usadas para esta classificação são: velocidade, relação entre altura/largura do objeto e área do objeto. Este passo do sistema é de suma importância para o módulo de gerenciamento de eventos pois para que o sistema possa reconhecer eventos entre os objetos é preciso que estes objetos estejam classificados corretamente. Por exemplo, quando uma pessoa se aproxima de um veículo e entra nele, caso o primeiro objeto, o humano, não esteja classificado como *humano*, o evento "humano entrando dentro de um veículo" não será reconhecido.

Ao final deste módulo o sistema tem cada objeto em movimento na cena segmentado e classificado. Estes objetos serão o ponto de partida dos outros módulos que compõem o sistema de vídeovigilância: módulo de monitoramento e módulo de gerenciamento de eventos. Estes outros dois módulos serão detalhados nos capítulos seguintes. No Capítulo 7 serão apresentados alguns resultados experimentais obtidos neste módulo.

Capítulo 5

Módulo de monitoramento

5.1 Introdução

Um dos problemas que deve ser resolvido pela análise de seqüências de vídeo é determinar objetos em movimento (feito no Capítulo anterior) e estudar a evolução espacial e/ou temporal destes objetos em diferentes imagens. A primeira etapa consiste em uma análise espaço-temporal da imagem original e a segunda etapa consiste no monitoramento dos objetos segmentados na seqüência.

Neste Capítulo, estamos interessado particularmente na etapa de monitoramento. Esta etapa consiste em monitorar (seguir) múltiplos objetos simultaneamente, de formas e texturas variadas, que podem assumir qualquer tipo de trajetória na seqüência de imagens infravermelhas. Isto é feito neste trabalho através de um ciclo de duas fases: previsão e correção. Nas secções seguintes, explicaremos o módulo de monitoramento usado no presente trabalho.

5.2 Algoritmo proposto

Um dos maiores problemas na realização de um sistema de vídeovigilância inteligente é o seguinte: como obter a partir de uma longa seqüência de quadros informações confiáveis, coerentes e de forma rápida, a fim de identificar e compreender eventos pertinentes sobre a evolução espaço temporal dos objetos em movimento na área supervisionada. A dificuldade é maior pois diferentes objetos podem entrar em movimento ao mesmo tempo, pois a informação visual depende não apenas do movimento relativo de um objeto em relação ao sensor (a câmera), mas também dos outros objetos em movimento naquela região. A análise deste tipo de movimento pode ser efetuada graças a diferentes técnicas de monitoramento.

O algoritmo de monitoramento desenvolvido nesta parte do projeto é baseado no monitoramento das coordenadas espaço-temporais dos centros de massa dos objetos (caixas

delimitadoras). Os objetivos visados nesta etapa são os seguintes:

- Determinar o momento de entrada de um novo objeto no campo de visão do sistema de vídeovigilância e inicializar o modelo de movimento para o monitoramento;
- Determinar a correspondência entre as regiões do *foreground* detectadas pelo método de segmentação de objetos em movimento e os objetos monitorados atualmente;
- Determinar a trajetória completa de cada objeto durante toda a sua presença na cena.

Os resultados da realização destes objetivos serão posteriormente utilizados para a classificação dos objetos e para a identificação de interações entre os mesmos. É importante então nesta etapa que tenhamos um algoritmo rápido e eficiente.

Como já foi dito nesta Seção, este módulo do sistema é baseado no monitoramento espaço-temporal de centros de massa. A escolha desta coordenada é justificada pela sua robustez na avaliação na posição dos objetos e porque ela não é afetada pelos grandes movimentos das extremidades do objeto [Haritaoglu et al. 1998]. A estimativa deste movimento é realizada através de um modelo linear que supõe que o movimento realizado por um objeto no intervalo de dois quadros consecutivos é relativamente pequeno e este objeto possui velocidade constante (não acelerado). Esta suposição é feita para que se possa simplificar o cálculo dos parâmetros e medidas realizadas.

Um algoritmo simples de monitoramento é utilizado neste módulo. Ele também foi usado em [Maadi e Maldague 2007] e está estruturado em duas etapas distintas (a Figura 5.1 ilustra estas duas etapas):

- Uma fase de previsão, durante a qual uma ou mais hipóteses sobre a posição do objeto na imagem ou em relação à câmera são feitas.
- Uma fase de correção ou exploração da vizinhança que faz fronteira com o estado previsto em que a posição do objeto é refinada.

É importante dizer que no instante k , a região monitorada está supostamente segmentada de forma adequada no novo quadro no instante $k + 1$. Assim, resta agora a estimativa dos parâmetros de movimento relacionado com o deslocamento do objeto da posição que ele estava no instante k para a posição que ele está no instante $k + 1$. Isto é

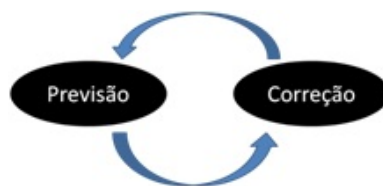


Figura 5.1: Diagrama de previsão e correção.

feito a fim de alimentar o modelo de monitoramento para que este último possa fornecer a nova previsão.

Entretanto, esta previsão não é sempre perfeita. A região obtida na previsão não corresponde sempre à região real do objeto. Por isso utilizamos um método de atualização baseado em um critério de sobreposição das caixas delimitadoras dos objetos previstos com os objetos realmente medidos.

5.2.1 Inicialização do monitoramento

Durante a inicialização, todas as regiões do *foreground* detectadas são consideradas como borrões. Estes são monitorados por uma certa quantidade de quadros (nosso sistema utiliza cinco quadros consecutivos) antes de serem passados para o módulo de monitoramento. Isto nos permite:

- Detectar falsas regiões segmentadas resultantes de uma segmentação de baixa qualidade, ou outras causas relacionadas com o sensor infravermelho, como um traço (ou pegada) térmico de baixa densidade deixado por um objeto que iniciou o movimento após permanecer um tempo parado. As impressões térmicas de alta densidade serão tratadas no decorrer desta Seção.
- Inicialização dos vetores de características onde a velocidade de deslocamento dos objetos são medidas para que o processo de previsão possa iniciar.

Esta fase é bastante crítica pois um erro que ocorra nesta inicialização talvez prejudique todo o restante do processo. Por isso o algoritmo passa por esta fase preliminar que analisa os borrões antes de criar o objeto propriamente dito. Para esta tarefa ele analisa os tamanhos e a distância entre os borrões presentes em quadros consecutivos. Quando a inicialização chega ao fim o algoritmo inicia a fase de previsão usando os dados recém adquiridos.

5.2.2 Fase de previsão

A fase de previsão corresponde à primeira etapa do algoritmo de monitoramento. O seu papel é tentar fornecer uma estimativa da próxima evolução espacial da região durante o seu movimento. Em outras palavras, nesta fase é efetuada uma projeção temporal das coordenadas do centro de massa do objeto da posição (x_k, y_k) no instante k para a posição (x_{k+1}, y_{k+1}) no instante $k + 1$. Para isso, o módulo de monitoramento realiza uma extrapolação linear utilizando o par velocidade (V_{x_k}, V_{y_k}) , como indicado na equação 5.1.

$$\begin{cases} x_{k+1} = x_k + V_{x_k} \\ y_{k+1} = y_k + V_{y_k} \end{cases} \quad (5.1)$$

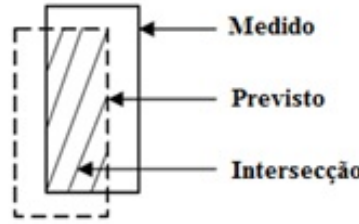


Figura 5.2: Estágio de correção realizado através da sobreposição de caixas delimitadoras.

onde V_{x_k} e V_{y_k} são os valores de velocidade de deslocamento no eixo x e eixo y respectivamente do centro de massa do objeto calculados usando o quadro precedente $k - 1$ através da equação 5.2.

$$\begin{cases} V_{x_k} = x_k - x_{k-1} \\ V_{y_k} = y_k - y_{k-1} \end{cases} \quad (5.2)$$

5.2.3 Fase de correção

Uma vez que a previsão é realizada, o estágio de correção é feito utilizando a posição real do objeto e a posição estimada na fase previsão. Um critério de correspondência baseado na sobreposição de caixas delimitadoras é computada entre o estado previsto e o medido. A Figura 5.2 ilustra este estágio.

Este critério de correspondência representa a porcentagem de interseção entre a caixa delimitadora do objeto em um instante qualquer e a caixa delimitadora do mesmo objeto previsto para este instante. O valor desta porcentagem depende da taxa de captura de imagens da câmera utilizada bem como a velocidade de deslocamento média dos diferentes objetos na cena. Neste projeto, o limiar para esta porcentagem foi fixado, experimentalmente, em 75% do tamanho da caixa delimitadora do objeto previsto. Isto significa que com um valor de correspondência maior que este limiar, a região que foi medida corresponde à região que havia sido prevista anteriormente.

Para aumentar a velocidade de processamento o algoritmo realiza uma redução no tamanho da janela de pesquisa em torno da zona prevista. O novo espaço de pesquisa é dado em função do comprimento da diagonal principal do retângulo formado pelas coordenadas da caixa delimitadora obtida no estágio de previsão.

Na prática, para buscar um possível candidato de correspondência, o algoritmo considera apenas objetos com centro de massa localizados a uma distância euclidiana menor que um certo limiar. Na nossa implementação o valor deste limiar é de 55% do valor do comprimento da diagonal principal da caixa delimitadora do objeto previsto. Assim para os candidatos cuja distância mínima é verificada e o critério de correspondência for obedecido o objeto previsto sofre uma atualização dos valores de seu vetor de característica (coordenadas do centro de gravidade, velocidade de deslocamento, tamanho, etc.).

5.2.4 Gestão de oclusões

O monitoramento de objetos em movimento em ambientes não controlados exige métodos mais complexos do que somente sobreposição de regiões. O algoritmo terá que lidar com problemas oriundos da segmentação, alto número de objetos a serem monitorados e principalmente da oclusão de objetos de diversas formas.

A respeito do problema de descontinuidade de movimento devido a oclusões, duas soluções são aplicadas neste trabalho: a primeira é prática e é feita no momento da captura das imagens, seu objetivo é reduzir ao máximo o risco de ocorrência de oclusões escolhendo-se um lugar adequado para a câmera de captura usada (altura e ângulo de inclinação); a segunda é algorítmica, onde as oclusões ocorridas são tratadas. Nosso módulo de monitoramento não considera que os objetos estejam parcialmente e temporariamente ocultos. Na verdade o que ocorre é que se um objeto "é perdido" durante alguns quadros devido a baixa qualidade de imagens ou uma oclusão total, o processo de monitoramento é parado e se faz necessário localizar novamente este objeto em um instante posterior.

Para aumentar a robustez do sistema a este tipo de fenômeno, é possível utilizarmos mais pontos do objeto para realizar o monitoramento. Entretanto, fazendo isto aumentaríamos consideravelmente o tempo de execução do algoritmo. Como estamos buscando um sistema com tempo de processamento em tempo real e visamos também o monitoramento de um número razoavelmente grande de objetos simultaneamente na cena essa abordagem fica inviável. Assim utilizamos somente o centro de massa do objeto para tratar as oclusões de objetos.

O algoritmo de oclusão é capaz de prever o acontecimento de uma oclusão com um quadro de antecedência. Para isto o algoritmo usa, em cada instante k , as coordenadas espaciais de cada objeto para prever eventuais interseções de trajetórias no instante $k + 1$. Desta forma, se o algoritmo detecta que uma interseção significativa ocorrerá no quadro seguinte, o algoritmo reconhece a presença de uma oclusão no quadro seguinte, e assim ele substitui no instante $k + 1$ o objeto medido pelo objeto que foi previsto no instante k . Em outras palavras, os objetos participantes da oclusão são substituídos pelos objetos previstos anteriormente até que haja a separação dos mesmos.

5.2.5 Gestão de impressões térmicas

Um dos problemas tratados por [Maadi e Maldague 2007] é a presença de ruído nas seqüências de imagens infravermelhas prejudicial ao algoritmo de monitoramento. Estes ruídos que aparecem nas imagens são traços ou impressões térmicas deixados por objetos que entram em movimento após estar parado durante um longo tempo (objetos que pertenciam ao *background*). Isso acontece, por exemplo, na saída de um veículo que estava estacionado durante horas. Uma impressão térmica se forma quando este veículo começa a deixar sua posição onde estava estacionado e esta impressão cresce à medida que o

veículo se move. Com isso o sistema detecta um único objeto: o veículo juntamente com a crescente impressão térmica deixada por ele.

Este fenômeno pode ser explicado por uma análise simples do algoritmo de segmentação utilizado. Como a equação de subtração de plano de fundo (equação 4.1) utiliza a diferença absoluta entre a imagem do quadro corrente e a imagem referência do *background*, os resultados não dão uma idéia de variação de nível de cinza para cada *pixel*.

Na prática, se aplicarmos este princípio no nosso exemplo da saída de um veículo, podemos ver fenômeno de impressão térmica pela diferença de temperatura do *background*, antes, durante e depois da saída do veículo. Desta forma, a imagem binária obtida após a segmentação contém uma massa única de tamanho crescente que na verdade é o resultado da segmentação do veículo em questão anexado a sua impressão térmica como podemos ver na Figura 5.3.

A Figura 5.3 mostra o problema que temos com as impressões térmicas por exemplo na saída de veículos. O algoritmo é incapaz de fazer a separação do objeto e da impressão térmica antes que o objeto saia do campo de visão da câmera e isto é um grande problema para o algoritmo de monitoramento. Nestes casos específicos utilizamos também uma taxa suplementar baseada no cálculo de densidade dos "borrões". Esta taxa é dada pela equação 5.3.

$$d = \frac{Ab}{Ac} \quad (5.3)$$

onde d é a densidade do borrão, Ab é a área do borrão e Ac é a área da caixa delimitadora do objeto.

À medida que o veículo sai do local estacionado o valor de d vai diminuindo. Se este valor decrescente atingir um limiar mínimo (neste trabalho este limiar está fixado em 0,5) o algoritmo realiza a separação do borrão em duas partes e analisando o sentido do movimento, ele considera uma das partes como sendo o objeto verdadeiro (veículo que continuará sendo monitorado) e a outra parte é automaticamente incorporada ao *background*.

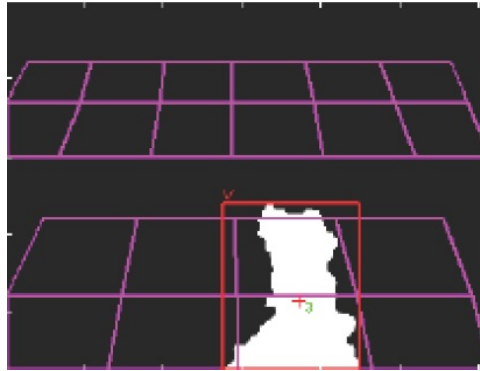


Figura 5.3: Impressão térmica anexada ao veículo no momento de sua saída após estar estacionado durante um tempo.

5.3 Considerações finais

Neste Capítulo apresentamos um algoritmo simples porém eficiente para o monitoramento de objetos em movimento em vídeo. Este algoritmo é baseado em um ciclo de previsão e correção. Ele trata as possíveis oclusões que possam ocorrer substituindo os objetos que participam da oclusão pelos objetos previstos e seus comportamentos até que a oclusão termine. O algoritmo trata também das impressões térmicas típicas de vídeos infravermelhos. No Capítulo 7 serão apresentados alguns resultados experimentais obtidos neste módulo.

