

Sistema de Rastreamento de Objetos Através de Múltiplas Câmeras

André Luís Beling da Rosa¹, Dalton Solano dos Reis¹,
Paulo César Rodacki Gomes^{1,3}, Mauricio Edgar Stivanello²

¹ Departamento de Sistemas e Computação, FURB - Universidade Regional de Blumenau
CAMPUS I / FURB, 89.012-900, Blumenau, SC, BRASIL

² Departamento de Automação e Sistemas, Universidade Federal de Santa Catarina
DAS / CTC / UFSC, 88.040-900, Florianópolis, SC, BRASIL

³ Unidade Urbana, Campus Rio do Sul, Instituto Federal Catarinense
IFC, 89.160-000, Rio do Sul, SC, BRASIL

andreluisbelingdarosa@gmail.com, {dalton, rodacki}@inf.furb.br, maustiva@das.ufsc.br

Resumo. *A automação de Sistemas de Inspeção Inteligentes tem se mostrado cada vez mais necessária, e as técnicas de visão computacional podem ser usadas para suprir esta demanda. Neste trabalho é apresentado o desenvolvimento de um sistema de rastreamento capaz de seguir objetos através de imagens providas por várias câmeras. O sistema foi desenvolvido empregando segmentação através do algoritmo NHD, rastreamento através de Fluxo Óptico, e para o tratamento de múltiplas câmeras é proposto um sistema global de controle. Para validação, um ambiente virtual tridimensional foi desenvolvido. Os resultados obtidos empregando o sistema descrito demonstram a efetividade da abordagem proposta, principalmente demonstrando a sua simplicidade conceitual e testes relacionados a performance em termos de tempo de processamento.*

1. Introdução

Os sistemas de vigilância vem sendo cada vez mais utilizados, seja no monitoramento de tráfego urbano, sistemas de segurança comercial, predial e doméstico. Através de câmeras instaladas ao longo de vias ou ambientes, sinais de vídeo são capturados e utilizados para monitoramento e extração de estatísticas diversas.

Na sua grande maioria, estes sinais de vídeo são monitorados por operadores humanos. Entretanto, esta tarefa mostra-se bastante repetitiva e exige atenção por longos períodos de tempo. Isso implica no aumento da probabilidade de falhas, como a falta de precisão nas estatísticas geradas, falta de monitoramento em todos os quadros e ainda a alta rotatividade destes operadores.

A alta demanda por sistemas de monitoramento automatizados têm motivado o desenvolvimento de um grande número de trabalhos científicos e o surgimento de vários produtos no mercado [Oliveira and Scharcanski 2010, Denman et al. 2006, Nair and Clark 2002]. No caso de controle de tráfego urbano, a utilização destes sistemas vai da leitura automática de placas de veículos ao levantamento de estatísticas de trânsito (Figura 1 - A). Já no caso de segurança comercial e predial, sistemas de

monitoramento inteligentes são empregados com o objetivo de identificar ações suspeitas de pessoas, como acesso não autorizado, transporte de mercadorias suspeitas, dentre outras situações (Figura 1 - B).

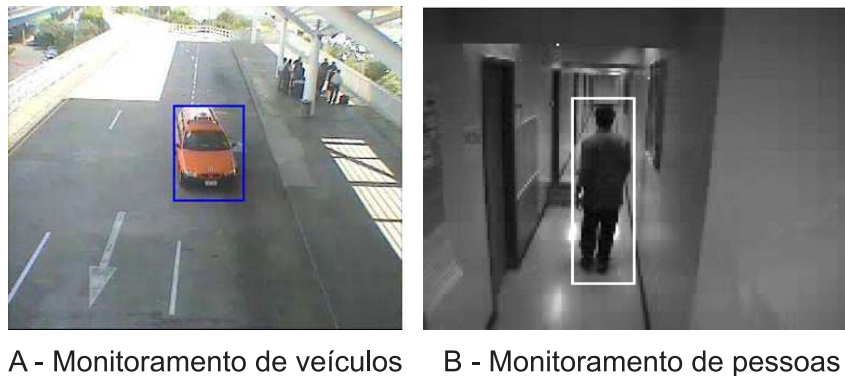


Figura 1. Exemplos de sistemas de monitoramento

Nestes sistemas de monitoramento automatizados, uma importante tarefa a ser realizada é o rastreamento dos elementos da cena em observação, sejam automóveis, pessoas, ou outros objetos. Deve-se considerar ainda que a grande maioria dos sistemas de vigilância são compostos por um grande número de câmeras. Desta forma, o rastreamento dos elementos de interesse deve considerar este cenário.

