

IDENTIFICAÇÃO DE PEDESTRES POR MEIO DE MAPAS DE DENSIDADE CONSTRUÍDOS COM ASIFT E FLUXO ÓPTICO

MARCUS VINÍCIUS FITZ LUCCHETTI*, PATRICK MARQUES CIARELLI†

** Universidade Federal do Espírito Santo
Departamento de Engenharia Elétrica
Vitória, Espírito Santo, Brasil*

Emails: marcus.lucchetti@aluno.ufes.br, patrick.ciarelli@ufes.br

Abstract— Pedestrian detection in digital images is a relevant paradigm in areas such as security and entities tracking. Recent worldwide events involving the investigation of suspicious actions in public places, such as airports, subway stations and crowd manifestations justify the direction of researches on this subject. Different clothing styles, colors, brightness, articulations, intra-inter people occlusions and similarities among boundaries of objects are some of the obstacles to be overcome by researchers. Under this perspective, this paper contributes with a proposal to detect people in crowd. The presented method combines a people detection technique with a density map generated by relevant keypoints detected using ASIFT and validated with optical flow. The experimental results indicate that this approach is more accurate than traditional methods.

Keywords— Optical Flow, Density Map, People Detection, Image Processing, ASIFT

Resumo— Detecção de pessoas em imagens digitais é um paradigma de interesse pelas oportunidades de aplicação em temas como segurança e rastreamento de entidades. Recentes acontecimentos no mundo envolvendo a investigação de ações suspeitas em locais públicos como aeroportos, estações de metrô e manifestações com grande aglomerado de pessoas justificam o direcionamento de pesquisas no tema. Os diferentes estilos de roupas, cores, iluminação, articulações, oclusões intra-entre pessoas e a similaridade do contorno de objetos são alguns dos obstáculos a serem vencidos nos estudos. Neste sentido, este artigo contribui com uma proposta para detecção de pessoas em grandes aglomerados de indivíduos. O método apresentado combina uma técnica de detecção de pessoas com um mapa de densidade da concentração de possíveis indivíduos, este obtido a partir de relevantes *keypoints* detectados com ASIFT e validados com fluxo óptico. Os resultados dos experimentos indicam que essa abordagem é mais precisa que os métodos tradicionais.

Palavras-chave— Fluxo Óptico, Mapa de Densidade, Detecção de Pessoas, Processamento de Imagens, ASIFT

1 Introdução

A demanda de estudos direcionados a detecção de objetos em imagens vem atraindo a atenção principalmente por questões de segurança pessoal e patrimonial, onde se observam a necessidade de monitoramento automatizado de pessoas em locais públicos. Recentes acontecimentos no mundo envolvendo a investigação de ações suspeitas em aeroportos e estações de trem ou ainda a estimativa de participantes em manifestações políticas justificam os investimentos na área.

Diversas são as aplicações da análise de densidade e monitoramento de pessoas como o acompanhamento da evolução da concentração de clientes em um estabelecimento, a medição do nível de conforto de transeuntes e a indicação de ameaças de tumultos, protestos violentos, brigas, pânico e agitação de massa (Junior et al., 2010).

Alguns dos desafios encontrados nos estudos voltados a detecção de pessoas em grandes aglomerações são: a oclusão de partes do corpo por outras pessoas e as detecções falsas, onde se confunde uma pessoa com um outro objeto (Zhang et al., 2014). Quanto às abordagens, observa-se uma maior preocupação sobre temas dirigidos ao reconhecimento, rastreamento e estimativa de localização de indivíduos, os quais variam em técnica, mas visam melhorias em precisão ou tempo

de processamento.