No Capítulo seguinte é apresentado o último módulo que compõe o sistema de vigilância proposto: o módulo de gerenciamento de eventos. Este módulo visa o reconhecimento de atividades suspeitas no local monitorado. Estas atividades reconhecidas serão reportadas ao operador do sistema para que o mesmo possa tomar as devidas atitudes para tratar tal evento.

Capítulo 6

Módulo de gerenciamento de eventos

6.1 Introdução

Este trabalho visa desenvolver técnicas que possam ser usadas na construção de um sistema automático de videovigilância de áreas externas. Tal sistema pode ser composto por três módulos: módulo de segmentação, módulo de monitoramento e módulo de gerenciamento de eventos. Nos capítulos anteriores discutimos como os dois primeiros módulos são desenvolvidos neste trabalho. No presente Capítulo iremos discutir o módulo de gerenciamento de eventos.

A detecção automática de eventos em uma seqüência de imagens tem ganhado mais e mais atenção dos pesquisadores na área de visão computacional ([Fernández et al. 2010], [Foresti et al. 2004], [Michelsoni et al. 2009]) devido a demanda por aplicações de videovigilância, interpretação e compreensão de interações entre objetos, dentre outras. Os tipos de eventos que podem ser reconhecidos vão de simples gestos feitos com as mãos até grandes ações encadeadas que têm várias interações físicas entre os objetos durante um longo período de tempo.

Neste trabalho iremos identificar eventos que possam ocorrer em uma área de um estacionamento. Estes eventos variam de uma simples ação de estacionar um veículo até uma ação mais complexa de suspeita de roubo de objetos do interior de um veículo estacionado. No restante deste Capítulo explicaremos como isto é feito neste trabalho.

6.2 Modelagem do evento

Realizar a detecção de evento não é uma tarefa fácil, na verdade é um grande desafio e requer a interpretação (alto nível) das características (baixo nível) extraídas do vídeo. Então é necessário realizar a ligação dos dados capturados pela câmera (*pixels*) com as descrições de alto nível que comporão as ações/eventos. Uma abordagem comum e utilizada neste trabalho consiste em primeiramente segmentar as regiões em movimento e

transformar estes *pixels* segmentados em características que possam representar de forma satisfatória os objetos em movimento na cena. A posterior análise destas primitivas permite a classificação destes objetos e o reconhecimento de interações espaço-temporais entre os mesmos. Entretanto, existem vários desafios para realizar esta tarefa:

- A detecção de movimento e o monitoramento de objetos em seqüências reais de vídeo são muita vezes instáveis por causa da baixa qualidade dos vídeos (oclusões, ruídos, etc.);
- A interpretação de características de baixo nível (por exemplo, aparência do objeto) pode depender do ponto de vista;
- Existem diferenças espaço-temporais em uma mesma ação realizada por diferentes autores fazendo que duas ações iguais sejam visualmente diferentes;
- Em uma mesma ação repetida pelo mesmo ator podem existir pequenas variações;
- Padrões de eventos semelhantes podem ser expressos por ações diferentes.

Conseqüentemente, é preciso que a representação do evento seja genérica capaz de descrever uma ação genérica bem como uma ação simples. E para que o sistema seja de fácil manutenção, também é necessário que ele seja facilmente extensível e fácil de modificar. Baseados nessas restrições escolhemos para realizar o reconhecimento dos eventos que acontecem na área monitorada, características como: posição espaço-temporal, velocidade de deslocamento e classe dos objetos.

Existem várias abordagens para a construção deste módulo de gerenciamento de eventos. Entretanto existem algumas que não são adequadas para este projeto pois violam restrições postas anteriormente ou porque não são adequadas ao tipo de imagem capturada ou ao ambiente monitorado. Como foi exposto no Capítulo 3 existem duas formas de modelar eventos: forma implícita e forma explícita. Neste trabalho escolhemos a abordagem explícita pois buscamos melhores resultados com respeito a falsos positivos e falsos negativos. Além disso, em nosso ambiente monitorado existe uma quantidade restrita de eventos que serão monitorados e por isso o nosso sistema não avalia novos eventos com frequência e não necessita ser auto-adaptativo.

Uma vez que o sistema utiliza imagens 2D de uma câmera estacionária, o reconhecimento de eventos é independente do ponto de vista. Os eventos tratados por este trabalho podem ser divididos em duas classes: eventos com um único ator e eventos com múltiplos atores, sendo que ator no contexto de eventos é aquele objeto que participa ativamente do evento e no nosso caso pode ser humano ou veículo. Na primeira categoria de eventos somente um objeto (ator) está envolvido no evento. Já na segunda categoria de eventos no mínimo dois objetos (atores) estão envolvidos no evento. Estes objetos podem ser de categorias diferentes e podem estar "chegando" ou "saindo" do *background* (exemplo: chegada ou saída de um veículo na área de estacionamento).

Esta estrutura de classificação representa a base da estratégia de classificação que o nosso algoritmo utiliza. A Figura 6.1 mostra como os tipos de eventos tratados por este algoritmo estão estruturados. Um evento pode envolver somente um humano, somente um veículo, a interação entre humanos, a interação entre humanos e veículos e a interação entre veículos.

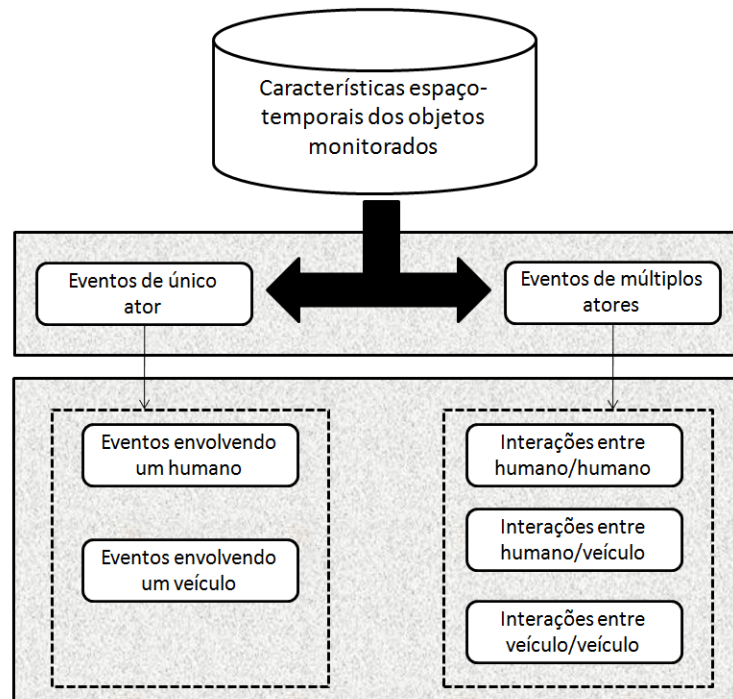


Figura 6.1: Classes de eventos que podem ser tratados pelo módulo de gerenciamento de eventos.

Diversos critérios podem ser usados para o reconhecimento de eventos. Em nosso trabalho as características usadas para criar os critérios de reconhecimento de eventos são apresentadas a seguir. Todos os eventos tratados em nosso sistema usam um ou uma combinação destes critérios.

- Critério de velocidade mínima;
- Proximidade espacial dos objetos;
- Proximidade temporal dos objetos;
- Análise de trajetória;
- Tempo parado em lugar em particular;
- Interseção de objetos.

Eventos do tipo único ator podem ser uma pessoa saindo de um veículo ou um veículo sendo estacionado em um lugar vago no estacionamento. Um evento do tipo múltiplos atores pode ser várias pessoas entrando dentro de um veículo. Nas próximas seções detalharemos cada um dos eventos tratados por este sistema.

6.3 Eventos de um único ator

Como realizado [Maadi e Maldague 2006], somente um objeto participa dos eventos pertencentes a esta classe. E como indicam as seções anteriores, somente humanos e veículos podem compor os eventos, ou seja, objetos não identificados não são tratados por este módulo. Quando um objeto não identificado aparece na cena o sistema somente extrai as características como velocidade, área e coordenadas e o monitora até que este saia do campo de visão da cena.

Para reconhecer eventos desta classe, utilizamos características como velocidade e coordenadas do retângulo que forma a caixa delimitadora do objeto. Com isso é possível caracterizarmos duas ações relacionadas com ações feitas por objetos na cena:

- Um movimento de desaceleração de um objeto até a sua total parada;
- Um movimento de aceleração de um objeto que pertencia ao *background* (início de um novo movimento).

Com estes dois casos, somos capazes de identificar ações relacionadas com o início e o fim de qualquer movimento. No primeiro caso, um limiar de velocidade mínima é utilizado para cada classe de objeto (humano e veículo). Os valores deste limiares são empíricos e não-nulos pois, valores de velocidade nulos fazem com que o objeto seja incorporado ao *background* nos quadros seguintes. Baseado nisto, a seguinte lista de eventos que podem ser identificados pelo algoritmo foi criada. No Capítulo 7 apresentamos exemplos destes eventos.

- **Parar caminhada**, válido para todos os objetos da classe *humano* onde a velocidade de deslocamento do centro de massa do objeto decresce abaixo de um limiar mínimo por mais de cinco quadros consecutivos. As características deste objeto são memorizadas para que o sistema possa seguir este objeto caso ele entre novamente em movimento;
- **Caminhar novamente**, válido também somente para objetos da classe *humano*. Este evento engloba objetos que já foram identificados anteriormente pelo sistema e estão agora retomando sua caminhada. Um casamento baseado na sobreposição de caixas delimitadoras é feito aqui para atribuir o movimento detectado ao objeto correto que voltará corretamente a ser monitorado pelo sistema;
- **Estacionar veículo**, válido somente para objetos da classe *veículo* onde cessaram seu movimento em uma vaga de estacionamento disponível. Como no primeiro tipo de evento, o valor da velocidade de deslocamento do objeto deve ser menor que um limiar mínimo por mais de cinco quadros consecutivos. Quando um evento deste é detectado o sistema consegue identificar o local vago onde o veículo parou (baseado

na sobreposição da vaga e da caixa delimitadora do objeto) e consegue atribuir um rótulo "ocupado" para aquela vaga;

- **Iniciar**, válido somente para veículos que iniciaram seu movimento deixando a área estacionada. O módulo de monitoramento começa a monitorar novamente este objeto e assim como no evento anterior, o sistema agora atribui um rótulo "desocupado" para a vaga de estacionamento deixada pelo veículo.

Além destes quatro eventos, os eventos de **entrada na cena** e **saída da cena** também são detectados pelo sistema. Estes eventos são trivialmente reconhecidos quando um objeto ultrapassa a fronteira da cena entrando ou saindo dela respectivamente.

6.4 Eventos de múltiplos atores

Como o próprio nome indica pelo menos dois atores estão envolvidos neste tipo de evento. Por isso é possível existir interações entre objetos da mesma classe (interações entre humano/humano e interações entre veículo/veículo) e interações entre objetos de classes diferentes (interações entre humano/veículo).

Particularmente neste projeto estamos interessados nos eventos do tipo humano/veículo. Para os dois outros tipos de eventos identificaremos somente os eventos de início de oclusão e fim de oclusão através do módulo de monitoramento.

Assim como os eventos de único ator tratados neste trabalho, os eventos de múltiplos atores tratados nesta Seção também foram estudados no trabalho [Maadi e Maldague 2006]. O nosso trabalho reconhece os mesmo tipos de eventos propostos em [Maadi e Maldague 2006] e propõe a extensão do módulo de gerenciamento de eventos para que este reconheça eventos mais complexos com mais de um ator. Nesta Seção primeiramente falaremos um pouco do que já foi feito em [Maadi e Maldague 2006] e em seguida passaremos a expor o que estamos propondo neste trabalho.

6.4.1 Eventos simples

Como iremos tratar somente eventos que envolvam humano e veículo simultaneamente, os eventos que serão tratados aqui são a entrada e saída de pessoas de veículos. Para realizar o presente reconhecimento iremos utilizar três critérios que já foram apresentados anteriormente no início desta Seção: a proximidade espacial entre objetos, a posição temporal dos objetos e a análise de trajetória.

A proximidade espacial entre objetos pode ser usada para reconhecer o evento **Entrar em um veículo** por um ou mais humanos através do cálculo de interseção da área onde o veículo está estacionado e a caixa delimitadora do humano. Este cálculo é feito um quadro antes do humano desaparecer da cena como imagem individual, ou seja, antes



Figura 6.2: Exemplo de caso onde pode ocorrer indecisão no momento da entrada do humano no veículo.

da sua imagem se fundir com a imagem do veículo. Desta maneira quando um humano está entre dois veículos estacionado é possível identificar em qual dos dois veículos ele irá entrar. A Figura 6.2 mostra essa situação.

A Figura 6.2 mostra duas situações onde o algoritmo pode se confundir ao decidir qual dos dois veículos o humano irá entrar. Para evitar esta confusão, consideramos que uma pessoa está mais próxima de seu veículo no momento que ela vai entrar nele (no momento da fusão dos objetos). Neste momento a interseção entre a caixa delimitadora do humano e a área delimitadora do lugar onde o veículo esta estacionado é máxima. Assim, analisando esta interseção é possível escolher o veículo correto.

O segundo evento de interesse é a **Saída do interior de um veículo** de um ou mais passageiros após o evento de estacionar um veículo. A restrição do evento estar condicionado ao acontecimento de outro evento é a utilização do critério temporal para o reconhecimento do evento. Além deste critério temporal, para reconhecer este evento, analisamos também a proximidade espacial do "humano" e do "veículo". Entretanto, algumas vezes, a identificação do passageiro como "humano" só se dá após o monitoramento do objeto por alguns quadros consecutivos por isso é possível que o objeto já esteja longe do "veículo".

Para evitar este problema, o algoritmo utilizado usa a análise da trajetória do objeto em movimento. Assim, olhando a primeira posição do objeto, ou seja, onde o objeto apareceu pela primeira vez, o sistema poderá identificar de qual veículo o objeto saiu através da análise da interseção das caixas delimitadoras. Desta forma o evento "saída do interior do veículo" é corretamente reconhecido.

6.4.2 Eventos suspeitos

Os eventos apresentados até agora são eventos simples de serem reconhecidos e por isso não necessitam de uma estrutura mais complexa de representação para serem reconhecidos. Os eventos tratados até agora já foram tratados anteriormente em [Maadi e Maldague 2006]. Agora em nosso trabalho estamos propondo o reconhecimento de even-

tos mais complexos. A estrutura de representação de eventos que modelamos aqui neste trabalho é adequada para o reconhecimento de qualquer evento mais complexo que possa acontecer na área monitorada. Por limitações de tempo para o desenvolvimento deste projeto e a falta de um banco de dados com eventos diversificados, optamos por trabalhar com somente um evento complexo sendo que este sistema pode ser alterado facilmente para o reconhecimento de outros eventos.

