

DISSERTAÇÃO DE MESTRADO

**SISTEMA VISUAL DE SEGMENTAÇÃO DE
OBJETOS ROBUSTO A ILUMINAÇÃO PARA
MONITORAMENTO AUTOMÁTICO DE TRÁFEGO
DE VEÍCULOS TERRESTRES**

Gabriel Iribarem Soares Ruas

Brasília, julho de 2009

UNIVERSIDADE DE BRASÍLIA

FACULDADE DE TECNOLOGIA

UNIVERSIDADE DE BRASÍLIA
Faculdade de Tecnologia

DISSERTAÇÃO DE MESTRADO

**SISTEMA VISUAL DE SEGMENTAÇÃO DE
OBJETOS ROBUSTO A ILUMINAÇÃO PARA
MONITORAMENTO AUTOMÁTICO DE TRÁFEGO
DE VEÍCULOS TERRESTRES**

Gabriel Irribarem Soares Ruas

*Relatório submetido ao Departamento de Engenharia
Elétrica como requisito parcial para obtenção
do grau de Mestre em Engenharia Elétrica*

Banca Examinadora

Prof. Marcus Vinicius Lamar, CiC/UnB
Orientador

Prof. Hani Camille Yehia, DELT/UFMG
Examinador externo

Prof. Ricardo Lopes de Queiroz, ENE/UnB
Examinador interno

FICHA CATALOGRÁFICA

RUAS, GABRIEL IRRIBAREM SOARES

Sistema Visual de Segmentação de Objetos Robusto a Iluminação para Monitoramento Automático de Tráfego de Veículos Terrestres [Distrito Federal] 2009.

viii, 38 p., 210 x 297 mm (ENE/FT/UnB, Mestre, Engenharia Elétrica, 2009)

Dissertação de Mestrado - Universidade de Brasília. Faculdade de Tecnologia.

Departamento de Engenharia Elétrica.

- | | |
|---------------------------------|-------------------------|
| 1. Visão Computacional | 2. Segmentação |
| 3. Subtração de Fundo | 4. Supressão de Sombras |
| 5. Controle automático de Ganho | |

I. ENE/FT/UnB

II. Título (série)

REFERÊNCIA BIBLIOGRÁFICA

RUAS, G. I. S. (2009). Sistema Visual de Segmentação de Objetos Robusto a Iluminação para Monitoramento Automático de Tráfego de Veículos Terrestres. Dissertação de Mestrado, Publicação PPGENE.DM - 384/09, Departamento de Engenharia Elétrica, Universidade de Brasília, Brasília, DF, 38 p.

CESSÃO DE DIREITOS

NOME DO AUTOR: Gabriel Iribarem Soares Ruas

TÍTULO DA DISSERTAÇÃO DE MESTRADO: Sistema Visual de Segmentação de Objetos Robusto a Iluminação para Monitoramento Automático de Tráfego de Veículos Terrestres.

GRAU / ANO: Mestre em Engenharia Elétrica / 2009

É concedida à Universidade de Brasília a permissão para reproduzir cópias desta dissertação de mestrado e para emprestar ou vender tais cópias somente para propósitos acadêmicos e científicos. O autor reserva outros direitos de publicação e nenhuma parte desta dissertação de mestrado pode ser reproduzida sem a autorização por escrito do autor.

Gabriel Iribarem Soares Ruas

gabrielruas@gmail.com

Dedicatória

Dedico este trabalho à minha mãe, Ana Maria, meus avós, Ieda e Travasso, e minha namorada Fabiane

Gabriel Irribarem Soares Ruas

Agradecimentos

Gostaria de agradecer à minha mãe Ana, meus avós Ieda e Travasso, e minha namorada Fabiane pelo apoio e paciência. Aos amigos Ticiano e Augusto, por toda a ajuda. Ao Professor Lamar, por ir muito além do papel de orientador. Ao CEFTRU e toda a equipe, pela existência do projeto e todo o apoio. E, finalmente, à cafeína em todas as suas formas, por permitir que este trabalho fosse completado.

Gabriel Irribarem Soares Ruas

RESUMO

O presente trabalho apresenta uma proposta de sistema de segmentação de objetos móveis robusto a variações de iluminação e capaz de detectar e suprimir sombras móveis. Utilizou-se como base a técnica de subtração de fundo por misturas de gaussianas. Para lidar melhor com variações de iluminação, a técnica foi adaptada para utilizar o espaço de cores HSV. Foram analisadas as influências das sombras sobre o modelo de fundo e desenvolveu-se um modelo para reduzir sua influência sobre a segmentação de objetos. Observou-se também o efeito do ajuste de ganho causado por variações na iluminação da cena. Utilizando o conceito de estabilidade da cena, propõe-se um método de modelagem e controle da influência do ganho da câmera sobre a cena.

Utilizando-se a supressão de sombras, obteve-se uma taxa de acerto de 36,63 %, similar à técnica original proposta por Stauffer. A taxa de falsos positivos foi reduzida de 71,38 % para 42,65 %. A metodologia proposta também foi capaz de reduzir significativamente a influência de variações bruscas de iluminação no resultado da subtração de fundo.

ABSTRACT

This study presents a proposal for a moving object segmentation system that is resistant to changes in lighting conditions and can detect and suppress moving shadows. The technique of background subtraction using mixtures of gaussians provided a basis for this study. To better deal with illumination changes, that technique was adapted to use the HSV color space. The influence of shadows over the the background model was analyzed. Furthermore, a model to reduce their influence over the object segmentation process was developed. The effects of automatic gain control caused by illumination changes in the scene were also studied. Using the concept of scene stability, a method to model and control the effects of camera gain on the scene was proposed.

Shadow suppression yielded a correct rate of 36.63 %, which is similar to results obtained using the original technique proposed by Stauffer. The false positive rate was reduced from 71.38 % to 42.65 %. In addition, the proposed methodology was also able to significantly reduce the influence of sudden illumination variations on the background subtraction results.

SUMÁRIO

1	INTRODUÇÃO	1
1.1	CONTEXTUALIZAÇÃO	1
1.2	DEFINIÇÃO DO PROBLEMA	1
1.3	OBJETIVOS	2
1.4	APRESENTAÇÃO DO MANUSCRITO	2
2	REVISÃO BIBLIOGRÁFICA	3
2.1	INTRODUÇÃO	3
2.2	SUBTRAÇÃO DE FUNDO	3
2.3	SUPRESSÃO DE SOMBRAS	6
2.4	INVARIÂNCIA À ILUMINAÇÃO	7
3	DESENVOLVIMENTO	9
3.1	INTRODUÇÃO	9
3.2	METODOLOGIA PROPOSTA	10
3.2.1	SEGMENTAÇÃO DE OBJETOS	10
3.2.2	SUPRESSÃO DE SOMBRAS	14
3.2.3	CORREÇÃO DE VARIAÇÕES DE ILUMINAÇÃO	20
4	RESULTADOS EXPERIMENTAIS	26
4.1	INTRODUÇÃO	26
4.2	RESULTADOS	28
4.2.1	SUPRESSÃO DE SOMBRAS	28
4.2.2	ROBUSTEZ À VARIAÇÃO GLOBAL DE ILUMINAÇÃO	32
5	CONCLUSÕES	34
	REFERÊNCIAS BIBLIOGRÁFICAS	36

LISTA DE FIGURAS

2.1	Exemplo da subtração de fundo com uma referência estática.	4
2.2	Objetos unidos pela sombra.	6
3.1	Posicionamento ideal da câmera sobre a pista.	9
3.2	Posicionamento real da câmera ao lado da pista.	10
3.3	Exemplo do modelo de mistura de Gaussianas.	12
3.4	Exemplo de ofuscamento.	13
3.5	Representação cônica do espaço de cores HSV.	14
3.6	Classificação das sombras em uma imagem.	15
3.7	Modelo de fundo obtido a partir de uma sequência típica de vídeo de monitoramento.	17
3.8	Componentes HSV do modelo de fundo obtido a partir de uma sequência típica de vídeo de monitoramento.	18
3.9	Efeitos da sombra projetada sobre a cena no espaço de cores HSV.	19
3.10	Máscara binária de sombras.	19
3.11	Nível médio da componente V da imagem, quadros 1 a 1000.	21
3.12	Efeitos do controle automático de ganho.	22
3.13	Pontos da cena avaliados.	23
3.14	Comportamento da componente V da cena ao longo do tempo em três pontos distintos.	23
3.15	Valor de V, normalizado pela média do valor de V no pixel.	24
4.1	O suporte utilizado para realizar as filmagens.	26
4.2	A câmera utilizada em seu suporte.	27
4.3	Veículos segmentados manualmente.	29
4.4	Histogramas de acerto das duas técnicas testadas.	30
4.5	Histogramas de vazamento das duas técnicas testadas.	31
4.6	Exemplo de uma máscara gerada pela subtração de fundo afetada pelo ruído de aquisição.	31
4.7	Exemplo de filtragem do ruído de aquisição presente na máscara obtida da subtração de fundo.	32
4.8	Quadro de vídeo após codificação.	32
4.9	Compensação de mudanças de iluminação durante a passagem de um caminhão.	33
4.10	Compensação de mudanças de iluminação após a passagem do móvel.	33

1 INTRODUÇÃO

1.1 CONTEXTUALIZAÇÃO

Com o aumento do número de veículos particulares em circulação, a expansão e criação de vias urbanas passou a ter um papel importante no planejamento urbano. Para que este planejamento possa ser feito de maneira eficiente são necessárias informações sobre o uso das vias. Uma das ferramentas utilizadas para obter estas informações é a contagem volumétrica de veículos, a qual permite ao engenheiro de transportes avaliar quais os pontos de maior demanda de uma determinada malha viária. Atualmente, este estudo é geralmente realizado através do preenchimento de formulários por pesquisadores posicionados ao redor da via a ser analisada (PINTO et al., 1996). Dependendo do fluxo aparente da via, apenas carros de determinadas cores são contados, e o fluxo é inferido através de métodos estatísticos. Este método, por depender de contadores humanos, apresenta certas limitações, como a necessidade de se obter pessoal suficiente para realizar as contagens, e a impossibilidade de executar um monitoramento por grandes períodos de tempo.

Para superar essas limitações, sistemas automatizados de contagem podem ser utilizados. Uma maneira de se realizar esta automação é a utilização de sensores indutivos fixados ao pavimento, de maneira temporária ou permanente. Isto permite a contagem da quantidade de veículos que atravessam um seção da via em um intervalo de tempo, bem como, de acordo com a tecnologia, a classificação dos veículos, baseado na sua assinatura magnética (MAGALHÃES, 2008). Porém, esta técnica exige a instalação de sensores na pista. O uso de sensores permanentes restringe o monitoramento a determinadas vias de maior interesse, uma vez que o equipamento não pode ser reutilizado. Existem sensores portáteis, que podem ser reutilizados, porém, eles devem ser fixados na via, tornando o processo oneroso, já que exige a interrupção parcial do tráfego durante a instalação e retirada do equipamento. Nenhum destes métodos permite a reidentificação de veículos, portanto não é possível relacionar o fluxo em diferentes pontos de medição. Isto impede a realização de estudos mais avançados do fluxo de veículos.

Outra solução é a utilização de câmeras e técnicas de processamento de vídeo para a contagem automática de veículos. Isto permite a contagem de veículos em qualquer trecho de uma via, inclusive em cruzamentos. Desta maneira é possível realizar a discriminação da distribuição do fluxo por diferentes vias, o que é conhecido na Engenharia de Transportes como a Matriz Origem/Destino. Entretanto, a utilização de visão computacional em ambientes não controlados também apresenta seus desafios, sendo um dos principais a variação das condições ambientais ao longo do dia.

1.2 DEFINIÇÃO DO PROBLEMA

O monitoramento em ambientes abertos está sujeito às condições ambientais, além de suas variações ao longo da filmagem. Isto traz uma série de problemas para o reconhecimento automático dos veículos e seu rastreamento, uma vez que essas condições não podem ser controladas.

Um dos problemas introduzidos pela iluminação natural são as sombras projetadas sobre a cena. Como

as sombras apresentam uma diferença em relação ao restante da cena, elas podem ser consideradas como objetos em sistemas automáticos. Dependendo da posição da fonte de iluminação(geralmente, o sol) diferentes objetos podem ser considerados erroneamente como um único, por serem ligados através de suas sombras. A variação da iluminação durante o monitoramento, tanto pela presença de nuvens como pela alteração na posição do sol, introduz diversas dificuldades ao processamento automático. Para um sistema de processamento de vídeo, as variações da luminosidade podem ser interpretadas como a introdução de novos objetos na cena. Portanto, para diminuir a taxa de erros do sistema, é necessária a utilização de métodos de reconhecimento capazes de se adaptar a estas variações.

Outro problema advém do equipamento utilizado para a obtenção dos vídeos a serem processados, isto é, as câmeras. Os sensores utilizados em câmeras digitais mapeiam uma grandeza contínua, a quantidade de luz refletida pelos objetos, em valores discretos a serem armazenados e processados digitalmente. Este mapeamento envolve uma perda de precisão devido ao processo de quantização necessário ao armazenamento digital, o que significa que os vídeos possuem uma faixa de valores limitada para representar as imagens de interesse. Caso a cena apresente brilhos maiores ou menores do que aqueles que o sensor está ajustado para captar, ocorre o fenômeno da saturação, isto é, estes brilhos são mapeados aos valores limites da faixa dinâmica da câmera, ocasionando perda de informação. Para evitar isto, as modernas câmeras digitais dispõem de mecanismos de ajuste automático de ganho, exposição e outros parâmetros. Estes mecanismos, embora eficientes para a visão humana, interferem com a visão computacional de maneira similar às mudanças de iluminação, representando mais um desafio a ser resolvido. Apesar desta interferência, este processo de auto-ajuste é importante na automação do sistema, visto que, se não fosse utilizado, os parâmetros da câmera deveriam ser ajustados manualmente de maneira regular, exigindo constante intervenção humana no monitoramento.

1.3 OBJETIVOS

Esta dissertação tem por objetivo o desenvolvimento de um sistema de detecção de veículos por visão computacional robusto às variações das condições ambientais. O foco deste trabalho é a etapa de segmentação dos objetos, fundamental para a extração de informações úteis da sequência de vídeo. Esta segmentação será realizada através de técnicas de subtração de fundo, uma vez que a posição da câmera é fixa em relação à cena. Serão analisadas técnicas para realizar a supressão de sombras projetadas na cena de forma robusta, a fim de tornar esta segmentação mais precisa. Também serão pesquisadas e avaliadas maneiras de se corrigir a influência das variações de iluminação introduzidas na filmagem, tanto de forma natural, devido ao sol e às nuvens, como artificial, devido ao auto-ajuste da câmera.

1.4 APRESENTAÇÃO DO MANUSCRITO

No Capítulo 2 é feita uma revisão bibliográfica sobre o tema de estudo. Em seguida, o Capítulo 3 descreve a metodologia proposta no desenvolvimento do projeto. Resultados experimentais são discutidos no Capítulo 4, seguido das conclusões no Capítulo 5.

2 REVISÃO BIBLIOGRÁFICA

2.1 INTRODUÇÃO

Este trabalho estuda a viabilidade do uso da união de técnicas para o desenvolvimento de um sistema de medição de fluxo de veículos robusto a variações das condições ambientais. Para isto, primeiramente são analisados diferentes métodos de subtração de fundo para realizar a segmentação de objetos móveis da cena. A seguir, métodos de supressão de sombras projetadas são avaliados, com o objetivo de aumentar a robustez e a precisão da segmentação. Finalmente, são pesquisadas maneiras de corrigir a influência das variações da iluminação no sistema proposto.

