

People Counting - Estado da Arte

Suellen Silva de Almeida
Victor Hugo Cunha de Melo

8 de setembro de 2010

1 Introdução

Nosso projeto consiste em implementar um sistema automático para contagem de pessoas de modo que possamos identificar a quantidade de passageiros que entrou em um ônibus.

Neste relatório apresentamos resumos de publicações relevantes para esta pesquisa.

1.1 Counting people in crowds with a real-time network of simple image sensors

No artigo [18] é proposto um método para estimar o número de pessoas em uma multidão sem utilizar contagem individual. Dispondo-se de diversos sensores em uma rede sem fio, os autores utilizam uma técnica para calcular os limites do número de objetos em uma região. Cada nó sensor extrai o primeiro plano do fundo e envia as silhuetas resultantes para os outros sensores. Os dados de todos os sensores são agregados para calcular uma projeção planar da cena. Essa projeção é utilizada para limitar o número e possíveis localizações de pessoas. Segundo os autores, o sistema executa em tempo-real e o modelo utilizado é ideal para detecção em pequenas áreas. Também afirmam que o número de sensores e seu posicionamento afeta a precisão da contagem.

1.2 A People-Counting System Using a Hybrid RBF Neural Network

Os autores de [10] descrevem um sistema de contagem de pessoas utilizando uma novo tipo de *rede neuronal RBF híbrida*. De forma a diminuir a complexidade para operação online, utilizam um novo método de extração de *features*, onde a informação de cada imagem é representada enxutamente na forma de *data-curves*. A rede RBF é utilizada como classificadora devido a sua excelente capacidade de classificação e auto-aprendizado, como sustentam os autores. Estes desenvolveram um método híbrido de rede RBF para lidar com distribuições complexas. A camada escondida da rede RBF híbrida é uma combinação de termos polinomiais e funções *radial basis*. Com o algoritmo proposto de clusterização supervisionado, o número de neurônios ocultos é automaticamente determinado de acordo com uma distribuição

de dados e informações de categorização. Os resultados obtidos mostram que o método RBF híbrido pode obter resultados promissores em contagem de pessoas em passagens abertas.

1.3 Automatic Counting of Interacting People by using a Un-calibrated Camera

O sistema proposto em [16] pode contar múltiplas pessoas interagindo em uma região de interesse (*region of interest, ROI*) utilizando uma única câmera. Há dois níveis hierárquicos de rastreamento. Para os casos que não envolvem junções ou divisões, um método *blob*¹ de localização é utilizado. Para lidar com interações de pessoas na ROI (que pode ser definida pelo usuário), o sistema utiliza o algoritmo de localização *mean-shift*. A taxa de precisão do sistema segundo os autores é de 98.5% para vídeos com junções, e de 95% para divisões. Atribuem os erros à remoção do segundo plano (*background subtraction, BGS*) quando as pessoas usam roupas com cores similares ao chão.

1.4 Counting pedestrians in video sequences using trajectory clustering

Em [2] apresenta-se um estudo comparativo entre diferentes abordagens para contagens de pessoas, baseado em técnicas de clusterização. Primeiramente, as bases de imagens são analisadas baseadas no tamanho e posição inicial de suas trajetórias. Em seguida, são utilizadas representações de dados diferentes e medidas de distância/similaridade. Inicialmente, são aplicadas a distância de Hausdorff e a similaridade LCSS para os ICA e representações de séries de tempo. Os resultados mostram que o espaço ICA provê uma representação mais dinâmica com relação ao domínio espaço-tempo original, reduzindo a presença de *outliers*.

1.5 A Viewpoint Invariant Approach for Crowd Counting

Em [12] é proposto um sistema para contar pedestres em multidões com uma única câmera. Neste sistema, ao invés de utilizar *features* simples como soma das bordas ou *blob pixels*, adota-se *feature histograms*. Esta representação, segundo os autores, é mais poderosa que *simple features* para tratar ruídos e pode ser mais precisa em contar grupos em movimento. Também escrevem que, diferentemente da abordagem baseada em aprendizado, este método leva em consideração a normalização das *features* para lidar com projeção perspectiva e orientação da câmera. Afirmam que esse sistema é treinado para ser invariante a perspectiva e pode ser empregado com configuração mínima.