No presente trabalho é apresentado o desenvolvimento de um sistema de monitoramento automático que realiza o rastreamento (*tracking*) de objetos de interesse identificados nos sinais de vídeo providos por múltiplas câmeras. Especificamente, o sistema proposto foi desenvolvido com o objetivo de acompanhar veículos que trafegam em uma via monitorada.

Na seção 2 são discutidas diferentes técnicas de visão computacional empregadas na detecção e rastreamento de objetos. Na seção 3 é descrita a metodologia utilizada no desenvolvimento do sistema apresentado. Na seção 4 são expostos os resultados experimentais obtidos pelo emprego do sistema. Finalmente, na seção 5, são apresentadas as conclusões e sugestões para trabalhos futuros.

2. Rastreamento de Objetos em Movimento

O processo de rastreamento pode ser resumido em duas fases principais. A primeira corresponde em detectar objetos alvo presentes em um dado quadro de uma sequência de vídeo. A segunda fase corresponde em acompanhar o objeto ao longo da sequência, ou seja, localizar ou corresponder um mesmo objeto alvo nos vários quadros subseqüentes [Denman et al. 2006, Denman et al. 2007, Alkanhal et al. 1999]. Esta pode se mostrar uma tarefa bastante complexa, e apresentar um alto custo computacional devido a grande quantidade de dados contida em um vídeo.

A correspondência, em especial, pode se mostrar bastante difícil no caso onde os objetos estão se movendo relativamente rápido em relação à taxa de aquisição. Outra situação que pode aumentar a complexidade do problema é quando os objetos rastreados mudam de orientação durante o tempo. Oclusões são outra fonte de problemas muito comum.

Também em relação à correspondência, a busca por alvos correspondentes em cada quadro de uma sequência de vídeo, sem o uso de qualquer conhecimento específico, é relativamente lenta. Por este motivo, para minimizar a busca entre as imagens, comumente os processos de rastreamento empregam de algum modo o conhecimento sobre o movimento do objeto que está sendo rastreado. Alguns trabalhos, por exemplo, limitam o espaço de busca adotando a suposição de que o objeto tem uma trajetória de movimento contínua, e esperando que ele se encontre nas proximidades da sua localização anterior [Denman et al. 2006].

Como descrito, o rastreamento de objetos alvo em uma única câmera já representa um desafio. Entretanto, visto que muitos dos sistemas de monitoramento são compostos por múltiplas câmeras, são necessárias técnicas para o tratamento desta condição. Neste caso, o rastreamento deve ocorrer através de uma cadeia de câmeras, em que um objeto, após ser identificado por uma das câmeras, é acompanhado por essa. No momento que o objeto for para a área de visão de outra câmera o rastreamento continua, reconhecendo-o como sendo o mesmo que estava sendo acompanhado na outra visão [Tyagi et al. 2007].

3. Metodologia Empregada

Este trabalho foi desenvolvido considerando como requisitos funcionais principais a detecção e o rastreamento de veículos que transitam em uma pista monitorada. Este tipo de informação é extremamente relevante visto que possibilita, por exemplo, a detecção de colisões, a identificação de veículos se locomovendo em sentido proibido, e ainda o ajuste em tempo real de semáforos com base na situação momentânea do trânsito.

A configuração considerada neste trabalho é apresentada na Figura 2. Nesta configuração, são utilizadas duas câmeras, posicionadas e orientadas de modo a observar a pista sobre a qual se quer realizar o monitoramento. Adicionalmente, as câmeras são instaladas de modo que ocorra uma intersecção parcial de seus campos de visão.

