

Impacto da Resolução do Quadro na Qualidade da Segmentação

Impact of Frame Resolution on Segment Quality

RESUMO

Fagner Henrique Santos Onesko
fagneronesko@alunos.utfpr.edu.br
Universidade Tecnológica Federal do Paraná, Cornélio Procópio, Paraná, Brasil

Claiton de Oliveira
claitonoliveira@utfpr.edu.br
Universidade Tecnológica Federal do Paraná, Cornélio Procópio, Paraná, Brasil

Este trabalho tem como objetivo o levantamento de informações da segmentação de imagens de diversos algoritmos, mostrando em como a resolução da imagem afeta essa ação. Para isso fez-se uso de 38 algoritmos e calculou-se diversas métricas utilizadas para avaliar a qualidade da segmentação de vídeos 4K, sendo um referente a pedestres e outro a veículos, os quais foram reduzidos suas resoluções até 10% do valor original. Com as métricas calculadas fez-se uso delas para o cálculo da classificação média para cada resolução e verificou-se que a resolução da imagem tem influência nos resultados da segmentação dos algoritmos.

PALAVRAS-CHAVE: Segmentação. Resolução. Impacto.

Recebido: 19 ago. 2019.

Aprovado: 01 out. 2019.

Direito autoral: Este trabalho está licenciado sob os termos da Licença Creative Commons-Atribuição 4.0 Internacional.



ABSTRACT

THIS PROJECT AIMS TO GATHER INFORMATION ON THE SEGMENTATION OF IMAGES OF AN ALGORITHM, SHOWING AN IMAGE RESOLUTION. FOR THIS, 38 ALGORITHMS WERE USED AND AVERAGED SEVERAL METRICS IN THE SEGMENTATION OF 4K VIDEOS, ONE REFERRING TO PEDESTRIANS AND THE OTHER TO VEHICLES, WHICH REDUCED THEIR RESOLUTIONS UP TO 10% OF THE ORIGINAL VALUE. WITH THE CALCULATED METRICS, THEY WERE USED TO CALCULATE METRICS FOR EACH RESOLUTION AND IT WAS FOUND THAT IMAGE RESOLUTION IMPACT ON THE QUALITY OF SEGMENTATION CLASSIFICATION.

KEYWORDS: Segmentation. Resolution. Impact.

INTRODUÇÃO

Sistemas voltados para vigilância por vídeo são utilizados com o objetivo de oferecer melhor qualidade de vida para a sociedade. A maioria das câmeras responsáveis pela captura do vídeo utilizado nessas aplicações produzem imagens de média ou baixa resolução devido, entre outros fatores, ao econômico. Câmeras mais sofisticadas, no entanto, são capazes de capturar vídeos com resoluções mais altas. A qualidade dessas imagens pode influenciar positivamente a qualidade da segmentação obtida desses algoritmos, no entanto, são necessários experimentos nesse sentido para verificar essa hipótese. Não há clareza do quanto impactante a resolução da imagem afeta sua segmentação (Dinakaran, Sexton, Seker, Bouridane, & Jiang, 2017).

O processamento de imagens ou vídeos de alta resolução é demorado e, por esse motivo, a resolução apropriada que alcança a confiabilidade que determinadas aplicações exigem juntamente com um tempo de processamento aceitável deve ser escolhida (Romić, Galić, & Leventić, 2016). Neste trabalho será analisado o impacto da resolução na qualidade da segmentação utilizando dois vídeos: um referente a pedestres e outros a veículos, por meio da execução de 38 algoritmos e o cálculo de diversas métricas necessárias para se observar o resultado.

MATERIAL E MÉTODOS

Inicialmente foram selecionados 38 algoritmos de segmentação utilizados em aplicações de vigilância por vídeo. Esses algoritmos foram executados sobre dois vídeos com resolução 4K. Em um deles, o conteúdo exibe pessoas em ambiente externo e o outro que exibe veículos passando por uma rodovia. O vídeo referente a pessoas foi capturado na UTFPR- Campus Cornélio Procópio e o referente aos veículos foi capturado na cidade de Marília-SP.