2.2 SUBTRAÇÃO DE FUNDO

A primeira etapa para o desenvolvimento de um sistema de contagem de objetos é a detecção e segmentação dos objetos de interesse na sequência de vídeo. A subtração de fundo é uma técnica que busca gerar uma imagem de referência que represente a porção estática ou semi-estática da cena, ou seja, a imagem do fundo. Para gerar esta referência, é necessário estimar um modelo da cena, a partir de imagens previamente obtidas. Uma vez que o fundo esteja modelado, objetos móveis podem ser definidos como as regiões que não se adequem ao modelo gerado do fundo.

O método mais simples de subtração de fundo é a utilização de uma imagem escolhida manualmente (NORIEGA; BERNIER, 2006). Apesar de sua simplicidade, ocorre uma acumulação de erros devido a mudanças na cena que não são modeladas pelo fundo estático. Isto introduz a necessidade de reinicialização periódica do modelo, fazendo com que este método seja útil apenas em situações específicas nas quais o ambiente seja totalmente controlado, ou que a duração do experimento seja pequena.

A Figura 2.1 apresenta imagens representando um exemplo da aplicação deste método. A imagem 2.1(a) representa a imagem escolhida como referência. As figuras 2.1(b), 2.1(d) e 2.1(f) são três quadros do vídeo, e 2.1(c), 2.1(e) e 2.1(g) são as máscaras binárias referentes a esses quadros. Estas máscaras representam o resultado da subtração de fundo. Caso a diferença entre o quadro e a referência seja maior que um determinado limiar, o pixel correspondente na máscara será definido branco(verdadeiro), caso contrário será preto(falso). Pode-se perceber que, com a passagem de apenas alguns quadros, a máscara já incorpora erros devido a pequenas variações na posição da câmera, gerando objetos que não correspondem à realidade. Com a passagem do dia e mudanças de iluminação, todo o quadro tende a ser considerado como um único objeto, tornando necessária uma redefinição da imagem de referência.

Devido a estas limitações, as pesquisas se concentram em métodos adaptativos de estimação do fundo. Uma maneira de se obter um modelo adaptativo do fundo é a utilização da média temporal dos pixels (STAUFFER; GRIMSON, 1999). Este método permite que a imagem de referência se mantenha atualizada com as mudanças graduais que ocorram na cena, porém é necessário que os objetos de interesse se mantenham em movimento constantemente, para que não sejam considerados parte do fundo. Além disso, o fundo

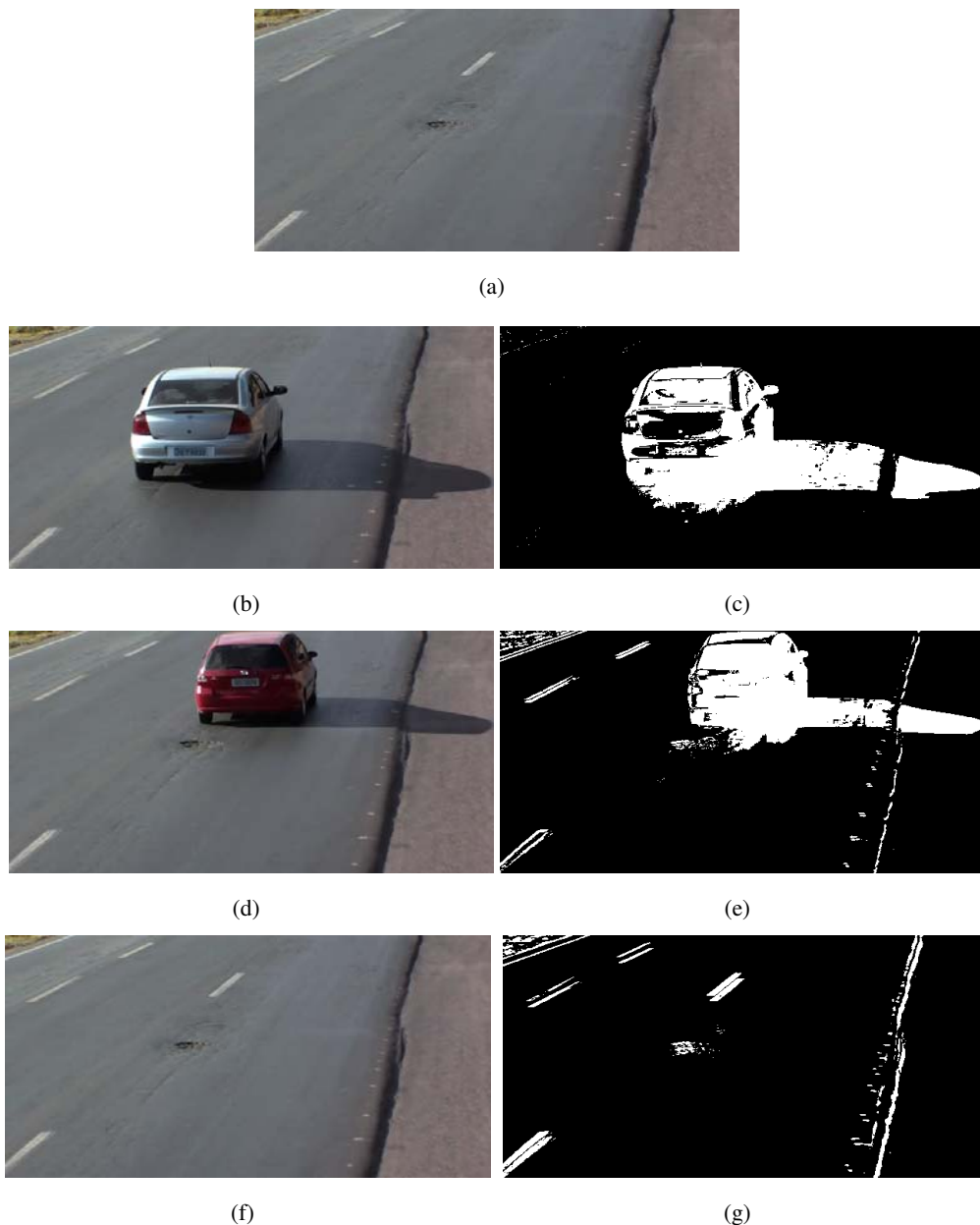


Figura 2.1: Exemplo da subtração de fundo com uma referência estática. (a) Figura escolhida como referência (b)Quadro 260 (c)Máscara do quadro 260 (d)Quadro 370 (e)Máscara do quadro 370 (f)Quadro 600 (g)Máscara do quadro 600.

deve ser visível a maior parte do tempo para que o modelo seja confiável, o que diminui sua utilidade em cenas com maior movimento, isto é, com um grande número de objetos.

Visando superar estas limitações, porém mantendo a simplicidade da média, Lai e Yung (LAI; YUNG, 1998) propõem um algoritmo de subtração de fundo baseado em placar(*Scoreboard*). Este algoritmo utiliza dois filtros por pixel, a média móvel e a moda móvel. A quantidade de movimento presente na cena é registrada em um placar e, dependendo da pontuação, um dos dois filtros é utilizado para decidir se um pixel pertence ao fundo ou não. Em regiões de pouco movimento, a média é utilizada, enquanto que em regiões de maior movimento utiliza-se a moda do pixel. Este método possui baixa complexidade computacional e é mais robusto que a utilização somente da média, uma vez que objetos não são adicionados ao modelo do

fundo quando o filtro de moda considerar a região como parte do fundo. Isto faz com que, em situações de movimento normal, apenas mudanças graduais sejam absorvidas pelo modelo do fundo. Porém, é necessário o ajuste manual dos parâmetros de ambos os filtros, além dos limiares do placar. Isto dificulta o uso desta técnica em situações em que a dinâmica da cena possa se alterar de maneira mais profunda, como é o caso em monitoramentos de longa duração.

Outras abordagens modelam o comportamento dos pixels de maneira mais detalhada. Ridder *et al* (RIDDER; MUNKELT; KIRCHNER, 1995) modelam cada pixel através de um filtro de Kalman, permitindo que o modelo se adapte bem a mudanças de iluminação. Esta técnica, porém, apresenta baixa capacidade de recuperação do fundo quando objetos móveis são temporariamente assimilados pelo modelo. Este modelo também é incapaz de representar um fundo multi-modal, isto é, em que mais de um valor possa ser considerado como fundo em uma determinada região. Esta é uma situação que pode ocorrer, por exemplo, com luzes piscando ou com o balançar de uma árvore ao vento. Um método que apresenta um modelo multi-modal do fundo é o proposto por Friedman e Russel (FRIEDMAN; RUSSELL, 1997). Os autores classificam cada pixel em três distribuições pré-determinadas, representando as cores da rua, da sombra e dos veículos. Devido ao uso de distribuições de probabilidade pré-definidas, este método apresenta uma grande dificuldade de adaptação a diferentes situações em que possa ser aplicado, exigindo uma calibração manual para cada cena.

Em (STAUFFER; GRIMSON, 1999), Stauffer e Grissom propõem uma metodologia multi-modal adaptativa. Nesse trabalho, cada pixel é tratado como um processo estocástico, com uma distribuição de probabilidade modelada por uma mistura de Gaussianas. Os parâmetros das gaussianas são ajustados através de um algoritmo de *expectation maximization*(EM) *online*. Neste modelo, os pixels são classificados como fundo se possuírem a probabilidade maior que um determinado limiar, de acordo com a distribuição das respectivas gaussianas. Este modelo permite que mais de um estado do pixel seja classificado como fundo, de acordo com a atividade na cena. Além disso, mudanças graduais no fundo são assimiladas com poucas perturbações na detecção de objetos. A representação por múltiplas gaussianas garante ao modelo uma memória, isto é, mudanças no modelo de fundo não apagam o modelo anterior. As gaussianas que deixam de ser consideradas como fundo continuam presentes no modelo. Caso a cena retorne a um estado anterior, o modelo pode restaurar o fundo de maneira mais rápida, uma vez que este estado anterior já está armazenado. A mistura de gaussianas não é capaz, porém, de ignorar mudanças bruscas de iluminação, como as causadas pela passagem de nuvens em um dia ensolarado. Em (KAEWTRAKULPONG; BOWDEN, 2001) são apresentadas modificações do sistema proposto em (STAUFFER; GRIMSON, 1999), visando melhorar seu desempenho. Uma fraqueza identificada no artigo original (STAUFFER; GRIMSON, 1999) é o elevado tempo requerido para a inicialização do modelo de fundo. No início de uma sequência de vídeo, não há amostras suficientes para suportar qualquer modelo de fundo. Isto faz com que, durante os quadros iniciais, o sistema não seja capaz de detectar qualquer movimento, até que um modelo inicial acumule evidências suficientes. Os autores propõem em (KAEWTRAKULPONG; BOWDEN, 2001) uma inicialização acelerada, utilizando apenas uma janela de tamanho igual ao número de quadros já obtidos para realizar a estimativa inicial do modelo. Isto permite que um modelo funcional do fundo seja obtido nos primeiros quadros da sequência, mesmo correndo o risco de absorver informações espúrias muito rapidamente no modelo. A dinâmica de atualização do modelo garante que essas informações não correspondentes ao fundo da imagem sejam rapidamente ignoradas pelo modelo, enquanto mais evidências são coletadas e o modelo se torna mais estável. Em (KAEWTRAKULPONG; BOWDEN, 2001), também é proposto um modelo simples de



Figura 2.2: Objetos unidos pela sombra.

supressão de sombras baseado no espaço de cores RGB.

2.3 SUPRESSÃO DE SOMBRAS

De acordo com (SALVADOR; CAVALLARO; EBRAHIMI, 2004), uma sombra ocorre quando um objeto oclui total ou parcialmente a iluminação direta de uma fonte de luz. A presença de sombras torna a segmentação de cenas sob iluminação natural mais complexa, uma vez que se confundem com os objetos de interesse, podendo causar a junção de vários objetos distintos no sistema de rastreamento e distorcendo suas formas (PRATI et al., 2001a). A Figura 2.2 apresenta um exemplo onde dois objetos distintos podem ser detectados como um único devido à presença de sombras.

Por esse motivo, é necessário diferenciar os objetos em movimento de suas sombras. Segundo (CUCCHIARA et al., 2001), sombras podem ser classificadas em dois tipos: sombras próprias e sombras projetadas. Sombras próprias correspondem às regiões do objeto que não são iluminadas diretamente devido à geometria do próprio objeto. Sombras projetadas são as regiões de sombra presentes na cena devido à presença do objeto, na direção da fonte de luz bloqueada. Em (KUMAR; SENGUPTA; LEE, 2002) é apresentado um estudo sobre os efeitos de diferentes espaços de cores na detecção de objetos e sombras. Cinco espaços de cores são avaliados quanto ao seu desempenho para subtração de fundo e supressão de sombras: RGB, HSV, XYZ, $YCrCb$ e RGB normalizado. O fundo é modelado por uma distribuição normal, com a média e o desvio padrão obtidos através do monitoramento de uma cena sem objetos por alguns segundos. As sombras são modeladas como uma constante aditiva ao fundo, com os critérios exatos dependendo do espaço de cores. Os autores concluem que o melhor espaço de cores, em termos da taxa de detecção e taxa de falsos positivos, é o espaço $YCrCb$, enquanto que o espaço de cores com o pior desempenho é o HSV. Por outro lado, outros estudos apontam o HSV como o espaço de cores ideal para realizar a eliminação de sombras, devido a este aproximar a percepção humana das cores (CUCCHIARA et al., 2001).

Em (PRATI et al., 2003), (PRATI et al., 2001b) e (PRATI et al., 2001a) são analisadas diferentes propostas de detecção de sombras em sequências de vídeo. As propostas são divididas em quatro classes diferentes: métodos estatísticos paramétricos, estatísticos não-paramétricos, determinísticos não-baseados em modelo

e determinísticos baseados em modelo. Os métodos são ainda avaliados quanto à utilização de informações espectrais, espaciais e temporais. Os artigos concluem que métodos determinísticos não-baseados em modelos são melhores para sistemas de uso geral, em que não se pode assumir as condições da cena. Para a detecção de sombras em uma situação específica, sistemas determinísticos oferecem um resultado melhor. Para ambientes internos, uma abordagem estatística é mais eficiente, uma vez que o ambiente é estável, permitindo uma boa descrição estatística. Em (NADIMI; BHANU, 2004) um modelo que leva em conta as diferenças cromáticas entre o sol e o céu é utilizado para realizar a detecção e remoção de sombras. A detecção de sombras é realizada pela comparação com um modelo de fundo representado por uma mistura de Gaussianas. Regiões candidatas a sombras são aquelas em que a luminosidade do quadro analisado é menor que a do modelo de fundo. Então critérios de uniformidade e taxa de azul são aplicados para identificar as sombras. No trabalho, os autores notam que em cenas externas, sombras projetadas em superfícies de cores neutras apresentam uma cor mais azulada, e utilizam esta informação para refinar o detector de sombras.