Em [Maadi e Maldague 2006] os autores reconhecem os eventos: Parar caminhada, Caminhar novamente, Estacionar veículo, Iniciar (iniciar funcionamento de um veículo), Entrar em um veículo e Saída do interior de um veículo. Usando o que foi proposto em [Maadi e Maldague 2007] e [Maadi e Maldague 2006] propomos neste trabalho o reconhecimento de eventos suspeitos que podem ocorrer na área monitorada. Alguns eventos considerados suspeitos para um ambiente de um estacionamento são:

- Humano abandonando uma mochila em algum lugar do estacionamento e depois este humano sai da área monitorada (este cenário representa o abandono de um explosivo);
- Um veículo parado em uma área não permitida;
- Um humano andando no estacionamento com uma trajetória suspeita/incomum (circulando entre os carros);
- Um humano removendo objetos do interior de um veículo;
- Uma pessoa ficando do lado de um veículo estacionado por um tempo, depois entrando dentro deste veículo, depois demorando um tempo dentro deste veículo e depois saindo do estacionamento dirigindo este veículo (cenário que representa o roubo do próprio veículo).

Eventos reconhecidos em [Maadi e Maldague 2006] como A : "Veículo chegando ou deixando uma posição x " e B : "Humano entrando ou saindo de um veículo" são eventos simples e como descritos anteriormente são de fácil reconhecimento. Neste trabalho propomos o reconhecimento do evento suspeito C : "suspeita de roubo de objeto(s) do interior de um veículo estacionado em uma posição x por um humano". Mesmo que este evento C seja um evento mais difícil de ser reconhecido que A e B , podemos reconhecê-lo utilizando a abordagem de sub-eventos discutida em [Lavee et al. 2009]. Podemos dividir o evento C em quatro sub-eventos da seguinte forma:

- $C1$: humano se aproximando de um veículo estacionado na posição x ;
- $C2$: humano ficando próximo deste veículo por um tempo;
- $C3$: humano entrando dentro do veículo;
- $C4$: humano se afastando deste veículo.

Neste trabalho desenvolvemos procedimentos e realizamos testes somente para o evento suspeito C embora como dito anteriormente este sistema pode ser estendido para a identificação de outros eventos suspeitos. Assim podemos dizer que o evento que queremos reconhecer é o evento C e os sub-eventos que o compõe são C_s onde $s = 1, 2, 3, 4$. Para criar o evento C_{s+1} o evento C_s deve existir para os mesmos atores. Ou seja, os sub-eventos têm entre si restrições temporais. As características analisadas para o reconhecimento de cada sub-evento são: proximidade espacial entre os objetos, tempo de um objeto parado em um lugar específico, intersecções entre objetos e velocidade de deslocamento dos objetos.

Como para reconhecer o evento C é preciso antes reconhecer, seqüencialmente, os sub-eventos $C1$, $C2$, $C3$ e $C4$, se faz necessário a utilização de uma estrutura de dados adequada para o armazenamento destes sub-eventos e os dados de cada um. Um *array* é utilizado para realizar este armazenamento. Nele são guardados o tempo em que o sub-evento ocorreu, o tipo do sub-evento e os dados de cada um dos atores do sub-evento. Na implementação deste trabalho os atores envolvidos nestes sub-eventos são dois e somente dois: um humano e um veículo. Todos os quadros do vídeo são analisados e quando uma restrição para a criação de um sub-evento é alcançada este sub-evento é criado. Quando os quatro sub-eventos são criados, o evento C então é criado.

O reconhecimento do sub-evento $C1$ (humano se aproximando de um veículo estacionado na posição x) é feito verificando um critério de proximidade espacial entre o humano e o veículo estacionado e um critério de velocidade mínima do humano. Para o critério de proximidade espacial, analisamos a área de interseção das caixas delimitadoras do humano e do veículo. Caso esta área de interseção seja maior que um limiar de 5% da área do veículo o critério de proximidade é atendido. Para o critério de velocidade mínima a velocidade do humano deve ser menor que um limiar de 0,2. Todos estes limiares foram obtidos experimentalmente após a execução de vários testes com diferentes valores. Quando um humano atende estes requisitos um sub-evento do tipo $C1$ é criado para cada um dos veículos os quais o humano está próximo (mais de um sub-evento é criado quando uma pessoa está entre dois carros estacionados). Um sub-evento é criado para cada veículo que atende o critério de proximidade pois neste momento não sabemos qual é o veículo "alvo" do humano. Precisamos de ter um sub-evento criado para cada veículo pois para criar o próximo sub-evento é necessário que este primeiro sub-evento exista para o humano e o veículo corretos.

Para o sub-evento $C2$ (humano ficando próximo deste veículo por um tempo) além da verificação da existência do sub-evento $C1$ para os mesmos atores do sub-evento que está sendo reconhecido, o tempo de ocorrência entre os sub-eventos $C1$ e $C2$ são comparados e um critério de velocidade mínima é aplicado na velocidade de deslocamento do humano. Este critério é aplicado para que se possa verificar se o humano está parado ou andando bem devagar (como se estivesse andando observando o alvo que escolheu anteriormente) perto do veículo. Este critério é verificado quando a velocidade do humano está abaixo

de um limiar escolhido experimentalmente. Em nossa implementação este limiar é de 0,8. Além disso, é verificado se a diferença de ocorrência entre os sub-eventos $C1$ e $C2$ é maior que um limiar, ou seja, o humano está perto do carro por um período de tempo considerável. Se estas duas situações ocorrem então o sub-evento $C2$ é criado para os atores envolvidos.

Já para os sub-eventos $C3$ (humano entrando dentro do veículo) devemos verificar a existência dos sub-eventos $C2$ e $C1$. Além desta primeira verificação, para o sub-evento $C3$ verificamos se a área ocupada pelo humano é menor que um limiar (em nossa implementação este limiar tem valor igual a 700 *pixels*) e a porcentagem de interseção das caixas delimitadoras do humano e do veículo deve ser maior que 10% da área total do veículo. Verificando-se estes dois critérios temos a indicação que o humano deve ter entrado dentro do veículo de alguma forma.

No caso do sub-evento $C4$ (humano se afastando deste veículo) é verificada a existência dos sub-eventos anteriores ($C3$, $C2$ e $C1$). Além disso, é verificado se o humano está se afastando do veículo. Isto é feito comparando a distância euclidiana dos dois centros de massa, do humano e do veículo, a um limiar pré-definido obtido experimentalmente. O valor deste limiar em nossa implementação é de 8,2. Outra forma de verificar se o humano está se afastando do veículo é analisando sua velocidade de deslocamento. Normalmente um assaltante foge com uma velocidade alta da cena do crime. Assim caso a velocidade de afastamento do humano seja maior que 4, também consideramos que ele está se afastando do veículo. Se uma destas duas condições é verificada e os sub-eventos anteriores existem, criamos então o sub-evento $C4$.

Após a criação do sub-evento $C4$ o sistema cria automaticamente um evento suspeito do tipo C e avisa o operador do sistema para que este possa tomar as devidas providências para tratar este acontecimento. Não consideramos o evento C como um evento "fora da lei" e sim como um evento suspeito pois existem casos em que os sub-eventos são verificados, mas na verdade não é um evento "fora da lei". Por exemplo, no caso em que o motorista do veículo estava fazendo compras em um *shopping center* e resolveu guardar algumas sacolas em seu veículo e depois retornar para fazer mais compras. Normalmente o motorista do veículo não gasta muito tempo do lado de fora do veículo antes de entrar dentro dele. Na maioria das vezes, o motorista se aproxima do veículo e rapidamente o abre e coloca as sacolas no seu interior. Mas pode acontecer do motorista ter esquecido onde colocou a chave do veículo gastando mais tempo para abri-lo. Desta maneira, a ação de motorista será caracterizado pelo nosso sistema como uma ação suspeita. A Figura 6.3 mostra uma cena de cada um dos quatro sub-eventos que compõem o evento suspeito C .

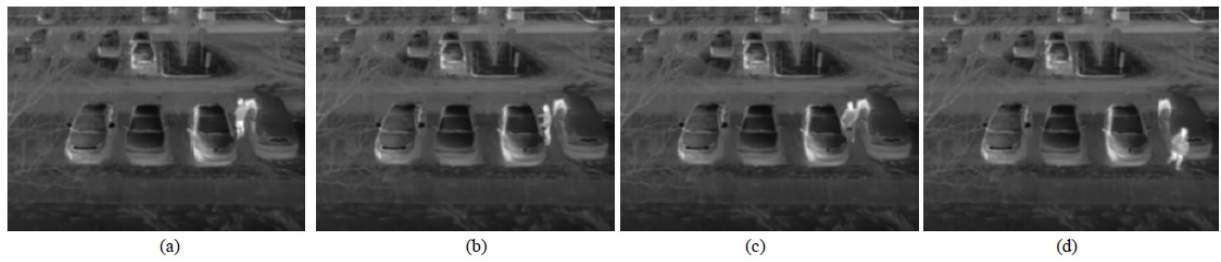


Figura 6.3: Evento suspeito do tipo C: (a) humano se aproximando de um veículo estacionado na posição x, (b) humano ficando próximo deste veículo por um tempo, (c) humano entrando dentro do veículo, (d) humano se afastando deste veículo.

6.5 Considerações finais

O módulo de gerenciamento de eventos utiliza características básicas dos objetos para poder reconhecer eventos que ocorram na área monitorada. As características usadas são: posicionamento espaço-temporal, velocidade de deslocamento, análise de trajetória e tempo que o objeto ficou parado em um determinado local.

Como foi feito em [Maadi e Maldague 2006] este trabalho também reconhece eventos de um único ator e de vários atores. Além dos eventos já tratados em [Maadi e Maldague 2006] este trabalho propõe o reconhecimento de eventos suspeitos na área monitorada. Apesar de serem eventos mais complexos que os eventos reconhecidos anteriormente, adotando uma abordagem que considera que um evento é composto por vários sub-eventos, podemos identificar um evento destes identificando sequencialmente cada um dos sub-eventos que o compõem. Por limitação de tempo, neste trabalho identificamos somente o evento "suspeita de roubo de objetos do interior de um veículo estacionado". Mas o algoritmo proposto pode ser facilmente adaptado para o reconhecimento de outro evento suspeito.

No próximo Capítulo mostraremos alguns resultados experimentais para os três módulos que compõem este sistema (segmentação, monitoramento e gerenciamento de eventos).

Capítulo 7

Resultados experimentais

7.1 Sistema de aquisição

Como foi exposto anteriormente, este trabalho visa desenvolver técnicas que possam ser usadas na construção de um sistema de vídeo vigilância de uma área externa utilizando uma câmera estacionária capaz de identificar eventos suspeitos que possam ocorrer na área monitorada. O desenvolvimento deste sistema foi inspirado nos trabalhos de [Maadi e Maldague 2007] e [Maadi e Maldague 2006] que também monitoram através de uma câmera infravermelha estacionária um estacionamento externo.

Como descrito no Capítulo 2, a banda de frequência ideal para trabalhar com avaliação de objetos na temperatura ambiente é a banda LWIR que vai de 8 a 14 μm . Por isso escolhemos a câmera Photon 640 produzida pela empresa FLIR Systems, Inc. A especificação completa desta câmera pode ser encontrada no site do fabricante (<http://www.flir.com> acessado em 21/10/2011). A Figura 7.1 mostra uma foto desta câmera e o gráfico na Figura 7.2 mostra a curva de resposta da câmera.



Figura 7.1: Câmera infravermelha utilizada neste projeto.

Esta câmera não possui um mecanismo de armazenamento interno de imagens. Na verdade ela funciona somente como um capturador de imagens, ou seja, ela fornece como saída um *streaming* de vídeo. Assim, precisamos conectá-la a um computador para que este possa fazer o armazenamento das imagens capturadas. A Figura 7.3 mostra o esquema de aquisição de imagens utilizado. Suas especificações básicas são:

- **Tipo do sistema:** câmera infravermelha não resfriada LWIR;
- **Tipo do sensor:** microbolômetro VOx;
- **Formato da imagem:** 640x512;
- **Espectro de frequência:** 7.5 a 13.5 μm ;
- **Faixa de temperatura de operação:** $-40^{\circ}C$ a $+80^{\circ}C$.

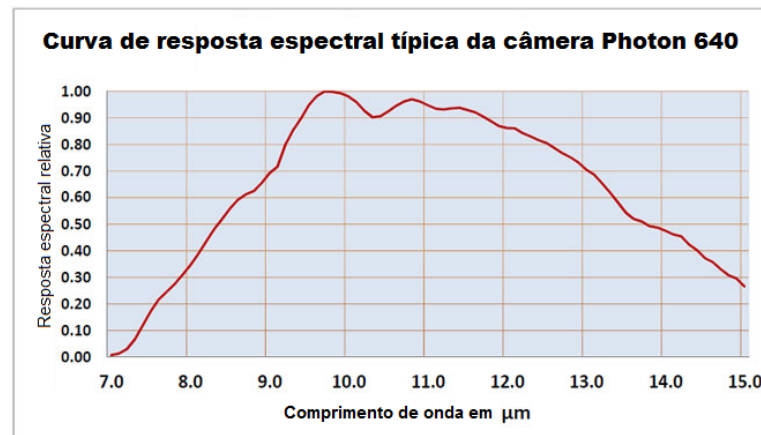


Figura 7.2: Curva de resposta espectral típica da câmera Photon 640.

Outra característica que deve ser destacada aqui é a taxa de captura de imagens utilizadas. Escolhemos uma taxa de 5 quadros/segundo. A escolha deste valor foi feita baseada em dois fatos:

- A necessidade de armazenamento de seqüências de vídeo mais longas possíveis a fim de simular um sistema real em diferentes condições. A baixa taxa de captura permite o armazenamento de um tempo determinado de vídeo em metade do espaço quando usamos uma taxa de captura duas vezes maior;
- Normalmente os objetos em um estacionamento se deslocam com uma velocidade reduzida.

O computador utilizado para o armazenamento dos vídeos foi um laptop que possui 4GB de memória secundária, 450GB de memória principal e um processador Intel(R)



Figura 7.3: Esquema do sistema de aquisição.



Figura 7.4: Sistema de aquisição usado para as gravações.

Core(TM) i5 CPU M460 2.53GHz. Neste laptop estava instalado o programa *Virtualdub* que pode ser encontrado no site virtualdub.org (acessado em 20/10/2011). Este programa realiza captura e edição de vídeos e é de uso gratuito. Os algoritmos usados neste trabalho foram implementados em *Matalab* mas podem ser portados para outra linguagem procedural como *C*.

A câmera usada neste trabalho captura imagens com tamanho de até 640x512. Porém para diminuir o tempo de processamento das imagens capturadas, as imagens que usamos possuem tamanho de 320x240. A Figura 7.4 mostra como ficou montado o esquema de captura durante as gravações dos vídeos usados neste trabalho.

7.2 Banco de dados

Em [Maadi e Maldague 2007] os autores usaram uma câmera infravermelha estacionária para monitorar um estacionamento externo. Entretanto, os vídeos utilizados em [Maadi e Maldague 2007] e [Maadi e Maldague 2006] não possuem eventos suspeitos. Como um dos objetivos deste trabalho é o reconhecimento de eventos suspeitos, o uso deste banco de dados fica inviável. Além disso, só tivemos acesso a partes deste banco.

Por causa disso foi decidido que criaríamos um novo banco com cenas infravermelhas de um estacionamento. Na criação deste novo banco, tentamos reproduzir fielmente as principais características do banco usado em [Maadi e Maldague 2007] e [Maadi e Maldague 2006]: altura da câmera, inclinação da câmera em relação ao solo, distância da câmera para os objetos da cena e tamanho da área monitorada. A Figura 7.5 mostra uma imagem de cada banco respectivamente.

