

PROPOSTA DE MODELO PARA A SEGMENTAÇÃO DE ATIVIDADES EM VÍDEOS EGOCÊNTRICOS PARA A CRIAÇÃO DE MANUAIS DE INSTRUÇÃO EM REALIDADE AUMENTADA

ANDRIANI, Fabrício²; WINKLER, Ingrid³; SENNA, Valter de⁴

¹ Projeto de pesquisa de dissertação de Mestrado

² Mestrando do Programa de Pós-Graduação em Modelagem Computacional e Tecnologias Industriais do SENAI CIMATEC (PPG MCTI/ CIMATEC), Salvador, BA, fcandriani@gmail.com;

³ Docente do Programa de Pós-Graduação em Modelagem Computacional e Tecnologias Industriais do SENAI CIMATEC (PPG MCTI/ CIMATEC), Salvador, BA, ingrid.winkler@fieb.org.br;

⁴ Docente do Programa de Pós-Graduação em Modelagem Computacional e Tecnologias Industriais do SENAI CIMATEC (PPG MCTI/ CIMATEC), Salvador, BA, senna@fieb.org.br

RESUMO

O uso de manuais de instrução em realidade aumentada traz consigo o benefício de reduzir despesas advindas de erros cometidos e o tempo gasto ao executar tarefas, mas encontra o difícil, especializado, demorado e custoso processo de autoria de conteúdo. Uma alternativa explorada para reduzir custos é o uso de vídeos gravados por especialistas na visão egocêntrica, realizando procedimentos e apresentá-los divididos em atividades em um ambiente em realidade aumentada. As técnicas atuais não são capazes de entregar o nível de precisão comparado a uma segmentação manual, ou necessitam de ajustes finos para seu bom funcionamento para diferentes casos. Este trabalho propõe um modelo preliminar para a segmentação de atividades em vídeos egocêntricos de procedimentos de manutenção industrial. O modelo tem como base a extração de características visuais, de movimento e egocêntricas para a identificação de trechos sem atividade e a classificação do vídeo em atividades de forma não supervisionada.

PALAVRAS-CHAVE: Realidade Aumentada, Manuais de Instrução, Visão Egocêntrica, Manutenção Industrial.

1. INTRODUÇÃO

O uso de manuais de instrução em realidade aumentada na indústria traz consigo o benefício de reduzir despesas advindas de erros cometidos e tempo gasto ao executar tarefas.^{1,2,3} No entanto, a adoção de manuais em realidade aumentada encontra o difícil, especializado, demorado e custoso processo de autoria de conteúdo.^{3,4,5,6}

Uma alternativa, explorada por alguns autores para reduzir custos, é o uso de vídeos gravados por especialistas na visão egocêntrica, também conhecida como visão em primeira pessoa, realizando procedimentos e apresentá-los divididos em atividades em um ambiente em realidade aumentada, para trabalhadores que precisam de instruções, seja para treinamento ou para consulta.^{6,7}

Entretanto essa abordagem necessita da segmentação dos procedimentos em atividades menores para melhor entendimento e as técnicas atuais não são capazes de entregar o nível de precisão necessário comparado a uma segmentação manual,⁷ ou necessitam de ajustes finos para seu bom funcionamento para diferentes casos.⁶

Este trabalho pretende então propor um modelo preliminar para a segmentação de atividades em vídeos egocêntricos de procedimentos de manutenção industrial. Para a construção do modelo foi feita uma revisão da literatura sobre segmentação, reconhecimento e classificação de atividades em vídeos egocêntricos, foram selecionadas características e técnicas para a segmentação, estas foram correlacionadas e analisadas com base em outros estudos em um modelo para a segmentação de atividades em vídeos egocêntricos.