Em (XU et al., 2005), (ZHANG et al., 2007) e (WANG; TANG; ZHU, 2006) são utilizadas metodologias baseadas em filtragem de bordas para supressão de sombras. Em (ZHANG et al., 2007), os autores sugerem uma nova medida invariante à iluminação para detecção de sombras, chamada razão de bordas (*Ratio Edge*). Em (XU et al., 2005), as regiões de sombra são detectadas através do mapa de bordas de Canny. Regiões de sombra apresentam bordas mais fracas que os objetos de interesse, sendo então eliminadas. Para lidar com regiões com grande quantidade de texturas, os autores propõem o uso de integração temporal das bordas. Desta maneira, é possível diferenciar bordas estáticas, já presentes na cena, e bordas em movimento, causadas pelos objetos de interesse e suas sombras. Em (WANG; TANG; ZHU, 2006), possíveis sombras são detectadas pela redução da luminosidade em relação ao fundo, e também pelo ângulo da diferença entre as cores do objeto detectado e do fundo no espaço de cores RGB. As regiões de sombra detectadas são então refinadas utilizando o mapa de bordas da imagem. Em (LEONE; DISTANTE; BUCCOLIERI, 2005), características de textura são utilizadas para realizar a detecção de sombras. Em (JACQUES; JUNG; MUSSE, 2005), sombras são detectadas em sequências de vídeo em escala de cinza. Um modelo de fundo estatístico é inicialmente montado, durante um período de treinamento de alguns segundos. O modelo é então mantido atualizado utilizando informações das regiões estáticas da cena. Sombras são inicialmente detectadas pela correlação cruzada normalizada, verificando se regiões detectadas não são apenas versões do fundo com menor luminosidade. Os resultados desta etapa são refinados, verificando-se se a relação entre a luminosidade da possível sombra e da imagem de fundo é aproximadamente constante em uma determinada vizinhança em torno dos pixels candidatos a sombra.

2.4 INVARIÂNCIA À ILUMINAÇÃO

Em (LEE; KO; HAN, 2002) e (CUCCHIARA; PICCARDI, 1999) é analisado o rastreamento de veículos em cenas noturnas. Em (CUCCHIARA; PICCARDI, 1999), veículos são tratados diferentemente de acordo com a hora do dia. Durante o dia, com a cena bem iluminada, a detecção de objetos móveis é feita utilizando as diferenças entre três quadros consecutivos. À noite, quando a iluminação é proveniente apenas dos próprios veículos, os objetos móveis são caracterizados como pares de faróis. Como o histograma de luminosidade apresenta uma forte bimodalidade nesta situação, é possível separar os pontos pertencentes aos faróis dos

pontos devidos à luz refletida. O artigo não indica, porém, sob que condições é realizada a transição de um modelo para o outro, durante a passagem do dia. Nam em (NAM; RHEE, 2004) apresenta uma proposta de pré-processamento para aumentar a robustez de aplicações de reconhecimento de faces. Este processamento consiste de uma equalização de histograma das imagens de faces, seguida da aplicação de um bloco de filtragem adaptativo, baseado em um sistema de aprendizado não-supervisionado. Em (KYO et al., 1999) é apresentado um sistema de compensação para condições chuvosas de monitoramento. De acordo com os autores, a chuva se apresenta como um ruído de alta frequência nos quadros de vídeo. Para reduzir seus efeitos, é utilizado um filtro passa-baixas, ajustado para as características espaciais da chuva. Em (MATSUYAMA et al., 2006), é proposta a utilização de propriedades invariantes à iluminação, especificamente a distância vetorial normalizada. Os autores mostram que essa distância não depende da luminosidade, sendo um bom estimador para a diferença entre o quadro e uma imagem de referência (fundo). Para lidar com as instabilidades dessa distância quando a luminosidade é muito pequena, é proposta a utilização de limiares adaptativos e características espaciais. A dispersão da distância vetorial normalizada é analisada em janelas sobrepostas dentro de blocos da imagem. Variações de brilho devido ao ruído são uniformemente distribuídas, com pequena dispersão. Já variações devido à presença de objetos apresentam uma maior dispersão.

Em (LIM et al., 2005), são utilizadas duas câmeras para obter invariância à iluminação. A classificação da cena em frente e fundo é realizada utilizando a diferença de cor entre pares conjugados de pixels. Pixels representando o fundo devem variar de maneira similar, enquanto que, caso em uma imagem haja um objeto, as variações serão diferentes. Utilizando esta metodologia, não há necessidade de realizar a modelagem exata do fundo de maneira *online*.

Em (NORIEGA; BERNIER, 2006), histogramas de *kernel* locais são utilizados para modelar o fundo. Histogramas de *kernel* locais permitem uma modelagem estatística espacial e temporal do fundo. Utilizando o gradiente da imagem como característica, em conjunto com as cores, os autores afirmam obter robustez a variações de iluminação.

3 DESENVOLVIMENTO

3.1 INTRODUÇÃO

Este trabalho é parte de uma proposta de sistema automático de contagem classificada de fluxo de veículos. A proposta consiste de um sistema de contagem de veículos portátil, isto é, que seja independente da posição relativa da câmera e da pista. O sistema de contagem consiste de três diferentes etapas: segmentação de objetos móveis, rastreamento e contagem dos veículos, e classificação dos veículos.

Na primeira etapa, é realizada a separação de objetos móveis presentes na cena. Como o sistema deve ser portátil, há limitações no posicionamento da câmera. A posição ideal para este tipo de estudo seria perpendicularmente sobre a via, eliminando qualquer tipo de oclusão dos objetos de interesse. Porém, isto somente é possível utilizando-se uma infra-estrutura fixa ou de difícil montagem. Na Figura 3.1, pode-se ver um exemplo deste posicionamento.

Portanto, a câmera só pode ser posicionada de maneira fácil e portátil ao lado da pista. Isto torna mais importante que haja uma etapa de segmentação precisa. Caso sombras e outros ruídos sejam segmentados juntamente com os veículos, existe o risco de vários objetos serem detectados como um único.

O rastreamento é a etapa responsável por associar temporalmente os objetos, definindo também suas rotas, e não será foco desse trabalho. Esta etapa é realizada tendo por base os resultados da segmentação de objetos. Para que a contagem seja feita de maneira correta, é fundamental que a segmentação seja feita precisamente, evitando a junção de objetos distintos e a separação de um objeto em vários.

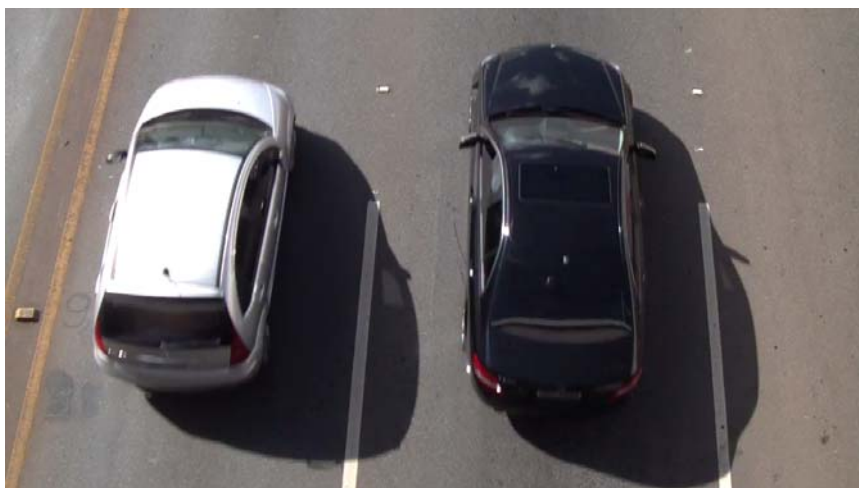


Figura 3.1: Posicionamento ideal da câmera sobre a pista.



Figura 3.2: Posicionamento real da câmera ao lado da pista.

3.2 METODOLOGIA PROPOSTA

Um dos requerimentos do sistema proposto neste trabalho é a portabilidade, isto é, o sistema deve ser utilizável em diferentes vias com o mínimo de adaptações. Para capturar o vídeo de um ponto de vista adequado ao monitoramento, um tripé metálico desmontável é utilizado para posicionar a câmera a até 5,20 m de altura, de acordo com as necessidades de cada via. O tripé com a câmera é posicionado ao lado da pista, de maneira que a câmera aponte numa direção quase paralela ao fluxo da via, filmando de uma altura suficiente para minimizar os efeitos de oclusões entre os veículos, conforme mostrado na Figura 3.2.

A metodologia proposta para o processamento do vídeo consiste, inicialmente, da segmentação dos objetos móveis da cena através do uso de técnicas de subtração de fundo. Técnicas de supressão de sombras serão usadas para aumentar a precisão da segmentação. Todo o processamento será realizado de forma a diminuir a influência das mudanças de iluminação devido às condições ambientais, naturais ou geradas pela câmera.

3.2.1 Segmentação de objetos

O sistema proposto busca segmentar veículos a partir de um ponto de vista fixo da câmera. Nesta situação, os objetos de interesse se movem sobre uma cena relativamente estática, representando a via. Pequenos movimentos devido a vibrações e à influência do vento são as principais dificuldades em considerar a cena estática. Para realizar esta tarefa de segmentação, foi escolhido como base o sistema proposto por (STAUFFER; GRIMSON, 1999), adequando-o com algumas modificações propostas em (KAEWTRAKULPONG; BOWDEN, 2001). Para que o sistema possa ser utilizado para monitoramentos por longos períodos, é necessário que seja adaptativo, alterando seu modelo de acordo com as variações de iluminação causadas pela passagem do dia. O modelo multimodal permite que alterações temporárias na cena, como carros estacionados, sejam modeladas sem que o modelo anterior seja descartado completamente. Desta maneira, o modelo pode recuperar seu estado anterior rapidamente. Este modelo também permite que alterações

oscilantes, tais como árvores balançando ou luzes piscando, sejam modeladas como parte do fundo.

No modelo proposto por (STAUFFER; GRIMSON, 1999), cada pixel é considerado um processo estocástico independente. O valor de cada pixel representa a soma das integrais da função refletância das superfícies presentes na cena, ao longo de todo o espectro das fontes de iluminação (SALVADOR; CAVALLARO; EBRAHIMI, 2004). Como a composição da cena e a intensidade da iluminação podem variar simultaneamente com o tempo, o modelo de fundo deve ser construído dinamicamente. As distribuições das cores dos pixels são aproximadas por um modelo de mistura de Gaussianas.

Neste modelo, a função densidade de probabilidade é aproximada como uma soma de Gaussianas, de maneira que a probabilidade de observação um determinado valor de pixel é dada por

$$P(X_t) = \sum_{k=1}^K \omega_{k,t} \eta(X_t, \mu_{k,t}, \Sigma_{k,t}), \quad (3.1)$$

onde K é o número de distribuições, e $\omega_{k,t}$, $\mu_{k,t}$ e $\Sigma_{k,t}$ são o peso, a média e a matriz de covariância da Gaussiana k no tempo t . A função η é a função densidade de probabilidade Gaussiana, dada por

$$\eta(X_t, \mu_{k,t}, \Sigma_{k,t}) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma_{k,t}|^{\frac{1}{2}}} e^{-\frac{1}{2}(X_t - \mu_{k,t})^T \Sigma_{k,t}^{-1} (X_t - \mu_{k,t})}, \quad (3.2)$$

onde n corresponde à dimensão de espaço utilizado. O peso $\omega_{k,t}$ de uma Gaussiana pode ser estimado como a quantidade de amostras representadas por essa Gaussiana em um determinado instante de tempo. Para diminuir a carga computacional, a matriz de covariância simplificada é definida como uma matriz diagonal

$$\Sigma_{k,t} = \sigma_{k,t}^2 I, \quad (3.3)$$

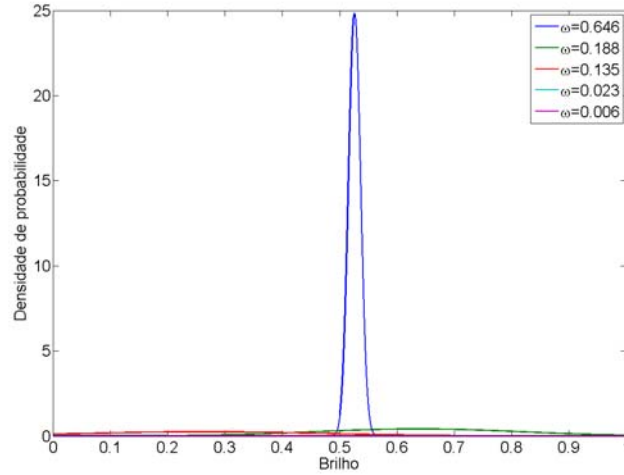
de modo que as componentes de cor do pixel são consideradas independentes, e com a mesma variância $\sigma_{k,t}^2$.

A Figura 3.3 exemplifica um modelo de mistura de Gaussianas de um pixel em uma sequência de vídeo em escala de cinza. Pode-se notar que, neste exemplo, há uma Gaussiana principal, com grande peso e pequeno desvio padrão. Isto significa que, na posição deste pixel, a maior parte dos quadros apresenta luminosidade dada pela sua média, sendo este o valor com maior probabilidade de representar o fundo da cena.

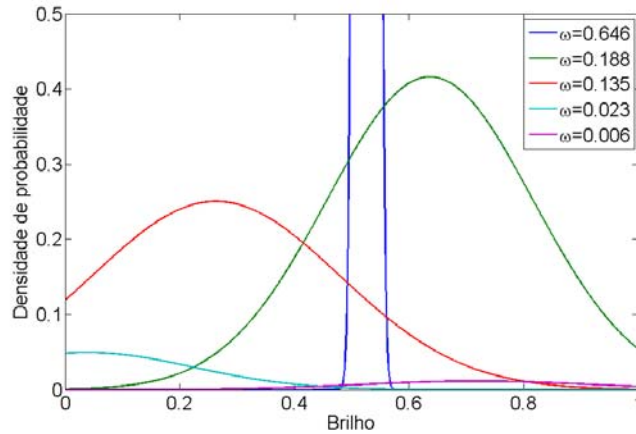
Para atualizar o modelo, é utilizada uma aproximação *K-means online* do algoritmo de *expectation maximization* (DEMPSTER et al., 1977). Cada nova amostra de pixel é verificada contra todas as Gaussianas correspondentes, até que haja um casamento. O casamento ocorre quando o valor do pixel esteja a menos de 2,5 desvios padrão da média de uma distribuição. Isto gera, efetivamente, um limiar de decisão para cada pixel, adaptado às condições de cada região da imagem. Os pesos das K distribuições são ajustados conforme

$$\omega_{k,t} = (1 - \alpha)\omega_{k,t-1} + \alpha(M_{k,t}) \quad (3.4)$$

onde α é a taxa de aprendizado e $M_{k,t}$ é 1 para as distribuições em que houve casamento, e 0 para as restantes. Caso não haja casamento, a distribuição menos provável, isto é, aquela com menor peso, é substituída. A nova Gaussiana do modelo possui a média igual ao valor do pixel recebido, grande desvio padrão e peso pequeno. Após estes ajustes, os pesos são renormalizados para que a probabilidade total dos pixels permaneça unitária. A Equação 3.4 representa um filtro passa-baixas IIR de primeira ordem, com



(a)



(b)

Figura 3.3: Exemplo de modelo de mistura de Gaussianas, para um pixel de uma sequência em escala de cinza: (a) modelo completo, (b) detalhe do modelo de mistura de gaussianas.

constante de tempo τ dada por $1/\alpha$. A constante de tempo τ define o ritmo de atualização do modelo. A média μ e o desvio padrão σ das distribuições que não casam permanecem inalterados. Para as distribuições que correspondem às observações, seus parâmetros são atualizados conforme

$$\mu_t = (1 - \alpha)\mu_{t-1} + \alpha X_t \quad (3.5)$$

$$\sigma_t^2 = (1 - \alpha)\sigma_{t-1}^2 + \alpha (X_t - \mu_t)^T (X_t - \mu_t) \quad (3.6)$$

onde α é a taxa de aprendizado, μ_t e σ_t^2 são a média e a variância da gaussiana no momento t e X_t é o valor observado do pixel. Uma vez modelada a densidade de probabilidade dos pixels, é necessário determinar quais gaussianas representam o fundo. Primeiramente, as gaussianas são ordenadas de acordo com o valor de ω/σ , em ordem decrescente. Desta maneira, as distribuições que apresentam mais evidência (maior peso) e maior consistência (menor desvio padrão) ficam posicionadas no topo da lista ordenada, enquanto que distribuições geradas por movimentos transientes permanecem no final da lista. Para cada pixel, as



Figura 3.4: Exemplo de ofuscamento: (a) Imagem de fundo (b) Diminuição do ganho devido ao ofuscamento.

primeiras B gaussianas são selecionadas como representantes do fundo, onde B é dado por

$$B = \underset{b}{\operatorname{argmin}} \left(\sum_{k=1}^b \omega_k > T \right) \quad (3.7)$$

onde T é um limiar definido empiricamente.