A criação deste novo banco de dados foi realizada pelo aluno durante o estágio que ele fez na *Universidade de Laval* na cidade de Québec no Canadá. Parte do estacionamento da universidade foi monitorado no final do outono de 2010 durante três dias consecutivos. Durante as gravações a câmera foi colocada em uma das salas de estudo que fica no

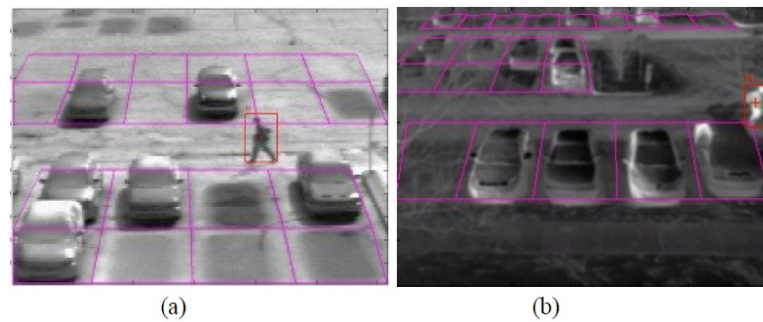


Figura 7.5: (a) imagem retirada do banco de dados usados em [Maadi e Maldague 2007] e (b) imagem retirada do banco de dados construído para este trabalho.

segundo andar de uma das residências estudantis da universidade. A câmera ficou a aproximadamente 10 metros do solo e monitorou a área do estacionamento através de uma janela aberta como pode ser visto na Figura 7.4. Neste trabalho, o número de quadros por segundo (*fps*) dos vídeos é igual a 5.

As gravações dos vídeos foram realizadas durante três dias consecutivos. A temperatura média do ambiente durante estes dias foi de $-1,5^{\circ}\text{C}$ com diversas condições climáticas. Durante as gravações houve cenas com o céu limpo e ensolarado, com o céu nublado, com a presença de neve no chão e sem a presença de neve, durante o dia e durante a noite. Os vídeos gravados nos três dias totalizam uma hora e meia de gravações onde temos a passagens de pedestres, ciclistas, grupos de pessoas, carros e caminhões pela área monitorada. Ocorreram também durante as gravações diversas interações entre humanos e veículos como entrada/saída do humano no/do veículo.

Entretanto, o Canadá é um dos países onde a taxa de criminalidade é baixa. Além disso históricos da universidade mostram que dificilmente um assalto ocorre dentro do campus. Por isso o evento suspeito que é alvo deste trabalho foi encenado no estacionamento pois mesmo que ficassemos um mês realizando gravações muito dificilmente presenciariamos uma cena criminoso.

A encenação do evento suspeito "suspeita de roubo de objetos do interior de um veículo estacionado" foi feita por dois alunos da universidade. Um deles ficou operando o sistema de gravação e o outro encenou o evento. As ações feitas na encenação do evento foram compatíveis com a divisão em sub-eventos do evento suspeito. Ou seja, primeiramente o ator entrava na cena andando e parava próximo a um veículo ficando um tempo próximo a este veículo. Em seguida o ator simulava o arrombamento do veículo por meio de uma porta que havia sido deixada destrancada ou uma janela que havia sido deixada aberta. Depois disto ele retira uma mochila do interior do veículo e se afasta correndo de perto dele.

Sendo assim, o banco construído contém a passagem pela cena de 265 humanos e 137 veículos enquanto que no banco usado em [Maadi e Maldague 2007] haviam 183 humanos e 58 veículos. Além disso, o nosso banco possui 14 eventos do tipo "suspeita de roubo de

objetos do interior de um veículo estacionado" que foram encenados pelos dois alunos.

Vários experimentos foram realizados para testar o desempenho de nosso sistema. A apresentação e análise destes resultados será feita nas próximas secções.

7.3 Módulo de segmentação e classificação de objetos

O módulo de segmentação é a primeira etapa para a construção do sistema proposto. Através de uma técnica dinâmica de subtração de plano de fundo o sistema é capaz de segmentar objetos em movimento na cena.

Podemos dividir os resultados deste módulo em duas etapas: segmentação sem restrição ao fluxo e segmentação com restrição ao fluxo. Na primeira etapa deste projeto utilizamos um modelo de subtração de plano de fundo onde temos Th_k e α_k calculados dinamicamente para cada instante k (vide equações 4.1 e 4.2). A implementação feita nesta etapa foi inspirada em [Maadi e Maldague 2007]. Na segunda etapa do desenvolvimento deste projeto propomos a utilização da análise de fluxo combinada com o método usado na primeira etapa para realizar a segmentação de objetos em movimento como foi explicado na Seção 4.3. Para as duas etapas os critérios de classificação de objetos foram os mesmos. O objeto, após ser segmentado, é classificado em uma das três classes (humano, veículo ou não-identificado) seguindo os critérios expostos anteriormente no Capítulo 4.

7.3.1 Segmentação não restrita ao fluxo

Primeiramente apresentaremos os resultados obtidos na primeira etapa do desenvolvimento deste trabalho. Realizamos diversos testes em nosso banco de dados com o método de segmentação sem a restrição de fluxo. A Figura 7.6 mostra a segmentação e classificação de alguns humanos. Nas Figuras 7.6(a) e 7.6(b) temos a segmentação de um humano caminhando na calçada onde a Figura 7.6(a) mostra o vídeo original e a Figura 7.6(b) mostra o resultado da segmentação onde o objeto foi classificado corretamente. Nas as Figuras 7.6(c) e 7.6(d) temos a segmentação de um grupo de pessoas (que é classificado como não identificado) e de um humano, classificado corretamente como humano, que acabou de entrar no canto inferior da cena. As Figuras 7.6(e) e 7.6(f) mostram a segmentação e classificação correta de um humano no topo da cena.

Como pode ser visto o sistema é capaz de segmentar e classificar corretamente humanos em diferentes situações no vídeo. Além de humanos, o sistema também é capaz de identificar veículos. A Figura 7.7 mostra a segmentação sem restrição de fluxo e a classificação de alguns veículos em diferentes cenas do banco de dados. Nas figuras (a) e (b) temos a segmentação e classificação de um veículo com movimento da direita para esquerda. Nas figuras (c) e (d) temos a segmentação de outro veículo que está fazendo uma curva. Nas figuras (e) e (f) temos a segmentação e classificação de outro veículo que

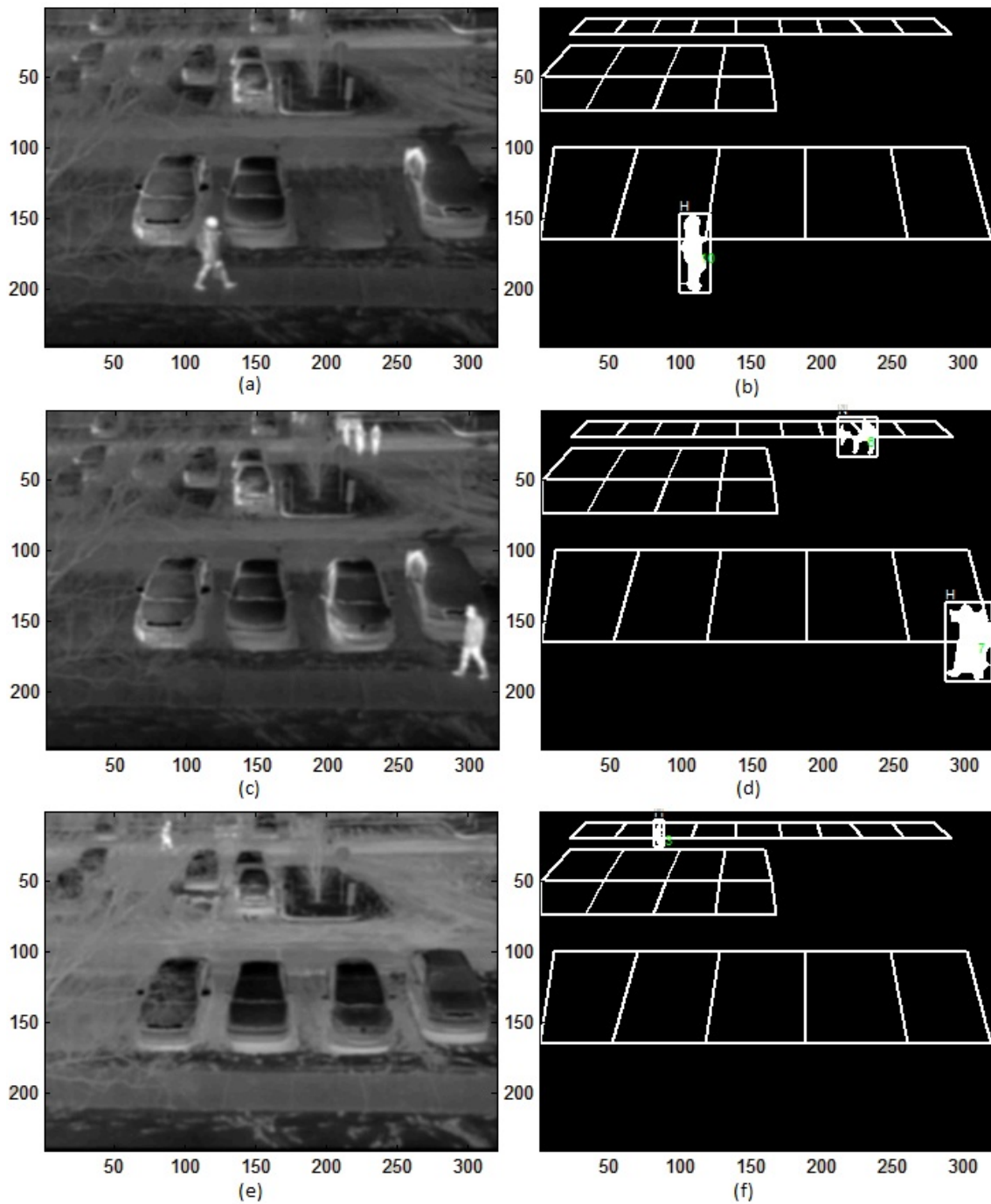


Figura 7.6: Segmentação sem restrição de fluxo e classificação de humanos em três diferentes cenas do banco de dados: (a) cena original e (b) resultado da segmentação da cena em (a); (c) cena original e (d) resultado da segmentação da cena em (c); (e) cena original e (f) resultado da segmentação da cena em (e).

tem movimento da esquerda para a direita.

Existem outros objetos que passam pela cena que não são identificados ou classificados como não identificados. Dentre estes objetos estão ciclistas, grupos de pessoas e animais. A Figura 7.6(d) mostra a segmentação e classificação de um grupo de humanos nessa classe.

Os resultados numéricos da classificação dos objetos de interesse segmentados sem

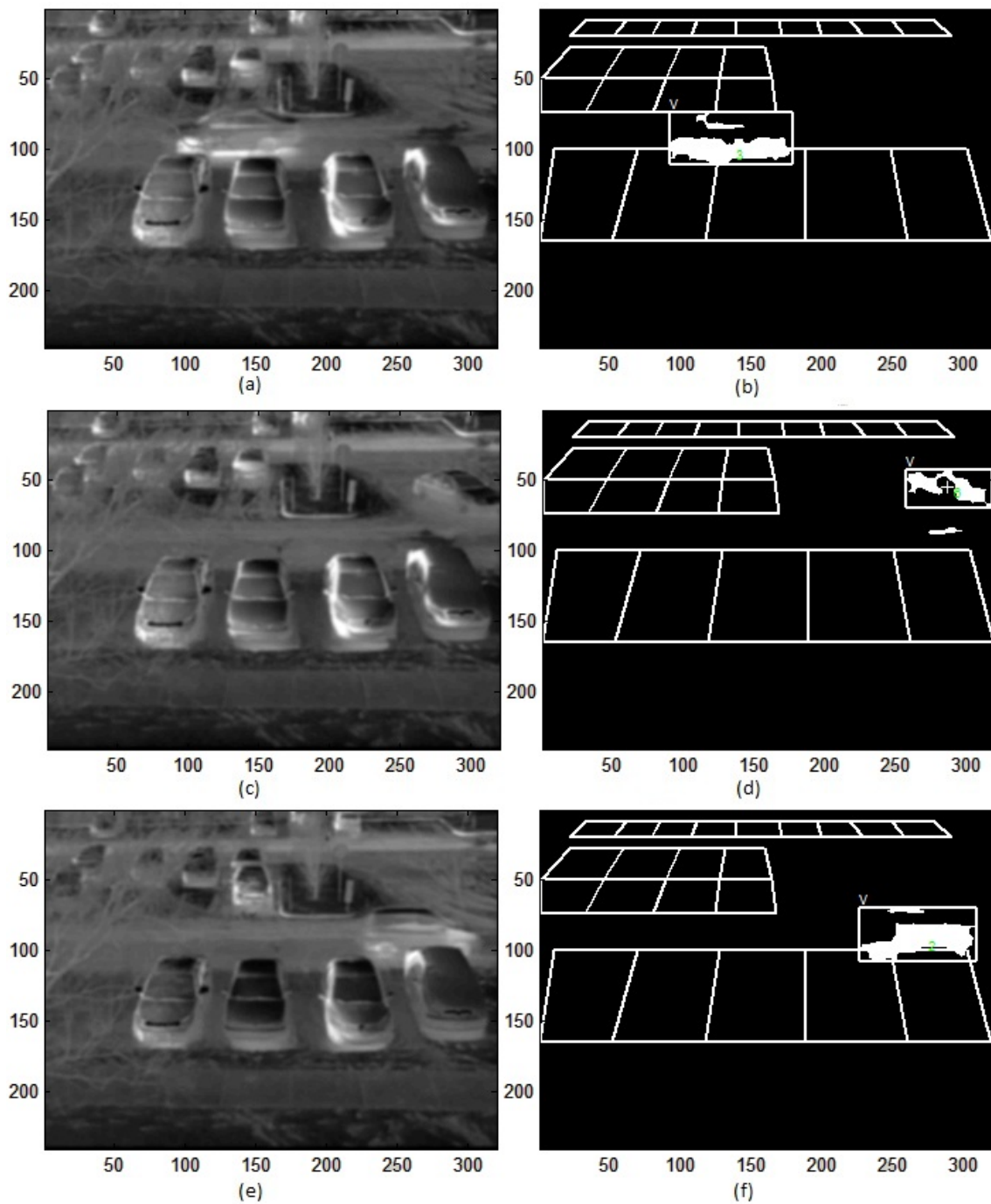


Figura 7.7: Segmentação sem restrição de fluxo e classificação de veículos em três diferentes cenas do banco de dados: (a) cena original e (b) resultado da segmentação da cena em (a); (c) cena original e (d) resultado da segmentação da cena em (c); (e) cena original e (f) resultado da segmentação da cena em (e).

restrição ao fluxo são apresentados na Tabela 7.1. Dos 265 objetos pertencentes a classe humano o sistema foi capaz de classificar corretamente 245 objetos sendo que 20 humanos ou foram classificados como veículos ou como não identificados. Isso nos dá uma taxa de 92,45% de acerto para objetos da classe humano. Dos 137 objetos pertencentes à classe veículo o sistema foi capaz de classificar corretamente 119 objetos sendo que 21 veículos ou foram classificados como humanos ou como não identificados. Isso nos dá uma taxa

Classe	N. total de obj.	N. de obj. classificados corretamente	N. de obj. não classificados corretamente	Taxa de acerto
Humano	265	245	20	92,45%
Veículo	137	116	21	84,77%
Total	402	361	41	89,8%

Tabela 7.1: Resultado da classificação usando a segmentação não restrita ao fluxo utilizando o banco de dados construído neste trabalho.

