

Ferramenta de Extração de Fundo de Vídeos

Henrique Schmitz^a

^aEscola Politécnica, Universidade do Vale do Rio dos Sinos - UNISINOS, São Leopoldo, Brasil

ARTICLE INFO

Keywords:
extração de fundo
vídeo estático
opencv
pytorch

ABSTRACT

Neste artigo, é detalhado o desenvolvimento e teste de uma ferramenta de extração de fundo para vídeo estático. O processo faz uso da mediana aplicada sobre quadros aleatórios do vídeo para determinar o fundo deste. O sistema foi desenvolvido utilizando a biblioteca *pytorch* e tensores para possibilitar otimização e aceleração por GPU. Esta ferramenta foi aplicada a um vídeo de competição de robótica, obtendo velocidades de 66 FPS sem aceleração e 808 FPS com aceleração por GPU para apenas o processo de remoção de fundo.

1. Introdução

No campo de detecção de objetos em vídeos, atualmente existem diversas ferramentas com diferentes níveis de complexidade, qualidades de resultados e tempos de execução. Enquanto existem modelos capazes de detectar objetos complexos em imagens com diversas formas, estes ainda podem apresentar confusão com áreas do ambiente contendo formas similares. Além disto, para obter os melhores resultados possíveis, modelos com execução mais lenta são necessários.

Uma característica presente em múltiplos exemplos de vídeo para processamento de imagem é o uso de câmeras paradas. Nestas, o ambiente observado se mantém constante enquanto algumas pessoas e objetos que não fazem parte do ambiente se movem por este. Isto faz com que, mesmo que objetos móveis estejam presentes em todos os quadros do vídeo, seja altamente provável para um vídeo que todas as partes do ambiente de fundo sejam visíveis em parte dos quadros. Com conhecimento de todo o ambiente de fundo, é possível reconstruir este com base em quadros do vídeo.

Para simplificar o processo de detecção de objetos em primeiro plano, uma maneira de diferenciação entre áreas de fundo e objetos móveis é desejável. Este trabalho propõe a criação de um sistema de obtenção de fundo para vídeos de câmera estática e uso de imagens de fundo para extrair apenas as zonas diferentes do fundo para quadros de vídeos. Idealmente este processo deve ser feito de forma veloz para que este possa ser aplicado em vídeos.

2. Materiais e métodos

A ferramenta proposta deve obter um vídeo com apenas os objetos em primeiro plano. Para isto, os seguintes passos são necessários: obtenção do plano de fundo, cálculo de pixels diferentes ao plano de fundo e remoção dos pixels iguais a este.

Em vídeos com câmera estática é esperado que o fundo seja visível em grande parte do tempo. Os objetos móveis bloqueiam a visão da área de fundo apenas transitoriamente. Desta forma, na maioria dos casos, para cada parte da imagem, o fundo estar bloqueado pode ser considerado um

valor extremo. Assim, o fundo pode ser obtido aplicando a mediana [2], uma medida de centralidade resistente a valores extremos, sobre cada pixel utilizando várias imagens de diferentes momentos do vídeo. Esta técnica pode considerar objetos que ficam parados na visão da câmera por tempos muito longos como parte do fundo, mas deve trazer benefícios na maior parte das situações. Desta forma, a extração de fundo segue as seguintes equações:

$$r_{fundo}(x, y) = med(r(x, y))_0^n$$

$$g_{fundo}(x, y) = med(g(x, y))_0^n$$

$$b_{fundo}(x, y) = med(b(x, y))_0^n$$

Com posse da imagem do plano de fundo a próxima operação é a criação de uma imagem de diferença. Esta imagem é a representação da diferença, calculada pelo valor absoluto da subtração, entre determinado quadro de vídeo e o fundo estático para cada pixel. Esta imagem é uma representação em escala de cinza da intensidade com a qual cada região difere do fundo. A equação para cálculo de cada pixel desta se dá por:

$$d_{pixel} = \frac{(|r_{quadro} - r_{fundo}| + |g_{quadro} - g_{fundo}| + |b_{quadro} - b_{fundo}|)}{3}$$

Cada quadro pode então ser avaliado em sua diferença total em relação ao fundo. Este valor é calculado como a média de valores de todos pixels de diferença. Esta informação é útil para determinar se a imagem do ambiente estático continua presente ou a visão de câmera foi trocada ou obstruída. Sabendo isto, o índice de diferença é utilizado para definir se o processo de extração de fundo será aplicado para uma imagem. Este é definido como:

$$d_{quadro} = \frac{d_1 + d_2 + d_3 + \dots + d_n}{n}$$

Com a decisão de se um quadro é passível a remoção de fundo, o último passo é a substituição de todos os pixels de fundo por pixels pretos. Esta operação requer o uso da imagem de diferença como uma máscara sobre a imagem original. A operação de máscara inicialmente necessita da definição de quais pixels serão considerados e quais serão removidos. Esta definição é feita aplicando um *threshold*, ou limite, sobre a imagem de diferença. O limite escolhido é o valor de d_{quadro} , desta forma, apenas os pixels com maior diferença em relação ao fundo serão válidos. O resultado da operação de *threshold* deve resultar em uma imagem com apenas