Esta modelagem possui diversas vantagens sobre técnicas mais simples. Uma delas é que, quando um objeto é modelado como pertencente ao fundo, ele não destrói o modelo anterior. Isto permite a recuperação rápida do modelo original do fundo, mesmo depois que um objeto inicialmente móvel fique estacionário durante um período. Um exemplo desta ocorrência seria um carro estacionando no campo de visão da câmera, e depois partindo novamente. Apenas quando uma gaussiana torna-se a gaussiana menos provável, entre as K utilizadas no modelo, e uma nova cor é observada é que o modelo anterior é perdido.

O resultado da subtração de fundo é uma máscara binária, onde 1 representa os objetos móveis, ou seja, a frente (*foreground*), e 0 representa o fundo (*background*). Todos os pixels que não casaram com nenhuma distribuição existente são pixels de frente. Pixels que casaram com alguma distribuição são classificados como fundo se o casamento tiver ocorrido com uma das B primeiras gaussianas do modelo naquela posição. Caso contrário, também serão classificados como frente.

O modelo da mistura de gaussianas permite uma representação robusta de uma cena, porém apresenta certas fraquezas. Como este modelo possui uma constante de tempo fixa, não reage a mudanças rápidas de iluminação, como a causada pela passagem de nuvens em um dia bem iluminado, ou o ofuscamento causado pela reflexão do sol nos veículos, como pode ser visto na Figura 3.4.

A Figura 3.4 mostra um exemplo real da influência do ofuscamento gerado pelo reflexo do sol, notando-se que ocorre um escurecimento do fundo quando da presença do veículo.

A alteração na cena é intensificada pelos sistemas de ajuste automático de ganho das câmeras digitais. Esses sistemas de compensação, apesar de prejudiciais à segmentação, são necessários para que a faixa dinâmica limitada dos sensores de imagem possam acomodar as diferentes intensidades luminosas presentes durante o dia. Portanto, para realizar o monitoramento por longos períodos sem supervisão, é necessário adicionar ao modelo parâmetros referentes à iluminação da cena. Nas seções seguintes serão apresentadas propostas para tornar o sistema aqui descrito robusto aos problemas introduzidos pela variação da iluminação da cena.

3.2.2 Supressão de Sombras

O modelo descrito anteriormente, proposto em (STAUFFER; GRIMSON, 1999) utiliza o espaço de cores RGB para realizar a segmentação de objetos. No espaço de cores RGB, uma imagem é representada pela sua intensidade em três faixas de cores (vermelho, verde e azul). Desta maneira, as informações de luminosidade e de cor estão presentes em todas as dimensões da representação. Para que o modelo de fundo possa se adaptar a diferentes condições luminosas como sombras e clarões, é necessário separar as duas informações. Um espaço de cores que realiza esta separação é o espaço YC_bC_r . Neste espaço, a informação do brilho(Y) é armazenada separadamente das componentes de diferença de cor do azul(C_b) e vermelho(C_r). O espaço de cores HSV representa uma imagem com base em sua tonalidade(H), saturação(S) e luminosidade(V). Este espaço de cores é baseado na percepção humana das cores. A tonalidade (*hue*) e a saturação representam a cor. A tonalidade indica a composição espectral aproximada da cor, enquanto que a saturação representa a pureza da cor. Uma saturação baixa indica uma cor neutra, com uma maior quantidade de branco. Já uma saturação alta indica uma cor mais forte, sem misturas. O espaço de cores HSV é usualmente representado em coordenadas cilíndricas, com a tonalidade representando o ângulo, a saturação o raio, e o brilho a altura.

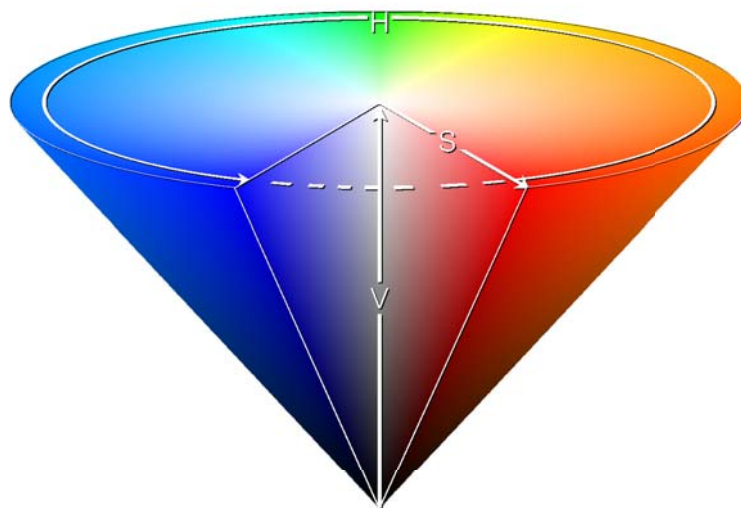


Figura 3.5: Representação cônica do espaço de cores HSV. Fonte: http://commons.wikimedia.org/wiki/File:HSV_cone.png

A Figura 3.5 mostra a representação do espaço de cores HSV como um cone, indicando a diminuição do espectro de cores à medida que a iluminação, representada pela componente V, aproxima-se de zero.

Neste espaço, há uma separação clara entre a cor, representada pelas componentes H e S, e a luminosidade, representada pela componente V.

Sombras são causadas pela obstrução parcial ou total da luz incidente em uma superfície, devido à presença de um outro objeto entre esta e uma fonte de luz. Uma região de sombra pode ser dividida, de acordo com suas características luminosas, em umbra e penumbra, conforme apresentado na Figura 3.6.

Na umbra, há uma ausência quase total de luz. É uma região de sombra com limites bem definidos. Já a região de penumbra é onde a luz é apenas parcialmente bloqueada pelo objeto, Figura 3.6(c). Como é

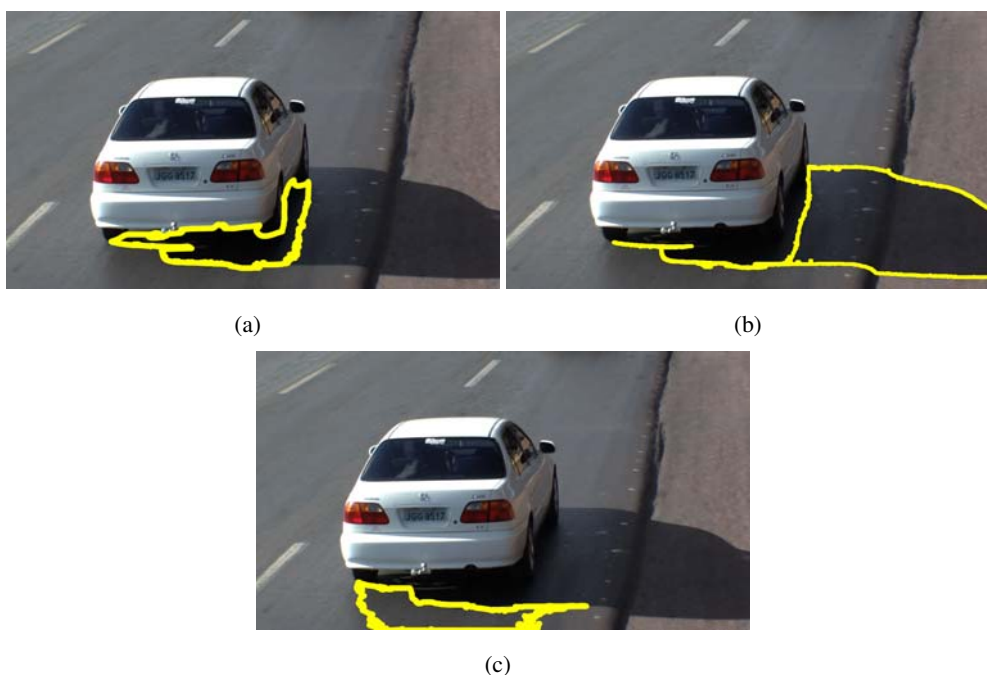


Figura 3.6: Classificação das sombras em uma imagem: (a) Umbra profunda (b) Umbra rasa (c) Penumbra.

parcialmente iluminada, não apresenta limites bem definidos, apresentando um gradiente entre a escuridão da umbra e a região iluminada. No caso específico do monitoramento de veículos, a umbra pode ser subdividida em umbra profunda e umbra rasa (WANG; TANG; ZHU, 2006). A umbra profunda é aquela que tipicamente se forma logo abaixo dos veículos, resultado do bloqueio tanto da luz do sol quanto da luz devida à refração atmosférica, Figura 3.6(a). A umbra rasa, apesar de apresentar limites definidos e a uniformidade da umbra, possui uma luminosidade um pouco maior devido à iluminação ambiental, Figura 3.6(b). A umbra rasa é a região da sombra projetada para fora da área do veículo, quando a luz solar incide em um ângulo mais agudo sobre o solo. A sombra de um objeto pode ainda ser classificada em dois tipos: sombras próprias e sombras projetadas. Sombras próprias são aquelas presentes no objeto devido à sua própria geometria. Sombras projetadas são aquelas causadas por um objeto em outros objetos da cena. A imagem percebida pelo sensor da câmera é o resultado da convolução da luz emitida pelas fontes de luz da cena com a refletância das superfícies dos objetos. A refletância de uma superfície é função do espectro luminoso da fonte de luz, porém não varia com a intensidade luminosa incidente. Dessa maneira, a presença de uma sombra causa uma mudança na iluminação recebida, em relação aos objetos já presentes na cena. Em regiões de penumbra, a fonte de iluminação é apenas parcialmente bloqueada. Portanto, pode-se considerar que apenas a intensidade luminosa varia significativamente, enquanto que a composição espectral da luz refletida sofre apenas pequenas alterações. Em cenas externas, pode haver um pequeno desvio no espectro, uma vez que a luz refratada pela atmosfera e refletida pelas nuvens forma uma fonte secundária de iluminação, com um espectro que tende ao azul (NADIMI; BHANU, 2004).

Para realizar a identificação e supressão das sombras, o modelo de fundo deve levar em conta a iluminação da cena. Para isso, o modelo de fundo deve ser gerado em um espaço de cores que diferencie entre luminância e crominância. Neste trabalho, foi escolhido o espaço de cores HSV, devido à sua aproximação da percepção humana. Para utilizar o modelo de mistura de Gaussianas com o espaço de cores HSV, algumas adequações são propostas. As três componentes do espaço HSV representam grandezas distintas.

Portanto, não é razoável supor que as três componentes possuam o mesmo desvio padrão. A matriz de covariância da k -ésima gaussiana passa a ser escrita como

$$\Sigma_{k,t} = \begin{bmatrix} \sigma_H^2 & 0 & 0 \\ 0 & \sigma_S^2 & 0 \\ 0 & 0 & \sigma_V^2 \end{bmatrix}, \quad (3.8)$$

onde σ_H , σ_S e σ_V são os desvios padrão das componentes H, S e V. A atualização dos desvios padrão é dada por

$$\sigma_{H_{k,t}} = \sqrt{(1 - \alpha)\sigma_{H_{k,t-1}}^2 + \alpha (X_{H_{k,t}} - \mu_{H_{k,t}})^2}, \quad (3.9)$$

$$\sigma_{S_{k,t}} = \sqrt{(1 - \alpha)\sigma_{S_{k,t-1}}^2 + \alpha (X_{S_{k,t}} - \mu_{S_{k,t}})^2}, \quad (3.10)$$

$$\sigma_{V_{k,t}} = \sqrt{(1 - \alpha)\sigma_{V_{k,t-1}}^2 + \alpha (X_{V_{k,t}} - \mu_{V_{k,t}})^2}, \quad (3.11)$$

onde $\sigma_{H_{k,t}}$, $\sigma_{S_{k,t}}$ e $\sigma_{V_{k,t}}$ são os desvios padrão correspondentes as componentes H, S e V da gaussiana k no tempo t , respectivamente. Da mesma maneira, $X_{H_{k,t}}$, $X_{S_{k,t}}$ e $X_{V_{k,t}}$ são os valores observados dos pixels, e $\mu_{H_{k,t}}$, $\mu_{S_{k,t}}$ e $\mu_{V_{k,t}}$ são as médias e α é a taxa de aprendizagem. Como os desvios padrão não são iguais, as gaussianas passam a ser ordenadas pelo valor de ω/σ_m , onde ω é o peso da gaussiana e σ_m é o desvio padrão médio, dado por

$$\sigma_{m_{k,t}} = \frac{\sigma_{H_{k,t}} + \sigma_{S_{k,t}} + \sigma_{V_{k,t}}}{3}, \quad (3.12)$$

onde $\sigma_{H_{k,t}}$, $\sigma_{S_{k,t}}$ e $\sigma_{V_{k,t}}$ são os desvios padrão correspondentes as componentes H, S e V da k -ésima no tempo t . As figuras 3.7 e 3.8 apresentam imagens representando o modelo de fundo de uma sequência de vídeo, após 3000 quadros.

Na Figura 3.7(a), estão representadas as médias de um modelo com 5 gaussianas. A imagem ao topo representa as médias das primeiras gaussianas de cada pixel, isto é, aquelas com o menor valor de ω/σ . As imagens abaixo desta representam as médias da segunda, terceira, quarta e quinta gaussiana, de cima para baixo. Na Figura 3.7(b) estão representados os pesos das gaussianas, organizados da mesma maneira. Os pesos estão representados em escala de cinza, com a cor preta representando o peso zero, e branco representando o peso um.

A Figura 3.8 apresenta a separação da média em suas componentes H, S e V. A média da tonalidade está colorida considerando as componentes S e V de valor igual a 1.0.