1.6 Towards a robust solution to people counting

O artigo [9] investiga a possibilidade de desenvolver um modelo estatístico robusto para contagem de pessoas. Primeiramente, avaliam diversos métodos existentes na literatura para correção na distorção da perspectiva. Em seguida, investigam a

¹“connected sets of pixels detected by means of various segmentation techniques”

robustez de um método simples baseado em uma extração mais confiável dos objetos de primeiro plano, em uma correção de perspectiva e na taxa de confiança que dirige um filtro da mediana ponderada para refinar a contagem. Conforme mostram os resultados, o filtro da mediana ponderada é mais eficaz que o filtro da mediana convencional. O método apresentado pelos autores apresenta resultados promissores, mas pode melhorar em: (1) remoção de sombras e (2) estratégias de ponderamento para contornar oclusões significantes. Além disso, reparam que alguns problemas podem ser resolvidos adicionando mais níveis de *features*, como detecção de pessoas.

1.7 A Vision-Based Method to Estimate Passenger Flow in Bus

O artigo [20] descreve um procedimento usando processamento de imagens para estimar, em tempo-real, o fluxo de passageiros em um complexo plano de fundo de um ônibus. Identificam cada passageiro pelo contorno da cabeça, que é obtido por uma modificação da Transformada de Hough e uma medida *fuzzy*. Os resultados mostram que o método consegue obter uma precisão de até 93.6%, com a câmera localizada dentro do ônibus e durante a noite, e de 82.6% com a câmera fora do ônibus durante o dia.

1.8 A Statistical Method for People Counting in Crowded Environments

O objetivo do método proposto em [5] é estimar o número de pessoas que passam por um portão em uma área pública. Requer apenas uma câmera de baixo-custo, montada no teto e próxima da região de interesse, de forma que seja possível observar o topo das cabeças. Dessa maneira, eles evitam problemas de oclusão e de privacidade. Porém, esta posição não resolve o problema de segmentação causado por pessoas que se tocam ou por quem é adjacente as portas. Sua abordagem de *motion detection* é uma combinação de um modelo “*running average-like background*” aplicado as bordas das imagens, além de remoção do plano de fundo da imagem atual de bordas. Para contar elementos em movimento, é preciso saber a média do vetor do número de movimentos correspondente a cada pessoa (*VPR*). Os resultados deste trabalho atingem a precisão máxima de 98.88% e mínima de 93.05%.

1.9 Estimating pedestrian counts in groups

É proposto em [11] um método que pode ser utilizado para estimar com grande precisão o número de pessoas em uma cena sem a necessidade de contar indivíduo por indivíduo. A abordagem é baseada em grupos, onde a quantidade de pessoas em um grupo em movimento é estimado como um todo. A segmentação é feita por uma técnica adaptativa de mistura de Gaussianas. São apresentados dois métodos para estimar o número de pessoas em um grupo: um baseado em heurística e outro baseado no formato. Eles são robustos com relação a movimentação, orientação e distribuição espacial do grupo. Os resultados mostram que os métodos são eficazes na identificação. Entretanto, precisa ser estendido de modo que possa lidar melhor

com pessoas estacionárias em uma cena. Isto requer uma outra forma de segmentação não baseada na movimentação. Além disso, alguns carros em baixa velocidade foram identificados como grupos de pessoas.

