

Proposição de método para identificar e manipular regiões de dificuldade na segmentação de quadros de vídeos

Proposition of a method to identify and manipulate regions of difficulty in segmenting video frames

RESUMO

Katharina Akemi Ikeda Rosa
katharina@ufpr.alunos.edu.br
Universidade Tecnológica Federal do Paraná, Cornélio Procopio, Paraná, Brasil

Silvio Ricardo Rodrigues Sanches
silviosanches@utfpr.edu.br
Universidade Tecnológica Federal do Paraná, Cornélio Procopio, Paraná, Brasil

A avaliação de desempenho de um novo algoritmo deve mostrar a superioridade deste algoritmo em relação ao estado da arte existente. Para avaliar um algoritmo de *change detection* é necessário executá-lo com o objetivo de identificar áreas de mudança (segmentação) em quadros de vídeo do conjunto de dados e comparar os resultados com um *ground truth*. É possível comparar desempenhos quando vários algoritmos segmentam vídeos do mesmo *dataset*. Apesar da forma tradicional de avaliação de desempenho de algoritmos de *change detection* ser aceita pela comunidade científica, existem outras informações relevantes a partir do resultado da segmentação. O nível de dificuldade para classificar um pixel é um exemplo. Neste artigo, é apresentado um método para identificar a dificuldade em classificar pixels por meio de algoritmos de detecção de alterações. O método inclui o uso de resultados de segmentação de vários algoritmos de última geração, bem como o armazenamento de valores que representam níveis de dificuldade em matrizes denominadas “mapas de dificuldade”. Os resultados mostraram que o mapa de dificuldade pode ser usado como (i) uma nova medida para avaliação de desempenho de um algoritmo de *change detection* e (ii) uma medida do potencial de avaliação de vídeos de um *dataset*.

Recebido: 19 ago. 2020.

Aprovado: 01 out. 2020.

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



PALAVRAS-CHAVE: Detecção de mudanças. Avaliação de desempenho. Mapa de dificuldade.

ABSTRACT

The performance evaluation of a new algorithm should show the superiority of this algorithm to the existing state-of-the-art. In order to *evaluate* a change detection algorithm, it is required to execute it aiming to identify changing areas (segmentation) in dataset video frames, and to compare the results with a ground truth. It is possible to compare performances when various algorithms segment videos of the same dataset. Despite of the traditional way of evaluation performance of change detection algorithms being accepted by the scientific community, exist another relevant informations from the result of the segmentation. The difficulty level to classify a pixel is an example of such information. In this paper, a method to identify the difficulty in classifying pixels by using change detection algorithms is presented. The method includes the use of segmentation results from various state-of-the-art algorithms, as well as storing such values that represent difficulty levels in matrices named “difficulty maps”. The results showed that difficulty map can be used (i) as

a new measure to performance evaluation of a change detection algorithm and (ii) as a measure of the potential for evaluating videos of a dataset.

KEYWORDS: Change Detection. Performance Evaluation. Difficulty Map.

1 INTRODUÇÃO

Vigilância por vídeo, ambientes inteligentes e recuperação de conteúdo são exemplos de sistemas que utilizam algoritmos de detecção de mudança *change detection* (SANCHES, 2019). Tais algoritmos identificam regiões (conjunto de *pixels*) que sofrem modificações ou se movem em relação a uma imagem que representa o modelo do plano de fundo da cena (GOYETTE, 2012). A detecção de mudança é pré-requisito para muitas aplicações de visão computacional e processamento de vídeo (GOYETTE, 2012).

A avaliação de desempenho de um algoritmo de detecção de mudanças é essencial para que se demonstre a superioridade um novo algoritmo em comparação ao encontrados na literatura. Ela consiste na execução do algoritmo para segmentar um conjunto de vídeos, *datasets*, e na comparação dos resultados com um *ground truth*, conjunto de quadros rotulados manualmente que permite identificar o resultado ideal para a segmentação. Com base no resultado da comparação são calculadas métricas que representam o desempenho do algoritmo.

A forma tradicional de avaliação de algoritmos de *change detection*, que consiste na comparação entre desempenhos de algoritmos tomando como base as mesmas métricas e *datasets*, é bem aceita pela comunidade acadêmica. Contudo, por meio dos resultados da segmentação de vídeos podemos obter outras informações relevantes. O nível de dificuldade para classificar um determinado pixel, por exemplo, pode ser utilizado para identificar algoritmos promissores, capazes de classificar *pixels* difíceis – ponto em que a maioria dos algoritmos falha (SANCHES, 2019).. Além do mais, o nível de dificuldade de cada pixel pode ser importante para identificar os vídeos de um *dataset* que possuem potencial para avaliar algoritmos. *Datasets* com a maioria dos *pixels* de fácil classificação podem não diferenciar algoritmos muito eficientes de algoritmos pouco eficientes (SANCHES, 2019)..