Classe	N. total de obj.	N. de obj. classificados corretamente	N. de obj. não classificados corretamente	Taxa de acerto
Humano	183	146	37	79,88%
Veículo	58	51	7	88,93%
Total	241	197	44	82,74%

Tabela 7.2: Resultado da classificação e segmentação obtidos em [Maadi e Maldague 2007]. Método e banco de dados usados naquele trabalho.

de 84,77% de acerto para objetos da classe veículo. A taxa de acerto global do sistema para a classificação de objetos nas duas classes, humano e veículo, foi 89,8%.

Os resultados obtidos em diferentes situações mostram que a capacidade do algoritmo em segmentar e classificar objetos em movimento é bem satisfatória. Isso mostra que o método proposto é adequado para realizar segmentação em imagens infravermelhas capturadas com uma câmera estacionária em um ambiente externo. Além disso, mostra também que os critérios usados para classificação dos objetos segmentados também é adequado.

Os resultados numéricos obtidos em [Maadi e Maldague 2007] utilizando o método e banco propostos por aquele trabalho são apresentados na Tabela 7.2.

Uma observação que deve ser feita neste ponto é que não usamos o banco de dados usado em [Maadi e Maldague 2007] e [Maadi e Maldague 2006] para testar todo o nosso sistema. Só executamos testes nesse banco em alguns vídeos pois não tivemos acesso ao banco completo e além disso o referido banco não possui eventos suspeitos. Por isso construímos nosso próprio banco de dados. Assim não realizamos uma comparação direta entre os métodos. Entretanto, embora esta comparação direta não tenha sido feita, algumas análises podem ser feitas.

Existem diversas dificuldades encontradas nas cenas do banco de dados usado neste trabalho quando comparado às cenas do banco de dados usado em [Maadi e Maldague 2007] e [Maadi e Maldague 2006]. Primeiramente o banco de dados usado neste trabalho é maior que o banco de dados usado em [Maadi e Maldague 2007] (402 e 241 objetos de

interesse respectivamente). Além disso, parte de uma árvore aparece no lado esquerdo da cena e ocasionalmente a presença de vento no ambiente balança os galhos desta árvore. Entretanto, este movimento não é detectado durante o processo de segmentação (isto indica a robustez do método proposto). Outra dificuldade encontrada nos vídeos é a presença de uma curva em um dos caminhos possíveis para os veículos. Neste caminho, os veículos passam devagar nesta curva e em seguida atrás da árvore (ou o sentido oposto, atrás da árvore em seguida devagar na curva). Por causa disso, veículos que passam por este caminho (atrás da árvore/curva ou curva/atrás da árvore) algumas vezes são classificados de maneira incorreta usando os critérios propostos em [Maadi e Maldague 2006]. Os critérios propostos no Capítulo 4 melhoram a classificação nestes cenários. Mesmo com estas dificuldades extras presente no nosso banco conseguimos ter um resultado melhor no geral em comparação com os resultados apresentados em [Maadi e Maldague 2007] (89,8% a 82,74% respectivamente). Isso se deve à alteração nos critérios de classificação e a um ajuste mais fino dos parâmetros que são definidos estaticamente.

7.3.2 Segmentação restrita ao fluxo

Mesmo com este bom resultado obtido da segmentação e classificação dos objetos existem situações onde objetos não são classificados corretamente. Por exemplo, quando um humano está no topo da cena ficando assim bem distante da câmera, ele aparece bem pequeno no vídeo e com isso a sua segmentação ficará comprometida. Isto se dá pelo fato de existir um número muito maior de *pixels* estacionários influenciando no cálculo dos limiares dinâmicos usados no processo de subtração de plano de fundo (este cenário foi apresentado na Seção 4.3). A não classificação de alguns destes objetos influi no módulo de gerenciamento de eventos pois o resultado do módulo de segmentação alimenta o módulo de gerenciamento de eventos. Se o módulo de segmentação fornecer resultados errados (por exemplo, não identificar/classificar um humano) o módulo de gerenciamento de eventos também fornecerá resultados errados. Considere um exemplo em que uma ação suspeita está ocorrendo no estacionamento. Se o humano que pratica esta ação não for corretamente segmentado pelo módulo de segmentação o módulo de gerenciamento de eventos não conseguirá reconhecer a ação (evento suspeito) praticada por este humano.

Visando obter melhores resultados em situações como a que acabamos de descrever, propomos na Seção 4.3 a utilização da análise de fluxo combinada com o método dinâmico de segmentação de plano de fundo. Na Figura 4.6 podemos ver duas situações que ilustram o problema e como a análise de fluxo pode favorecer o método proposto. Além disso, na Tabela 7.3 podemos ver também o resultado numérico da segmentação e classificação dos objetos de interesse utilizando a segmentação dinâmica restrita ao fluxo. A taxa de acerto para humanos subiu de 92,45% para 94,34%, a taxa de acerto para veículos subiu de 84,77% para 88,32% e o total subiu de 89,80% para 91,33%.

Classe	N. total de obj.	N. de obj. classificados corretamente	N. de obj. não classificados corretamente	Taxa de acerto
Humano	265	250	15	94,34%
Veículo	137	121	16	88,32%
Total	402	371	31	91,33%

Tabela 7.3: Resultado da classificação usando a segmentação restrita ao fluxo utilizando o banco de dados construído neste trabalho.

A Figura 7.8 mostra um gráfico comparando a taxa de acerto das três propostas de segmentação e classificação separadas em humanos, veículos e total. Estes resultados mostram que o método proposto tem potencial para realizar eficientemente a tarefa de vigilância de um estacionamento externo. Além disso, podemos falar que o nosso método é melhor que o método usado em [Maadi e Maldague 2007] pelo fato de que o nosso banco possui maiores dificuldades (árvore e curva presentes na cena) para a segmentação e o monitoramento que o banco usado em [Maadi e Maldague 2007] e mesmo assim ele apresentou resultados melhores (humano e total) ou próximos (veículo) aos resultados obtidos em [Maadi e Maldague 2007]. Não tivemos um melhor resultado para a classe veículo pois nossos vídeos foram gravados no inverno e os vídeos usados em [Maadi e Maldague 2007] foram gravados no verão. Essa diferença faz com que os veículos, feitos de aço, tenham uma temperatura mais baixa no inverno do que no verão, ficando assim mais parecidos com o plano de fundo do vídeo, enquanto que esta diferença de temperatura não acontece para os objetos da classe humano.

Nas duas próximas seções apresentaremos os resultados obtidos nos testes dos dois outros módulos que compõem este sistema.

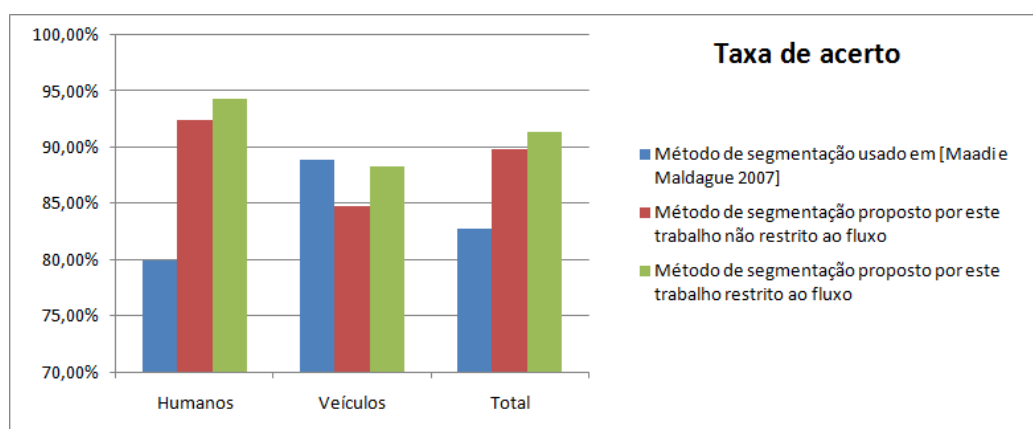


Figura 7.8: Comparação entre os três métodos de segmentação e classificação: método usado por [Maadi e Maldague 2007], método usado neste trabalho sem restrição de fluxo e método usado neste trabalho com restrição de fluxo. Podemos ver que o último método usando restrição de fluxo é o que apresentou melhor desempenho.

7.4 Módulo de monitoramento

Nessa Seção apresentamos alguns resultados de experimentos realizados a fim de testar a robustez do módulo de monitoramento proposto. O objetivo de se monitorar humanos e veículos em um estacionamento é conhecer a trajetória destes objetos e ser capaz de identificar interações entre eles.

Mesmo que o objeto tenha um modelo de movimento simples (velocidade constante), ou tenha uma aceleração, ou tenha um movimento linear ou não, o sistema é capaz de seguir este objeto ao longo de sua trajetória de forma satisfatória. As Figuras 7.9 e 7.10 mostram o resultado de um monitoramento em uma sequência de 15 quadros consecutivos. Na primeira figura temos os quadros originais e na segunda o respectivo resultado de monitoramento para cada quadro.

Estas figuras mostram a capacidade do nosso algoritmo de monitoramento em acompanhar e atualizar corretamente a posição de vários objetos simultaneamente. Um ponto fraco chamou nossa atenção durante a execução dos testes. O algoritmo é incapaz de monitorar objetos com velocidades muito altas. Isso ocorre por causa da taxa de aquisição de imagens que é de 5 quadros/segundo. A escolha deste valor foi explicada no início deste Capítulo.

Neste caso, em dois quadros consecutivos um mesmo objeto ocupa posições espaciais distantes. Desta forma os centros de massa dos objetos nos dois quadros consecutivos estarão distantes. Se esta distância for maior que o limiar definido no Capítulo 5, o sistema concluirá que são dois objetos diferentes atribuindo um rótulo diferente para cada um.

O sistema também é eficiente para gerir descontinuidade de movimentos. Como indicamos no Capítulo 5, o algoritmo somente trata os casos em que existe oclusão parcial e temporária. A Figura 7.11 ilustra o reconhecimento de uma oclusão parcial com um quadro de antecedência. No quadro seguinte, onde ocorre a oclusão, o objeto medido neste quadro é substituído pelo objeto previsto anteriormente. No final da oclusão o objeto previsto é substituído pelo objeto real novamente. Podemos ver na Figura 7.11 que o sistema gere de forma satisfatória as oclusões. A oclusão se deu quando um humano passou na frente de outro.

Os diferentes testes realizados também provam a eficácia do módulo de monitoramento para desenhar as trajetórias do centro de massa do objeto em toda a evolução espacial do mesmo pela cena. Cada trajetória indica o sentido do movimento e cobre praticamente todas as posições onde o objeto esteve na cena do momento que este começou a ser monitorado até o momento em que ele sai do campo de visão da câmera. A Figura 7.12 mostra dois humanos e um veículo com suas respectivas trajetórias por alguns quadros do vídeo.

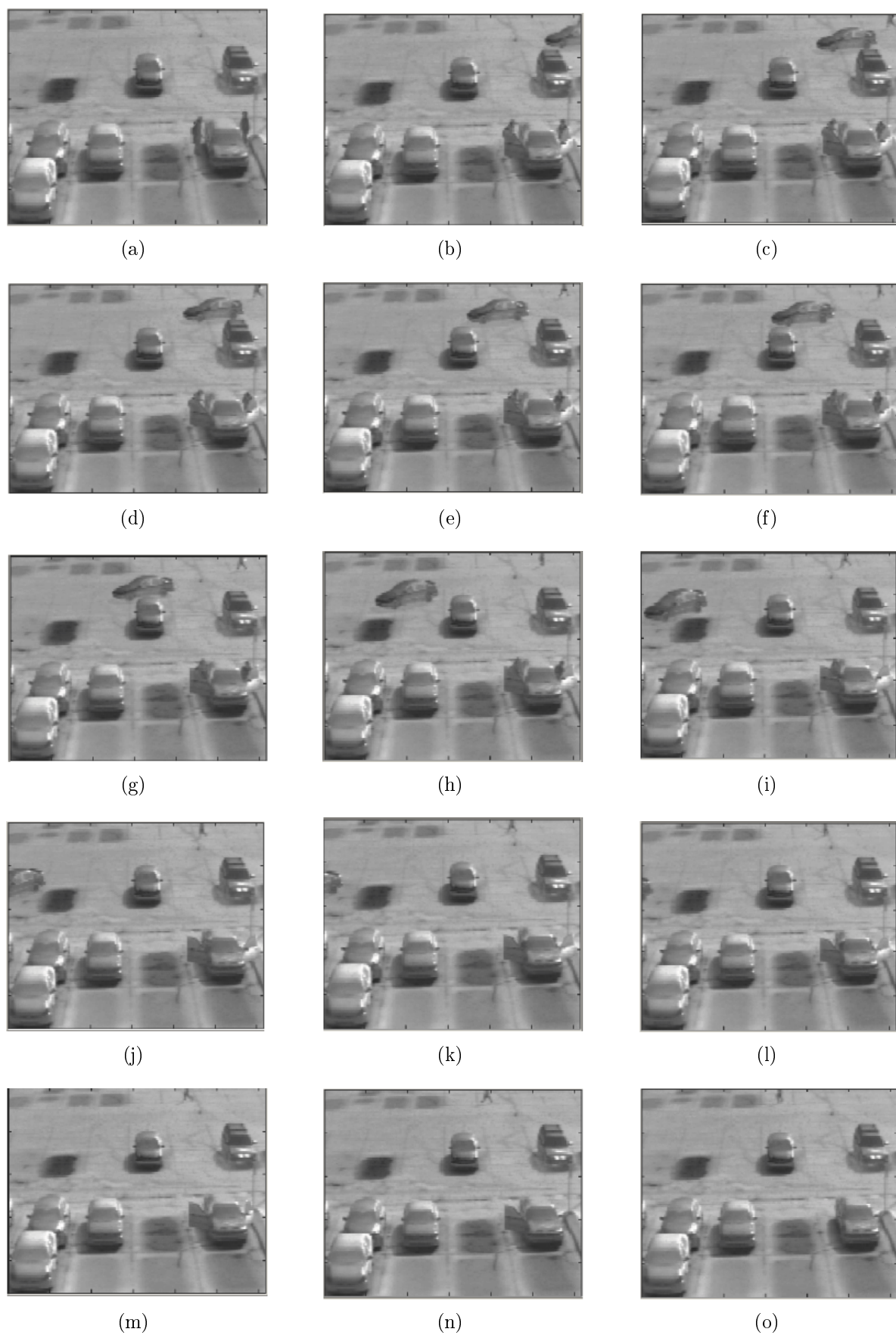


Figura 7.9: Exemplo de uma seqüência de imagens originais: os quadros (a) a (o) são consecutivos e contém diversos objetos em movimento.