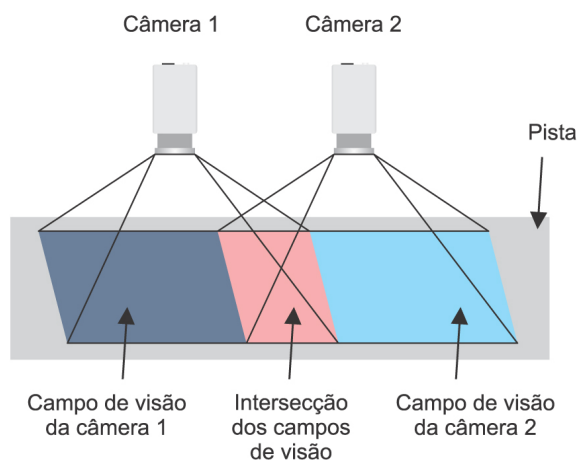


Figura 2. Configuração do sistema

O fluxograma das principais etapas que descrevem o processamento presente no sistema desenvolvido é apresentado na Figura 3. No fluxograma apresentado,

além da etapa inicial de preparação do sistema, podem ser facilmente identificadas etapas clássicas de processamento de imagens muito comuns a sistemas de inspeção e monitoramento automatizados [Gonzalez and Woods 2000].

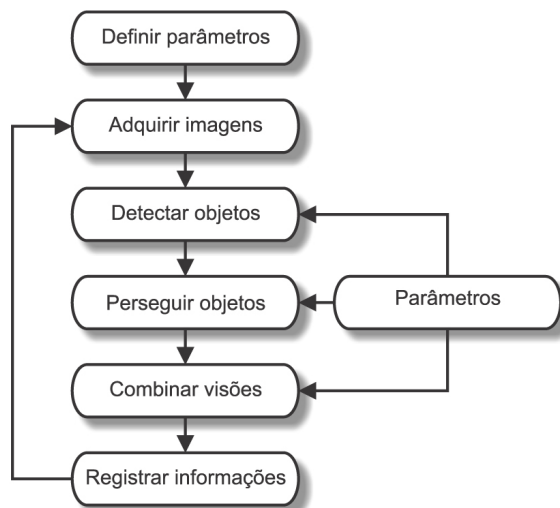


Figura 3. Fluxograma do sistema

Uma vez configurados os parâmetros do sistema, é iniciado um ciclo de processamento com o objetivo de identificar os elementos presentes na cena observada e em seguida realizar o rastreamento. Neste sentido, para cada imagem capturada, são identificados componentes conexos relevantes através de uma etapa de segmentação. Os elementos que aparecem em uma sequência de imagens capturadas ao longo do tempo são então correspondidos. Através da avaliação da posição relativa de um mesmo elemento, observado nas várias imagens de uma sequência, é estimada a trajetória realizada. Finalmente, com base neste deslocamento, podem ser levantadas as estatísticas desejadas.

Na sequência, as abordagens utilizadas para cada uma destas etapas são apresentadas.

3.1. Aquisição e Pré-Processamento

Na aquisição de imagens são obtidos os sinais que descrevem a cena observada. Para o caso específico de sistemas de monitoramento de tráfego de veículos, a aquisição consiste em capturar os sinais provenientes de diferentes câmeras, posicionados em locais específicos da pista de rodagem. Os sinais obtidos são codificados em diferentes formatos de vídeo [Santos et al. 2009].

Nesta etapa também é comum ser realizado um pré-processamento, com o objetivo melhorar a qualidade da imagem adquirida e, desta forma, torná-la mais adequada ao processamento subsequente. Dentre processamentos comumente empregados nesta etapa podem ser citados os filtros de realce e restauração. O emprego destes filtros, implementados no domínio da frequência ou espaço, têm como objetivo eliminar ruídos introduzidos durante a aquisição das imagens, bem como destacar características de interesse.