Este trabalho apresenta uma abordagem de detecção de características de interesse usando ASIFT (*Affine Scale Invariant Feature Transform*) (Yu e Morel, 2011) cuja validação dessas características é realizada pelo fluxo óptico CBRLOF (*Cross-Based Robust Local Optical Flow*) (Senst et al., 2014). As características mais relevantes são utilizadas para estimar, em cada *frame*, um mapa de densidade de indivíduos. O mapa de densidade é integrado a um detector de pedestres que permite assim localizar a posição de cada pessoa em cada *frame*. Neste trabalho foi utilizado um detector baseado em aprendizado a partir da mistura de modelos de partes deformáveis multiescalares (Forsyth, 2014).

A utilização da estratégia proposta possibilitou resultados melhores do que a simples utilização de detectores de pedestres e de outras técnicas de construção de mapa de densidade, como o uso de FAST, apresentado em (Fradi e Dugelay, 2015).

2 Trabalhos Relacionados

Alguns dos desafios a serem superados por pesquisas dedicadas a imagens com concentração de pessoas estão nas inter-occlusões ou auto-occlusões existentes nas partes dos corpos das pessoas e nas falsas detecções, onde confunde-se a detecção de

uma pessoa com um objeto ou com parte do corpo de uma outra pessoa, como um ombro ou braço (Zhang et al., 2014).

As imagens apresentadas na Figura 1 ilustram alguns desafios encontrados no reconhecimento de pessoas em imagens com maior concentração de pedestres.

Em (Ali e Dailey, 2012) é combinado o detector Viola e Jones (Viola e Jones, 2001) com a eliminação de falsos positivos pelo método de mínimos quadrados não linear, e um classificador Adaboost cascadeado para o rastreamento de pessoas em ambientes de grandes oclusões a partir do mapeamento de cabeças de indivíduos detectados nas imagens.

Na tendência de trabalhos com extração de características locais, em (Yang et al., 2012) é utilizado o código SST-LBP (*Sparse Spatio-temporal Local Binary Parttern*) como extrator. Este extrator separa as informações em duas categorias, temporais e espaciais, que são relacionadas ao final do processo por meio de um SVM (*Support Vector Machine*), cuja saída apresenta o nível de densidade de pessoas por região da imagem.

Soluções que utilizam fluxo óptico na determinação do mapa de densidade de pessoas em *frames* sequenciais, como em (Rao et al., 2014) e (Fradi e Dugelay, 2015), não necessitam de treinamento prévio para detecção de características distintas e aplicação de técnicas para eliminação de ruídos inerentes à captura das imagens. Abordagens recentes ainda justificam a utilização de redes neurais convolucionais, como em (Fu et al., 2015), que traz a possibilidade de aprendizado de características pela alteração dos pesos das conexões da rede.

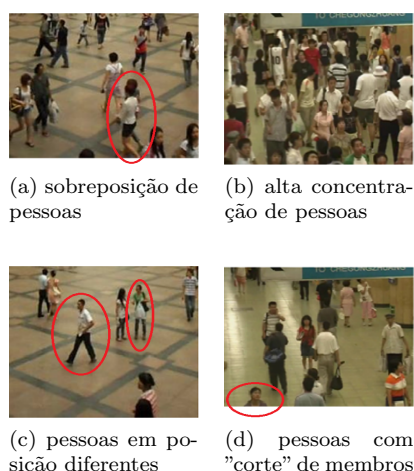
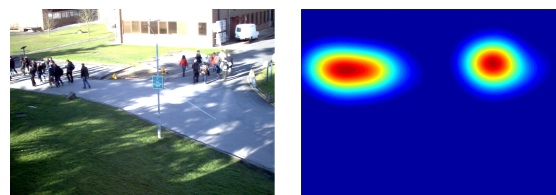


Figura 1: Exemplo do posicionamento de pessoas em diferentes cenas.



(a) Imagem com grupo de pessoas (b) Mapa de Densidade da Figura 2a

Figura 2: Mapa de Densidade Visual de uma imagem da base de dados PETS2009.¹

3 Mapa de Densidade

O uso de mapas de densidade auxilia na identificação das regiões mais prováveis de presença de pessoas, sendo portanto um método para melhorar a precisão de um sistema de detecção de pedestres.

