https://eventos.utfpr.edu.br//sicite/sicite2019

Base de dados para reconhecimento de ações humanas e interface web para extração de características

Human action recognition dataset and web interface for feature extraction

Wagner Rodrigues Ulian Agostinho
wagner.rua@hotmail.com
Universidade Tecnológica Federal do Paraná - UTFPR, Curitiba, Paraná, Brasil

Heitor Silvério Lopes
hslopes@utfpr.edu.br
Universidade Tecnológica Federal do Paraná - UTFPR, Curitiba, Paraná, Brasil

RESUMO

Com o aumento de imagens sendo geradas por câmeras em vários lugares no mundo e consequentemente o crescimento no uso de aprendizagem de máquina e aprendizado profundo no processamento de imagens a falta de dados bem documentados e organizados para o uso dessas tecnologias tem se tornado um problema para os pesquisadores dessa área, além disso o processamento de grandes quantidades de informação pode ser dificultado se feito em máquinas que não utilizam placas de vídeo e tem baixo poder de processamento. O trabalho intitulado "Base de dados para reconhecimento de ações humanas e interface web para extração de características" teve duas frentes de desenvolvimento da pesquisa. Na primeira, tem-se a criação de uma base de dados documentada, organizada e de fácil acesso aos pesquisadores da universidade com propósito de ser usado em aprendizado profundo. Na segunda tem-se a criação do *front-end* e do *back-end* de uma aplicação web que facilita a extração de características e processamento de base de dados em unidades de processamento gráfico(GPU).

PALAVRAS-CHAVE: . Aprendizagem profunda, Base de dados, Aplicação Web.

Recebido: 19 ago. 2019.
Aprovado: 01 out. 2019.

ABSTRACT

With the increase of images being generated by cameras in places around the world and consequently the growth in the use of machine learning and deep learning in image processing the lack of well documented and organized data for the use of these technologies has become a problem for researchers in this area, moreover, can process large amounts of information if made on machines that do not use video cards and have low processing power. The work entitled "Database for human action recognition and web interface for feature extraction" had two fronts of research development. In the first, there is the creation of a database, documented, organized and easily accessible to university researchers with the purpose of being used in deep learning. In the second there is the creation of the front end and the back end of a web application that facilitates feature extraction and database processing in graphics processing units (GPU).

KEYWORDS: Deep Learning, Datasets, Web Applications.

INTRODUÇÃO

O ramo da Visão computacional (VC) na inteligência artificial e na ciência da computação tem como ênfase dar ao computador, através de imagens digitais, o entendimento visual do mundo. A combinação de diversas áreas como aprendizado de máquina e aprendizado profundo geram a VC que conduz o caminho para compreensão semântica da imagem (SCHALKOFF, 1989). O atual estado-da-arte já conquistou espaço em diversas tarefas que antes só eram realizadas por humanos, como: classificação de imagens, condução de veículos, controle de robôs, automatização de indústrias, entre outros (SHIH, 2017).

Apesar do constante avanço e grande desempenho demonstrado nessa área, ainda existem limitações que precisam ser resolvidas, como por exemplo a falta de base de dados organizadas e de forma acessível para uso (YU, SEFF, ZHANG, SONG, FUNKHOUSER, XIAO, 2015), dentre os problemas temos anotações ruins, servidor de *download* ruim e falta de permissão para acessar os dados. Entretanto, mesmo com essas limitações as possibilidades futuras que a VC oferece são de valor inestimável para a sociedade, possibilitando a redução de trabalhos manuais exaustivos, a automação em grande escala, entre outras vantagens.

Uma das utilizações da VC é o reconhecimento automático de ações humanas em vídeos e essa área tem recebido uma quantidade significativa de atenção da comunidade científica, principalmente em grande escala (KUEHNE, H.; JHUANG; GARROTE; POGGIO; SERRE,2011). Em comparação com a classificação de imagens fixas, o componente temporal dos vídeos fornece uma pista adicional (e importante) para o reconhecimento, pois várias ações podem ser reconhecidas com confiabilidade com base nas informações de movimento.