3. RESULTADOS E DISCUSSÃO

A partir da revisão da literatura, três grupos de características para a segmentação foram identificados: as características visuais, as de movimento e as egocêntricas como deslocamento das mãos, olhar e câmera. A combinação dessas diversas características melhora os resultados para classificação de atividades.^{8,9,10}

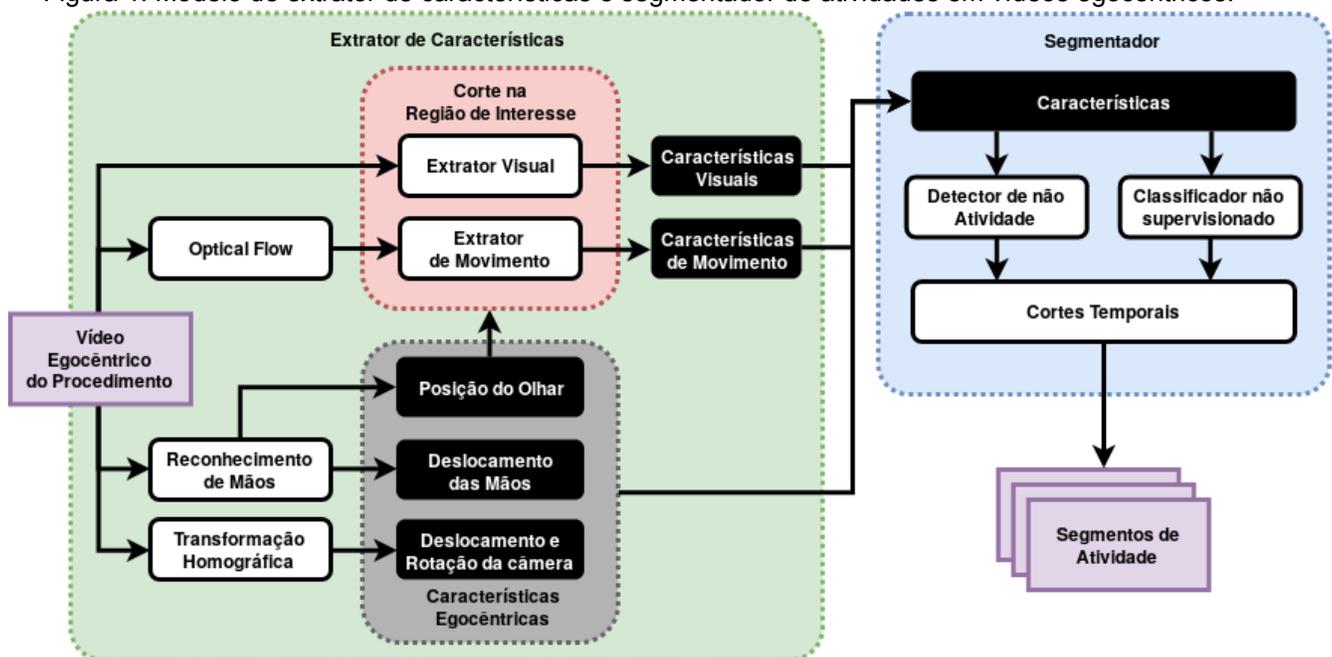
A abordagem mais explorada atualmente para extração de características visuais e de movimento é a do uso de redes neurais convolucionais de dois fluxos (TS-CNN),^{9,10} sendo que o primeiro fluxo recebe os quadros do vídeo e o segundo recebe um conjunto de 10 imagens representando a movimentação dos pixels entre os próximos quadros.

Em atividades manuais a posição das mãos está relacionada com a posição do olhar, essa é uma aproximação que é suficiente para não depender de sensores extras de rastreamento de olhar.¹¹ Dentro do nosso modelo a posição da mão é extraída a partir de uma rede neural recorrente (RNN) para a detecção de mãos.¹² Um corte na região mais próximas a onde a pessoa está olhando permite extrair as características mais importantes para classificação de atividades, desconsiderando outros elementos no plano de fundo do vídeo.¹³

Como as mãos tem um papel fundamental na realização de atividades a detecção de padrões temporais dão pistas importantes para a segmentação e classificação de atividades.⁸ Com o intuito de remover a necessidade de sensores inerciais dedicados e a sincronização destes, a proposta é estimar o deslocamento das mãos a partir da posição das mãos nos quadros do vídeo. O movimento da câmera também é importante, pois indica se a pessoa está focando, ou não, a sua atenção em alguma atividade.⁷ Para estimar o movimento da câmera a proposta é estimar com base na transformação homográfica entre quadros obtida através da comparação de características similares.