Baseado em (Fradi e Dugelay, 2015), neste trabalho é utilizado uma metodologia que prevê a extração de características locais da imagem indicando a presença do conteúdo de interesse. Em seguida, rastreia-se o conteúdo encontrado aplicando fluxo óptico, o qual é responsável por eliminar características desprezíveis, como regiões que aparecem e desaparecem na imagem, por um processo conhecido como *forward-backward*. Ao final desse processo, o mapa de densidade de pessoas é calculado por uma função de soma de gaussianas fornecendo os valores de contribuição da densidade por *pixel* da imagem. O resultado desse processo pode ser visto na Figura 2.

3.1 Reconhecimento de Características - ASIFT

Para a extração local das características, este trabalho propõe o uso do extrator ASIFT apresentado em (Yu e Morel, 2011). O método ASIFT faz uso da Transformada Afim (*Affine Transform*), a qual representa a imagem digital em outra forma como uma combinação de rotações, escalonamento, desvio e translação em n dimensões espaciais.

A técnica SIFT (*Scale Invariant Feature Transform*), já utilizada em outros trabalhos para a extração de características (Zhang e Zhang, 2015; Fradi e Dugelay, 2015), utiliza quatro parâmetros: rotação, translação (x, y) e escala. O ASIFT, por meio de simulações de variação do eixo de orientação de câmera da imagem, conhecidos como ângulos de latitude e longitude, detém mais dois novos parâmetros não previstos pelo SIFT. Dessa forma, o ASIFT consegue capturar muito mais informações relevantes do que o SIFT. A Figura 3 ilustra a extração e uma imagem usando ASIFT e os diferentes métodos avaliados em (Fradi e Dugelay, 2015). Como pode ser

¹<http://www.cvg.reading.ac.uk/PETS2009/>

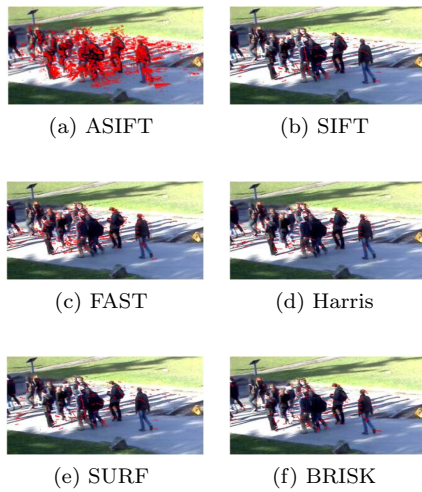


Figura 3: Extração de características por diferentes métodos.

observado, o ASIFT é superior a todos os outros métodos apresentados.

A saída do extrator neste trabalho é um vetor y_0 representando as coordenadas (x,y) de localização das características encontradas.

3.2 Projeção Forward-Backward

Neste trabalho propõe-se o uso do fluxo óptico CBRLOF (Senst et al., 2014), um método mais preciso e de melhor performance computacional do que outros métodos clássicos.

Para cada coordenada de *pixel* do vetor y_0 do *frame* I_k verifica-se, pelo fluxo óptico, a sua projeção sobre o *frame* seguinte I_{k+1} , gerando o vetor y_1 de coordenadas (x,y) com as projeções encontradas (passo de *forward*). De maneira similar, faz-se o caminho inverso, verificando as projeções do conteúdo de y_1 , do *frame* I_{k+1} sobre I_k gerando o vetor y_2 (passo de *backward*).

As coordenadas resultantes da operação $y'_0 = y_0 \cap y_2$ representam os *pixels* de saída deste processo. O vetor y'_1 recebe então os valores das coordenadas de y_1 que projetaram em y'_0 . A Figura 4 exemplifica este processo pelo fluxo óptico de *pixels* hipotéticos.