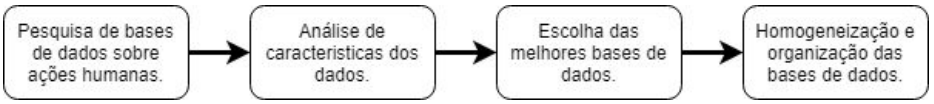
Dessa forma a existência de base de dados suficientemente grandes e bem anotadas tem se mostrado importante nos últimos tempos, principalmente para a utilização em aprendizado profundo. A partir disso umas das frentes do presente trabalho foi criar uma base de dados grande, anotada, organizada e de fácil acesso para uso em futuros trabalhos envolvendo VC.

Além disso existem processos comumente utilizados para extração de *features* e até mesmo processamento de base de dados, tanto na área de ações humanas quanto nas outras áreas de VC, que podem ser feitos de forma mais rápida e fácil em servidores com máquinas de processamento com placas de vídeo , então a segunda frente desse trabalho foi a criação do *front-end* e de parte do *back-end* para extração de características e processamento em aprendizado profundo de base de dados.

MATERIAL E MÉTODOS

Para a criação da base de dados foram seguidas algumas etapas que podem ser vistas na Figura 1.

FIgura 1 - Abordagem ao Problema



Fonte: Própria

A primeira etapa consistiu em pesquisar bases de dados sobre ações humanas que já existiam no meio acadêmico para que fosse possível, através de várias bases pequenas, criar uma bem maior.

Depois de encontrados um total de vinte base de dados veio a segunda etapa do processo onde foram analisadas e escolhidas as bases que tinham mais características em comum, ou seja, conjunto de ações (número de classes), número de quadros por segundo (FPS), cor e resolução.

Após essa análise sobraram um total de sete base de dados, Weizmann Event-Based Analysis (IRANI, ZELNIKI-MANOR, 2001), Weizmann Actions as Space-Time Shapes (BLANK, GORELICK, SCHECHTMAN, IRANI, BASRI, 2007), KTH recognition of human actions (SCHULDT, LAPTEV, CAPUTO, 2004), IXMAS (WEILAND, RONFARD, BOYER, 2004), CASIA (WANG, HUANG, TAN, 2006), UCF101 (SOOMRO, ZAMIR, SHAH, 2011) e HMDB51 (KUEHNE, JHUANG, GARROTE, POGGIO, SERRE, 2011). Na Tabela 1 pode ser visto o total de quadros e de tempo que foram utilizados de cada um.

Tabela 1 –Bases de dados utilizadas

Base de Dados	Tempo Total Utilizado (s)	Total de *Frames*
Weizmann Event-Based Analysis	374	9350
Weizmann Actions as Space-Time Shapes	227,9	5697,5
KTH recognition of human actions	11628,4	290710
IXMAS	5758	143950
CASIA	4651,8	116295
UCF101	3050	76250
HMDB51	350	8750

Fonte: Própria

Por último, com as bases de dados já escolhidos, foi necessário fazer uma normalização dos dados para que fosse possível a real utilização desses dados, ou seja, as anotações e a estrutura de organização dos arquivos de cada base foram

refeitas, portanto todas ficaram organizadas de forma homogênea para que pudessem ser usadas como uma única grande base de dados.

Na segunda frente do trabalho onde foram feitos o *front-end* e parte do *back-end* para extração de características e processamento em aprendizado profundo de bases de dados em unidades de processamento gráfico, foram primeiramente escolhidos os principais modelos de extração de característica, além dos principais parâmetros necessários para execução no servidor. Da mesma forma foram levantados dados para o aprendizado profundo, fazendo com que dessa forma fosse possível criar uma página web que solicita as informações necessárias do usuário, sendo que para sua criação foram utilizados HTML, CSS e Javascript. Além da página foi necessário todo um mecanismo de gerenciamento e de filas no *back-end,* onde foram usados PHP e Shell Script.

RESULTADOS E DISCUSSÃO

O resultado obtido ao final da pesquisa e de todo processo descrito anteriormente foi uma base de dados com mais de 6 horas de duração total e com mais de 500 mil quadros. As ações contidas na base de dados, assim como suas respectivas durações e total de quadros podem ser vistas na Tabela 2 a seguir.