Pode-se notar que as médias das primeiras Gaussianas formam uma representação quase exata da estrada sendo monitorada. O pesos destas gaussianas é alto, indicando sua elevada relevância no modelo. As médias das segundas Gaussianas também formam uma imagem similar à pista, porém a representação é mais ruidosa. Seus pesos também são altos, em alguns lugares até mais altos do que os peso das primeiras Gaussianas. Isto indica que o modelo está realizando uma transição, possivelmente devido a uma pequena mudança na posição da câmera. As outras Gaussianas possuem médias representando apenas objetos móveis que passaram pela cena.

Com o modelo adaptado para separar as informações de crominância e luminância, é possível realizar a detecção e supressão de sombras. Conforme exposto anteriormente, a sombra causa uma mudança significativa no brilho(V), porém sem causar mudanças significativas na cor (H e S).

Na Figura 3.9 pode-se verificar o efeito de uma sombra projetada sobre o modelo de fundo HSV.

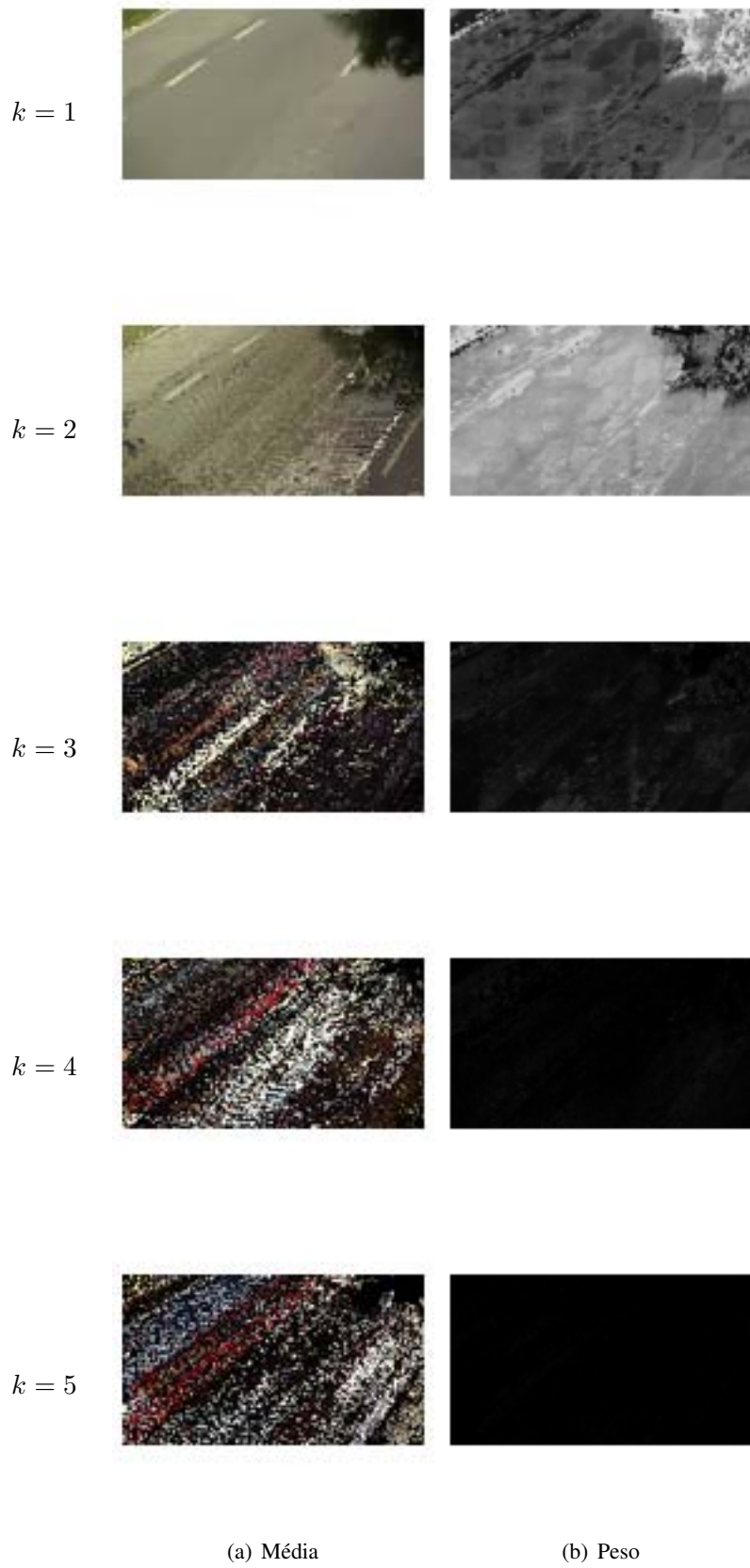


Figura 3.7: Modelo de fundo obtido a partir de uma sequência típica de vídeo de monitoramento. (a) $\mu_{k,t}$ (b) $\omega_{k,t}$.

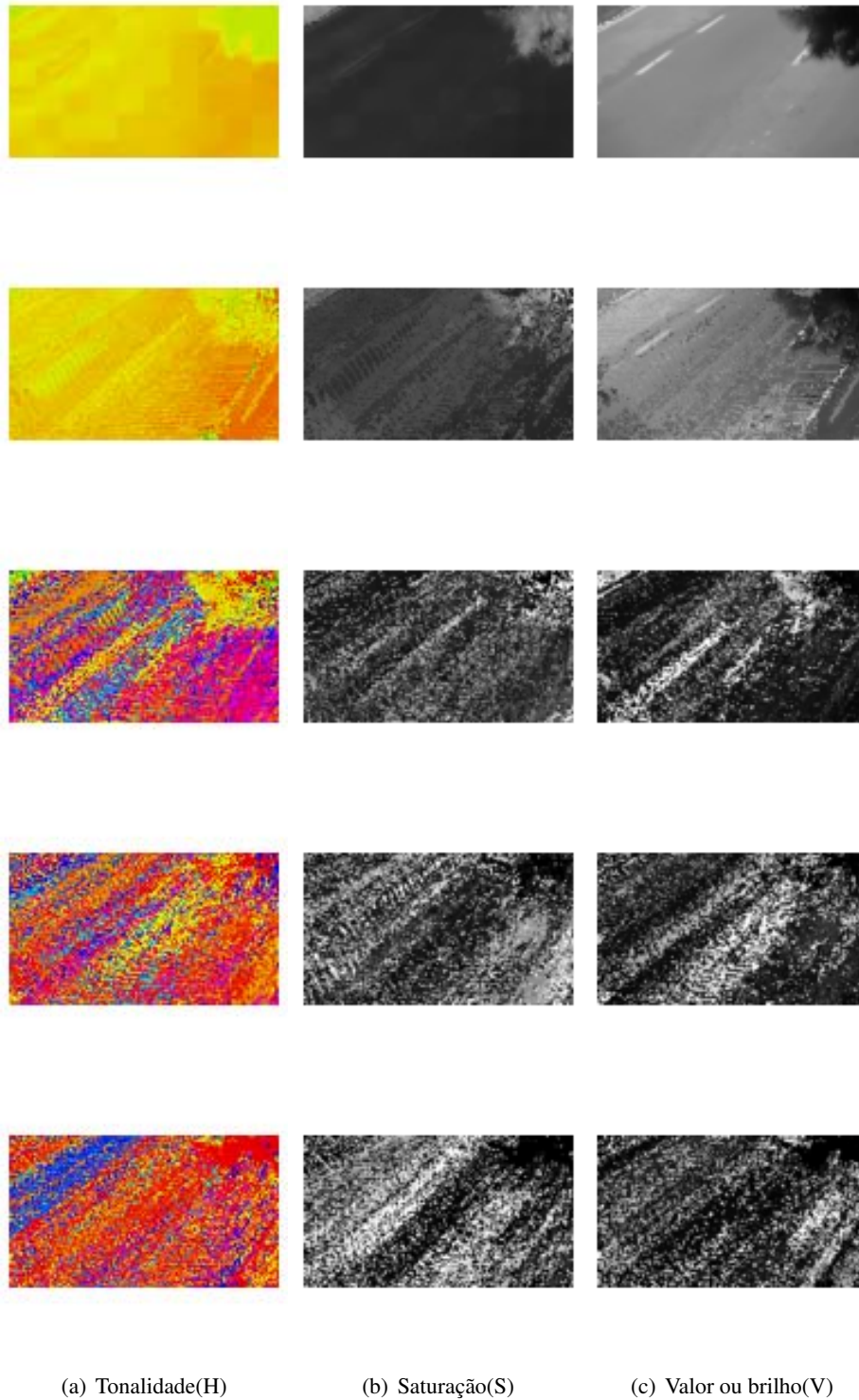


Figura 3.8: Modelo de fundo obtido a partir de uma sequência típica de vídeo de monitoramento.

A primeira característica a ser analisada é que sombras causam uma redução no brilho, em relação ao modelo de fundo (CUCCHIARA et al., 2001). Essa redução é analisada pela razão entre o Valor do quadro atual e as componentes de brilho das B Gaussianas que representam o fundo. A região só é considerada como uma possível sombra caso essa razão possua um valor entre dois limiares, o limite inferior V_l e o

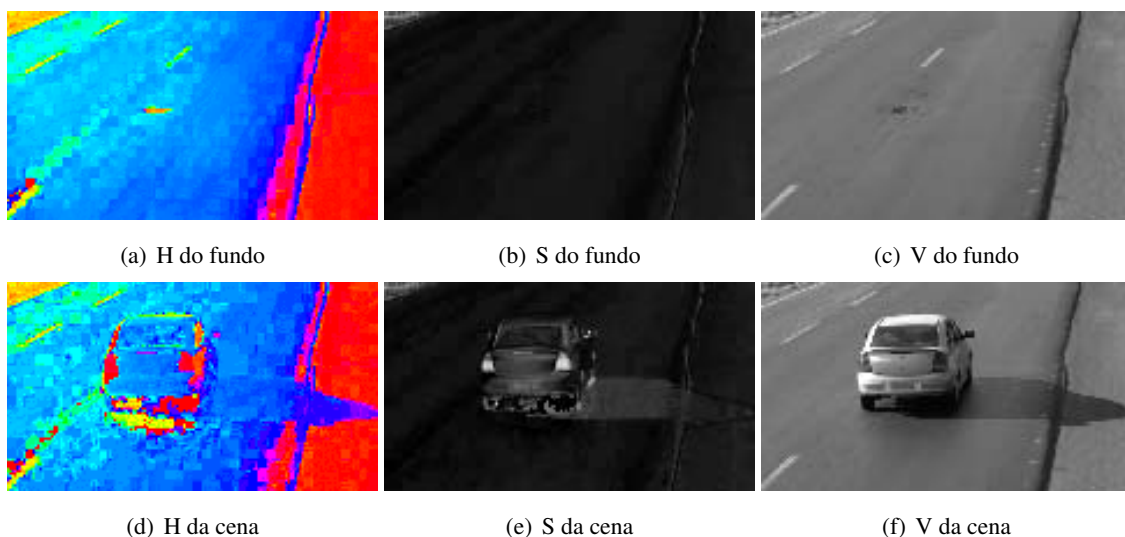


Figura 3.9: Efeitos da sombra projetada sobre a cena no espaço de cores HSV

limite superior V_h . Estes limiares, com valores entre 0 e 1, devem ser definidos de acordo com a cena. O limite superior V_h impede que pequenas variações naturais do brilho sejam classificadas como sombras. O limite inferior V_l leva em conta a intensidade e posição relativa da fonte de iluminação na cena.

O segundo critério analisado é a constância da cor. As componentes H dos possíveis pontos de sombra são comparados com os valores de tonalidade das Gaussianas de fundo. Caso se ajuste ao modelo, o candidato é então considerado como uma sombra válida.

É gerada então uma máscara binária, onde os pontos de sombra são representados por 1 (verdadeiro), e os outros pontos por 0 (falso), como pode ser visto na Figura 3.10.

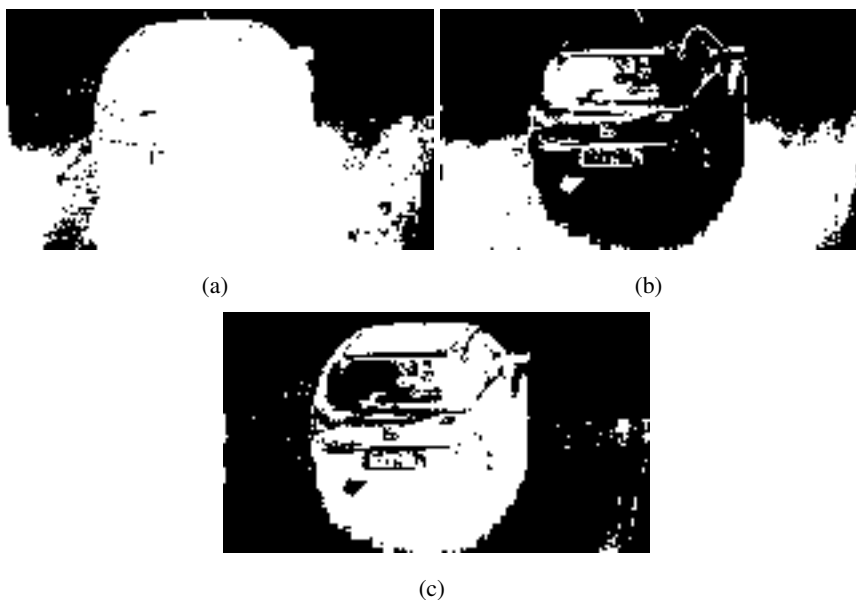


Figura 3.10: Máscara binária de sombras (a) Mascara resultante da subtração de fundo (b) Máscara resultante do detector de sombras (c) Resultado da combinação das duas máscaras

Esta máscara é então negada e combinada com a máscara resultante da subtração de fundo, através da

operação booleana AND. O resultado é uma máscara binária delimitando os objetos de interesse, porém sem as sombras que foram detectadas com sucesso, conforme pode ser visto na Figura 3.10(c).

3.2.3 Correção de Variações de Iluminação

As modernas câmeras digitais apresentam dispositivos para realizar o ajuste automático dos parâmetros da filmagem (ganho, balanço de branco, foco) de acordo com as condições de filmagem. Estes ajustes, pouco perceptíveis ao sistema visual humano, podem interferir nos sistemas automáticos de monitoramento, como o sistema proposto neste trabalho. Um dos sistemas de compensação cujos efeitos são mais perceptíveis é o ganho automático.

Os sensores das câmeras digitais são compostos de diferentes etapas (LITWILLER, 2001). Inicialmente, a irradiação luminosa incidente sobre o sensor é convertida em um sinal elétrico, proporcional à intensidade luminosa. Esse sinal é então amplificado, antes de ser convertido em um sinal digital. O processo de conversão do sinal analógico para digital envolve perdas, devido às limitações da representação numérica digital com um número finito de bits. Este fenômeno, chamado quantização, implica em uma perda de precisão e uma limitação da faixa dinâmica que o sinal digital pode representar. Essa limitação implica que em uma mesma cena, a câmera não é capaz de representar com exatidão intensidades luminosas muito grandes ou muito pequenas. O controle automático de ganho atua sobre a etapa de amplificação do sinal do sensor, buscando manter a amplitude do sinal de saída do sensor dentro da faixa dinâmica do conversor analógico-digital da câmera, e ocupando uma grande região desta faixa.