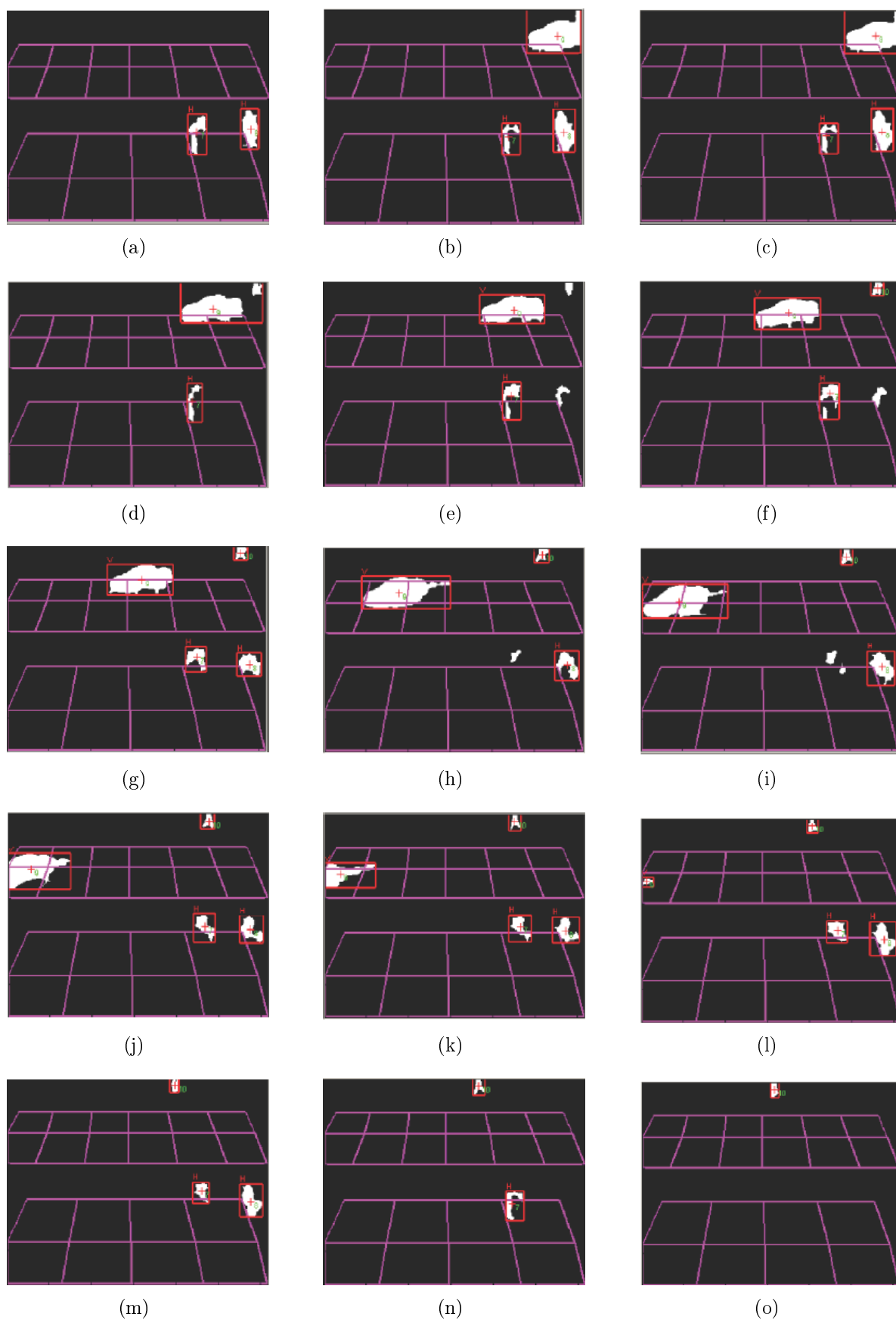


Figura 7.10: Exemplo de resultado de monitoração: os quadros (a) a (o) são uma sequência de quadros binários que representam o monitoramento de diversos objetos simultaneamente.

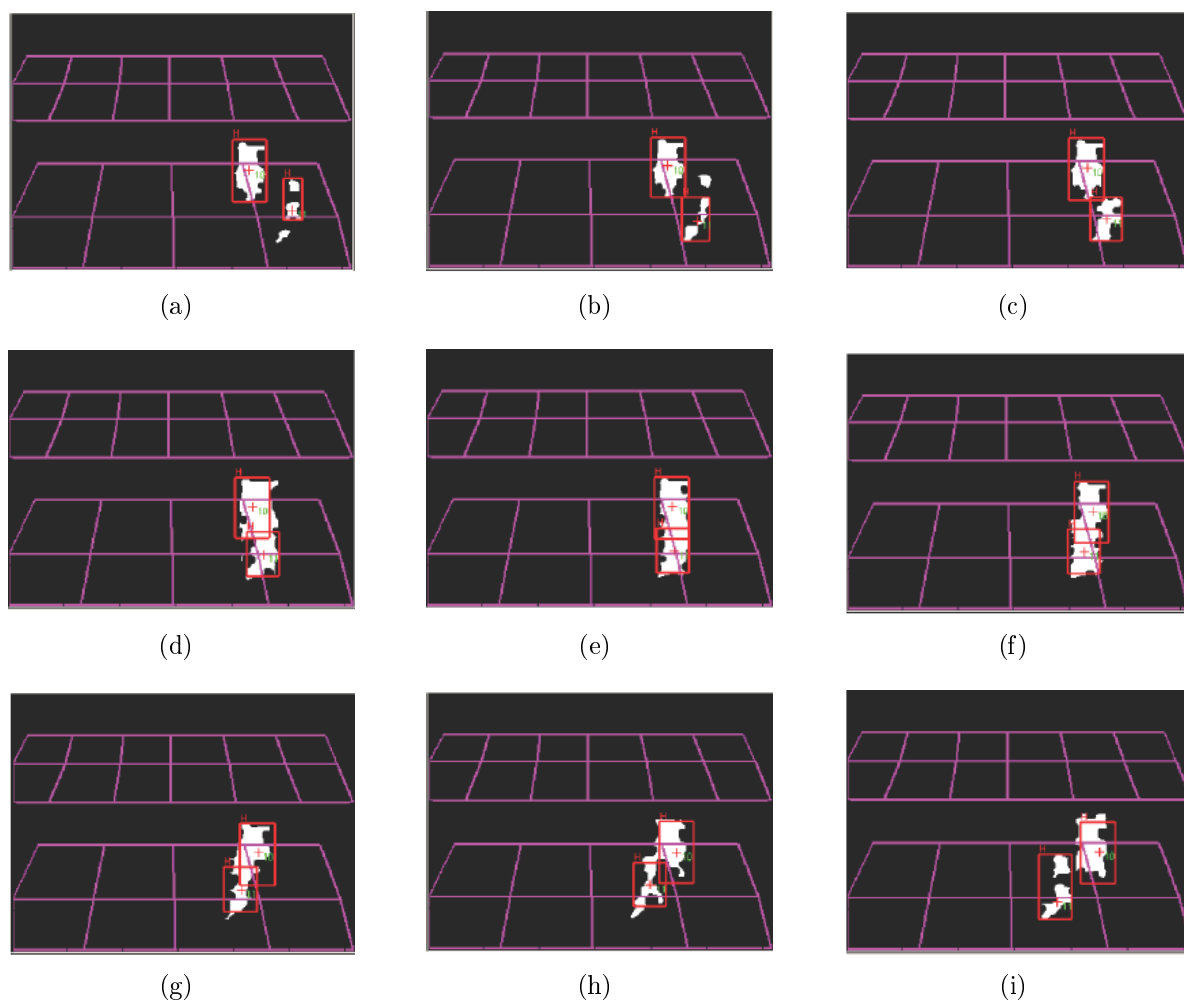


Figura 7.11: Exemplo de uma sequência de imagens que ilustra o monitoramento de dois humanos durante uma oclusão parcial e temporária.

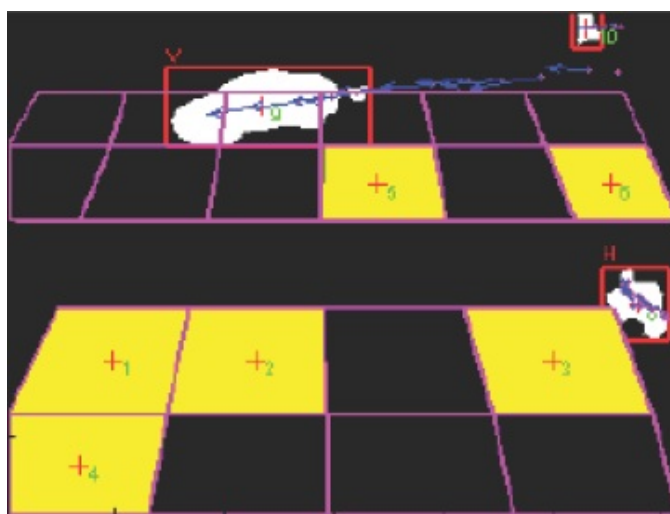


Figura 7.12: Exemplo de trajetórias calculadas durante a monitoração de objetos. Cada trajetória contém as informações de posição e sentido de movimento do centro de massa do objeto.

7.5 Módulo de gerenciamento de eventos

Esta Seção ilustra alguns resultados obtidos na avaliação do método proposto para identificar eventos de interesse na área monitorada. Baseado no algoritmo de classificação que foi proposto no Capítulo 6, mostramos primeiramente resultado de reconhecimento de eventos de um único autor, em seguida de múltiplos autores e por fim de eventos suspeitos.

7.5.1 Eventos de um único ator

Durante o monitoramento espaço-temporal do objeto, o algoritmo calcula o módulo da velocidade de deslocamento do dentro de massa do objeto para detectar uma eventual parada deste objeto. A Figura 7.13 mostra um humano que entra na cena e caminha até parar em frente um veículo estacionado. Neste instante graças a baixa velocidade de deslocamento do seu centro de massa (abaixo do limiar fixado) o algoritmo identifica o evento "Parar caminha" para este humano.

Em caso de parada durante vários quadros consecutivos, o objeto parado fará parte do *background*. No entanto o algoritmo salvaguarda todas as informações necessárias para efetuar a retomada do monitoramento deste objeto quando este voltar a se movimentar.



Figura 7.13: Exemplo da detecção do evento "Parar caminha" para a classe *humano*.



Figura 7.14: Exemplo da detecção do evento "Caminhar novamente" para a classe *humano*.

O algoritmo consegue monitorar novamente um objeto que estava parado e voltou a se movimentar como é mostrado na Figura 7.14 que é a continuação do cenário ilustrado pela Figura 7.13. Assim que o humano retoma a caminha, o algoritmo consegue fazer a correspondência entre o novo movimento e o objeto correto que havia parado anteriormente. Isso quer dizer que a técnica de monitoramento consegue identificar o evento "Retomar caminha" fazendo a correspondência correta entre o objeto que entrou em movimento e objeto que havia parado.

Para o reconhecimento de eventos envolvendo um veículo, o algoritmo também não encontrou dificuldade em reconhecê-los. Os resultados foram satisfatórios em todos os testes realizados. Exemplos deste tipo evento podem ser encontrados nas Figuras 7.15 e 7.16. A Figura 7.15 ilustra um evento de "Estacionar veículo" realizado por um veículo que estava sendo monitorado pelo sistema e parou em uma zona reservada para estacionamento. A parada do veículo é reconhecida porque a velocidade de deslocamento de seu centro de massa é menor que o limiar estabelecido. Como a caixa delimitadora do objeto cobre a maior parte da área reservada para o estacionamento o sistema atualiza a posição do veículo (esta estacionado naquela vaga específica) e reconhece o evento "Estacionar veículo".



Figura 7.15: Exemplo da detecção do evento "Estacionar veículo" para a classe *veículo*.

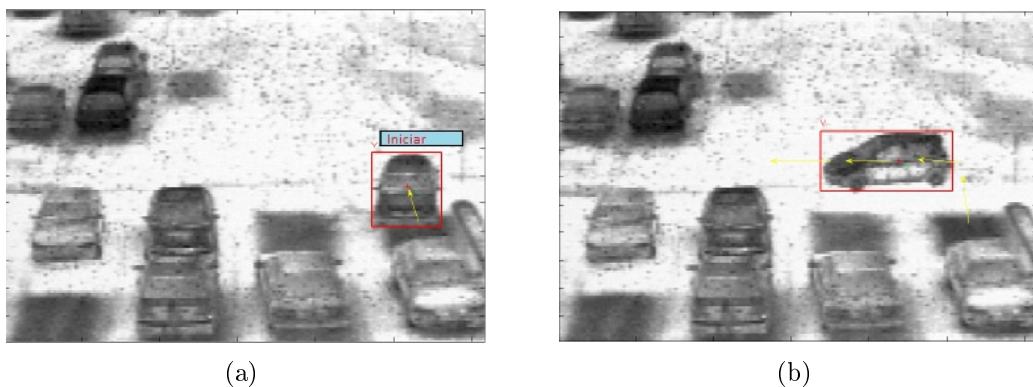


Figura 7.16: Exemplo da detecção do evento "Iniciar" para a classe *veículo*. Cena gravada sob forte chuva.

Cada zona reservada para estacionamento possui um rótulo contendo o seu número, suas coordenadas espaciais e seu estado atual: "livre" ou "ocupada". No caso de uma zona "ocupada" o algoritmo associa ao seu rótulo o vetor de características do veículo correspondente que está ocupando aquela zona de estacionamento. Quando o sistema é iniciado, um operador humano marca quais zonas estão ocupadas, ou seja, onde existem veículos estacionados.

Baseados nestes princípios, o algoritmo também é capaz de reconhecer um veículo que está deixando uma zona que ele ocupava. A Figura 7.16 mostra o reconhecimento do evento "Iniciar" de um veículo sob forte chuva. O veículo estava estacionado durante um tempo e agora esta saindo do estacionamento. Com o passar do tempo o sistema gerencia corretamente as zonas "Ocupadas" e "Livres" de acordo com que os veículos vão estacionado e/ou deixando o estacionamento.

7.5.2 Eventos de múltiplos atores

Resultados satisfatórios também foram obtidos para o reconhecimento dos eventos deste tipo. Estes eventos incluem algumas interações entre humanos e veículos os quais já foram classificados anteriormente.

Primeiramente, a Figura 7.17 mostra o reconhecimento de um evento "Entrar em um veículo". O reconhecimento deste evento começa quando o objeto humano e o veículo estacionado começam a se fundir e termina quando o humano desaparece completamente da cena. Neste último momento temos o reconhecimento completo de um evento do tipo "Entrar em um veículo".

A Figura 7.18 mostra o reconhecimento do evento de saída de um humano de dentro de um veículo estacionado. Esta evento normalmente é reconhecido logo depois que um evento de "Estacionar veículo" é reconhecido. Depois que o veículo estaciona, o algoritmo procura ao redor do veículo se existe algum objeto do tipo humano. Se tal objeto é encontrado, é feita uma análise de trajetória para decidir se este humano saiu do veículo que acabou de estacionar.



(a)



(b)



(c)

Figura 7.17: Exemplo da detecção do evento "Entrar em um veículo" envolvendo um humano e um veículo.