Para descartar informações de pequenas trajetórias, normalmente associadas a pequenos ruídos, imperfeições do sistema de aquisição ou movimentos irrelevantes, um procedimento adicional de limiar de trajetória é realizado. A partir das coordenadas em y'_0 e y'_1 calcula-se as distâncias euclidianas dos *pixels* relacionados. Coordenadas com distâncias d' menores que um dado limiar ζ são eliminadas, e as coordenadas restantes são representadas pelo vetor y''_0 .

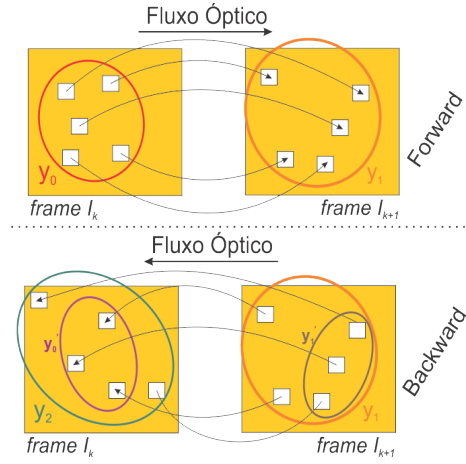


Figura 4: Processo Forward-Backward.

3.3 Construção do Mapa de Densidade

A etapa de cálculo do mapa de densidades assume que as características remanescentes dos processos das Seções de 3.1 a 3.2 possuem igual contribuição na imagem, sendo que regiões com maior concentração de características representam uma maior densidade de pessoas e regiões com menor concentração representam menor densidade de pessoas em um *frame* I_k . O mapa de densidade C_k é estimado por uma soma de funções de densidade de probabilidade (pdf) Gaussianas representada na Equação 1 a partir das características locais em y''_{0k} , considerando a contribuição de cada *pixel* (x,y) do *frame* I_k .

$$C_k(x, y) = \frac{1}{\sqrt{2\pi}\sigma} \sum_{i=1}^{y''_{0k}} e^{-\left(\frac{(x-x_i)^2 + (y-y_i)^2}{2\sigma^2}\right)}, \quad (1)$$

onde σ é a largura de banda definida para uma sequência de vídeo. A matriz $C_k(x, y)$ é então normalizada para valores entre 0 e 1. Um resultado visual da Equação 1 pode ser visto na Figura 2b.

O fluxograma da Figura 5 apresenta de forma resumida a interação dos processos até o momento de cálculo do mapa de densidade.

4 Detetor de Pessoas

Para detecção de pessoas, este trabalho emprega o detetor genérico proposto por (Forsyth, 2014). Neste métodos, os objetos são descritos como uma mistura de modelos de partes deformáveis multi-escalares treinadas, a partir das informações marcadas por “caixas” ou *boxes* na base de dados de treino, pela técnica discriminativa LSVM (*Latent Support Vector Machines*).

O detetor foi treinado na base de dados INRIA¹, a qual fornece informação de pessoas isoladas em diferentes posições. Ao fazer detecções em

¹<http://lear.inrialpes.fr/data>

um *frame* I_k , um conjunto $D_k = \{d_1^k, \dots, d_{n_k}^k\}$ é gerado, onde d_j^k é a j -ésima detecção de D_k , e d_j^k os valores de localização e dimensões das *boxes* dado por $d_j^k = \{x_j^k, y_j^k, w_j^k, h_j^k, score_j^k\}$. O par x_j^k e y_j^k correspondem às localizações do canto esquerdo superior das *boxes*, w_j^k e h_j^k são a quantidade de *pixels* largura e altura das *boxes*, respectivamente, e o termo $score_j^k$ refere-se a avaliação daquela detecção calculada pelo detector. Quanto maior o valor de $score_j^k$ mais confiável foi aquela detecção. Existem ainda dois limiares ajustáveis no detector: t , limiar de *score*, e B , limiar de sobreposição de hipóteses, sendo eliminadas detecções que apresentarem valores menores que tais limiares.