Com esta compensação, a sequência de vídeo não sofre perda de detalhes devido às mudanças de iluminação naturais resultantes da passagem do dia e da sombra causadas pelas nuvens. Porém, certos eventos podem ocasionar uma alteração indevida no ganho. Por exemplo, a passagem de um grande objeto escuro cobrindo parcialmente o campo de visão da câmera causa uma diminuição do nível global da brilho da cena, como pode ser visto na Figura 3.11.

Entre os quadros 256 e 355, um caminhão passa pela cena, com o lado sombreado voltado para a câmera. Isto causa uma grande queda no valor médio da componente V do quadro, alcançando um mínimo no quadro 297. Mesmo após a saída do caminhão da cena no quadro 355, apenas após o quadro 451 é que a média do brilho retorna ao seu valor anterior. Efeito semelhante também ocorre entre os quadros 870 e 945. Esta diminuição da iluminação incita uma resposta do controle automático de ganho, aumentando o ganho da câmera, como pode ser visto na Figura 3.12(b).

O ganho permanece mesmo após a saída do objeto da cena, como pode ser visto na Figura 3.12(c). Isto pode fazer com que o algoritmo considere temporariamente a cena inteira como um objeto, fazendo com que objetos se movendo pela cena durante essa interferência não sejam detectados corretamente. Durante filmagens sob uma iluminação muito intensa, reflexos podem causar o ofuscamento parcial da câmera e causar uma diminuição do ganho, gerando os mesmos problemas, como pode ser visto observando-se a região correspondente à via na Figura 3.4.

O ofuscamento e as correções de ganho interferem principalmente com o brilho da cena, não distorcendo significativamente as cores. Desta maneira, uma maneira simples de corrigir os problemas causados por essas variações seria ignorar a componente V da representação da sequência, e realizar a subtração de fundo utilizando apenas as componentes H e S. Porém, desta maneira a máscara resultante fica menos

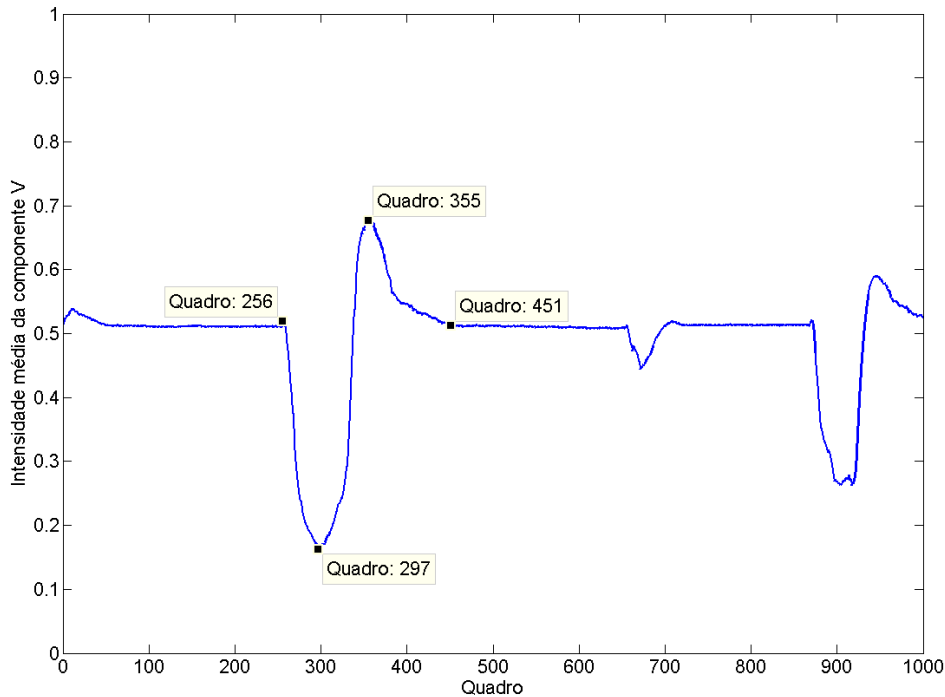


Figura 3.11: Nível médio da componente V da imagem, quadros 1 a 1000.

definida, e perde-se a principal referência para a supressão de sombras. Por esses motivos, propõe-se aqui um novo sistema para corrigir mudanças bruscas indevidas de iluminação.

As mudanças de iluminação causadas pelo ajuste automático de ganho são globais, isto é, afetam igualmente todos os pixels da cena. Portanto, é possível estimar essas mudanças observando-se qualquer região não saturada da imagem. Para escolher as regiões nas quais isto deve ser observado, deve-se mensurar a estabilidade da cena. Neste trabalho, a estabilidade de uma região refere-se à probabilidade de passagem de objetos por esta região. Regiões consideradas estáveis apresentam a imagem do fundo a maior parte do tempo, enquanto que regiões instáveis possuem uma grande presença de objetos móveis. As regiões estáveis são utilizadas para realizar o monitoramento das mudanças de brilho bruscas da cena. Uma equação para determinar a estabilidade da cena pode ser

$$S_{i,t} = (1 - \beta)S_{i,t-1} + \beta\overline{M_{i,t}}, \quad (3.13)$$

onde $\overline{M_{i,t}}$ é o complemento da máscara de movimento do quadro t na posição i desconsiderando a componente V, isto é, 1 para regiões de fundo e 0 para regiões de movimento, e β é a taxa de aprendizado da estabilidade, proporcional ao tempo esperado entre a presença de objetos. $S_{i,t}$ é o valor que representa a estabilidade do pixel i no tempo t . Para calcular esta estabilidade, a máscara de movimento é gerada verificando o casamento apenas das componentes H e S. Como estas componentes não são afetadas significativamente pelas mudanças do ganho, não há o risco de classificar erroneamente regiões onde ocorre apenas uma mudança de brilho como uma região de instabilidade. Aplicando-se um limiar simples, é possível obter uma máscara binária representando as regiões estáveis, isto é, que apresentam pouco ou nenhum movimento de objetos.

Para avaliar a influência do ganho na componente V do vídeo, foram acompanhados os valores desta



(a)



(b)



(c)

Figura 3.12: Efeitos do controle automático de ganho: (a) Quadro 52, anterior à passagem do objeto (b) Quadro 327, durante a passagem do objeto (c) Quadro 355, logo após a passagem do objeto.

ao longo do tempo em três pixels distintos: um pixel sob iluminação direta ao lado da pista (2), um pixel em uma região sombreada (1), e um pixel iluminado no meio da pista (3), em uma região de passagem de veículos, conforme mostrado na Figura 3.13.

Quando ocorrem mudanças na iluminação devido ao ajuste de ganho, o desvio da iluminação apresenta o mesmo comportamento em todos os pixels não-pertencentes a um objeto, como pode ser visto na Figura 3.14.

No quadro 259, ocorre o início da passagem de um caminhão pela cena. Este caminhão causa uma queda no nível global médio da componente V, conforme a Figura 3.11. Esta diminuição faz com que a câmera aumente o ganho, aumentando o valor dos pixels em que não há movimento. Pode-se notar que, em ambos os pixels a variação da componente V apresenta a mesma forma, variando apenas em amplitude. Após a passagem do caminhão pelo pixel da estrada, no quadro 341, este também apresenta uma variação



Figura 3.13: Pontos da cena avaliados

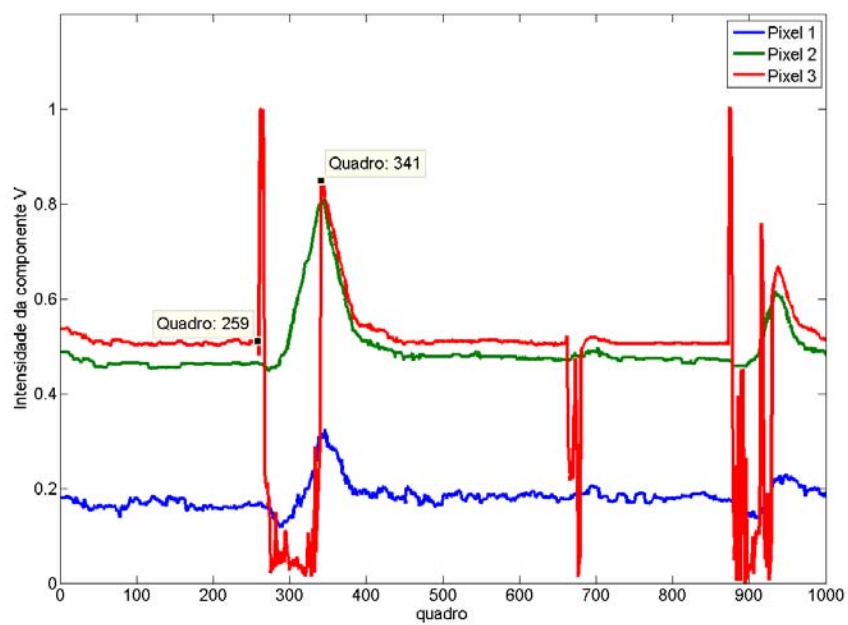


Figura 3.14: Comportamento da componente V da cena ao longo do tempo em três pontos distintos.

com o mesmo comportamento dos pixels estáveis, isto é, sem movimento. Normalizando os valores de V dos pixels pelas suas médias, percebe-se que a variação normalizada é aproximadamente a mesma para os pixels em que não há movimento, como pode ser visto na Figura 3.15.

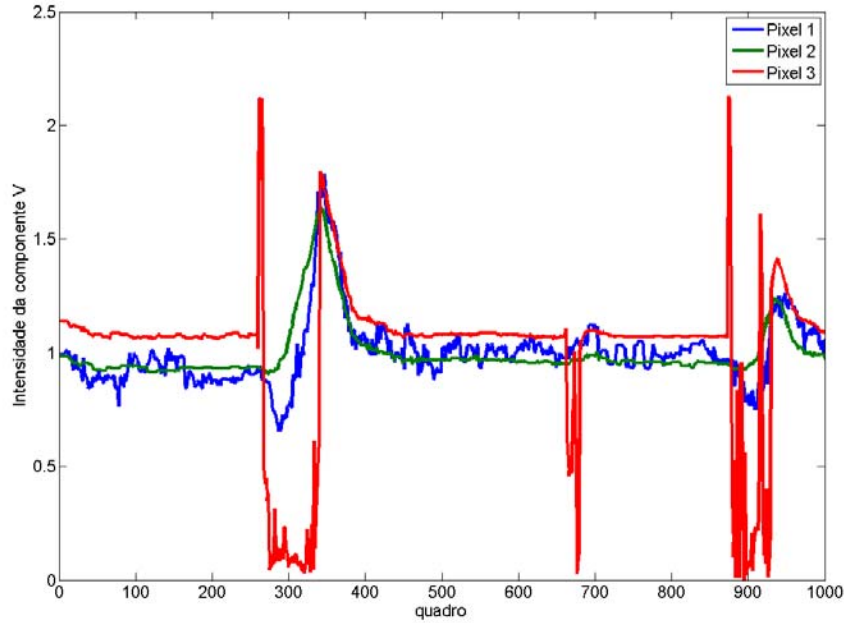


Figura 3.15: Valor de V , normalizado pela média do valor de V no pixel.

Dessa maneira, o fator de correção normalizado é obtido analisando-se o desvio da componente V em relação ao modelo de fundo, considerando apenas pixels estáveis. Uma mudança homogênea no desvio do brilho, normalizado pela média do brilho nos pixels, indica que a correção automática de ganho está atuando. Esta mudança pode ser detectada calculando-se a média da razão entre o valor do pixel e o fundo na região que é considerada estável.

Portanto, pode-se corrigir esta variação utilizando um fator de correção calculado a partir dos pixels estáveis, como

$$F_t = \frac{1}{n_E} \sum_E \frac{X_{V_{i,t}}}{\mu_{V_{i,t}}}, \quad (3.14)$$

onde F_t é o fator de correção de iluminação no tempo t , $X_{V_{i,t}}$ é o valor da componente V do pixel i no tempo t , $\mu_{V_{i,t}}$ é a média da primeira Gaussiana do pixel i no tempo t , E é o conjunto de pontos estáveis e n_E é quantidade de pontos pertencentes a este conjunto. O conjunto E é definido como

$$E = S_{i,t} | S_{i,t} > q, \quad (3.15)$$

onde $S_{i,t}$ é o valor da estabilidade do pixel i no tempo t , e q é o limiar de estabilidade. O valor corrigido da componente V do quadro é dado, portanto, por

$$X'_{V_{i,t}} = \frac{X_{V_{i,t}}}{F_t}, \quad (3.16)$$

onde $X'_{V_{i,t}}$ é o valor corrigido da componente V da imagem na posição i no instante t .

Como não há movimento nesta região, o fator de correção será aproximadamente igual a um durante a maior parte do tempo. Quando houver uma mudança global de iluminação, a média representará essa variação. Desta maneira, este valor pode ser utilizado como um fator de correção da luminância da cena. O modelo continua se adaptando ao valor não-corrigido do brilho, de maneira que, se a mudança de luminosidade for permanente (por exemplo, causada por uma nuvem), o modelo pode se adaptar às novas condições de luminosidade.

4 RESULTADOS EXPERIMENTAIS

4.1 INTRODUÇÃO

O sistema utilizado para obter as sequências de vídeo empregadas neste trabalho consiste de um suporte metálico, para posicionar a câmera de um ponto de vista privilegiado. A câmera utilizada é do modelo Samsung HMX 10A, com resolução 1280 x 720 pixels (HD, 720p) progressiva com captura a 60 quadros por segundo. Os vídeos são armazenados em cartões de memória SDHC (*Secure Digital High Capacity*) de 16 GB para processamento *offline*. Cada cartão é capaz de armazenar aproximadamente 3 horas de vídeo na qualidade mais alta. O suporte utilizado é um tripé metálico, construído especialmente para esta aplicação, mostrado na Figura 4.1.



Figura 4.1: O suporte utilizado para realizar as filmagens.

Este tripé é composto por uma base metálica de suporte, e seções tubulares de aproximadamente 1 metro de comprimento. Utilizando a base e quatro seções, o tripé permite realizar filmagens a uma altura de aproximadamente 5,20 m. A câmera é fixada ao topo do tripé através de uma chapa metálica, o que permite sua rotação em dois eixos, em relação ao eixo do tripé, conforme mostrado na Figura 4.2.

A subtração de fundo é uma ferramenta para a delimitação de objetos móveis em uma sequência de



Figura 4.2: A câmera utilizada em seu suporte.

vídeo. Como é, em geral, o primeiro passo em um sistema de visão computacional, não há necessidade que o processamento seja feito utilizando todo o nível de detalhamento espacial que o sistema de captura de imagens pode oferecer. Uma região móvel detectada grosseiramente em uma imagem de baixa resolução pode ser remapeada sobre a imagem original, reduzindo a carga computacional desta etapa de processamento. Por esse motivo, os testes da subtração de fundo foram realizados sobre versões reduzidas dos vídeos originais, redimensionados de 1280x720 para 160x90 pixels. A comparação entre os dois métodos de subtração de fundo, RGB sem correções originalmente proposto por (STAUFFER; GRIMSON, 1999) e o modelo ora proposto, utilizando HSV com correções, é realizada utilizando os mesmos parâmetros. A função densidade de probabilidade de cada pixel é representada pela mistura de 5 gaussianas. O limiar de probabilidade do fundo é definido como 0,7, e o limiar de casamento das gaussianas é de 2,5 desvios padrão, conforme originalmente propostos em (STAUFFER; GRIMSON, 1999). A constante de tempo τ é definida como 300 quadros, equivalente a 5 segundos de gravação, portanto a taxa de aprendizado α é igual a $1/300$. A constante de tempo τ representa a quantidade aproximada de dados de entrada que são considerados no modelo de fundo. Por exemplo, com a constante de tempo em 300 quadros, um objeto inicialmente móvel deve permanecer aproximadamente esta quantidade de tempo estático na cena antes de começar a ser considerado como parte do fundo. Para os experimentos deste trabalho, foram utilizadas filmagens realizadas na rodovia BR-450, em locais sem semáforos ou outros impedimentos ao fluxo de veículos. Dessa maneira 300 quadros, equivalentes a 5 segundos em uma filmagem a 60 quadros por segundo, é um tempo suficiente para evitar que os veículos sejam absorvidos pelo modelo de fundo indevidamente. Em monitoramentos urbanos, é esperado que os veículos permaneçam na cena por mais tempo, devido a manobras, sinalização e pedestres. Nesse caso, a constante de tempo do sistema deve ser ajustada para evitar alterações errôneas no modelo de fundo.