3.2. Detecção de Objetos de Interesse

Os objetos de interesse de uma cena podem ser detectados através de segmentação de imagens. Na etapa de segmentação procura-se reproduzir o processo da visão humana, onde são efetuados agrupamentos dos pontos presentes na imagem percebida. Este agrupamento, que pode ser baseado em proximidade, similaridade e continuidade, tem como objetivo isolar os diferentes elementos presentes na cena observada. A segmentação de imagens é muitas vezes o maior desafio de um sistema de processamento de imagens. O problema é ainda maior quando são tratados cenários dinâmicos, onde variações no ambiente devem ser consideradas.

Muitas das técnicas de segmentação existentes são baseadas nas informações presentes em uma única imagem, como é o caso da limiarização ou extração de contornos. Entretanto, em sistemas que utilizam vídeo é comum a utilização de técnicas de segmentação específicas para seqüência de imagens, cujos resultados tendem a ser superiores.

Para o caso do sistema desenvolvido, estamos considerando o monitoramento de uma cena dinâmica, onde desejamos classificar os diferentes elementos que transitam pela mesma. Um método simples e bastante empregado, que poderia ser utilizado, é baseado na subtração de fundo da cena. Esta abordagem mostra-se adequada para situações onde é desejado discriminar um objeto em movimento onde existe a presença de um fundo estático, como é o caso da aplicação considerada.

O procedimento geral é realizar uma operação de subtração de intensidades entre a imagem correntemente avaliada e outra imagem de referência que contenha apenas o fundo da cena, construída a partir de uma seqüência de imagens durante um período de treinamento. Desta operação é obtida uma imagem onde a intensidade dos pontos pertencentes aos elementos dinâmicos de interesse estarão destacados. Uma limiarização global é suficiente para se obter uma imagem binária contendo os elementos de interesse.

Apesar da simplicidade e desempenho do algoritmo de subtração de fundo, ele não se mostra robusto a variações de foco e iluminação, comumente presentes nos vídeos obtidos para o monitoramento de tráfego urbano. Por este motivo, para este trabalho, foi selecionado o algoritmo adaptativo de segmentação de fundo (NHD), que se adapta a novas condições de aquisição [Butler and Sridharan 2003, Santos et al. 2009]. Neste algoritmo de segmentação de fundo adaptativo, os componentes de cor e luminosidade de cada pixel são modelados como grupos de clusters, provendo uma distribuição multi-modal para cada pixel. Neste procedimento são utilizadas imagens no formato de cor YCbCr, com compressão 4:2:2.

Os pixels são correspondidos como ilustrado na Figura 4. Quando uma correspondência é feita, o centróide do cluster correspondido e os pesos de todos os clusters nos agrupamentos de pontos são adaptados para incorporar a informação do pixel atual. Se não é realizada a correspondência, então o cluster com menor peso é substituído por um novo cluster contendo o valor do pixel atual, e este é classificado como sendo em movimento [Denman et al. 2007].

Seguindo este procedimento, os clusters e seus pesos são alterados gradualmente na medida que novos quadros são processados, permitindo que o resultado da

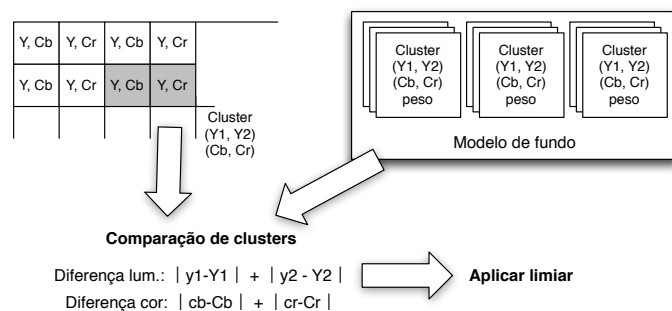


Figura 4. Modelo NHD baseado em clusters

segmentação seja adaptado a mudanças de fundo. Novos objetos podem então ser adicionados à cena, e ao longo do tempo eles podem ser incorporados ao modelo de fundo.

O algoritmo NHD retorna imagens binárias contendo os elementos em movimento detectados na cena. Estas imagens são então processadas por filtros morfológicos para eliminar a presença de pequenos elementos, derivados de ruído gaussiano ou outros artefatos presentes nas imagens processadas.