Os algoritmos escolhidos para realizar a segmentação foram citados e obtidos da biblioteca *bgslibrary* (Sobral, 2013). Sendo estes: *FrameDifference*, *StaticFrameDifference*, *WeightedMovingMean*, *WeightedMovingVariance*, *AdaptiveBackgroundLearning*, *AdaptiveSelectiveBackgroundLearning*, *DPAdaptiveMedian*, *DPGrimsonGMM*, *DPZivkovicAGMM*, *DPMean*, *DPWrenGA*, *DPPratiMediod*, *DPEigenbackground*, *DPTtexture*, *T2FGMM_UM*, *T2FGMM_UV*, *T2FMRF_UV*, *FuzzySugenoIntegral*, *FuzzyChoquetIntegral*, *MultiLayer*, *PixelBasedAdaptiveSegmenter*, *VuMeter*, *LBSimpleGaussian*, *LBFuzzyGaussian*, *LBMixtureOfGaussians*, *LBAAdaptiveSOM*, *LBFuzzyAdaptiveSOM*, *KDE*, *IndependentMultimodal*, *MultiCue*, *SigmaDelta*, *Subsense*, *Lobster*, *TwoPoints*, *ViBe*, *CodeBook*, *SWCD*, *CLVID*.

Ambos os vídeos possuem resolução 4K e um total de 51 quadros. Para cada quadro, foram rotulados manualmente 51 quadros correspondentes em que os pixels pertencentes ao fundo são da cor preta e os pertencentes ao elemento de interesse são da cor branca. Esse conjunto de quadros rotulados são conhecidos como *ground truth* e, por meio deles, é possível identificar tanto os pixels classificados de forma correta quanto os classificados de forma incorreta pelos algoritmos. Para isso, os quadros contendo o resultado da segmentação de determinado algoritmo são comparados com o *ground truth*.

Utilizando os vídeos com resolução de 4K (3840 x 2160) *pixels* e seus respectivos ground truths, foram gerados novos vídeos com resoluções mais baixas, reduzidas de 10% da resolução anterior. Foram gerados novos vídeos com resoluções (3456 x 1944) *pixels*, (3072 x 1728) *pixels*, (2688 x 1512) *pixels*, (2304 x 1296) *pixels*, (1920 x 1080) *pixels*, (1536 x 864) *pixels*, (1152 x 648) *pixels*, (768 x 432) *pixels*, (384 x 216) *pixels*. Todos os algoritmos foram executados sobre todos os vídeos utilizando uma máquina Intel Quad-Core i5-3330 64 bits (3.00GHz), 12GB RAM, Linux Ubuntu 16.04 (64 bits), gerando resultados na forma de quadros segmentados em que o elemento de interesse, segundo o algoritmo, apresenta-se na cor branca e fundo apresenta-se na cor preta.

Uma vez obtidos os resultados da segmentação, foram calculadas 7 métricas utilizando *scripts* Matlab (R2013b). Sendo assim, para os dois vídeos foram realizados as análises para cada resolução e para cada algoritmo, gerando uma tabela de 760 linhas com informações que serão discutidas na seção de resultados e discussões.