Os métodos testados foram implementados utilizando a ferramenta MATLAB® 7.8.0.347 (R2009a) 64 bits, devido à rapidez de desenvolvimento e a flexibilidade para análise do funcionamento interno dos algoritmos. Os testes foram executados em um computador com processador Intel®Core™2 Quad Q8200 2,33 GHz com 4 GiB de memória RAM.

4.2 RESULTADOS

4.2.1 Supressão de Sombras

A supressão de sombras permite que os objetos móveis sejam segmentados de maneira mais precisa, reduzindo a probabilidade de ocorrência de sobreposição entre as áreas detectadas de dois ou mais objetos móveis. Desta maneira, aumenta-se a precisão das etapas subsequentes, o rastreamento e a contagem de objetos.

Para a análise da supressão de sombras, foram utilizados 3 vídeos de 20 minutos de duração cada. Destes vídeos, foram escolhidos 412 quadros, contendo veículos inteiros. Os veículos foram segmentados manualmente. A partir desta segmentação, foram geradas máscaras binárias, representando as áreas de referência para a detecção de objetos. A Figura 4.3 mostra exemplos de veículos segmentados manualmente e suas máscaras de referência.

Foram escolhidas duas figuras de mérito para avaliação dos resultados. Primeiramente, mensurou-se a exatidão na definição do veículo. Isto foi feito analisando-se a área comum entre a máscara gerada pelo sistema e a máscara gerada manualmente, de acordo com

$$T = \frac{\sum_i M_i R_i}{\sum_i R_i} \times 100\%, \quad (4.1)$$

onde M_i representa a máscara gerada pelo sistema, R_i a máscara de referência, e T o percentual da área do objeto que foi efetivamente detectada pelo sistema. Também foi analisada a taxa de vazamento, isto é, quanto da área detectada pelo sistema não pertence ao veículo. Esta análise foi feita de acordo com

$$F = \frac{\sum_i M_i (1 - M_i R_i)}{\sum_i R_i} \times 100\%, \quad (4.2)$$

onde F é a área detectada pelo sistema que não pertence ao carro, relativa à área de referência do carro.

A Tabela 4.1 apresenta uma comparação dos resultados obtidos dos sistemas de detecção de objetos usando o método originalmente proposto por (STAUFFER; GRIMSON, 1999) e modificado por (KAEWTRAKULPONG; BOWDEN, 2001) e a metodologia com supressão de sombras proposta neste trabalho.

Tabela 4.1: Resultados da detecção de sombras

Método	Original	Proposto
Acerto(T)	34,76%	36,63%
Vazamento(F)	71,38%	42,65%

Da Tabela 4.1 podemos observar que ambas as técnicas apresentam taxas de acerto médias similares. Apesar das taxas de acerto serem baixas, esses valores não implicam em uma baixa qualidade de segmentação. Observando-se resultados da segmentação, pode-se ver que, apesar da máscara gerada apresentar falhas, a forma do objeto é preservada. Áreas internas de objetos que não são detectadas não afetam o processo de segmentação, porém diminuem a taxa de acerto aparente do sistema, como pode ser visto. A taxa

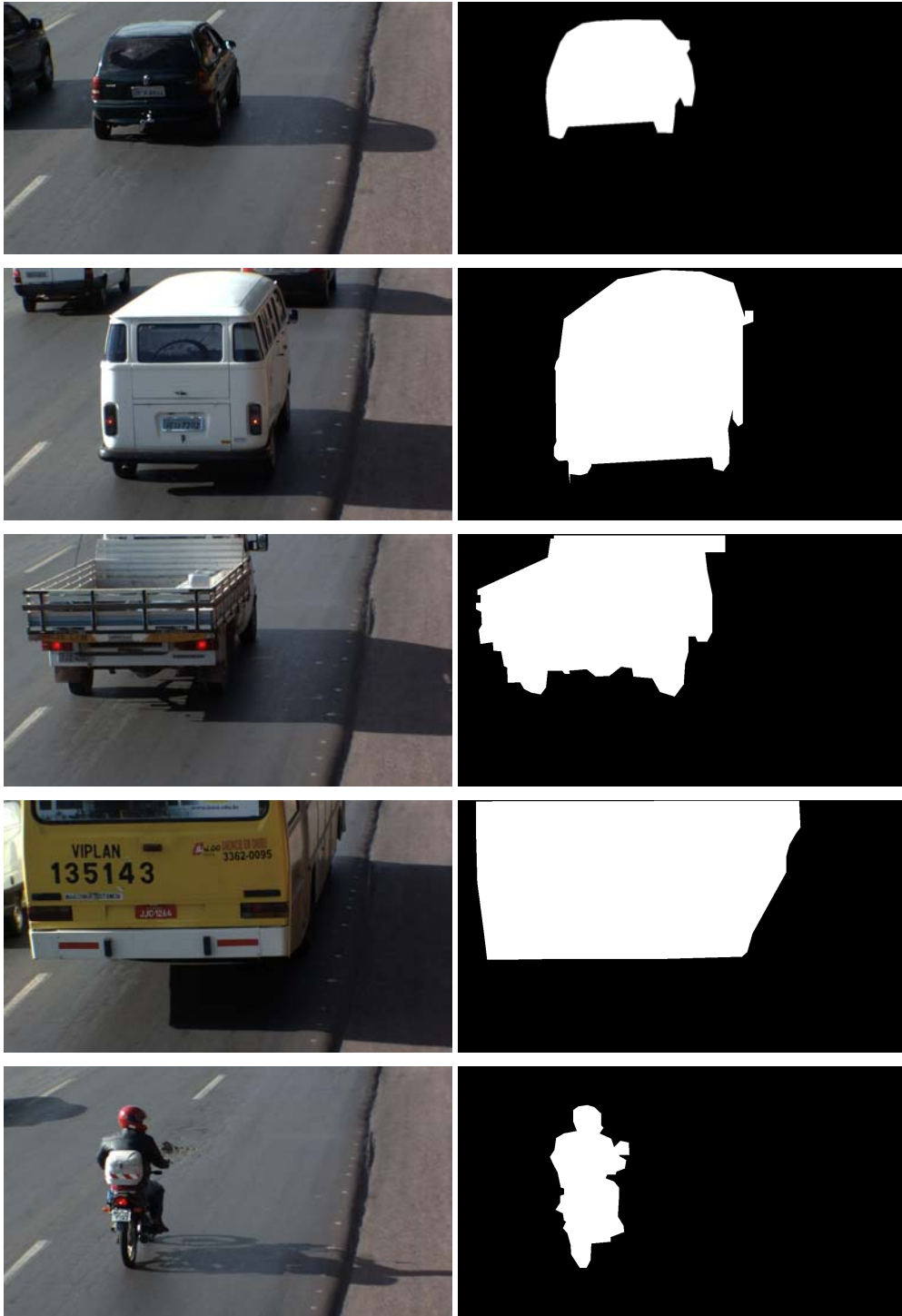


Figura 4.3: Vehículos segmentados manualmente.

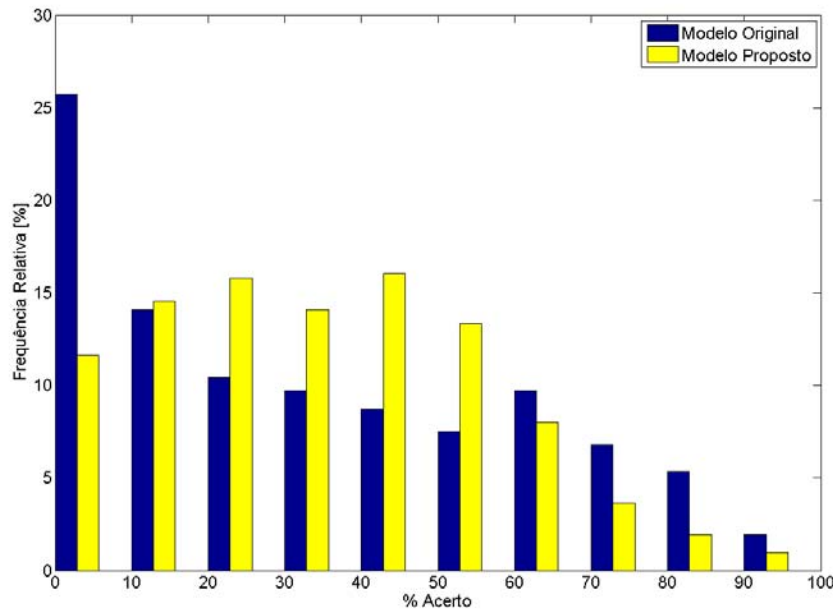


Figura 4.4: Histogramas de acerto das duas técnicas testadas. As barras do histograma representam a quantidade relativa de casos de teste com a taxa de acerto na faixa representada pelo eixo x.

de vazamento da metodologia proposta é significativamente menor que a do método original, mostrando a eficácia do método de supressão de sombras proposto.

Uma análise mais detalhada desses resultados pode ser feita através da análise dos histogramas representando a distribuição das taxas de acerto e vazamento, vistos nas figuras 4.4 e 4.5.

Pela Figura 4.4, pode-se notar que, apesar de possuírem médias de acerto similares, a proposta deste trabalho apresenta uma maior consistência, isto é, os resultados estão mais próximos da média. A técnica original apresenta uma grande quantidade de acertos entre 0 e 10%, ou seja, apenas uma pequena parte do carro é detectada.

A Figura 4.5 indica que o método proposto é capaz de diminuir o vazamento em relação a técnica original. O vazamento da técnica proposta encontra-se entre 0 e 50%, enquanto que a técnica original apresenta uma quantidade significativa de vazamentos acima de 100%.

A detecção de sombras é mais eficiente para delinear a umbra rasa, isto é, a sombra bem definida e parcialmente iluminada. Nas regiões de umbra profunda, há uma ausência quase total de iluminação, o que causa uma distorção na cor da região. Porém, como a umbra profunda só é projetada próxima ao objeto em cenas externas, sua detecção não é importante para a separação de objetos móveis. A penumbra próxima ao veículo também apresenta distorções cromáticas, que dificultam sua efetiva localização.

As áreas de objeto e sombra podem ser refinadas através do uso de operações morfológicas, como erosão e dilatação (GONZALEZ; WOODS, 2002), uma vez que existe uma correlação espacial entre os pontos de objeto e de sombra. A subtração de fundo está sujeita a ruídos aleatórios, advindos tanto da própria cena quanto devidos ao processo de aquisição e codificação do vídeo. O ruído natural, devido ao processo de aquisição de dados, é em sua maioria modelado pelo desvio padrão das gaussianas. Porém, caso o

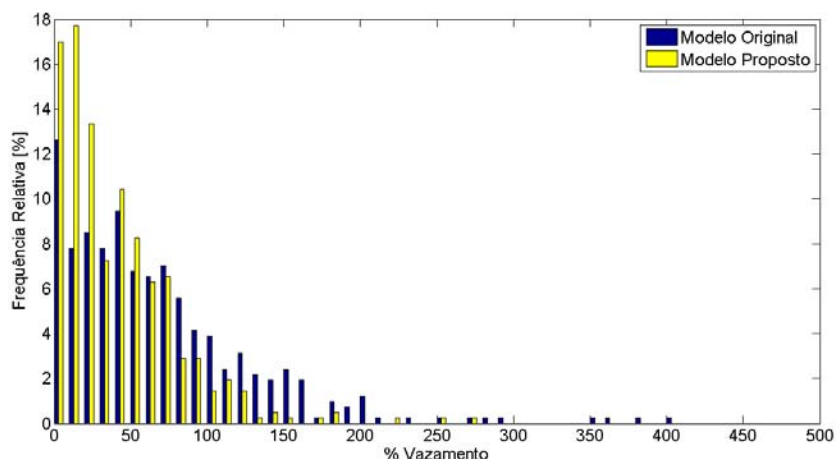


Figura 4.5: Histogramas de vazamento das duas técnicas testadas. As barras do histograma representam a quantidade relativa de casos de teste com a taxa de vazamento na faixa representada pelo eixo x.

ruído não seja gaussiano, é possível que ocasionalmente pontos sejam classificados como frente, mesmo na ausência de qualquer perturbação significativa na cena. Esse ruído se manifesta na máscara de saída como pontos desconexos, similar ao ruído *salt and pepper*, conforme mostrado na Figura 4.6.



Figura 4.6: Exemplo de uma máscara gerada pela subtração de fundo afetada pelo ruído de aquisição.

Como os pontos são isolados, a aplicação do filtro de erosão com um *kernel* quadrado elimina este ruído (GONZALEZ; WOODS, 2002). Nos testes realizados, foi utilizado um filtro de abertura (erosão seguida de dilatação) para diminuir o ruído, obtendo resultados semelhantes aos apresentados na Figura 4.7.

Um ruído mais destrutivo ao processo é o causado pelo efeito de blocagem dos processos de codificação de vídeo. A codificação de vídeo busca reduzir a quantidade necessária de dados ao armazenamento ou transmissão de uma sequência através da eliminação de redundâncias e de informações perceptivamente insignificantes. A visão humana possui uma maior capacidade de perceber diferenças de intensidade do que diferenças de coloração. Portanto, um passo comum em codificadores de vídeo, tais como MPEG2 e MPEG4 (SAYOOD, 2000), é sub-amostrar as componentes de crominância da imagem, preservando a resolução da componente de luminância. A consequência disso é o aparecimento do efeito de blocagem, isto é, a presença de blocos homogêneos de tamanho definido na imagem. Mesmo que os blocos não sejam perceptíveis na imagem codificada no espaço de cores RGB, sua presença é facilmente identificada quando



Figura 4.7: Exemplo de filtragem do ruído de aquisição presente na máscara obtida da subtração de fundo.

a imagem é transformada para o espaço HSV, principalmente na componente de tonalidade(H), como pode ser visto na Figura 4.8.