Tabela 2 – Ações com seus respectivos valores totais.

Ações	Tempo Total (s)	Total de *Frames*
Andar	6639,50	165988,00
Correr	3695,80	92395,50
Pular	532,20	13306,00
Acenar	2746,20	68653,75
Se abaixar	2237,80	55944,75
Bater	3541,40	88535,00
Bater Palmas	1706,70	42667,00
Interação com Objetos	1540,60	38515,25
Total	22640,20	566005,25

Fonte: Própria

Além disso tivemos o resultado final da aplicação web que possibilita a extração de características em um total de 20 modelos diferentes além do processamento por aprendizado profundo no servidor do Laboratório de Bioinformática e Inteligência Computacional (LABIC) da UTFPR que contém grande poder de processamento.

CONCLUSÃO

Com base no estudo realizado neste trabalho, notamos que o uso de aprendizado de máquina e aprendizado profundo na área de processamento de imagens está muito presente atualmente, trazendo consigo a necessidade alto poder de processamento e de bases de dados específicas, ou seja, com um grande número de informação organizada corretamente. Com isso foi observado a falta de bases de dados que contém essa estruturação, principalmente na área de reconhecimento ou descrição de ações humanas. Analisando ambas as necessidades da área podemos dizer que os resultados obtidos neste trabalho, sendo eles uma base de dados com 8 categorias diferentes de ações humanas totalizando mais de 6 horas de duração, sendo que todas categorias estão organizadas, fisicamente, em suas devidas pastas acompanhadas das suas anotações, e uma aplicação web para extração de características de bases de dados por meio de processamento em unidades gráficas, irão ajudar no avanço da área de visão computacional.

AGRADECIMENTOS

Agradeço meu Professor orientador Heitor Silvério Lopes pela oportunidade oferecida assim como a todos os meus colegas de laboratório pelo apoio no desenvolvimento desta pesquisa. Também agradeço ao CNPQ pelo apoio financeiro concedido.

REFERÊNCIAS

SCHALKOFF, R.J. **Digital image processing and computer vision** (Vol. 286). 1989. New York: Wiley.

SHIH, F. Y. **Image processing and mathematical morphology: fundamentals and applications.** CRC press, 2017.

YU, F.; SEFF, A.; ZHANG, Y.; SONG, S.; FUNKHOUSER, T.; XIAO, J. L. **Construction of a large-scale image dataset using deep learning with humans in the loop**. *arXiv preprint arXiv:1506.03365*. 2015.

KUEHNE, H.; JHUANG, H.; GARROTE, E.; POGGIO, T.; SERRE, T. **A large video database for human motion recognition.** In Proc. ICCV.

IRANI, M.; ZELNIKI-MANOR, L. **Event-Based Analysis of Video.** Dept. of Computer Science and Applied Math. 2001

BLANK, M. ; GORELICK, L. ; SCHECHTMAN, E.; IRANI, M. ; BASRI, R. **Actions as Space-Time Shapes .** IEEE. 2007.

SCHULDT, C.; LAPTEV, I.; CAPUTO, B.; **Recognizing Human Actions: A Local SVM Approach**. in *Proc. ICPR'04, Cambridge, UK.* 2004.

WEILAND, D.; RONFARD, R.; BOYER, E. **Free Viewpoint Action Recognition using Motion History Volumes, Computer Vision and Image Understanding**. 2006

WANG, Y.; HUANG, K.; TAN, T.. **Human Activity Recognition Based on R Transform. Computer Vision and Pattern Recognition**. 2007. CVPR '07.

SOOMRO, K.; ZAMIR, A., R.; SHAH, M. **UCF101: A Dataset of 101 Human Action Classes From Videos in The Wild.** CRCV-TR-12-01, November, 2012.

KUEHNE, H.; JHUANG, H.;GARROTE, E.; POGGIO, T.; SERRE, T.; **HMDB: A Large Video Database for Human Motion Recognition**. ICCV, 2011.