3.3. Rastreamento de Objetos

Na etapa de rastreamento os componentes conexos encontrados após a segmentação são associados a cada novo quadro pela proximidade dos centróides às posições dos componentes presentes no quadro anterior. Para realizar este processo foram implementados dois algoritmos. O primeiro deles ignora a informação sobre o movimento do alvo rastreado. Assim, a posição onde se inicia a busca por um dado ponto em um quadro corrente é a mesma posição onde ele estava no quadro anterior.

Um segundo algoritmo também foi implementado, considerando a informação de movimento para otimizar as buscas. Quando o movimento é detectado em um componente, a área a sua volta é examinada para determinar o Fluxo Óptico (*optical flow*) [Denman et al. 2007]. O tamanho da área examinada deve ser determinado pela máxima aceleração permitida para um componente conexo, tanto no eixo x quanto no eixo y . Estes valores devem ser determinados de acordo com as cenas a serem analisadas. Uma vez que o movimento de um componente foi determinado, sua nova posição pode ser prevista. Com um modelo de velocidade constante a localização do novo componente p , na próxima imagem, é dada por:

$$p_x^{n+1} = p_x^n + (p_x^n - p_x^{n-1}) \quad (1)$$

$$p_y^{n+1} = p_y^n + (p_y^n - p_y^{n-1}) \quad (2)$$

onde p_x^{n-1} e p_y^{n-1} são as posições do elemento conexo p na imagem anterior, p_x^n e p_y^n são as posições de p na imagem atual e p_x^{n+1} e p_y^{n+1} são as posições esperadas do componente p na próxima imagem. Caso o componente já se encontre classificado como em movimento, então a posição esperada é usada como posição inicial para a busca.

3.4. Combinação das Múltiplas Visões

As etapas de aquisição e pré-processamento de imagens, detecção de objetos em movimento e rastreamento de objetos detectados é feita separadamente, para cada uma das câmeras que compõe o sistema. Destas etapas, para cada câmera, é obtida uma lista de objetos, descritos através de suas coordenadas no sistema de coordenadas local, ou sistema de coordenadas da câmera.

De modo a obter uma visão global do ambiente monitorado, é necessário unir as informações providas por cada uma das câmeras que compõe o sistema. Assim, as coordenadas dos objetos detectados em cada câmera são convertidos para um sistema único, empregando as transformações geométricas que relacionam cada sistema de coordenadas local com o sistema de coordenadas global assumido.

Na Figura 5 é ilustrada a situação onde as operações de rotação e translação são utilizadas para relacionar um sistema de coordenadas local, de uma câmera, com o sistemas de coordenadas global.

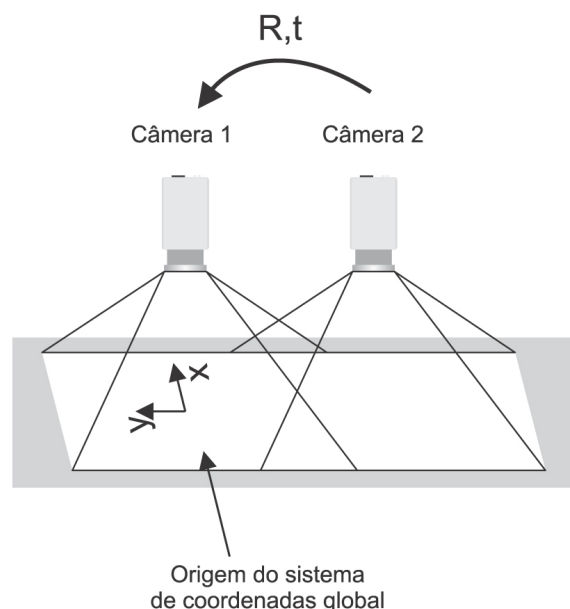


Figura 5. Transformação de coordenadas

Conforme ilustrado, o sistema de coordenadas global assumido pode, por exemplo, corresponder ao sistema de coordenadas de uma das câmeras que compõe o sistema.

