

Rastreamento de Pedestres utilizando Viola-Jones e Método Húngaro com múltiplas câmeras

FELIPE AGUIAR CORRÊA¹; PROF. MSC. ANDERSON PRIEBE FERRUGEM²

¹Universidade Federal de Pelotas – facorra@inf.ufpel.edu.br

²Universidade Federal de Pelotas – ferrugem@inf.ufpel.edu.br

1. INTRODUÇÃO

A área de Visão Computacional é uma subárea da Inteligência Artificial que tem como objetivo desenvolver soluções para o computador não apenas receber as imagens, mas também processá-las cognitivamente a ponto de adquirir algum tipo de informação ou conhecimento presente nelas. De acordo com PRINCE (2012), este processo comumente é feito através da aprendizagem de certos padrões invariantes em certas categorias de objetos, algo semelhante ao que o cérebro humano faz para aprender a distinguir objetos ao processar imagens provenientes dos olhos. Inspirando-se na forma com que o cérebro reconhece padrões e somando isso à utilização de otimizações estatísticas e matemáticas, a área de Visão Computacional procura dar a habilidade de processar imagens cognitivamente aos computadores.

De acordo com PRINCE (2012), são exemplos de problemas abordados pela área de Visão Computacional: a estimativa de movimento e segmentação, reconstrução de objetos para modelos tridimensionais, análise de segmentação, reconhecimento de faces e objetos, rastreamento de alvos, entre outros muitos desafios. Este trabalho trata sobre o problema de reconhecimento e rastreamento de objetos em vídeo.

Um tópico bastante abordado pela área de Visão Computacional é o rastreamento de pessoas. Isso ocorre devido à crescente necessidade de aplicação em diversos campos, desde sistemas de monitoramento de ruas através de câmeras até o uso de rastreamento em esportes para obtenção de estatísticas de rendimento e outras informações relevantes. De acordo com MASOUD;PAPANIKOLOPOULOS (1997), rastrear pessoas em ambiente urbano pode trazer valiosas informações, como por exemplo reconhecimento de eventos, contagem de pedestres e nível de tráfego. O problema em questão consiste em identificar uma pessoa através de uma ou mais câmeras e manter sua identidade, rastreando-a enquanto ela se movimenta mantendo a associação dela a uma espécie de rótulo, sempre cuidando para que a sua identidade esteja vinculada ao rótulo.

Rastrear pessoas em ambiente urbano prova-se um desafio, pois deve levar em consideração diversos problemas. O primeiro desafio é o reconhecimento em si, que consiste em reconhecer corretamente o que são pessoas dentro de um ambiente com muitos outros objetos e oclusões parciais causadas por obstáculos no campo de captura das câmeras. O segundo desafio seria rastrear os alvos, ou seja, segui-los enquanto movimentam-se. Estimar movimentação e seguir pessoas que sofrem oclusões é um problema difícil de ser resolvido. O terceiro e último desafio seria manter a associação de identidade dos alvos, uma vez que o alvo pode ter variações de ângulos e formato que não permitem a fixação de apenas um único ponto invariante, dificultando a identificação. Além disso, pode-se ter pessoas com um grande número de características em comum tais como cores e formato das roupas (por exemplo,

vendedores uniformizados de uma mesma loja), estatura, corte do cabelo, etc. Esse problema é grave, pois a perda de associação dos dados pode fazer com que pessoas sejam identificadas como pessoas diferentes entre câmeras diferentes ou entre períodos curtos de oclusão.