valores 0 ou 1. Para aplicar esta máscara sobre a imagem original, a imagem de máscara é multiplicada pela original, zerando os pixels de fundo e mantendo o mesmo valor para os pixels de primeiro plano. Matematicamente, esta operação é representada por:

$$\text{novaPixel} = \begin{cases} r = r_{\text{pixel}}, g = g_{\text{pixel}}, b = b_{\text{pixel}}, & \text{se } d_{\text{pixel}} \geq d_{\text{quadro}} \\ r, g, b = 0, & \text{caso contrário} \end{cases}$$

$$\text{novaPixel} = \begin{cases} r = r_{\text{pixel}}, g = g_{\text{pixel}}, b = b_{\text{pixel}}, & \text{se } d_{\text{pixel}} \geq d_{\text{quadro}} \\ r, g, b = 0, & \text{caso contrário} \end{cases}$$

Enquanto o cálculo de fundo só precisa ser feito uma vez por vídeo, a extração de fundo é executada para cada pixel de cada quadro do vídeo. O número de operações necessárias para executar esta transformação é excessivo para a aplicação de processamento linear utilizando uma CPU, mas possui o caráter de operações matemáticas simples e paralelizáveis que possibilita aceleração por GPU. Para fazer proveito destas características e manter a velocidade de execução alta, a extração de fundos foi implementada utilizando as capacidades de aceleração por GPU da biblioteca *pytorch* e do uso de tensores.

3. Resultados e discussão

A ferramenta de extração de fundo foi desenvolvida como um módulo separado para reuso e aplicada sobre um vídeo de partida de competição de robótica [1] com câmera estática. Esta foi aplicada para um quadro aleatório, apresentado na figura 1, para a visualização do funcionamento do processo de extração de fundo.

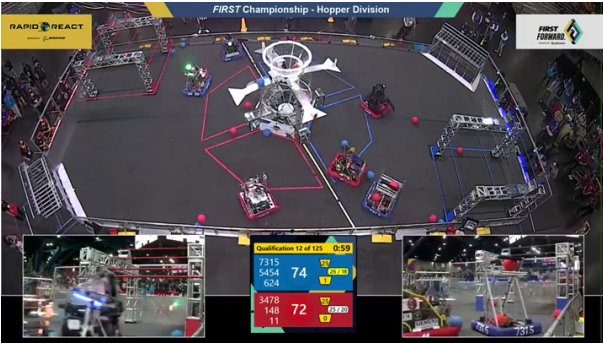


Figure 1: Quadro de vídeo utilizado para teste de extração de fundo.

Os testes de tempo de execução foram feitos tanto sem o uso de aceleração por GPU, utilizando apenas os benefícios de performance do uso de tensores, e com aceleração. A GPU utilizada para os testes foi a Nvidia Tesla T4. Estes testes foram executados em um ambiente do Google Colab.

O primeiro passo aplicado foi a definição da imagem de fundo por meio de mediana. Este é o processo mais custoso, devido tanto à necessidade de percorrer o vídeo para encontrar quadros aleatórios espaçados uniformemente pelo tempo deste quanto pela quantidade de quadros maior que a das outras operações.

Este processo foi aplicado no vídeo de exemplo, que possui 3min 11s de duração, e utilizando o valor padrão de 25 quadros para a operação. O tempo de execução foi de 5,23s sem aceleração por GPU e 4,89ms utilizando aceleração. O tempo necessário para transformar a imagem de fundo de um tensor para uma imagem comum depois desta já estar calculada foi de 1,83ms sem aceleração

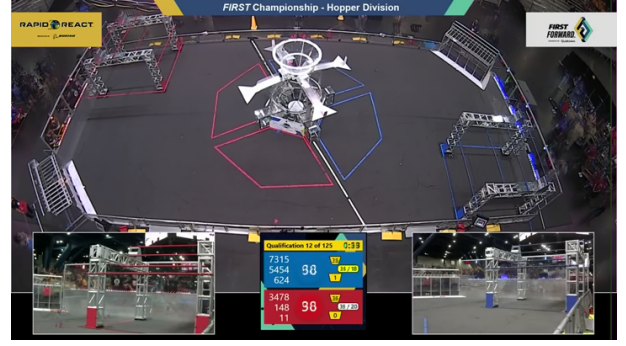


Figure 2: Fundo calculado para o vídeo de teste.

e 1,5ms com auxílio da GPU. O quadro de fundo resultante pode ser visto na figura 2.

A operação de cálculo de imagem de diferença foi inicialmente testada de forma separada. Este teste foi feito para possibilitar a visualização desta e para analisar a velocidade de execução do processo. O cálculo da imagem de diferença foi feito em 8,78ms com uso de CPU e 2,55ms com uso de GPU.