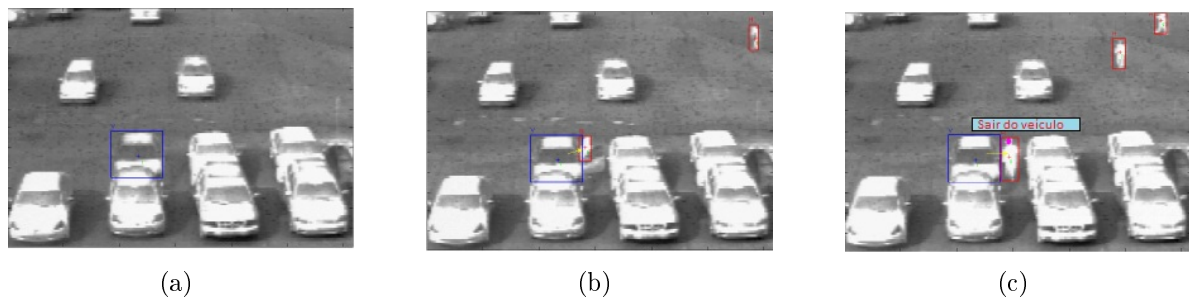


Figura 7.18: Exemplo da detecção do evento "Sair do interior do veículo" envolvendo um humano e um veículo.

7.5.3 Eventos suspeitos

Como assaltos são difíceis de acontecer no campus da Universidade de Laval, dificilmente testemunharíamos um evento desse durante as gravações. Por isso, decidimos que simulariamos o evento suspeito alvo de identificação deste trabalho. É claro que sabemos que para realizar uma validação mais precisa do sistema precisamos executar testes em cenas reais de assalto. Entretanto estas cenas encenadas já são suficientes para desenvolver e testar o sistema.

Dentre os 14 eventos do tipo C que foram gravados existem eventos realizados por diferentes atores, com veículo estacionado em diferentes lugares e diferentes maneiras do humano entrar dentro do veículo (usando uma janela que foi deixada aberta, simulando o "arrombamento" da porta do veículo e simulando a "quebra" da janela do veículo). Isto mostra a diversidade dos vídeos do nosso banco de dados.

Como mencionamos anteriormente, executamos teste e desenvolvemos procedimentos somente para o evento suspeito C. Mas a estrutura que criamos para o reconhecimento deste evento é adaptável para reconhecer outros eventos que possam também ser divididos em sub-eventos. A Figura 7.19 mostra os quatro sub-eventos que compõem o evento suspeito do tipo C sendo reconhecidos. Neste exemplo um humano se aproxima de um veículo estacionado no canto inferior da cena, fica observando o veículo e o ambiente que o cerca por aproximadamente 5 segundos, em seguida ele simula o arrombamento da porta do veículo abrindo a porta do mesmo e pegando uma mochila que estava no banco do motorista. Por fim o humano com a mochila nas costas sai correndo em direção ao lado direito da cena até sair do campo de visão da câmera.

Como fizemos na Seção anterior podemos dividir os resultados obtidos no reconhecimento de eventos suspeitos também em duas etapas. A primeira parte sendo o reconhecimento de eventos suspeitos a partir dos objetos segmentados com a técnica dinâmica de subtração de plano de fundo não restrita ao fluxo de movimento e a segunda parte sendo o reconhecimento de eventos suspeitos a partir dos objetos segmentados com a mesma técnica de subtração de plano de fundo com a restrição de análise das regiões onde temos fluxo de movimento.

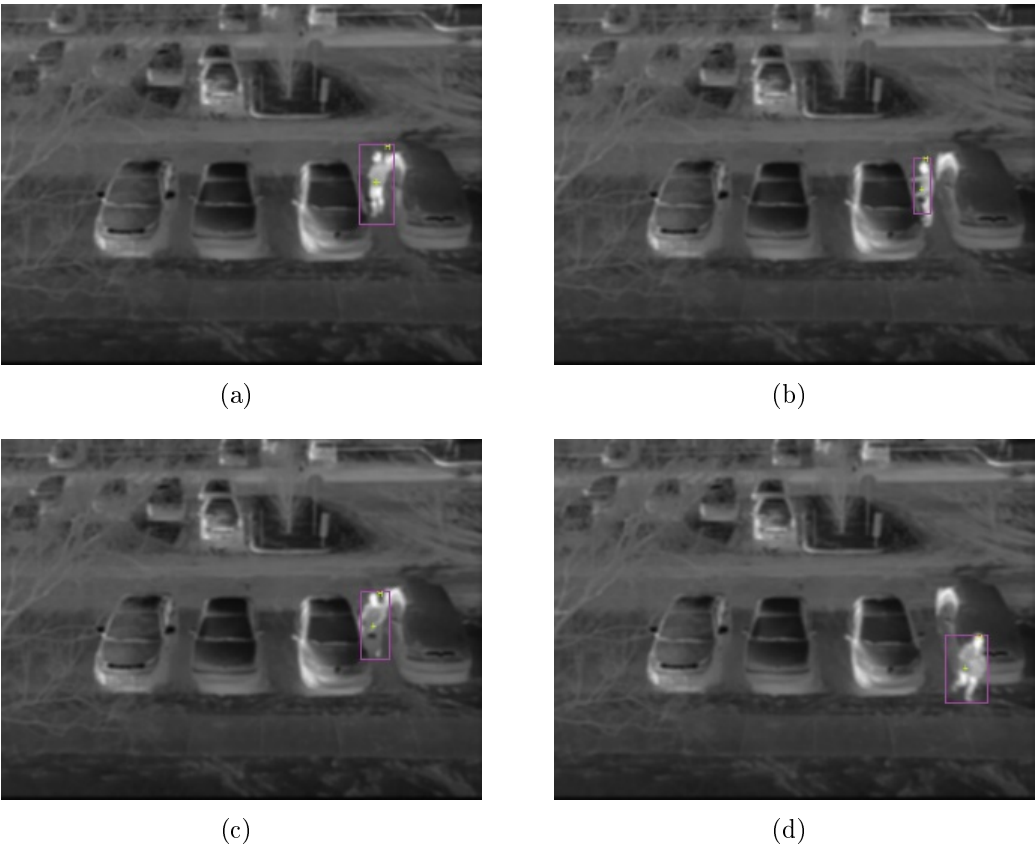


Figura 7.19: Evento suspeito alvo da análise deste projeto. Na seqüência os 4 sub-eventos que o compõem: (a) c1 - homem se aproximando do veículo estacionado, (b) humano ficando próximo deste veículo por um certo tempo, (c) humano entrando dentro do veículo e (d) humano se afastando do veículo. (Utilizando segmentação sem restrição ao fluxo).

	Tipo	Total de eventos	Número de eventos reconhecidos	Número de falsos negativos	Número de falsos positivos	
	C	14	11	3	2	

Tabela 7.4: Resultado da classificação de eventos suspeitos usando a segmentação restrita ao fluxo utilizando o banco de dados construído neste trabalho.

A Tabela 7.4 mostra os resultados numéricos do reconhecimento destes eventos na primeira etapa do projeto, ou seja, do reconhecimento de eventos a partir dos objetos segmentados sem a restrição de fluxo. Dos 14 eventos encenados existentes no banco, o sistema conseguiu reconhecer corretamente 11 deles sendo que 3 deles não foram reconhecidos. Além disso, 2 eventos que não havíamos considerado como eventos suspeitos foram reconhecidos pelo sistema e consideramos eles como falsos positivos.

Durante o desenvolvimento deste projeto notamos que estes três eventos que não foram identificados têm algo em comum. Em ambos o local onde o evento suspeito estava ocorrendo era no topo da cena, ou seja, mais distante da câmera do que eventos que ocorram na parte inferior da cena e conseqüentemente os objetos nesta região no topo são

	Tipo	Total de eventos	Número de eventos reconhecidos	Número de falsos negativos	Número de falsos positivos	
	C	14	13	1	3	

Tabela 7.5: Resultado da classificação de eventos suspeitos usando a segmentação restrita ao fluxo utilizando o banco de dados construído neste trabalho.

menores que os objetos na parte inferior.

Motivados pela idéia do fluxo proposta na Seção 4.3, fizemos os testes usando também os objetos segmentados pelo método dinâmico de segmentação proposto restrito ao fluxo de movimento para realizar o reconhecimento de eventos suspeito. Neste caso o fluxo restringe a análise e o cálculo dos coeficientes dinâmicos somente às áreas onde existe movimento e assim os pequenos objetos em movimento são mais facilmente detectados por não sofrerem influência dos demais *pixels* estacionários da cena.

Como foi mostrado na Seção anterior, a taxa de acerto na segmentação de objetos melhorou. Por exemplo, a taxa de acerto de reconhecimento de objetos da classe humano aumentou de 92,34% para 94,45%. Com isso a taxa de reconhecimento de eventos também aumentou. Anteriormente eram reconhecidos somente 11 eventos. Agora utilizando os objetos obtidos com a restrição de fluxo são reconhecidos 13 eventos. Esses novos resultados podem ser vistos na Tabela 7.5

Note também que o número de falsos positivos subiu de 2 para 3. Isto mostra que o método fica mais sensível quando utilizamos análise de fluxo. Entretanto este resultado não é ruim. Ora, em um sistema de vigilância o que queremos é reconhecer todos os eventos suspeitos, se um evento que não é suspeito for classificado como tal, o operador humano pode decidir se tal evento é ou não suspeito. Pelo menos o sistema chamou a atenção do operador. O importante em um sistema automático de vídeovigilância é que ele seja capaz de identificar todos os eventos suspeitos. Assim, a presença de falsos positivos é menos importante que a presença de falsos negativos pois, se um evento suspeito não for identificado pelo sistema, o sistema falha na sua principal função: vigilância.

7.6 Considerações finais

Neste Capítulo apresentamos e analisamos alguns resultados experimentais realizados ao longo do desenvolvimento deste projeto. Os resultados mostram que o sistema proposto é robusto para realizar a segmentação e monitoramento de objetos em movimento em vídeos infravermelhos capturados em um ambiente externo com uma câmera estacionária. Os resultados mostram também que quando utilizamos o método de segmentação restrito ao fluxo de movimento conseguimos melhores resultados. Isso se deve ao fato de analisarmos somente regiões onde existe movimento no momento da segmentação.

Dessa forma regiões onde não existe movimento não influenciaram no processo (estamos interessados somente em objetos em movimento na cena). Os resultados parciais obtidos com a segmentação não restrita ao fluxo foram publicados em [Fernandes et al. 2011].

Por fim os resultados mostram também que o sistema é adequado para o reconhecimento de eventos de forma explícita. Embora tenhamos realizado testes somente para um tipo de evento suspeito, o sistema proposto é adequado para reconhecer de forma explícita qualquer evento que possa ser dividido em sub-eventos e que possa ser descrito por características como as usadas neste trabalho (velocidade, posição, entre outras). Na próxima Seção apresentaremos nossas considerações finais e perspectivas futuras.

Capítulo 8

Conclusões

É notório o aumento do sentimento de insegurança vivido pela população nos dias atuais. Juntamente com esse crescimento também podemos notar o crescente avanço tecnológico nos diversos setores da sociedade. Devido a estes dois fatores, sistemas digitais de segurança pública e privada têm surgido constantemente. Com isso, se faz necessário o estudo de técnicas computacionais para catalisar o desenvolvimento de tais sistemas.

Um dos sistemas mais comuns é o sistema de vídeovigilância que utiliza uma ou mais câmeras para realizar o monitoramento de uma área externa ou interna. Este monitoramento é feito visando garantir a segurança, pessoal e patrimonial, das pessoas que transitam pelo local e seus arredores. Um sistema como este pode ser dividido basicamente em três partes: (i) um módulo de detecção de movimento, (ii) um módulo de monitoramento de objetos e (iii) um módulo de gerenciamento de eventos (interações entre os objetos na cena).

Este trabalho visou o desenvolvimento de técnicas de processamento de vídeo que possam ser usadas na construção de um sistema de vídeovigilância. Mais especificamente um sistema de monitoramento de um estacionamento externo utilizando uma câmera estacionária que captura imagens infravermelhas. Um sistema como este necessita operar 24 horas por dia. Durante o período da noite, câmeras normais não conseguiriam captar o que está acontecendo na cena pela falta de luz. Isso não ocorre quando utilizamos imagens infravermelhas pois elas captam a energia infravermelha emitida pelos corpos. Tal energia é emitida por humanos e veículos mesmo sem a presença da luz visível.

No primeiro módulo, utilizamos uma técnica de subtração de plano de fundo que se adapta dinamicamente às mudanças de iluminação ocorridas na cena (exemplo: surgimento repentino do sol). Além disso, neste módulo também utilizamos análise de fluxo para restringir o cálculo dos parâmetros dinâmicos somente as regiões onde existe movimento na cena. Isso foi feito para aumentar a qualidade da segmentação de objetos pequenos na cena. No segundo módulo utilizamos um ciclo de previsão e correção que consegue monitorar vários objetos simultaneamente como foi mostrado anteriormente. Este módulo também consegue lidar com oclusões e impressões térmicas de forma satis-

fatória. O último módulo é o módulo onde identificamos e interpretamos interações entre os objetos monitorados. Este módulo é super importante pois é nele que o sistema identificará ações suspeitas que ocorram na área monitorada e é claro que a identificação destas ações suspeitas é o principal objetivo de um sistema como este. No desenvolvimento deste trabalho nos propomos a reconhecer eventos simples que ocorram no estacionamento como entrada e saída de veículos. Além destes eventos simples foi proposto o reconhecimento do evento suspeito C: "suspeita de roubo de objeto(s) do interior de um veículo estacionado em uma posição x por um humano".

Os resultados experimentais mostram que a técnica de subtração de plano de fundo dinâmica restrita ao fluxo é adequada para a segmentação de objetos em movimento em cenas externas capturadas com uma câmera estacionária infravermelha. Os resultados obtidos e analisados neste trabalho mostram também que a nossa proposta é melhor que a proposta apresentada em [Maadi e Maldague 2007]. Realizamos ainda a identificação de eventos suspeitos utilizando imagens infravermelhas. Os resultados da identificação de eventos suspeitos mostram que a estrutura proposta é adequada para o reconhecimento explícito de eventos suspeitos. Por causa da restrição de tempo desenvolvemos e realizamos testes somente para o evento suspeito C. Entretanto o sistema pode ser adaptado para que outros eventos também sejam identificados. Tudo isto mostra que as técnicas estudadas aqui têm o potencial para serem usadas para criar um sistema de vídeovigilância comercial que ajude na atividade de vigilância.

8.1 Trabalhos Futuros

Para trabalhos futuros, sugerem-se alguns tópicos relacionados às possíveis extensões e validação dos métodos apresentados, tais como:

- utilizar vídeos reais de assaltos a veículos em estacionamento;
- utilizar conhecimento de especialistas em segurança para a modelagem dos eventos suspeitos e de seus sub-eventos;
- identificar outros eventos suspeitos que possam ocorrer em um estacionamento como um humano abandonar uma mochila em algum lugar do estacionamento e depois este humano sair da área monitorada (este cenário representa o abandono de um explosivo);
- utilizar outras características para descrever os objetos;
- utilizar perspectiva para classificação de objetos em diferentes posições da cena;
- calcular automaticamente o nível de detecção de fluxo m de tal forma a maximizar a qualidade da segmentação.