Tendo em vista as dificuldades encontradas atualmente para construir um sistema robusto de rastreamento de pessoas, neste trabalho foi desenvolvida uma proposta de solução híbrida para rastrear pessoas em ambiente urbano, utilizando o algoritmo de reconhecimento de objetos conhecido como Viola-Jones, um algoritmo robusto de reconhecimento de objetos em tempo real proposto por VIOLA; JONES (2004), que utiliza uma arquitetura em cascata em sua execução. Para realizar o treinamento do detector, foi utilizado um conjunto de dados de *tracklets* proposto por KAWANISHI et al. (2014). Este conjunto de dados consiste em vinte e duas mil imagens de pedestres capturadas nas mais diferentes posições e com diferentes intensidades de iluminação, com o objetivo de treinar reconhecedores através da obtenção de padrões invariantes presentes nestas imagens. Para manter a associação de dados entre as câmeras e curtos períodos de oclusão do alvo, foi utilizado o Algoritmo Húngaro proposto por KUHN, H. (1955), um método de otimização combinacional estatística que compara histogramas de cores candidatos, gerados a partir de alvos detectados, para associar identidades entre pessoas.

2. METODOLOGIA

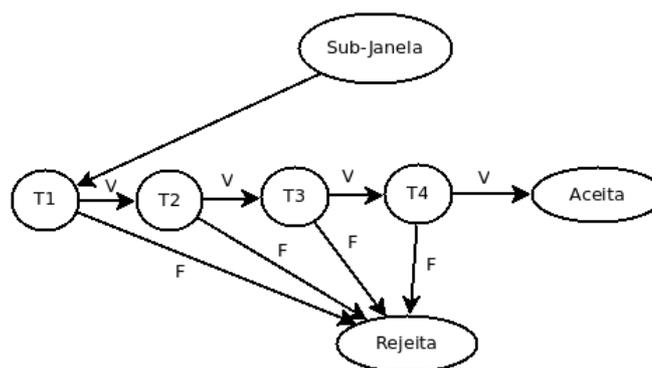
No capítulo anterior foi mostrado que detectar e rastrear objetos são problemas distintos apesar de que, para rastrear um objeto, é preciso primeiramente detectá-lo. Também no capítulo anterior, fora citado o modelo proposto por VIOLA; JONES (2004) para detectar faces ou quaisquer alvos desde que seja treinado um classificador para o mesmo previamente. A primeira etapa deste trabalho, então, consiste em treinar padrões de detecção para tornar o algoritmo de Viola-Jones capaz de detectar pedestres nas mais diferentes variações de ângulos e iluminação ambiente.

Para treinar um detector, usamos um conjunto de imagens de alvos os quais queremos detectar e um conjunto de imagens negativas. Imagens negativas consistem em fundos ou outras imagens que não são as que queremos detectar. Com isso, através de aprendizado de máquina e a combinação dessas imagens positivas e negativas, podemos treinar um classificador para o alvo em questão: pedestres. Tudo isso utilizando imagens positivas provenientes do conjunto de dados Shinpuhkan2014.

Após treinado, teve-se como produto final um arquivo classificador no formato XML contendo as características retiradas na etapa de treinamento. Essas características ficam devidamente organizadas em etapas de detecção (foi usado 20 etapas), o que justifica o fato do algoritmo de Viola-Jones ser um algoritmo em cascata. Essas etapas vão da precisão mais grossa até a mais fina, e um objeto só será detectado se passar por todas essas etapas de detecção. Na Figura 1, exemplifica-se essa estrutura através de um diagrama. Essa arquitetura em cascata dá-se ao fato de podermos eliminar áreas da imagem que não possuem chance de conter o alvo pretendido, afinal, essas áreas não passam pelas primeiras etapas do arquivo de características.

Assim, utilizando o arquivo de padrões obtido na etapa de treinamento com o algoritmo de Viola-Jones é possível detectar pedestres, mas não rastreá-los. A cada *frame* de vídeo, o nosso alvo seria detectado como um alvo totalmente diferente, sem manter a identidade.

Para resolver o problema de rastreamento, ou seja, associar um alvo a uma espécie de identidade, geramos histogramas de cores para descrever os alvos e comparamos os mesmos com Método Húngaro. De acordo com STOCKMAN;SHAPIRO (2001), um histograma de cores é um bom descritor de objeto, pois o mesmo é relativamente invariante com translação e rotação do eixo de visão podendo apenas variar lentamente conforme muda o ângulo de visão em relação ao alvo. Em vídeo, como no trabalho proposto, é comum usarmos histogramas de cores reduzidos, não contendo todas informações de cores pelo fato de variações na iluminação e calibramento das câmeras dar pequenas diferenças nos tons das cores.