Inserido neste contexto, este trabalho apresenta um método para identificar nos quadros dos vídeos de um *dataset* os *pixels* que são difíceis de serem classificados. O nível de dificuldade de classificar cada pixel é armazenado em uma estrutura denominada mapa de dificuldade. Desse modo, obteve-se um mapa de dificuldade para cada quadro de cada vídeo do *dataset* analisado. Os mapas de dificuldade podem ser utilizados para (i) avaliar novos algoritmos para identificar soluções promissoras e (ii) identificar os vídeos de um *dataset* que possuem maior potencial para avaliar algoritmos de detecção de mudança.

2 MATERIAIS E MÉTODOS

A abordagem proposta tem como objetivo identificar o nível de dificuldade para classificar cada pixel de cada quadro de vídeo de um *dataset*. Essas informações são armazenadas em um mapa de dificuldade e utiliza-se desta estrutura para avaliar novos algoritmos de detecção de mudanças e *datasets*.

Inicialmente, foi necessário o estabelecimento de um *dataset* de utilização unificada no estudo. Nos últimos anos, o *ChangeDetection.net* tem sido o *dataset* mais utilizado pela comunidade científica para avaliar algoritmos de detecção de mudanças. Ele possui duas versões: o *CDNet 2012*, com 31 vídeos e o *CDNet 2014*,

com 53 vídeos, ambos possuem quadros do *growth truth* rotulados e seus vídeos são distribuídos em 11 diferentes categorias. O *CDNet 2014* é o acervo utilizado no presente trabalho

As métricas empregadas como base para a avaliação de um novo algoritmo de detecção de mudanças e *datasets*, foram os valores:

- verdadeiros positivos (*TP*) – *pixels* classificados corretamente como pertencentes ao elemento de interesse;
- falsos positivos (*FP*) – *pixels* classificados como pertencentes ao elemento de interesse incorretamente;
- falsos negativos (*FN*) – *pixels* corretamente classificados como pertencentes ao plano de fundo;
- verdadeiros negativos (*TN*) – *pixels* incorretamente classificados como pertencentes ao plano de fundo.

Dada a norma da matriz $\|A\| = \sum_{i=1}^m \sum_{j=1}^n |a_{ij}|$ são definidas pelas Eq. (1), (2), (3) e (4).

$$TP = \|G \odot S\| \quad (1)$$

$$FP = \|(1 - G) \odot S\| \quad (2)$$

$$FN = \|G \odot (1 - S)\| \quad (3)$$

$$TN = \|(1 - G) \odot (1 - S)\| \quad (4)$$

Considerando valores normalizados, os algoritmos de detecção de mudanças geram uma máscara $S \in \{0,1\}^{n \times m}$ como resultado e o *ground truth* possui rótulos $G \in [0,1]^{n \times m}$, em ambos 0 é o rótulo de *pixels* do fundo, 1 é o rótulo de *pixels* do elemento de interesse e $n \times m$ é o tamanho do quadro. Além disso, existem rótulos para regiões como sombras, localizadas fora da área de interesse e desconhecidas (UNIVERSITÉ DE SHERBROOKE, 2019).

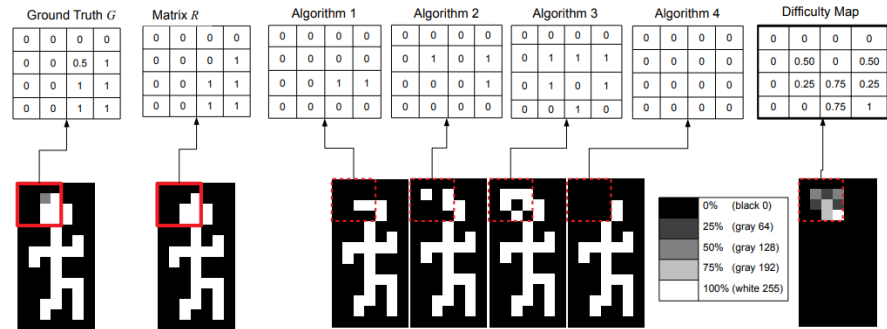
Logo, utilizando as métricas base calcula-se a Taxa de Falso positivo (*FPR*), Taxa de Falso Negativo (*FNR*), Precisão (*Pr*), Recall (*Re*), Especificidade (*Sp*) e Porcentagem de Classificações Erradas (*PWC*) e a *Fscore* (JODOIN, 2020).