Após todos os objetos detectados serem convertidos para o sistema global, estes são avaliados, de modo a identificar quais destes correspondem a um único objeto, observado sob dois pontos de vista.

4. Resultados Experimentais

Para demonstrar a aplicação da metodologia proposta, foi criado um ambiente virtual 3D que simula uma pista sobre a qual veículos transitam, e de onde pode-se adquirir imagens de câmeras virtuais.

O ambiente virtual e a configuração utilizada são apresentados na Figura 6, onde é possível obter uma visão geral da cena (A), obter uma visão da cena capturada pela câmera 1 (B) e uma visão da cena capturada pela câmera 2 (C).



Figura 6. Ambiente virtual 3D

Como pode ser observado, duas câmeras virtuais, posicionadas no ambiente conforme a configuração descrita na Seção 3, fornecem imagens da pista e dos veículos. Apesar de o ambiente virtual não simular todas as condições observadas em uma situação real, sobre ele é possível ter maior controle sobre variáveis como veículos, velocidade, número e posicionamento das câmeras.

Para a execução dos testes foram utilizados quatro cenários, sendo o primeiro com apenas um veículo transitando na pista, o segundo e terceiro com dois veículos em posições diferentes e o quarto com quatro veículos. Para cada cenário foi aplicado o algoritmo de rastreamento de três formas diferentes, sendo uma sem execução do rastreamento, de modo a medir o custo da execução do algoritmo de detecção (NHD), uma executando a detecção e o rastreamento sem otimização, e a última executando a detecção e o rastreamento com otimização por Fluxo Óptico.

Na Tabela 1 são apresentados os tempos médios de processamento de cada quadro obtidos para cada execução, em cada cenário.

Tabela 1. *Tempos médios dos testes de rastreamento (ms)*

Etapas	Cen. 1	Cen. 2	Cen. 3	Cen. 4
Detecção	125,73	124,03	122,99	122,83
Rastream.	128,10	127,82	128,33	128,72
Fluxo Ópt.	127,78	126,71	126,48	126,55

Para a execução dos testes foi utilizado um computador com processador Pentium dual core T2370 de 1.73GHz e 2GB de memória RAM. Os algoritmos citados foram implementados na linguagem C++.

Observando os tempos obtidos, percebe-se que o algoritmo de rastreamento, com ou sem otimização, tem pouca influência no tempo de execução geral do sistema. Para melhorar o desempenho é necessário otimizar a etapa de detecção (NHD).

Na Tabela 2 são apresentados os ganhos percentuais pelo emprego da otimização de Fluxo Óptico na etapa de rastreamento para cada cenário.

Tabela 2. *Ganho pela utilização de Fluxo Óptico (%)*

Etapas	Cen. 1	Cen. 2	Cen. 3	Cen. 4
Detecção	13,25	29,12	34,70	36,91

Observando-se os tempos obtidos é possível verificar que o uso deste recurso reduz consideravelmente o custo do rastreamento. Como é de se esperar, o ganho fica ainda mais evidenciado na medida que mais objetos alvo são rastreados.

5. Conclusões

Neste trabalho é proposta uma metodologia para rastreamento de veículos observados por múltiplas câmeras. A partir dos dados gerados pelo sistema desenvolvido, vários parâmetros podem ser estimados, como posição (localização atual), trajetória (seu movimento) e velocidade dos objetos tridimensionais da cena. A metodologia foi implementada e testada, sob diferentes aspectos, empregando um ambiente virtual tridimensional que simula as condições reais de uso.

A aplicação da técnica de Fluxo Óptico resultou em uma redução significativa do custo computacional específico da etapa de rastreamento, onde a melhora ficou em torno de 30%. A implementação desta técnica se mostrou simples, demonstrando uma boa relação entre custo e benefício. De qualquer modo, visto que a maior parcela do custo está associada à detecção de objetos, um maior esforço na otimização desta etapa mostra-se necessária.

A implementação do rastreamento através de várias câmeras se mostrou funcional, inclusive para vários objetos na cena. Embora os testes tenham sido feitos usando apenas duas câmeras, o acréscimo de outras pode ser feito de maneira relativamente simples.