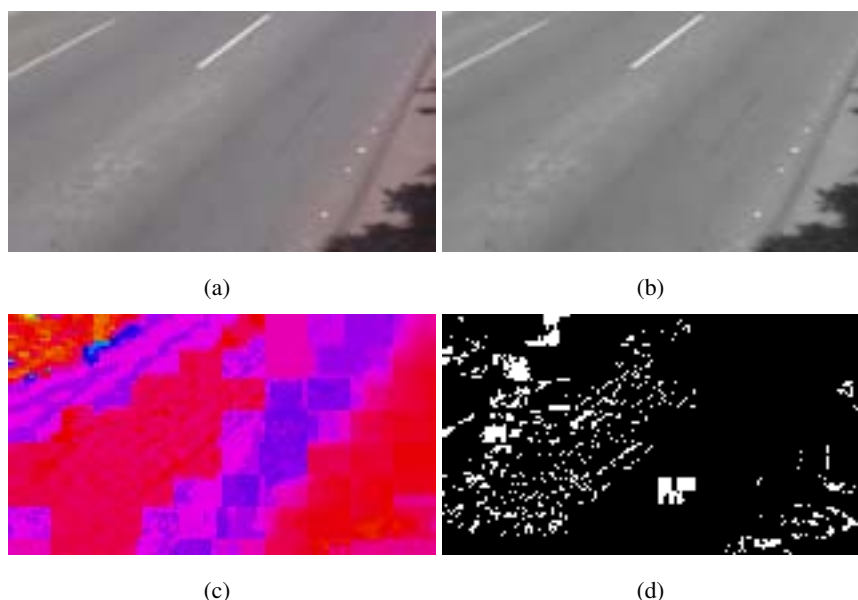


Figura 4.8: Quadro de vídeo após codificação (a) Imagem colorida (b) Brilho(V) (c) Tonalidade(H) (d) Exemplo de máscara sob efeito de bloqueio.

Este ruído, apesar de aleatório com relação ao tempo, possui uma grande correlação espacial, afetando blocos de pixels. A interferência causada na máscara por este tipo de ruído é mais difícil de ser compensada, uma vez que os falsos objetos apresentam-se como uma ou mais regiões conexas, ao invés de pontos espalhados pela imagem, conforme apresentado na Figura 4.8(d).

4.2.2 Robustez à Variação Global de Iluminação

Durante os testes do sistema, foram observadas várias situações em que atua o ajuste automático de ganho da câmera. A principal situação em que este ajuste ocorre é durante a passagem de objetos grandes, com relação ao campo visual da câmera. A variação resultante faz com que a imagem inteira seja

considerada como um objeto por alguns quadros, mesmo após a saída do objeto de cena. Dada a dificuldade de obtenção de uma medida objetiva deste fenômeno, é feita uma análise subjetiva dos resultados da compensação das variações de iluminação.

Foram analisadas duas sequências de vídeo em que observou-se a ocorrência frequente de variações do ganho automático. Na Figura 4.9 pode-se ver uma comparação do resultado da subtração de fundo sem e com a compensação das mudanças de iluminação.

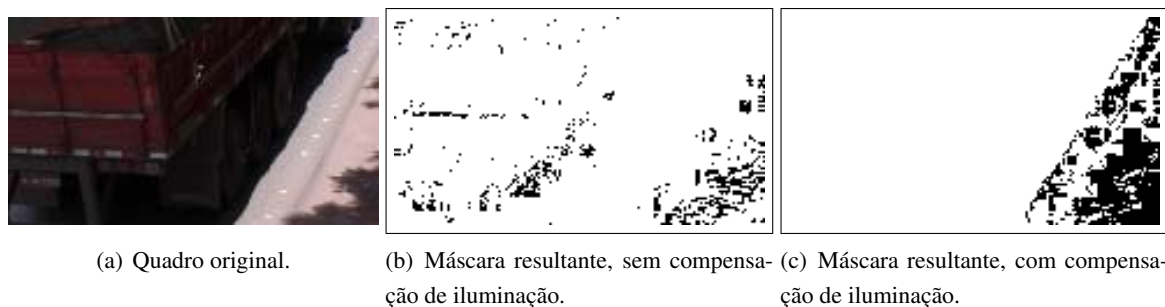


Figura 4.9: Compensação de mudanças de iluminação durante a passagem de um caminhão.

A Figura 4.9(a) apresenta um quadro em que ocorre a compensação automática de ganho. Nas figuras 4.9(b) e 4.9(c), estão representadas as máscaras resultantes da subtração de fundo sem e com a compensação de iluminação, respectivamente. Pode-se notar que em 4.9(c) apenas o objeto móvel está em destaque, enquanto que em 4.9(b) toda a cena é selecionada como frente.



Figura 4.10: Compensação de mudanças de iluminação após a passagem do móvel.

Na Figura 4.10, pode-se ver a recuperação da cena com a compensação ativada. Na Figura 4.10(b), apesar de o objeto já haver passado pela cena, o ganho da câmera ainda está elevado. Na Figura 4.10(c), pode-se ver que o fator de correção, apesar de não eliminar completamente este efeito, o diminui consideravelmente.

As sequências de vídeo originais e processados podem ser vistas através do site <http://www.cic.unb.br/~lamar/ftp/samtra>.

Utilizando as técnicas propostas, foi possível realizar a detecção e supressão de sombras móveis presentes na cena, sem afetar negativamente a detecção de objetos móveis. Além disso, foi possível detectar com sucesso variações no ganho da câmera, permitindo a compensação deste, aumentando a robustez do sistema a variações bruscas de iluminação.

5 CONCLUSÕES

Este trabalho apresenta parte de um sistema de rastreamento automático de veículos por visão computacional. O foco foi dado à etapa de segmentação, e como torná-la robusta às variações das condições ambientais. O sistema de segmentação utilizado se baseia em técnicas de subtração de fundo, uma vez que as filmagens são realizadas de uma posição fixa em relação à pista.

Neste trabalho foi proposto um sistema de subtração de fundo baseado na modelagem da cor dos pixels por mistura de gaussianas, invariante a iluminação e com supressão de sombras. O sistema foi adaptado para trabalhar com o espaço de cores HSV, o qual fornece uma separação entre luminosidade (V) e cor (H e S). Um sistema baseado em dois critérios: cor e intensidade, foi utilizado para classificar os pontos detectados pela subtração de fundo como sombra ou objeto. Também foi proposto um método para compensar o ajuste automático de ganho, necessário para filmagens em condições variáveis de iluminação. A influência da alteração de ganho na cena foi avaliada. Para poder separar variações devido ao ganho de variações causadas pela passagem de objetos, foi introduzido o conceito de estabilidade da cena. Baseado nesse conceito de estabilidade, percebeu-se que é possível estimar o ganho da câmera considerando o modelo já construído do fundo. Com base no ganho estimado foi possível realizar uma compensação dos seus efeitos sobre o processo de segmentação.

A eficiência da subtração de fundo com detector de sombras foi analisada de maneira objetiva, obtendo-se uma taxa de acerto de 36,63%, com uma taxa de vazamento (falsos positivos) de 42,65%. A saída do sistema sem e com o detector de sombras foi comparada com veículos segmentados manualmente. A taxa de vazamento de 42,65% com o sistema de supressão de sombras é bastante reduzida, quando comparada à taxa de 71,38% obtida sem o detector de sombras.

A robustez a variações de iluminação foi analisada de maneira subjetiva. Observando as sequências de vídeo processadas, percebe-se que a correção aplicada, mesmo não eliminando completamente os efeitos do ajuste automático de ganho, diminui consideravelmente suas consequências. Utilizando o fator de correção, é possível detectar, mesmo que sujeito a erros, objetos móveis que surjam durante o período de recuperação do ganho, isto é, o período entre a saída da causa do ajuste de ganho e a volta do ganho ao seu valor anterior. Foi demonstrado, portanto, que utilizando técnicas simples é possível mitigar as influências de variações da iluminação ambiente ao sistema de monitoramento.

Para continuação deste trabalho, existem alguns pontos que devem ser analisados mais profundamente. O espaço de cores HSV apresenta uma característica cilíndrica, porém foi utilizado no modelo proposto de subtração de fundo de forma retangular. Talvez, analisando-se as relações entre as diferentes componentes deste espaço de cores, seja possível montar um modelo mais robusto para detecção de sombras, e até para compensação de iluminação. Objetos e sombras apresentam visualmente uma grande correlação espacial interna. Este fato pode ser explorado para a montagem de um modelo de fundo mais completo, e para delimitar objetos e sombras de maneira mais exata. Para aumentar a robustez às variações de iluminação, devem ser buscados estimadores mais eficientes tanto para a estabilidade da cena quanto para a variação do brilho na região estável da cena. A subtração de fundo por misturas de gaussianas é robusta, porém depende da definição de diversos parâmetros, os quais são função das condições da cena e do objetivo do

monitoramento. Nos sistemas que utilizam a subtração de fundo, a análise da cena é feita posteriormente, em etapas com maior conhecimento sobre o que está sendo estudado. Utilizando estas informações, pode ser possível realizar o ajuste dos parâmetros da subtração de fundo de forma automática, diminuindo a necessidade de ajustes manuais do sistema.

REFERÊNCIAS BIBLIOGRÁFICAS

- CUCCHIARA, R. et al. Improving shadow suppression in moving object detection with hsv color information. *Intelligent Transportation Systems, 2001. Proceedings. 2001 IEEE*, p. 334–339, 2001.
- CUCCHIARA, R.; PICCARDI, M. Vehicle detection under day and night illumination. In: *Proceedings of the Third International Symposia on Intelligent Industrial Automation and Soft Computing*. [S.l.: s.n.], 1999.
- DEMPSTER, A. et al. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, Royal Statistical Society, v. 39, n. 1, p. 1–38, 1977.
- FRIEDMAN, N.; RUSSELL, S. Image segmentation in video sequences: A probabilistic approach. In: *Proc. of the Thirteenth Conference on Uncertainty in Artificial Intelligence (UAI 97)*. [S.l.]: Morgan Kaufmann, 1997. p. 175–181.
- GONZALEZ, R.; WOODS, R. *Digital image processing*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 2002.
- JACQUES, J. C. S.; JUNG, C. R.; MUSSE, S. R. Background subtraction and shadow detection in grayscale video sequences. *Computer Graphics and Image Processing, 2005. SIBGRAPI 2005. 18th Brazilian Symposium on*, p. 189–196, October 2005. ISSN 1530-1834.
- KAEWTRAKULPONG, P.; BOWDEN, R. An improved adaptive background mixture model for real-time tracking with shadow detection. In: *2nd European Workshop on Advanced Video Based Surveillance Systems*. [S.l.]: Kluwer Academic Publishers, 2001.
- KUMAR, P.; SENGUPTA, K.; LEE, A. A comparative study of different color spaces for foreground and shadow detection for traffic monitoring system. *Intelligent Transportation Systems, 2002. Proceedings. The IEEE 5th International Conference on*, p. 100–105, 2002.
- KYO, S. et al. A robust vehicle detecting and tracking system for wet weather conditions using the imap-vision image processing board. *Intelligent Transportation Systems, 1999. Proceedings. 1999 IEEE/IEEJ/JSAI International Conference on*, p. 423–428, 1999.
- LAI, A. H. S.; YUNG, N. H. C. A fast and accurate scoreboard algorithm for estimating stationary backgrounds in an image sequence. *Circuits and Systems, 1998. ISCAS '98. Proceedings of the 1998 IEEE International Symposium on*, v. 4, p. 241–244, 1998.
- LEE, I.; KO, H.; HAN, D. K. Multiple vehicle tracking based on regional estimation in nighttime CCD images. *Acoustics, Speech, and Signal Processing, 2002. Proceedings. (ICASSP '02). IEEE International Conference on*, v. 4, p. IV–3712–IV–3715, 2002. ISSN 1520-6149.
- LEONE, A.; DISTANTE, C.; BUCCOLIERI, F. A texture-based approach for shadow detection. *Advanced Video and Signal Based Surveillance, 2005. AVSS 2005. IEEE Conference on*, p. 371–376, 2005.

- LIM, S. et al. Fast illumination-invariant background subtraction using two views: Error analysis, sensor placement and applications. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005. CVPR 2005*. [S.l.: s.n.], 2005. v. 1.
- LITWILLER, D. CCD vs. CMOS: Facts and fiction. *Photonics Spectra*, Laurin publishing, Janeiro 2001.
- MAGALHÃES, H. A. *Análise em alta Resolução de Perfis Magnéticos de Sensores a laço indutivo e metodologia de fusão com características de Imagem para Classificação de Veículos Automotores*. Tese (Doutorado) — Universidade Federal de Minas Gerais, Belo Horizonte, 2008.
- MATSUYAMA, T. et al. Background subtraction under varying illumination. *Systems and Computers in Japan*, Wiley Subscription Services, Inc., A Wiley Company Hoboken, v. 37, n. 4, 2006.
- NADIMI, S.; BHANU, B. Physical models for moving shadow and object detection in video. *Transactions on Pattern Analysis and Machine Intelligence*, v. 26, n. 8, p. 1079–1087, August 2004. ISSN 0162-8828.
- NAM, M. Y.; RHEE, P. K. An efficient face recognition for variant illumination condition. *Intelligent Signal Processing and Communication Systems, 2004. ISPACS 2004. Proceedings of 2004 International Symposium on*, p. 111–115, November 2004.
- NORIEGA, P.; BERNIER, O. Real time illumination invariant background subtraction using local kernel histograms. *British Machine Vision Association (BMVC)*, 2006.
- PINTO, A. B. et al. Matriz de deslocamentos a partir de amostra de leitura de placas e contagem volumétrica em campo. In: *X ANPET - Congresso de Pesquisa e Ensino em Transportes, Anais do*. Brasília: ANPET (Associação Nacional de Pesquisa e Ensino em Transportes), 1996. v. 1, p. 179–188.
- PRATI, A. et al. Analysis and detection of shadows in video streams: a comparative evaluation. *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, v. 2, p. II–571–II–, 2001. ISSN 1063-6919.
- PRATI, A. et al. Shadow detection algorithms for traffic flow analysis: a comparative study. *Intelligent Transportation Systems, 2001. Proceedings. 2001 IEEE*, p. 340–345, 2001.
- PRATI, A. et al. Detecting moving shadows: algorithms and evaluation. *Transactions on Pattern Analysis and Machine Intelligence*, v. 25, n. 7, p. 918–923, July 2003. ISSN 0162-8828.
- RIDDER, C.; MUNKELT, O.; KIRCHNER, H. Adaptive background estimation and foreground detection using kalman-filtering. In: *Proceedings of International Conference on recent Advances in Mechatronics*. [S.l.: s.n.], 1995. p. 193–199.
- SALVADOR, E.; CAVALLARO, A.; EBRAHIMI, T. Cast shadow segmentation using invariant color features. *Computer Vision and Image Understanding*, v. 95, n. 2, p. 238–259, August 2004.
- SAYOOD, K. *Introduction to data compression*. 2ª. ed. San Diego: Morgan Kaufmann, 2000.
- STAUFFER, C.; GRIMSON, W. E. L. Adaptive background mixture models for real-time tracking. In: *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on*. [S.l.: s.n.], 1999. v. 2, p. 252 Vol. 2.

WANG, Y.; TANG, M.; ZHU, G. An improved cast shadow detection method with edge refinement. *Intelligent Systems Design and Applications, 2006. ISDA '06. Sixth International Conference on*, v. 2, p. 794–799, October 2006.

XU, D. et al. Insignificant shadow detection for video segmentation. *Circuits and Systems for Video Technology, IEEE Transactions on*, v. 15, n. 8, p. 1058–1064, August 2005. ISSN 1051-8215.

ZHANG, W. et al. Moving cast shadows detection using ratio edge. *Multimedia, IEEE Transactions on*, v. 9, n. 6, p. 1202–1214, October 2007. ISSN 1520-9210.