Um mapa de dificuldade é gerado utilizando os resultados das segmentações de diversos algoritmos. Por meio das máscaras geradas como resultados da segmentação dos algoritmos, é possível identificar o nível de dificuldade de um *pixel*, utilizando da quantidade algoritmos que erraram sua classificação. Dessa forma, para cada quadro de um vídeo é gerado um mapa de dificuldade correspondente, responsável por armazenar essas informações.

Consequentemente, o número de algoritmos n utilizado determina a quantidade de níveis de dificuldade ($n + 1$) contidos no mapa. Cada nível é representado no mapa por uma cor da escala cinza. Por exemplo, a Figura 1 apresenta um quadro de um *growth truth* G e quatro máscaras S . Portanto, o mapa de dificuldade utilizando os resultados dos quatro algoritmos é constituído por cinco níveis de dificuldade, sendo, nível 0 (sem dificuldade) e 1 (alta dificuldade). Ademais, a dificuldade atribuída a cada *pixel* é determinada pela porcentagem de

algoritmos que erram sua classificação. Por exemplo, na Figura 1 o *pixel* rotulado com nível 0.25 (25% e cor cinza 64) foi ponto de falha de um algoritmo.

Figura 1 – Um mapa de dificuldade com $n = 4$.



Fonte: O autor

Continuamente, um mapa de dificuldade pode ser utilizado como suporte para uma nova medida de avaliação do desempenho de um algoritmo de detecção de mudanças. Utilizando *grouth truth* G e as máscaras S é possível identificar se os erros da segmentação são FP , TP , FN ou TN . Com os níveis de dificuldades armazenados em um mapa de dificuldade D obtém-se FP_D , TP_D , FN_D e TN_D de acordo com as Eq. (5), (6), (7) e (8) (JODOIN, 2020).

$$TP_D = \|G \odot S \odot D \odot R\| \quad (5)$$

$$TN_D = \|(1 - G) \odot (1 - S) \odot D \odot R\| \quad (6)$$

$$FP_D = \|(1 - G) \odot S \odot D \odot R\| \quad (7)$$

$$FN_D = \|G \odot (1 - S) \odot D \odot R\| \quad (8)$$

Onde a matriz $R \in \{0,1\}^{n \times m}$ é responsável por armazenar a região de interesse do quadro, desconsiderando, sombras ou regiões indeterminadas.

Em sequência, realiza-se o cálculo das métricas FPR_D , FNR_D , Pr_D , Re_D , Sp_D , PWC_D e $Fscore_D$. Em suma, o conjunto de todas essas métricas determina o desempenho de um algoritmo de detecção de mudanças considerando apenas as regiões difíceis de classificar.

Um *dataset* pode ter seu potencial para avaliar um algoritmo determinado com o auxílio do mapa de dificuldade. Esta pesquisa propõe a utilização de uma medida chamada dificuldade média D_M de um vídeo, que define seu potencial para avaliar algoritmos, a qual também pode ser calculada considerando conjuntos de vídeos. Enfim, tomando um conjunto de quadros k a dificuldade média pode ser obtida segundo a Eq. (9).

$$D_M(k) = \sum_{j=start}^{end} \sum_{i=1}^{Nvp} \frac{d(i,j,D_k)}{Nvp_j} \quad (9)$$

Dado o número de *pixels* válidos Nvp para o j^{th} quadro de um *grouth truth* G normalizado. E onde *start* é o quadro inicial, *end* o quadro final, D_K é o mapa de dificuldade da sequência de quadros k e $d(i,j,D_K)$ é o nível de dificuldade armazenado considerando o *pixel* i do quadro j no mapa de dificuldade D_K .

3 RESULTADOS E DISCUSSÕES

Em um repositório como o do *CDNet 2014*, por exemplo, é possível encontrar (i) vídeos, (ii) *ground truths*, (iii) resultados da segmentação de vários algoritmos na forma de máscaras S e (iv) um ranking que mostra o desempenho de vários algoritmos de detecção de mudanças que utilizam o *CDNet 2014* para avaliação de seu desempenho. Neste trabalho utilizou-se os resultados de 30 algoritmos, na forma de máscaras S , obtidos no repositório *CDNet 2014*, escolhidos de forma aleatória, e pertencentes ao ranking dos 45 algoritmos de melhor desempenho.