5 Integração Mapa de Densidades e Detector

A partir das informações sobre o mapa de densidades gerado na Seção 3, (Fradi e Dugelay, 2015) propõem uma forma de otimização do comportamento dos detectores de pessoas pela aplicação de filtros e utilização de um limiar dinâmico (τ_{din}) calculado para cada uma das detecções. Um dos maiores benefícios do emprego desta rotina está relacionado com a redução de falsos positivos deixando o detector mais preciso.

O processo de integração compreende a uma etapa inicial de restrições geométricas que relacionam as razões de largura e altura das detecções e predições de detecções entre o *frame* I_{k-1} e I_k . Uma segunda etapa, chamada de NMS (*Non-Maximum Supression*), verifica a sobreposição das detecções e a sua relevância quanto à densidade daquela região.

5.1 Restrições Geométricas

As aplicação das restrições geométricas neste ponto é sugerida devido às características do modelo de partes deformáveis, que pode mesclar par-

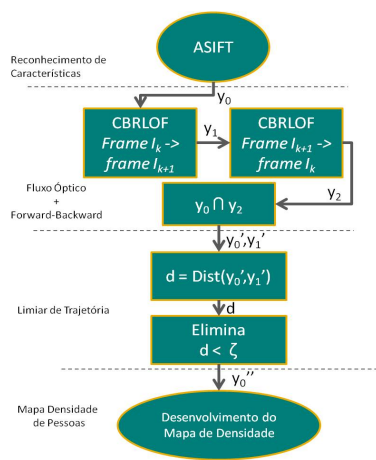


Figura 5: Procedimento de Cálculo do Mapa de Densidade.

tes de pessoas diferentes em uma única detecção. Essas restrições são aplicadas em duas etapas. Seja D_k o conjunto de detecções realizadas no *frame* I_k , a Equação 2, referente à primeira etapa, relaciona as alturas h_j^k e os pontos y_j^k das detecções de I_k , permitindo por regressão linear que os valores de α_{k-1} e β_{k-1} sejam encontrados:

$$h_j^k = \alpha_{k-1} \cdot y_j^k + \beta_{k-1} + \epsilon, \quad j \in \{1 \dots n_k\}, \quad (2)$$

apenas detecções que atenderem à restrição de erro de altura dada por $h_j^k \leq \tilde{h}_j^k \pm \Delta_h$ são aceitas, onde $\tilde{h}_j^k$ são as alturas projetadas a partir da Equação 2 pelos dados de y_j^k , e Δ_h é um certo limiar de aceite ajustado pelo. O erro ϵ é ocasionado pela imprecisão do modelo.

Uma segunda restrição geométrica é dada pelas razões de largura e altura das detecções. A Equação 3 introduz o parâmetro γ_{k-1} que permite a visualização dessa relação:

$$\gamma_{k-1} = \text{mediana} \left\{ \frac{w_j^i}{h_j^i} \right\}_{1 \leq i \leq (k-1), 1 \leq j \leq n_i}. \quad (3)$$

Observe que na Equação 3 os valores de i variam desde o primeiro *frame* da sequencia até o penúltimo *frame* I_{k-1} . O valor de γ_{k-1} encontrado é então comparado na inequação de erro $\left(\frac{w_j^i}{h_j^i}\right) \leq \gamma_{k-1} \pm \Delta_\gamma$ sendo que detecções em que a razão $\left(\frac{w_j^i}{h_j^i}\right)$ não atenderem à inequação são descartadas. Δ_γ é um escalar de aceite ajustado pelo usuário. Os parâmetros α_k, β_k e γ_k são atualizados a cada nova rodada do cálculo do mapa de densidade.

5.2 Restrições NMS

As restrições NMS permitem a eliminação de detecções sobrepostas D_k e cálculo dos limiares dinâmicos (τ_{din}) para cada detecção de D_k . A etapa inicial testa as superposições de todas as detecções d_j^k de um *frame* I_k entre si, sendo que na identificação de sobreposições que ultrapassem um dado valor Δ_0 , atribuída pelo usuário, as *boxes* de menores *scores* são eliminadas.