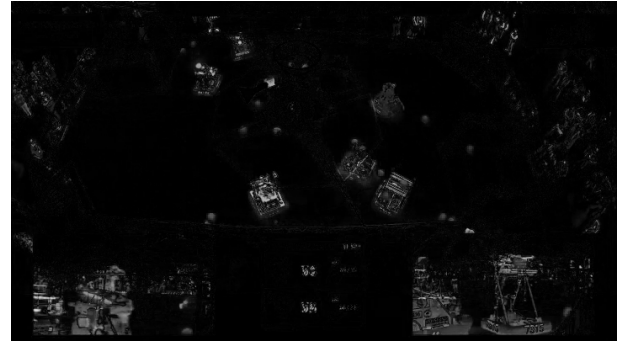


Figure 3: Representação de diferença pixel a pixel com o fundo calculada para a imagem testada.

O cálculo do índice de diferença para o quadro completo também foi executado de maneira separada. Esta execução consiste no cálculo da imagem de diferença juntamente da agregação desta em um índice. Este levou 11,9ms sem GPU e 2,83ms com esta. O seguinte valor foi obtido:

$$d_{\text{quadro}} = 0,03019443365905218$$

A execução do processo completo de extração de fundo resultou na figura 4. Este processo incluiu o cálculo da imagem de diferença, o cálculo do índice de diferença e a remoção de fundo com base nestes. Os tempos de execução foram 19ms e 3,77ms, sem e com uso de GPU respectivamente.

Na tabela 1, os tempos de execução tanto com aceleração por GPU quanto sem o uso desta podem ser observados de maneira agregada.

Por fim, o processo de extração de fundo foi executado para 1000 quadros do vídeo de teste. Este teste teve objetivo de obter uma imagem mais completa do tempo de execução em casos de uso real. Os tempos de execução foram calculados desconsiderando tempos de leitura de imagem e tempo necessário para o cálculo inicial do fundo. Estes consistem apenas na velocidade do processo de extração de fundo em si. Com uso de CPU para processamento, a

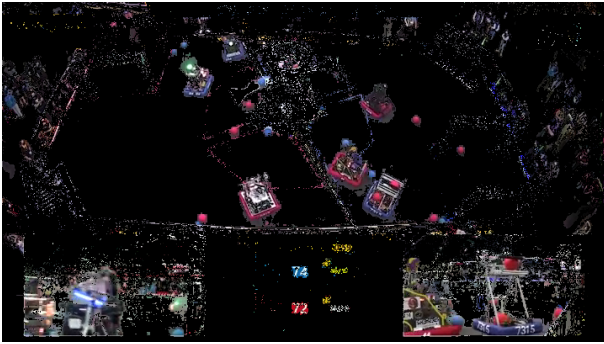


Figure 4: Imagem testada com fundo removido.

Table 1

Esta é uma tabela de exemplo.

Operação	Tempo CPU	Tempo GPU
Cálculo de Fundo	5,23s	4,89ms
Uso de Fundo	1,83ms	1,5ms
Imagem de Diferença	8,78ms	2,55ms
Índice de Diferença	11,9ms	2,83ms
Remoção de Fundo	19ms	3,77ms

velocidade de execução foi de 66,1 FPS (15,1ms), enquanto com o uso de aceleração por GPU a execução deste se deu com velocidade de 808,8 FPS (1,2ms).

4. Conclusão

O desenvolvimento desta ferramenta demonstra a funcionalidade do processo de reconstrução de imagem de fundo a partir de vídeo de câmera estática. Este processo requer que o fundo da imagem esteja visível em um percentual considerável de quadros, mas mesmo assim se apresentou robusto para a aplicação na qual este foi executado.

A ferramenta de extração de fundos tem os seguintes subprodutos: imagem de fundo, imagem de diferença do quadro ao fundo, índice de diferença do quadro ao fundo e quadro com fundo extraído. O índice de diferença pode ser utilizado como métrica para indicar mudança de câmera ou obstrução desta. A imagem final sem fundo pode ser utilizada para simplificar o processo de detecção de objetos em ambientes com fundo complexo e que possa causar confusão.

O uso da biblioteca *pytorch* e a aplicação de tensores possibilitaram grandes ganhos em velocidade de execução, tendo benefícios mesmo sem o uso de aceleração por GPU. As velocidades de extração de fundo foram de 66 FPS utilizando CPU e 808 FPS utilizando GPU.

O módulo de extração final foi disponibilizado no github em <https://github.com/HenriqueSchmitz/VideoBackgroundExtractor>, juntamente com exemplos de uso. Uma versão pública do documento do Google Colab utilizado para os testes de velocidade de execução foi disponibilizada juntamente no repositório da ferramenta.

References

- [1] FIRSTRoboticsCompetition. Qualification 12 - 2022 first championship - hopper division. Disponível em https://www.youtube.com/watch?v=mAaBmH_vgY4, 2022. Acesso em: 2022-11-23.

- [2] LearnOpenCV. Simple background estimation in videos using opencv (c++/python). Disponível em <https://learnopencv.com/simple-background-estimation-in-videos-using-opencv-c-python/>, 2022. Acesso em: 2022-11-23.