1.10 People-Tracking-by-Detection and People-Detection-by-Tracking

[1] combina as vantagens de detecção e rastreamento em um único sistema. A articulação aproximada de cada pessoa é detectada em cada *frame*, baseando-se em *features* locais que modelam a aparência do corpo de um indivíduo. Para identificar a trajetória dos indivíduos (*people-tracklets*), o sistema extrai a posição, escala e a articulação dos membros. Para modelar a articulação dos membros, utiliza um *hierarchical Gaussian process latent variable model* (hGPLVM). Isto permite identificar as pessoas com maior confiabilidade do que seria possível utilizando *frames* isolados. Em seguida, combinam estas informações com um *hidden Markov model* (HMM) que permite estender os *people-tracklets*. Os resultados mostram que o sistema consegue identificar vários casos de indivíduos parcialmente ocultos que não são localizáveis por um *frame* isolado. Também identifica corretamente a configuração dos membros de um indivíduo. Nos vídeos em que os autores executaram os testes, todos os indivíduos foram identificados, enquanto a abordagem por *frame* isolado falhou em diversos aspectos.

1.11 Estimating the Number of People in Crowded Scenes by MID Based Foreground Segmentation and Head-shoulder Detection

Em [13] é proposto um método para o problema de estimar o número de pessoas em cenas de segurança. Esse método consiste na combinação de dois algoritmos, MID (Diferença de Imagem em Mosaico) e HOG (Histograma de Gradientes Orientados). O MID é a segmentação para obter áreas ativas no cenário e o HOG detecta a região da cabeça e ombro de uma pessoa para contar o número de pessoas na área observada. O método, além de contar o número de pessoas na cena, também localiza a posição individual de cada uma. Os resultados dos testes mostram que o método funciona perfeitamente para o objetivo proposto.

1.12 A Preprocessing Method for Tracking and Counting Pedestrians in Bus Video Monitor

O artigo [19] apresenta um método de pré-processamento para monitoramento e contagem de pessoas em imagens em escala de cinza obtidas a partir de uma câmera dentro de um ônibus. O método proposto é para segmentação de imagem e inclui quatro passos: limiarização, estatística de valores cinza, seleção e binarização. No entanto, o método não é eficaz para diferentes cenários, além de altos custos com cálculos. O algoritmo em estudo não chegou aos resultados esperados e precisa ser melhorado em várias partes.

1.13 People Counting System for Getting In/Out of a Bus Based on Video Processing

Em [6] é apresentado um sistema automático para contagem de pessoas que entram ou saem de um ônibus baseado no processamento de um vídeo (obtido através de uma câmera zenital no ônibus). Cada frame (quadro) capturado é dividido em vários blocos e cada bloco é classificado de acordo com seu vetor de movimento. Se a quantidade de blocos com similares vetores de movimento é maior que um limiar, esses blocos pertencem ao mesmo objeto em movimento. Como resultado, o número de objetos em movimento é o número de passageiros entrando ou saindo do ônibus. O problema de agitação na câmera e variação da iluminação no ônibus é superado nesse método, que apresentou 92% precisão nos testes realizados.

1.14 A People Counting System Based on Dense and Close Stereovision

O sistema proposto em [17] conta passageiros nos ônibus baseado na estereovisão. O processamento correspondente a esse sistema de contagem envolve vários blocos, dedicados à detecção, segmentação, controle e contagem. A partir de imagens estereoscópicas originais, o sistema opera na informação contida em mapas de disparidade, previamente calculados por um algoritmo sugerido no artigo. Os resultados dos testes mostram que o método tem 99% de precisão em vários cenários diferentes.

1.15 People detection and tracking using the Explorative Particle Filtering

O método proposto em [15] apresenta uma abordagem para detecção de pessoas e monitoramento em 3D. Ele é baseado em um modelo da parte de cima do corpo humano em 3D e uma função para avaliar sua presença em certa região de uma cena. Depois disso, são encontrados os máximos desta função, usando um algoritmo de filtragem de partículas chamado Explorative Particle Filtering (ExPF). A técnica apresentada é genérica e simples como modelos não dinâmicos. Os resultados dos testes mostram sucesso em detectar as pessoas, porém algumas melhorias devem ser aplicadas para reduzir o tempo de processamento e aumentar a precisão de rastreamento.