Por sua vez, o ambiente virtual 3D implementado se mostrou bastante útil para a realização dos testes. Esta ferramenta permitiu, sem muito esforço, modificar o cenário e as variáveis do ambiente, câmeras e dos objetos observados, facilitando a validação da metodologia empregada.

Este trabalho pode ser estendido de várias maneiras, tanto para melhorar as funcionalidades existentes como para adicionar novas características.

No presente trabalho são descritos os resultados da avaliação do sistema utilizando um ambiente virtual. Será muito interessante validar o sistema em situações reais de uso, e assim fazer uma análise qualitativa dos resultados obtidos. O posicionamento das câmeras, por exemplo, deve ser flexibilizado. A configuração adotada para as câmeras simula posicionamento e orientação ideal, difícil de ser conseguido na prática. Assim, devem ser implementadas técnicas de calibração que permitam maior liberdade para a pose das câmeras utilizadas [Stivanello et al. 2008].

O próprio ambiente virtual pode ser melhorado, de forma a simular outras variáveis que influenciam na performance do sistema, como como ruído no sensor, variações de iluminação, sombras e distorção radial de lente, distorções de perspectiva.

É importante observar a possibilidade de redução do custo computacional obtido pela adaptação dos algoritmos utilizados, de modo que empreguem recursos específicos de hardware, como Computação paralela (*Single Instruction, Multiple Data*), processamento em Unidade gráfica (General-purpose computing on graphics processing units), dentre outros.

Finalmente, outra extensão interessante do sistema seria adicionar uma etapa de reconhecimento de objetos identificados na cena [Santos et al. 2009, Stivanello and Gomes 2006], classificando-os como pessoas, carros, motos e caminhões, por exemplo.

Referências

- Alkanhal, M., Turaga, D., and Chen, T. (1999). Correlation based search algorithms for motion estimation. In *Picture Coding Symposium*.
- Butler, D. E. and Sridharan, S. S. (2003). Real-time adaptive background segmentation. In *Proceedings of the 2003 International Conference on Multimedia and Expo - Volume 3 (ICME '03)*, volume 3, page 3.
- Denman, S., Chandran, V., and Sridharan, S. (2007). An adaptive optical flow technique for person tracking systems. *Pattern recognition letters*, 28(10):1232–1239.
- Denman, S., Fookes, C., Cook, J., Davoren, C., Mamic, A., Farquharson, G., Chen, D., Chen, B., and Sridharan, S. (2006). Multi-view intelligent vehicle surveillance system. In *IEEE International Conference on Video and Signal Based Surveillance*, page 26.
- Gonzalez, R. C. and Woods, R. E. (2000). *Processamento de Imagens Digitais*. Editora Edgard Blucher, 1 edition.
- Nair, V. and Clark, J. J. (2002). Automated visual surveillance using hidden markov models. In *VI02*, page 5.
- Oliveira, A. B. and Scharcanski, J. (2010). Vehicle counting and trajectory detection based on particle filtering. *Graphics, Patterns and Images, SIBGRAPI Conference on*, 0:376–383.
- Santos, D., Reis, D. S., Gomes, P. C. R., and Stivanello, M. E. (2009). Desenvolvimento de um sistema de visão computacional para monitoramento automatizado de veículos. In *Anais do IX Simpósio Brasileiro de Automação Inteligente*.
- Stivanello, M. E. and Gomes, P. C. R. (2006). Inspeção visual industrial automatizada por análise de forma com descritores de fourier e redes neurais artificiais. In *Anais do XV Semimco*, page 11.
- Stivanello, M. E., Leal, E. S., Palluat, N., and Stemmer, M. R. (2008). Desenvolvimento de uma biblioteca para sistemas de visão estereoscópica para robótica móvel. In *Anais da VIII Conferência Internacional de Aplicações Industriais*.
- Tyagi, A., Potamianos, G., Davis, J., and Chu, S. (2007). Fusion of multiple camera views for kernel-based 3d tracking. In *IEEE Workshop on Motion and Video Computing*.