Projeto de Formatura – Turmas 2018

PCS - Departamento de Engenharia de Computação e Sistemas Digitais

Engenharia de Computação

Tema: ColorMotion: colorização automática de vídeos

Proposta

Ainda não existe solução automática satisfatória para coloração de vídeos em escala de cinza. No ramo profissional, ferramentas de coloração requerem alto nível de experiência e muitas entradas manuais do operador. O custo de realização de restaurações fotorealísticas por um estúdio profissional passa de US$1000 por minuto.

O trabalho proposto realiza coloração automática ou semi-guiada de maneira satisfatória, para que um profissional seja necessário apenas para retoques finais.

Arquitetura

O trabalho utiliza redes neurais profundas (DNNs) convolucionais, em arquitetura *encoder-decoder*. Entradas e saídas usam espaço de cor *L*a*b**, escolhido pois o canal *L** reflete fielmente a luminância observada pelo olho humano, e mudanças nos canais de cor preservam a luminância observável do vídeo original.

A arquitetura suporta, opcionalmente, coloração semi-guiada por entrada de valores de *a*b**. O usuário pode intervir na escolha de cor, que é multimodal (mais de uma cor é plausível para alguns objetos, como roupas).

O principal objetivo da arquitetura é manter a coerência entre frames: redes sem estado gerariam cintilação (*flickering*) observável entre frames, pois mudanças pequenas na entrada geram oscilações na cor inferida. Nosso modelo mantém estado das *features* encodadas. Por meio de fluxo ótico denso, *features* são propagadas para próximos frames. Uma camada classificadora tenta prever se um pixel pertence a uma região sendo oculta/exposta, ou região que pode ser mapeada às cores do quadro anterior. A classificação gera uma máscara indicando quais *pixels* devem ser reaproveitados do último quadro e quais devem ser gerados novamente.

O uso de fluxo ótico denso pelo método de Lucas-Kanade inviabilizou que o modelo fosse treinável fim-a-fim, já que a função não é diferenciável. Realizamos parte do algoritmo em CPU, separado da DNN, com dificuldades na integração com a DNN, mas o resultado final foi muito satisfatório.

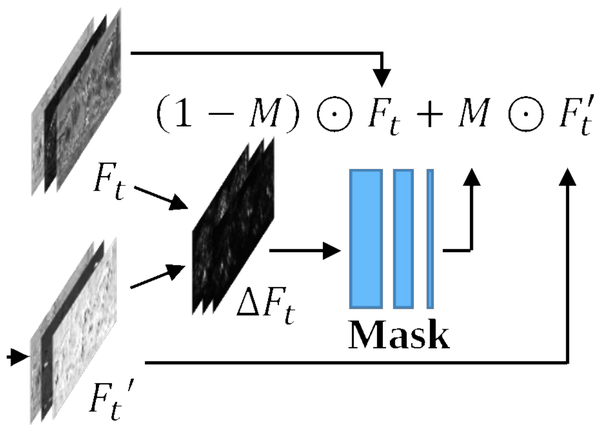

Figura 1: Máscara seletora entre novas cores (*Ft*) e cores propagadas por fluxo ótico (*Ft'*)

Integrantes: Henrique Cassiano Souza Barros
Tiago Koji Castro Shibata
Victoria Akina Tanaka

Professor Orientador: Bruno de Carvalho Albertini

Dataset

Criamos um dataset baseado em vídeos *open-source*, com 591780 quadros divididos em 1629 cenas distintas.

O dataset se mostrou pouco variado e propenso a overfitting, já que quadros da mesma cena são bastante similares. Decidimos utilizar parte do dataset *ImageNet* e realizar *data augmentation* para gerar artificialmente o fluxo que seria presente em quadros de vídeos. Usamos 895881 imagens da *ImageNet*.

Resultados

Figura 2: comparação dos resultados obtidos com o *ground-truth*.

O modelo final obteve 32ms/quadro em uma placa de nível de consumidor (NVIDIA GTX1080) com batches unitários, possuia 34051138 parâmetros treináveis, e após serializado ocupava 390MB. Após otimização (remoção de camadas e pesos pouco influentes pelo método Average Percentage of Zeros - APoZ) e uso de batches de 8 frames, obtivemos o desempenho de 28ms/quadro, com 28879168 parâmetros treináveis, e 111MB. O modelo é aplicável em fluxos de vídeo em tempo real.

Testes de usuário foram realizados para testar se o resultado é convincente. Trechos de 10 segundos foram extraídos de vídeos novos, nunca vistos pela rede, e foram convertidos para escala de cinza e coloridos artificialmente, em resolução 256x256 e 0.016% dos valores *a*b** originais dados como *guidance*. Voluntários assistiram aleatoriamente à versão original ou colorida pela rede e decidiram se a coloração parecia real ou gerada por computador:

	Escolhida		
Real	**Não**	**Sim**	
Não	40.4	7.9	48.3
Sim	6.7	44.9	51.7
	47.2	52.8	

Figura 3: resultados obtidos para o teste de reconhecimento da versão original *versus* a versão colorida pela rede.

Observamos que nos vídeos coloridos por computador, 7.9 / 40.4 = 19.6% dos usuários acreditaram que se tratasse de uma coloração real. Em outro teste, mostramos a novos usuários as versões original e colorida por computador lado a lado, simultaneamente, e pedimos para que a original fosse apontada. Em 12.9% dos testes o usuário foi enganado.