As 7 métricas citadas acima são apresentadas nas equações de 1 a 7

$$Recall = TP / (TP + FN) \quad (1)$$

$$Specificity = TN / (TN + FP) \quad (2)$$

$$FPR = FP / (FP + TN) \quad (3)$$

$$FNR = FN / (TP + FN) \quad (4)$$

$$PWC = 100 * (FN + FP) / (TP + FN + FP + TN) \quad (5)$$

$$Precision = TP / (TP + FP) \quad (6)$$

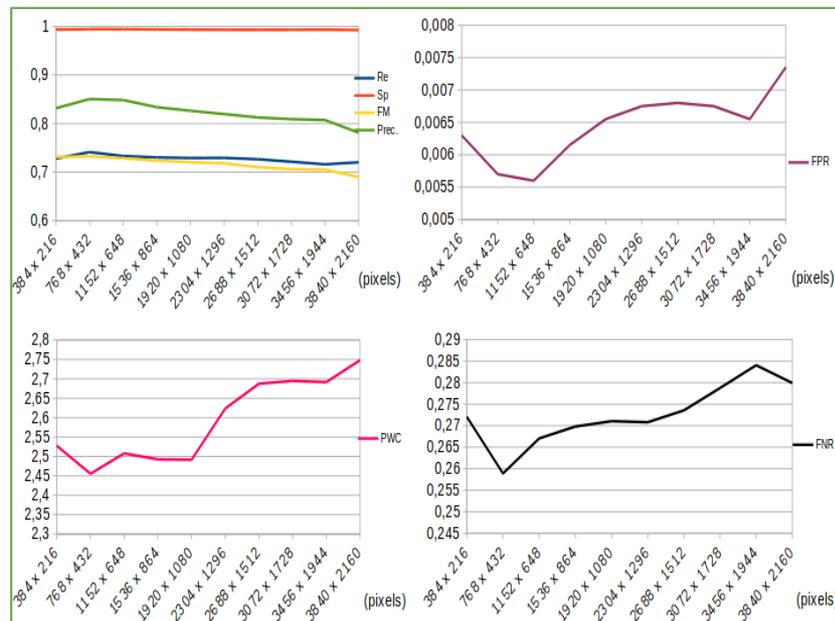
$$FMeasure = (2 * Precision * Recall) / (Precision + Recall) \quad (7)$$

onde FP são os *Falsos Positivos*, FN (*Falsos Negativos*), TP (*Verdadeiros Positivos*) e TN (*Verdadeiros Negativos*). Com a tabela, realizou-se a média das 7 métricas para cada uma das resoluções e para cada um dos vídeos, e obteve-se gráficos exibidos na seção seguinte.

RESULTADOS E DISCUSSÃO

Levantado os valores das 7 métricas descritas na seção anterior, realizou-se a média das métricas calculadas. Foram gerados 4 gráficos para cada um dos vídeos, facilitando a visualização devido a escala.

Figura 1 – Média das métricas para vídeo de pedestres



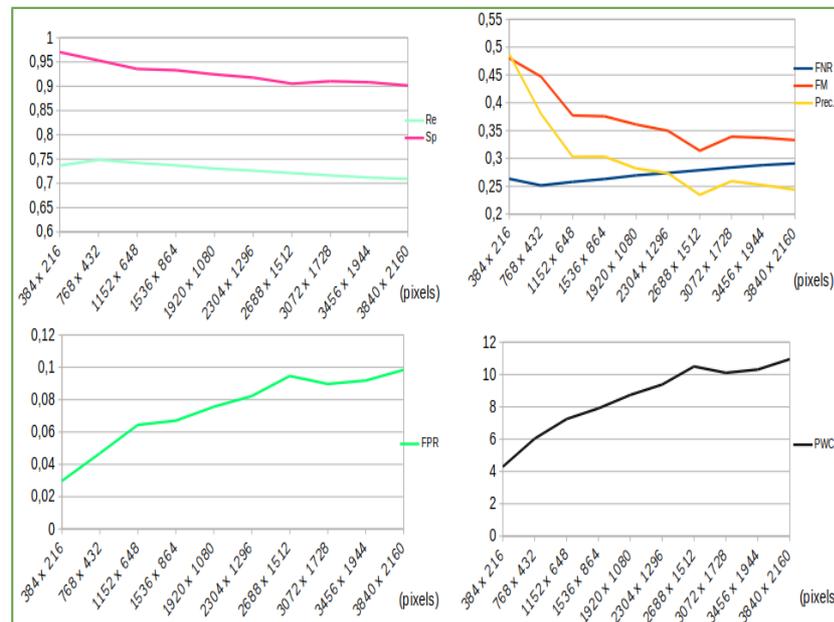
Fonte: elaborada pelo autor.

Observando a Figura 1 nota-se que a Precisão (Prec.) para as resoluções mais altas tendem a diminuir quando comparada a resoluções mais baixas, e que o valor mais alto é atingido quando a resolução é de (768 x 432) pixels. A partir dessa resolução, há uma perda na precisão. A métrica *Fmeasure*, que é calculada com base nas métricas Precisão e *Recall*, também tende a diminuir de acordo com o aumento da resolução, e seu valor maior pertence a resolução mínima trabalhada.