Uma segunda etapa de cálculo de limiar de detecção dinâmica (τ_{din}) para cada detecção d_j^k é dada pelas Equações 1, 4 e 5.

$$\hat{C}_k(d_j^k) = \frac{\sum_{p=0}^{h_j^k-1} \sum_{q=0}^{w_j^k-1} C_k(x_j^k + p, y_j^k + q)}{h_j^k \cdot w_j^k}, \quad (4)$$

$$\tau_{din} = \tau_{max} + (\tau_{min} - \tau_{max}) \cdot \hat{C}_k(d_j^k), \quad j \in \{1, \dots, n_k\}, \quad (5)$$

onde $\tau_{max} = -0.5$ e $\tau_{min} = -1.2$ foram empiricamente encontrados e fornecidos em (Fradi e Dugelay, 2015).

Tabela 1: Valores dos par metros utilizados neste trabalho.

Base de Dados	t	B	Δ_γ	Δ_0	Δ_h	σ	ζ
PETS2009-S1L1-1-(13-57)	-1	0.5	0.003	0.7	100	90	1
PETS2009-S1L1-2-(13-59)	-1	0.5	0.003	0.7	100	90	1
PETS2009-S1L2-2-(14-31)	-1.3	1	0.005	0.5	25	20	1
PETS2009-S2L3-(14-41)	-1	0.5	0.003	0.7	100	90	1

6 Resultados Experimentais

6.1 Base de Dados

Nos experimentos foi utilizada a base de dados PETS2009, que   composta por sequ ncias de imagens com resolu  o 768x576, capturadas por uma c mera em posi  o fixa, mas que variam em quantidade de pessoas e complexidade de cen rio.

As sequ ncias S1-L1-1 e S1-L1-2 representam n veis m dios de quantidade de pessoas bem distribu das com 221 e 241 *frames*, respectivamente. A sequ ncia S1-L2-2 possui uma maior quantidade de pessoas distribu das em 131 imagens contendo regi es de sombra e ilumina  o relevantes, tornando-a uma sequ ncia mais desafiadora. J  a sequ ncia S2-L3   composta por 240 imagens que apresentam tamb m um grande n mero de pessoas em grupos mais concentrados.

6.2 Indicadores de Avalia  o

Com o intuito de comparar os resultados desta pesquisa com outros trabalhos, foram empregados os indicadores MODP (*Multiple Object Detection Precision*) e MODA (*Multiple Object Detection Accuracy*), dispon veis em (Stiefelhaven et al., 2007), e NMAE (*Normalized Mean Absolute Error*) (Shani e Gunawardana, 2011). O indicador NMAE avalia a precis o do extrator de caracter sticas ASIFT em rela  o aos *pixels* de pessoas dispon veis nas anota  es (*ground truths*)². Os indicadores de desempenho MODP e a MODA revelam a precis o das detec  es relacionando, respectivamente, a  rea detectada sobre o n mero de detec  es realizadas, e a quantidade de falsos positivos somada  s perdas de detec  es (*misses*), tudo considerando os *ground truths*. Quanto mais pr ximo de 1 (um) estiverem os valores de MODP e MODA mais precisa   a detec  o. Quanto mais pr ximo de zero for NMAE, melhor.

²<http://milanton.de/data/>

Tabela 2: Valores de NMAE calculados para diferentes t cnicas de extra  o de caracter sticas na base PETS2009.

Base de Dados PETS2009	ASIFT	ASIFT sem Fluxo �ptico	FAST	SIFT	GFT	MSER	SURF
S1L1-1-(13-57)	0.02	0.09	0.07	0.07	0.08	0.07	0.07
S1L1-2-(13-59)	0.03	0.1	0.04	0.04	0.04	0.04	0.05
S1L2-2-(14-31)	0.02	0.04	0.09	0.09	0.10	0.07	0.07
S2L3-(14-41)	0.01	0.05	0.04	0.03	0.03	0.03	0.03

Tabela 3: Indicadores de MODP e MODA aplicados   base de dados PETS2009 utilizando diferentes t cnicas de extra  o de caracter sticas.