Onde Tx é uma etapa do classificador

Figura 1 – Etapas em cascata do Viola-Jones

Com a obtenção dos histogramas de cores dos alvos, é possível compará-los e manter a identidade dos alvos entre câmeras e até mesmo entre curtos períodos de oclusão. Usamos o Método Húngaro, um algoritmo de otimização combinacional para comparar histogramas de forma acelerada e determinar qual histograma pertence a qual alvo, assim sabendo quem é o alvo nos próximos frames de vídeo e entre as câmeras que estão captando um mesmo alvo.

Para implementação, utilizou-se a ferramenta *OpenCV* na linguagem de programação *Python*, uma biblioteca completa de Visão Computacional com seu próprio *plugin* para realizar treinamento de detectores e sua própria implementação do algoritmo de detecção Viola-Jones.

3. RESULTADOS E DISCUSSÃO

Até escrita deste trabalho, foi apenas realizada a etapa de treinamento dos classificadores. Nesta etapa, teve-se como produto final um classificador capaz de reconhecer pedestres em vários ângulos, diferente do classificador para corpos inteiros já existente na ferramenta *OpenCV* citada no capítulo anterior. Na Figura 2, exemplifica-se os formatos de detecção de alvos.



Figura 2 – *Tracklets* detectados de pedestres

Com a continuação do trabalho, pretende-se medir estatisticamente a taxa de acerto do novo detector comparado aos existentes na ferramenta utilizada. Depois disso, pretende-se seguir para a etapa 2: implementação do rastreamento. Com a conclusão do trabalho espera-se ter um rastreador de pedestres capaz de ter bons resultados e manter associadas as identidades aos alvos de forma correta.

4. CONCLUSÕES

Enfim, ao término do desenvolvimento deste projeto pretende-se chegar a um rastreador de pedestres robusto, que seja capaz de seguir alvos através de múltiplas câmeras e manter a identidade mesmo que um alvo fique parcialmente ocluído. A aplicação do produto final vai desde sistemas de segurança automatizados a sistemas de estatísticas de rendimento para esportes.

Fazer máquinas enxergarem e ter cognição semelhante à do ser humano ainda é um grande desafio para os próximos anos de pesquisa na área de Visão Computacional e da Inteligência Artificial como um todo. Grande parte disso deve-se ao fato da máquina ser muito determinística e carecer de “senso comum” e “intuição”, estratégias que os seres humanos utilizam-se muito para detectar padrões.

5. REFERÊNCIAS BIBLIOGRÁFICAS

KAWANISHI, Y.; WU, Y.; MUKUNOKI, M.; MINOH, M.; Shinpuhkan2014: A Multi-Camera Pedestrian Dataset for Tracking People across Multiple Cameras. **THE 20TH KOREA-JAPAN JOINT WORKSHOP ON FRONTIERS OF COMPUTER VISION, FCV2014**. Okinawa, 2014.

MASOUD, O.; PAPANIKOLOPOULOS N. P.; Robust pedestrian tracking using a model-based approach. **IEEE CONFERENCE IN INTELLIGENT TRANSPORTATION SYSTEM '97**, Boston, Nov. 1997: IEEE, 1997. pg. 338 – 343.

KUHN; H. The Hungarian Method for the assignment problem. **Naval Research Logistics Quarterly**. Washington D.C., v.2, pg. 83 – 97, 1955.

PRINCE, S. J. D. **Computer Vision: models, learning and inference**. Cambridge University: Prentice Hall, 2012.

STOCKMAN, G.; SHAPIRO, L. G.; **Computer Vision**. 1st.ed. Upper Saddle River, NJ, USA: Prentice Hall PTR, 2001.

VIOLA, P.; JONES, M.; Robust Real-time Face Detection. **International Journal of Computer Vision**. Hingham, v.57, n.2, p. 137 – 154, 2004.