1.16 K-Means based Segmentation for Real-Time Zenithal People Counting

O artigo [4] apresenta uma abordagem eficiente e confiável para segmentação, monitoramento e contagem automática de pessoas, criada para um sistema com uma câmera que captura a parte de cima da cabeça das pessoas. Após da subtração do fundo da cena, o método de clusterização K-Means é utilizado para permitir a segmentação de pessoas sozinhas no cenário. O número de pessoas na cena é estimado como o número máximo de clusters que pertencem a um limite de separação dos clusters. Foram utilizados vídeos para testes e o resultado foi o esperado, a contagem de pessoas foi realizada com grande precisão.

1.17 Multi-camera people tracking using evidential filters

Em [14] é proposto uma abordagem de filtragem que pode ser considerada uma extensão dos filtros de partículas Bayesiano. O algoritmo proposto foi projetado para monitorar múltiplos alvos através da fusão da informação de vários sensores confiáveis. Ele é utilizado para fornecer uma solução para o problema de monitoramento de pessoas com multi-câmeras. Os alvos são monitorados combinando informações de cor e de forma. O filtro de partículas proposto é adequado para modelar as oclusões freqüentes, e por causa disso há um mapa da ocupação usado para detectar oclusões. Os testes de desempenho mostraram que o algoritmo proposto consegue estimar a localização de pessoas sendo monitoradas mesmo com condições de oclusão.

1.18 Robust Head-shoulder Detection by PCA-Based Multi-level HOG-LBP Detector for People Counting

O artigo [7] propõe um método robusto e rápido que detecta cabeça e ombros para contagem de pessoas. Esse método combina o HOG (Histogramas de Gradientes Orientados) com LBP (Padrão Local Binário) como o conjunto de características. Dessa forma é possível detectar cabeça e ombros das pessoas de forma robusta. Para melhorar o desempenho da detecção, foi utilizado PCA (Principal Components Analysis) para reduzir o conjunto de características HOG-LBP. Para estimar o fluxo de pessoas, o detector também foi incorporado em um filtro de partículas de monitoramento, o que aumentou a precisão do método. Segundo o autor, o detector é melhor que os algoritmos segundo o estado da arte, levando em consideração a precisão da detecção.

1.19 Support Vector Machine-Based Human Behavior Classification in Crowd through Projection and Star Skeletonization

Em [3] é apresentado um sistema de segurança para classificar comportamentos normais e anormais de humanos. Para detectar os comportamentos anormais, características de esqueletos são extraídas e utilizadas para alimentar a aprendizagem do SVM. Então os dados processados são colocados em um vetor para classificar os comportamentos em duas classes. O sistema proposto detectou com sucesso os comportamentos humanos e pode ser bastante útil no caso em estudo que é a detecção e contagem de pessoas em ambientes com multidões.

1.20 A Method for Counting Moving People in Video Surveillance Videos

O método proposto em [8] é para contagem de pessoas em vídeos de segurança. O problema é tratado estabelecendo um mapa entre alguma característica da cena e o número de pessoas (evitando o complexo problema de detecção). O método é baseado no uso de características SURF e um regressor E-SVR fornece a estimativa da contagem. O algoritmo funciona bem com oclusões parciais e perspectiva. Para

avaliação experimental, o método proposto foi comparado com o algoritmo vencedor do PETS 2009, no contexto de contagem de pessoas. Os resultados mostraram que o método proposto aumentou a exatidão e continuou com a robustez do algoritmo vencedor.