Referências Bibliográficas

- [Amici et al. 2010] Amici, S. D., Sanna, A., Lamberti, F., e Pralio, B. (2010). A Wii remote-based infrared-optical tracking system. *Entertainment Computing*, 1(3-4):119–124.
- [Arandjelovic e Cipolla 2004] Arandjelovic, O. e Cipolla, R. (2004). An illumination invariant face recognition system for access control using video. In *Proc. British Machine Vision Conference*.
- [Avdelidis et al. 2003] Avdelidis, N., Hawtin, B., e Almond, D. (2003). Transient thermography in the assessment of defects of aircraft composites. *NDT and E International*, 36(6):433 – 439.
- [Balaras e Argiriou 2002] Balaras, C. e Argiriou, A. (2002). Infrared thermography for building diagnostics. *Energy and Buildings*, 34(2):171–183. TOBUS - an European method and software for office building refurbishment.
- [Barreira e Freitas 2007] Barreira, E. e Freitas, V. P. d. (2007). Evaluation of building materials using infrared thermography. *Construction and Building Materials*, 21(1):218–224.
- [Baumberg 1995] Baumberg, A. M. (1995). *Learning deformable models for tracking the human motion*. PhD thesis, The University of Leeds.
- [Benkhalil et al. 2000] Benkhalil, A., Ipson, S., e Booth, W. (2000). Real-time video surveillance system using a field programmable gate array. *International Journal of Imaging Systems and Technology*, 11(2):130–137.
- [Chen et al. 2010] Chen, L., Liang, X., e Wang, X. (2010). An effective Traffic Monitoring video segmentation method. In *2010 International Conference on Computer Design and Applications (ICCDA)*, volume 5, pp. 593–597.
- [Dai et al. 2007] Dai, C., Zheng, Y., e Li, X. (2007). Pedestrian detection and tracking in infrared imagery using shape and appearance. *Computer Vision and Image Understanding*, 106(2-3):288–299. Special issue on Advances in Vision Algorithms and Systems beyond the Visible Spectrum.
- [Deng et al. 2010] Deng, H., Liu, J., e Chen, Z. (2010). Infrared small target detection based on modified local entropy and EMD. *Chinese Optics Letters*, 8:24–28.
- [Diamantopoulos e Spann 2005] Diamantopoulos, G. e Spann, M. (2005). Event detection for intelligent car park video surveillance. *Real-Time Imaging*, 11:233–243.

- [Fernandes et al. 2011] Fernandes, H., Maldague, X., Batista, M. A., e Barcelos, C. Z. (2011). Suspicious event recognition using infrared imagery. In *2011 IEEE International Conference on Systems, Man, and Cybernetics, October 9-12, 2011, Anchorage, Alaska*.
- [Fernández et al. 2010] Fernández, C., Baiget, P., Roca, F. X., e González, J. (2010). Determining the best suited semantic events for cognitive surveillance. *Expert Systems with Applications*, 38:4068–4079.
- [Fernández-Caballero et al. 2010] Fernández-Caballero, A., Castillo, J. C., Serrano-Cuerda, J., e Maldonado-Bascón, S. (2010). Real-time human segmentation in infrared videos. *Expert Systems with Applications*, 38:2577–2584.
- [Foresti et al. 2004] Foresti, G. L., Micheloni, C., e Snidaro, L. (2004). Event classification for automatic visual-based surveillance of parking lots. In *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004: Cambridge, England, UK*, volume 3, pp. 314–317.
- [Galata et al. 2011] Galata, A., Johnson, N., e Hogg, D. (2011). Learning variable-length Markov. *Computer Vision and Image Understanding models of behavior*, 81(3):398–413.
- [Grimson et al. 1998] Grimson, W. E. L., Stauffer, C., Romano, R., e Lee, L. (1998). Using adaptive tracking to classify and monitor activities in a site. In *1998 Conference on Computer Vision and Pattern Recognition (CVPR'98), June 23-25, 1998, Santa Barbara, CA, USA*, pp. 22–29. IEEE Computer Society.
- [Haering et al. 2008] Haering, N., Venetianer, P., e Lipton, A. (2008). The evolution of video surveillance: an overview. *Machine Vision and Applications*, 19:279–290.
- [Hamdoun et al. 2008] Hamdoun, O., Moutarde, F., Stanciulescu, B., e Steux, B. (2008). Person re-identification in multi-camera system by signature based on interest point descriptors collected on short video sequences. In *Second ACM/IEEE International Conference in Distributed Smart Cameras. September 2008*, pp. 1–6.
- [Haritaoglu et al. 1998] Haritaoglu, I., Harwood, D., e Davis, L. (1998). W4: Who, when, where, what: A real-time system for detecting and tracking people. In *IEEE 3rd International Conference on Face and Gesture Recognition, April 14-16, 1998, Nara, Japan*, pp. 222–227.
- [Huang et al. 2008] Huang, K., Wang, L., Tan, T., e Maybank, S. (2008). A real-time object detecting and tracking system for outdoor night surveillance. *Pattern Recognition*, 41(1):432–444.
- [Hudson 1969] Hudson, R. D. J. (1969). *Infrared system engineering*. Wiley-Interscience, New York and London, 1 edition.
- [Ibarra-Castanedo 2005] Ibarra-Castanedo, C. (2005). *Quantitative subsurface defect evaluation by pulsed phase thermography: depth retrieval with the phase*. PhD thesis, Laval University.
- [Ichihashi et al. 2009] Ichihashi, H., Notsu, A., Honda, K., Katada, T., e Fujiyoshi, M. (2009). Vacant parking space detector for outdoor parking lot by using surveillance camera and FCM classifier. In *2009 IEEE International Conference on Fuzzy Systems. FUZZ-IEEE 2009. ICC Jeju, Jeju Island, Korea. August 20-24, 2009*, pp. 127–134.

- [Ivanov et al. 1999] Ivanov, Y., Stauffer, C., Bobick, A., e Grimson, W. (1999). Video surveillance of interactions. In *Second IEEE Workshop on Visual Surveillance, 1999. (VS'99)*, pp. 82–89.
- [Javed et al. 2009] Javed, O., Rasheed, Z., Alatas, O., e Shah, M. (2009). KNIGHT/spl trade/: a real time surveillance system for multiple and non-overlapping cameras. In *2003 International Conference on Multimedia and Expo - Volume 1 (ICME '03). July 06-09, 2009. Baltimore, MD, USA*, volume 1, pp. 649–652.
- [Lavee et al. 2009] Lavee, G., Rivlin, E., e Rudzsky, M. (2009). Understanding video events: A survey of methods for automatic interpretation of semantic occurrences in video. In *IEEE Transactions on Systems, Man, and Cybernetics - Part C: applications and reviews*, volume 39, pp. 489–504.
- [Li et al. 2010] Li, J., Gong, W., Li, W., e Liu, X. (2010). Robust pedestrian detection in thermal infrared imagery using the wavelet transform. *Infrared Physics and Technology*, 53(4):267 – 273.
- [Lipton et al. 1998] Lipton, A. J., Fujiyoshi, H., e Patil, R. S. (1998). Moving target classification and tracking from real-time video. *4th IEEE Workshop on Applications of Computer Vision (WACV'98). October 19-21, 1998. Princeton, NJ , USA.*, 0:8.
- [Maadi e Maldague 2006] Maadi, A. E. e Maldague, X. P. V. (2006). Classifying tracked objects and their interactions from infrared imagery. In *Canadian Conference on Electrical and Computer Engineering, 2006. CCECE'06. May 7-10, 2006 - Ottawa Congress Centre, Ottawa, Canada.*, pp. 2194–2198.
- [Maadi e Maldague 2007] Maadi, A. E. e Maldague, X. P. V. (2007). Outdoor infrared video surveillance: A novel dynamic technique for the subtraction of a changing background of IR images. *Infrared Physics and Technology*, 49:261–265.
- [Maldague 2001] Maldague, X. P. V. (2001). *Theory and practice of infrared technology for nondestructive testing*. Wiley-Interscience, 1st edition.
- [Mandellos et al. 2011] Mandellos, N. A., Keramitsoglou, I., e Kiranoudis, C. T. (2011). A background subtraction algorithm for detecting and tracking vehicles. *Expert Systems with Applications*, 38:1619–1631.
- [Marinetti et al. 1997] Marinetti, S., Maldague, X., e M., P. (1997). Calibration procedure for focal plane array cameras and noise equivalent material loss for quantitative thermographic NDT. *Materials evaluation*, 55:407–412.
- [Masoud e Papanikolopoulos 2001] Masoud, O. e Papanikolopoulos, N. (2001). A novel method for tracking and counting pedestrians in real-time using a single camera. *IEEE Transactions on Vehicular Technology*, 50(5):1267–1278.
- [Meyer e Denzler 1997] Meyer, D. e Denzler, J. (1997). Model based extraction of articulated objects in image sequences for gait analysis. *International Conference on Image Processing, 1997. (ICIP'97). Washington, DC, USA.*, 3:78.
- [Micheloni et al. 2009] Micheloni, C., Snidaro, L., e Foresti, G. L. (2009). Exploiting temporal statistics for events analysis and understanding. *Image and Vision Computing*, 27:1459–1469.

- [Mohan et al. 2001] Mohan, A., Papageorgiou, C., e Poggio, T. (2001). Example-based object detection in images by components. *IEEE Transactions on Pattern Recognition and Machine Intelligence*, 23:349–361.
- [Nanda e Davis 2002] Nanda, H. e Davis, L. (2002). Probabilistic template based pedestrian detection in infrared videos. In *IEEE Intelligent Vehicle Symposium, 2002. June 17-21, 2002. Versailles, France.*, volume 1, pp. 15–20.
- [Peterfreund 2002] Peterfreund, N. (2002). Robust tracking of position and velocity with Kalman snakes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21:564–569.
- [Petrushin et al. 2006] Petrushin, V. A., Wei, G., Shakil, O., Roqueiro, D., e Gershman, V. (2006). Multiple-sensor indoor surveillance system. *3rd Canadian Conference Computer and Robot Vision. (CRV'06). June 07-09, 2006. Quebec City, Quebec, Canada.*, 1:53–60.
- [Piotrowski e Rogalski 2004] Piotrowski, J. e Rogalski, A. (2004). Uncooled long wavelength infrared photon detectors. *Infrared Physics and Technology*, 46(1-2):115–131.
- [Polana e Nelson 1994] Polana, R. e Nelson, R. (1994). Low level recognition of human motion. In *IEEE Workshop Motion of Non-Rigid and Articulated Objects. October 11-12, 1994. Austin, TX, USA.*, pp. 77–82.
- [Qihao e Weng 2009] Qihao e Weng (2009). Thermal infrared remote sensing for urban climate and environmental studies: Methods, applications, and trends. *ISPRS Journal of Photogrammetry and Remote Sensing*, 64(4):335–344.
- [Rafael C. González 2008] Rafael C. González, R. E. W. (2008). *Digital image processing*. Prentice Hall, 2008, 3, illustrated edition.
- [RNCOS 2009] RNCOS (2009). Global CCTV Market Analysis (2008-2012). Technical report, RNCOS Industry Research Solutions.
- [Roller et al. 1993] Roller, D., Daniilidis, K., e Nagel, H. H. (1993). Model-based object tracking in monocular image sequences of road traffic scenes. *International Journal of Computer Vision*, 10:257–281.
- [Silva e Scharcanski 2010] Silva, L. e Scharcanski, J. (2010). Video segmentation based on motion coherence of particles in a video sequence. *IEEE Transactions on Image Processing*, 19(4):1036–1049.
- [Stauffer e Grimson 1999] Stauffer, C. e Grimson, W. (1999). Adaptive background mixture models for real-time tracking. In *Conference on Computer Vision and Pattern Recognition (CVPR '99), June 23-25, 1999, Ft. Collins, CO, USA.*, volume 2, pp. 637–663.
- [Tingjun et al. 2010] Tingjun, L., Fuguang, Z., Xinju, C., Qilai, H., e Qiang, G. (2010). The model of visual attention infrared target detection algorithm. *2010 International Conference on Communications and Mobile Computing. 12-14, April 2010. Shenzhen, China.*, 3:87–91.

- [V. Kastrinaki e Kalaitzakis 2003] V. Kastrinaki, M. Z. e Kalaitzakis, K. (2003). A survey of video processing techniques for traffic applications. *Image and Vision Computing*, 21:359–381.
- [Wang 2010] Wang, P. (2010). Moving object segmentation algorithm based on edge detection. In *2010 International Conference on Computational Intelligence and Software Engineering (CiSE). December 10-12, 2010. Wuhan, China.*, pp. 1–4.
- [Welch e Bishop 1995] Welch, S. e Bishop, G. (1995). An introduction to the Kalman filter. Technical report, University of North Carolina at Chapel Hill. (Online at <http://www.cs.unc.edu/~welch/kalman>).
- [Wong et al. 2009] Wong, W. K., Tan, P. N., Loo, C. K., e Lim, W. S. (2009). An effective surveillance system using thermal camera. *International Conference on Signal Acquisition and Processing. April 03-05, 2009. Kuala Lumpur, Malaysia.*, pp. 13–17.
- [Wright et al. 2010] Wright, D., Friedewald, M., Gutwirth, S., Langheinrich, M., Mordini, E., Bellanova, R., Hert, P. D., Wadhwa, K., e Bigo, D. (2010). Sorting out smart surveillance. *Computer Law and Security Review*, 26(4):343–354.
- [Wu e Huang 2001] Wu, Y. e Huang, T. S. (2001). A co-inference approach to robust visual tracking. In *Eighth IEEE International Conference on Computer Vision. ICCV 2001. July 07-14, 2001. Vancouver, BC, Canada*, volume 2, pp. 26–33.
- [Xiang e Gong 2008] Xiang, T. e Gong, S. (2008). Video behavior profiling for anomaly detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(5):893–908.
- [Xu et al. 2005] Xu, F., Liu, X., e Fujimura, K. (2005). Pedestrian detection and tracking with night vision. *IEEE Transactions on Intelligent Transportation Systems*, 6(1):63–71.
- [Xue et al. 2010] Xue, Z., Ming, D., Song, W., Wan, B., e Jin, S. (2010). Infrared gait recognition based on wavelet transform and support vector machine. *Pattern Recognition*, 43(8):2904–2910.
- [Yilmaz et al. 2003] Yilmaz, A., Shafique, K., e Shah, M. (2003). Target tracking in airborne forward looking infrared imagery. *Image and Vision Computing*, 21(7):623–635.
- [Younus e Yang 2011] Younus, A. M. e Yang, B.-S. (2011). Intelligent fault diagnosis of rotating machinery using infrared thermal image. *Expert Systems with Applications*, 39(0):2082–2091.
- [Zhang et al. 2009] Zhang, F., qian Liu, S., bao Wang, D., e Guan, W. (2009). Aircraft recognition in infrared image using wavelet moment invariants. *Image and Vision Computing*, 27(4):313 – 318.
- [Zhang et al. 2006] Zhang, W., Chen, F., Xu, W., e Zhang, E. (2006). Real-time video intelligent surveillance system. In *IEEE International Conference on Multimedia and Expo.: July 9-12, 2006. Toronto, Ontario, Canada.*, pp. 1021–1024.

- [Ziou e Tabbone 1998] Ziou, D. e Tabbone, S. (1998). Edge detection techniques - An overview. *International Journal of Pattern Recognition and Image Analysis*, 8:537–559.