Base de Dados PETS2009	ASIFT		FAST		{sem mapa}	
	MODP	MODA	MODP	MODA	MODP	MODA
S1L1-1-(13-57)	0.64	0.57	0.63	0.63	0.61	0.56
S1L1-2-(13-59)	0.68	0.60	0.68	0.6	0.66	0.56
S1L2-2-(14-31)	0.63	0.48	0.63	0.47	0.69	-17.53
S2L3-(14-41)	0.65	0.47	0.57	0.35	0.63	0.46

6.3 Resultados do Mapa de Densidade de Pessoas

Os experimentos realizados nas Se  es 6.3 e 6.4 utilizam os par metros detalhados na Tabela 1 para cada v deo.

A Tabela 2 apresenta os resultados da m trica NMAE para o ASIFT e fluxo  ptico, calculado neste trabalho, e os resultados de outras t cnicas de extra  o de caracter sticas em conjunto com fluxo  ptico fornecidos em (Fradi e Dugelay, 2015). O melhor resultado para cada v deo est  real ado em negrito. Como pode ser observado, o ASIFT foi superior em todos os casos.

Ainda pela Tabela 2, os maiores valores de NMAE verificados em experimentos com ASIFT sem a utiliza  o do fluxo  ptico e suas valida  es, caracterizam o importante papel deste est gio no processo.

6.4 Resultados da Detec  o de Pessoas ap s a Integra  o com o Mapa de Densidades

Os resultados da integra  o s o apresentados na Tabela 3, onde   feita para cada v deo a compara  o dos resultados usando o ASIFT e o FAST (que foi o melhor extrator de caracter sticas usado em (Fradi e Dugelay, 2015) e seus resultados est o na Tabela 3). Tamb m foram adicionadas para compara  o os resultados obtidos sem o uso do mapa de densidade. O melhor resultado para cada m trica e v deo est  real ado em negrito. Como pode ser observado, a abordagem usando o ASIFT mostra ganhos quando comparada com o uso do FAST. Tamb m observa-se que os resultados do uso do mapa de densidade, em especial ao constru do com ASIFT, s o superiores ao m todo sem us -lo. A  nica exce  o foi para o v deo S1L2-2, em que MODP foi maior sem usar o mapa, por m o baixo valor de MODA evidencia uma grande detec  o de falsos positivos.

Percebe-se pela Tabela 1 que foi necess rio o arranjo de par metros exclusivos para a base de dados PETS2009-S1L2-2-(14-31) para obten  o de melhores resultados. A utiliza  o dos mesmo par metros das demais sequ ncias n o ofereceram bons resultados. O motivo para esta diferencia  o est  nos maiores desafios inerentes a essa base de dados devido  s abruptas altera  es de ilumina  o que acaba comprometendo mais ainda a precis o do detector.

7 Conclus es

Este trabalho teve como objetivo investigar o uso da t cnica ASIFT junto com o fluxo  ptico CBR-LOF para a constru  o de uma mapa de intensidade de forma a tornar mais precisa a identifica  o de pessoas em grandes concentra  es.

Os resultados obtidos mostraram que o extrator de caracter sticas ASIFT consegue identificar mais pontos relevantes e o fluxo  ptico ajuda a torn  -lo mais preciso, sendo superior a todas as outras abordagens comparadas, tornando o sistema mais preciso.

As respostas encontradas com a integra  o do mapa de densidades com a detec  o de pedestres indicou que o uso do ASIFT contribui com resultados mais precisos do que o uso do FAST. Al m disso foi observado a contribui  o do mapa de densidade para a detec  o de pedestres.