3.1 AVALIAÇÃO DE ALGORITMOS

A primeira etapa de validação da abordagem proposta se destinou a avaliar o desempenho de algoritmos utilizando mapas de dificuldade. Sendo assim, 9 algoritmos, listados no ranking do *CDNet 2014* e não utilizados na construção do mapa de dificuldade, tiveram seu desempenho avaliado.

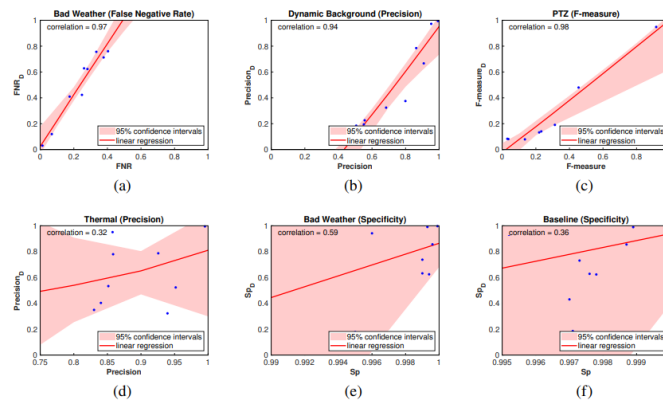
O desempenho de cada um dos algoritmos foi calculado pelas métricas TP_D , TN_D , FP_D , FN_D , FPR_D , FNR_D , Pr_D , Re_D , Sp_D , PWC_D e $Fscore_D$. Em uma primeira análise, realizou-se uma comparação entre o desempenho da avaliação utilizando o mapa de dificuldade e a avaliação tradicional, destacaram-se dois grupos de algoritmos em que (i) o desempenho obtido pela nova abordagem e pela abordagem tradicional (*ground truth*) são diferentes e (ii) o desempenho obtido pela nova abordagem e pela abordagem tradicional são similares.

Por exemplo, considerando apenas os vídeos da categoria *Bad Weather* o algoritmo *CL_VID* apresenta $Sp = 0,9950$ e $Sp_D = 0,1788$ e o algoritmo *M4CDVersion2_0* apresenta $PWC = 1,4478$ e $PWC_D = 40,1373$ considerando a categoria *CameraJitter*. Estes exemplos de comparação de desempenho mostram que alguns algoritmos têm bom desempenho quando avaliados segundo os mapas de dificuldade, ou seja, suas melhorias são mais fáceis, uma vez que, devem superar desafios já superados por outros algoritmos. Entretanto, algoritmos que obtiveram um desempenho baixo precisam enfrentar desafios ainda não superados pelo estado-da-arte, ou seja, sua melhoria é mais difícil.

Enfim, em uma segunda análise, agrupou-se os resultados das 7 métricas (calculadas dos 9 algoritmos) de acordo com as categorias apresentadas no *CDNet 2014* (11 categorias), com o objetivo de mostrar que nem todas as soluções compartilham as mesmas dificuldades de classificação de *pixel*. Isso ocorre quando os valores das avaliações tradicional e utilizando o mapa de dificuldade apresenta resultados significativamente diferentes. Exemplos da correlação entre os resultados obtidos da avaliação de desempenho utilizando o *ground truth* e o mapa de dificuldade podem ser visualizados na Figura 2.

Em síntese, nesta primeira etapa os resultados das análises apresentadas indicam que a abordagem baseada em mapas de dificuldade pode ser utilizada como uma nova medida de desempenho de algoritmos de detecção de mudanças.

Figura 2 – Correlação entre os resultados obtidos pela avaliação tradicional e com o mapa de dificuldade



Fonte: O autor

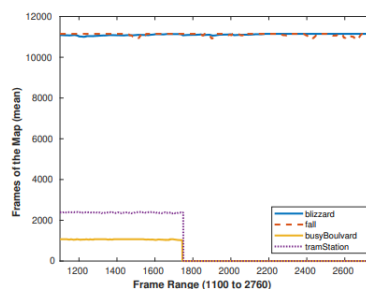
3.2 AVALIAÇÃO DE DATASETS

A segunda etapa da validação da abordagem proposta utilizou os 30 mapas de dificuldades gerados para avaliar o potencial de vídeos, contidos no acervo CDNet 2014, para avaliar algoritmos de detecção de mudanças.

Realizando uma varredura quanto ao percentual da frequência de ocorrência de cada nível de dificuldade (exceto o nível 0) dos *pixels* válidos do CDNet 2014 detectou-se que a soma dos *pixels* válidos que apresentam níveis de dificuldade no intervalo de 1 até 30 representam 10,52% do total de *pixels* válidos. Além do mais, os *pixels* rotulados com níveis de dificuldade mais alto representam menos de 1%. Em outras palavras, a maior parte dos vídeos do *dataset* não representam um desafio para a maioria dos algoritmos de detecção de mudanças.