Referências

- [1] M. Andriluka, S. Roth, and B. Schiele. People-tracking-by-detection and people-detection-by-tracking. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8, 2008.
- [2] G. Antonini and J. Thiran. Counting pedestrians in video sequences using trajectory clustering. *IEEE Transactions on Circuits and Systems for Video Technology*, 16(8):1008 – 1020, 2006.
- [3] M. Archana B. Yogameena, E. Komagal and S. Raju Abhaikumar. Support vector machine-based human behavior classification in crowd through projection and star skeletonization. *JCS*, pages 1008–1013, 2010.
- [4] Dubravko Čulić, Borislav Antić, Dragan Letic and Vladimir Crnojević. K-means based segmentation for real-time zenithal people counting. *ICIP*, pages 2537–2540, 2009.
- [5] Massimiliano Bozzoli and Luigi Cinque. A statistical method for people counting in crowded environments. *Image Analysis and Processing, International Conference on*, 0:506–511, 2007.
- [6] Tsong-Yi Chen, Chao-Ho (Thou-Ho) Chen, Yin-Chan Chang and Da-Jinn Wang. People counting system for getting in/out of a bus based on video processing. *ISDA*, pages 565–569, 2008.
- [7] Huadong Ma, Chengbin Zeng. Robust head-shoulder detection by pca-based multilevel hog-lbp detector for people counting. *ICPR*, 2010.
- [8] Donatello Conte, Pasquale Foggia, Gennaro Percannella, Francesco Tufano, and Mario Vento. A method for counting moving people in video surveillance videos. *EURASIP*, 12(23):40–50, 2010.
- [9] Hasan Elik, Alan Hanjalic, and Emile Hendriks. Towards a robust solution to people counting. In *International Conference on Image Processing (ICIP)*. IEEE Computer Society, 2006.
- [10] D. Huang and Tommy W. S. Chow. A people-counting system using a hybrid rbf neural network. *Neural Processing Letters*, 18:97–113, 2003. 10.1023/A:1026226617974.
- [11] Prahlad Kilambi, Evan Ribnick, Ajay J. Joshi, Osama Masoud, and Nikolaos Papanikolopoulos. Estimating pedestrian counts in groups. *Comput. Vis. Image Underst.*, 110(1):43–59, 2008.

- [12] Dan Kong, Doug Gray, and Hai Tao. A viewpoint invariant approach for crowd counting. In *ICPR '06: Proceedings of the 18th International Conference on Pattern Recognition*, pages 1187–1190, Washington, DC, USA, 2006. IEEE Computer Society.
- [13] Kaiqi Huang Min Li, Zhaoxiang Zhang and Tieniu Tan. Estimating the number of people in crowded scenes by mid based foreground segmentation and head-shoulder detection. *ICPR*, pages 1–4, 2008.
- [14] Rafael Muñoz-Salinas, R. Medina-Carnicer, F.J. Madrid-Cuevas, and A. Carmona-Poyato. Multi-camera people tracking using evidential filters. *IJAR*, pages 732–749, 2009.
- [15] Jamal Saboune and Robert Laganier. People detection and tracking using the explorative particle filtering. *ICCV*, pages 1298–1305, 2009.
- [16] Senem Velipasalar, Ying li Tian, and Arun Hampapur. Automatic counting of interacting people by using a single uncalibrated camera. *Multimedia and Expo, IEEE International Conference on*, 0:1265–1268, 2006.
- [17] Tarek Yahiaoui, Cyril Meurie, Louahdi Khoudour, and François Cabestaing. A people counting system based on dense and close stereovision. *ISP - LNCS*, 5099:59–66, 2008.
- [18] Danny B. Yang, Héctor H. González ba nos, and Leonidas J. Guibas. Counting people in crowds with a real-time network of simple image sensors. In *In Proc. IEEE International Conference on Computer Vision*, pages 122–129, 2003.
- [19] Qiang Wu Ying Xin, Guangmin Sun. A preprocessing method for tracking and counting pedestrians in bus video monitor. *INDIN*, pages 1689–1693, 2008.
- [20] Haibin Yu, Zhiwei He, and Jilin Liu. A vision-based method to estimate passenger flow in bus. In *International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)*, 2007.