A utiliza  o da base de dados PETS2009 mostrou-se desafiadora para o detector de partes deform veis utilizada, principalmente pelas diferen as de luminosidade, as oclus es existentes entre os indiv duos e a semelhan a entre objetos do cen rio gerando perdas de detec  es e falsos positivos. Foi verificada a dificuldade para encontrar uma calibra  o  nica para todos os v deos, sendo que diferentes condi  es de ilumina  o e de ambientes podem necessitar de diferentes calibra  es, como foi o caso para o v deo PETS2009-S1L2-2- (14-31).

Uma vez que as t cnicas utilizadas neste trabalho podem ser empregadas a qualquer outro detector, um desafio ainda previsto est  relacionado a utiliza  o de detectores ainda mais robustos para tornar a tarefa de identifica  o de pessoas em grandes concentra  es mais precisas. A experimenta  o de limiares que podem melhorar ainda mais as precis es de MODA e MODP no que tange a remo  o de falsos positivos e a precis o das  rea de detec  o   uma outra  rea que pode ser investigada em trabalhos futuros.

Agradecimentos

Agradecemos ao Programa de P s-Gradua  o em Engenharia El trica da UFES por disponibilizar a estrutura e recursos para o desenvolvimento desta pesquisa.

Refer ncias

Ali, I. e Dailey, M. N. (2012). Multiple human tracking in high-density crowds, *Image and Vision Computing* **30**(12): 966 – 977.

Forsyth, D. (2014). Object detection with discriminatively trained part-based models, *Computer* **47**(2): 6–7.

Fradi, H. e Dugelay, J.-L. (2015). Towards crowd density-aware video surveillance applications, *Information Fusion* **24**: 3 – 15.

Fu, M., Xu, P., Li, X., Liu, Q., Ye, M. e Zhu, C. (2015). Fast crowd density estimation with convolutional neural networks, *Engineering Applications of Artificial Intelligence* **43**: 81 – 88.

Junior, J. C. S. J., Musse, S. R. e Jung, C. R. (2010). Crowd analysis using computer vision techniques, *IEEE Signal Processing Magazine* **27**(5): 66–77.

Rao, A. S., Gubbi, J., Marusic, S. e Palaniswami, M. (2014). Estimation of crowd density by clustering motion cues, *The Visual Computer* **31**(11): 1533–1552.

Senst, T., Borgmann, T., Keller, I. e Sikora, T. (2014). Cross based robust local optical flow, *21th IEEE International Conference on Image Processing*, Paris, France, pp. 1967–1971.

Shani, G. e Gunawardana, A. (2011). *Recommender Systems Handbook*, Springer US, Boston, MA, chapter Evaluating Recommendation Systems, pp. 257–297.

Stiefelhagen, R., Bernardin, K., Bowers, R., Garofolo, J., Mostefa, D. e Soundararajan, P. (2007). *Multimodal Technologies for Perception of Humans, CLEAR 2006, Southampton, UK, April 6-7, 2006, Revised Selected Papers*, Springer Berlin Heidelberg, Berlin, Heidelberg, chapter The CLEAR 2006 Evaluation, pp. 1–44.

Viola, P. e Jones, M. (2001). Robust real-time object detection, *International Journal of Computer Vision*.

Yang, H., Cao, Y., Su, H., Fan, Y. e Zheng, S. (2012). The large-scale crowd analysis based on sparse spatial-temporal local binary pattern, *Multimedia Tools and Applications* **73**(1): 41–60.

Yu, G. e Morel, J.-M. (2011). Asift: An algorithm for fully affine invariant comparison, *Image Processing On Line* **1**.

Zhang, S., Bauckhage, C. e Cremers, A. B. (2014). Informed haar-like features improve pedestrian detection, *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pp. 947–954.

Zhang, S. e Zhang, X. (2015). Crowded pedestrian detection and density estimation by visual words analysis, *International Journal of Multimedia and Ubiquitous Engineering* **10**(3): 99–108.