O cálculo da dificuldade média (D_M) e da porcentagem de *pixels* com alto nível de dificuldade foi realizado para todos os vídeos do acervo. Observando a Figura 3 encontramos os valores D_M , calculados quadro a quadro, no intervalo 1161 até 2760. Foram utilizados 4 vídeos dos CDNet 2014 nesta análise. Os vídeos *blizzard* e *fall* possuem altas D_M e os vídeos *busyBoulevard* e *tramStation* possuem muitos *pixels* com o nível de dificuldade mais alto.

Figura 3 – D_M obtida quadro a quadro dos vídeos *blizzard*, *fall*, *busyBoulevard* e *tramStation*



Fonte: O autor

Pode-se perceber que, vídeos que possuem a maior porcentagem de *pixels* com níveis altos de dificuldade não necessariamente apresentam maior dificuldade de classificação de *pixels*. Um vídeo mesmo não contendo *pixels* com o maior nível de dificuldade, pode conter muitos *pixels* com outros níveis altos de dificuldade e de forma contínua durante todo o vídeo.

Por fim, como últimas conclusões: os algoritmos que possuem D_M baixas possuem *pixels* que foram classificados corretamente pelos algoritmos que geraram o mapa ou a maioria dos quadros possui muitos *pixels* não válidos. Outrossim, a D_M de uma categoria de vídeo é totalmente afetada pela dificuldade média dos vídeos que a constitui, ou seja, uma categoria com diversos vídeos de dificuldade média alta terá uma D_M da alta.

4 CONCLUSÃO

Na avaliação do desempenho de um algoritmo de detecção de mudança deve-se mostrar claramente que o novo algoritmo apresentado é superior aos encontrados na literatura. Apesar da forma tradicional de avaliar o desempenho de algoritmos ser bem aceita pela comunidade científica, novas medidas podem ser utilizadas para comparar algoritmos.

Uma vez que os algoritmos de detecção de mudança se baseiam em diferentes abordagens, uma solução pode ser capaz de classificar corretamente *pixels* difíceis de serem classificados, mas pode falhar em regiões que são classificados corretamente pelos algoritmos do estado-da-arte. Tais algoritmos podem ser considerados promissores, na medida em que conseguem superar desafios já resolvidos por outros algoritmos da literatura. Este trabalho apresentou um método de avaliação de algoritmos de detecção de mudança, baseado em mapas de dificuldade, capaz de identificar algoritmos promissores.

Além de auxiliar a avaliação de algoritmos, os mapas de dificuldade gerados também se mostraram úteis na geração de uma medida capaz de identificar vídeos de um *dataset* que possuem potencial para avaliar algoritmos. Esse potencial é uma característica desejável aos *datasets* uma vez que vídeos que possuem muitos quadros em que a maioria dos *pixels* são de fáceis de classificar podem não diferenciar algoritmos muito eficientes de algoritmos pouco eficientes. Os resultados das análises realizadas mostraram que nossa nova abordagem para avaliação de algoritmos e *datasets* se mostrou promissora quando aplicada utilizando os vídeos do *dataset* CDNet 2014.

AGRADECIMENTOS

Os autores agradecem à Universidade Tecnológica Federal do Paraná, pela participação no Programa de Voluntariado em Iniciação Científica e Tecnológica (Edital PROPPG 02/2019— PVICT) da acadêmica Katharina Akemi Ikeda Rosa.

REFERÊNCIAS

SANCHES, S. R. R. et al. **Challenging situations for background subtraction algorithms.** Applied Intelligence, Massachusetts, Estados Unidos, v. 49, n. 5, p. 1771–1784, maio 2019.

GOYETTE, N.; JODOIN, P.-M.; PORIKLI, F; KONRAD, J.; ISHWAR, P.
Changedetection.net: A new change detection benchmark dataset. In: 2012 IEEE COMPUTER SOCIETY CONFERENCE ON COMPUTER VISION AND PATTERN COGNITION WORKSHOPS, 10, 2012, Providence, RI, Estados Unidos. Providence, RI, Estados Unidos: IEEE, 2012, p. 1–8.

JODOIN, P.-M.; GOYETTE, N.; PORIKLI, F; KONRAD, J.; ISHWAR, P.
ChangeDetection.NET: A video database for testing change detection.
Disponível em: <http://www.changedetection.net>. Acesso em: 22 jul. 2020.