Outra métrica com diferenças notórias é o PWN, ou seja, a porcentagem de erros de classificações, que nesse caso tende a aumentar de acordo com o aumento das resoluções. Tais valores podem ser compreendidos observando que o FNR e FPR, ou seja, a taxa de falsos negativos e falsos positivos respectivamente, tende a aumentar nas resoluções mais altas. A maior taxa de falso positivo corresponde a resolução máxima trabalhada, e a para a taxa de falso negativo o máximo valor foi para a resolução de (3456 x 1944) pixels. Já os valores mínimos dessas métricas foram para as resoluções de (768 x 432) pixels e para (1152 x 648) pixels quando observado o FNR e FPR, respectivamente.

Na Figura 2 está ilustrado as médias do comportamento das 7 métricas para o vídeo referente a veículos.

Figura 2 – Média das métricas para vídeo de veículos



Fonte: elaborada pelo autor.

Na Figura 2 observa-se que a Precisão (Prec.) também tende a diminuir de acordo com o aumento da resolução, e que para o vídeo de veículos a precisão foi maior quando a resolução era de (384 x 216) pixels. A taxa de falso positivo (FPR) e falso negativo (FNR) também aumentaram para resoluções mais altas, e com isso a resolução máxima possuiu maior porcentagem de erros de classificação. Tal dado pode ser observado quando analisado o PWN.

O *Fmeasure* para esse vídeo também possuiu maior valor quando as imagens de resolução mínima eram segmentadas. O *Recall* (Re) e *Specificity* (Sp), tiveram um comportamento de regressão quando as resoluções se aproximam da máxima, e seus maiores valores eram das resoluções (384 x 216) pixels e (768 x 432) pixels.

Com isso, verificou-se que resoluções, a partir de determinado valor, tendem a errar mais e resoluções menores possuíram melhor precisão e acurácia nos testes realizados, desde que sejam acima de determinado valor. Tal fato pode se dar ao excesso de detalhes na imagem que atrapalham os algoritmos para segmentação de imagens com maiores resoluções, pois segundo (Romić et al., 2016) a alta resolução não necessariamente fornece melhores acurácias, pois o excesso de detalhes podem fornecer falsos positivos que baixam o desempenho.

CONCLUSÃO

O estudo da resolução de imagem e seu impacto é de extrema relevância para diversos estudos que envolvem o uso de algoritmos de segmentação de vídeos, seja para segmentação de pedestres, veículos ou para qualquer outro objeto de interesse. É de grande importância a seleção da menor resolução

possível que atenda as necessidades envolvidas, equilibrando o custo x benefício do sistema.

Como trabalho futuro pode ser realizado o estudo do impacto causado da qualidade da imagem na segmentação destas para outras diversas condições, como vídeo gravado com tempo chuvoso ou a noite, uma vez que neste trabalho os vídeos foram gravados em condições climáticas favoráveis ao algoritmo.

AGRADECIMENTOS

Agradeço primeiramente a Deus por me dar saúde, força e condições de superar todas as dificuldades. A minha família por todo apoio, pois sem isto não seria possível a conclusão deste trabalho. A Claiton de Oliveira e Silvio Rodrigues Sanches por toda instrução no desenvolvimento deste projeto.

REFERÊNCIAS

CDNET. **A video database for testing change detection for algorithms**. Disponível em: <http://changedetection.net/>. Acesso em: 10 ago. 2019.

DINAKARAN, R., SEXTON, G., SEKER, H., BOURIDANE, A., & JIANG, R. **Image resolution impact analysis on pedestrian detection in smart cities surveillance**. p. 1-8, 2007. Disponível em: <https://doi.org/10.1145/3109761.3109797>. Acesso em: 10 ago. 2019.

ROMIĆ, K., GALIĆ, I., & LEVENTIĆ, H. **Influence of the input image resolution on the staircase detection**. Proceedings Elmar - International Symposium Electronics in Marine. p. 177-180, set. 2016. Disponível em: <https://doi.org/10.1109/ELMAR.2016.7731781>. Acesso em: 10 ago. 2019.

SOBRAL, A. BGSLibrary: An opencv c++ background subtraction library. In: IX Workshope Visão Computacional (WVC'2013). Rio de Janeiro, Brazil: [s.n.], 2013. Disponível em: <<https://github.com/andrewssobral/bgslibrary>>