A partir das características extraídas, existem duas tarefas a serem realizadas, sendo a primeira identificar períodos sem atividades, isso é feito avaliando se a diferença entre quadros em um intervalo de tempo é menor que um limiar.^{7,14} A segunda é classificar os quadros em atividades de forma não supervisionada para encontrar os pontos de corte entre atividades consecutivas. Os cortes temporais são feitos suavizando as classes de atividade e períodos sem ação de forma a reduzir o ruído, por exemplo, um quadro de uma classe entre uma sequência de quadros de outra classe. A figura 1 descreve o fluxo de informações dentro do modelo proposto.

Figura 1. Modelo de extrator de características e segmentador de atividades em vídeos egocêntricos.



4. CONSIDERAÇÕES FINAIS

Este trabalho teve como objetivo propor um modelo preliminar para a segmentação de atividades em vídeos egocêntricos a partir de uma revisão da literatura, agrupando diversas características. O modelo pode ser utilizado para gerar trechos de vídeo referente a um procedimento a fim de gerar manuais de instrução para realidade aumentada. Este modelo será validado em trabalhos futuros utilizando um *dataset* de atividades em vídeos egocêntricos.

AGRADECIMENTOS

Os autores agradecem ao Grupo de Pesquisa em Realidade Aumentada e Visão Computacional para Inovação na Indústria (CNPQ) e à EMBRAPAII pelo auxílio financeiro da bolsa de formação e desenvolvimento tecnológico concedida.

5. REFERÊNCIAS

- ¹ CAUDELL, Thomas P.; MIZELL, David W. Augmented reality: An application of heads-up display technology to manual manufacturing processes. In: **Proceedings of the twenty-fifth Hawaii international conference on system sciences**. IEEE, 1992. p. 659-669.
- ² WEBEL, Sabine et al. An augmented reality training platform for assembly and maintenance skills. **Robotics and Autonomous Systems**, v. 61, n. 4, p. 398-403, 2013.
- ³ PALMARINI, Riccardo et al. A systematic review of augmented reality applications in maintenance. **Robotics and Computer-Integrated Manufacturing**, v. 49, p. 215-228, 2018.
- ⁴ SANTOS, Marc Ericson C. et al. Augmented reality learning experiences: Survey of prototype design and evaluation. **IEEE Transactions on learning technologies**, v. 7, n. 1, p. 38-56, 2014.
- ⁵ RAMIREZ, Hector et al. Authoring software for augmented reality applications for the use of maintenance and training process. **Procedia Computer Science**, v. 25, p. 189-193, 2013.
- ⁶ PETERSEN, Nils; STRICKER, Didier. Cognitive augmented reality. **Computers & Graphics**, v. 53, p. 82-91, 2015.
- ⁷ LEELASAWASSUK, Teesid; DAMEN, Dima; MAYOL-CUEVAS, Walterio. Automated capture and delivery of assistive task guidance with an eyewear computer: the GlaciAR system. In: **Proceedings of the 8th Augmented Human International Conference**. ACM, 2017. p. 16.
- ⁸ SPRIGGS, Ekaterina H.; DE LA TORRE, Fernando; HEBERT, Martial. Temporal segmentation and activity classification from first-person sensing. In: **2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops**. IEEE, 2009. p. 17-24.
- ⁹ SINGH, Suriya; ARORA, Chetan; JAWAHAR, C. V. First person action recognition using deep learned descriptors. In: **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition**. 2016. p. 2620-2628.
- ¹⁰ VERMA, Sagar et al. Making Third Person Techniques Recognize First-Person Actions in Egocentric Videos. In: **2018 25th IEEE International Conference on Image Processing (ICIP)**. IEEE, 2018. p. 2301-2305.
- ¹¹ LI, Yin; FATHI, Alireza; REHG, James M. Learning to predict gaze in egocentric video. In: **Proceedings of the IEEE International Conference on Computer Vision**. 2013. p. 3216-3223.
- ¹² DIBIA, Victor. Real-time Hand Tracking Using SSD on Tensorflow. GitHub repository, 2017. Disponível em: <<https://github.com/victordibia/handtracking>>. Acesso em: 10 abr. 2019.
- ¹³ ZHANG, Yun C.; LI, Yin; REHG, James M. First-person action decomposition and zero-shot learning. In: **2017 IEEE Winter Conference on Applications of Computer Vision (WACV)**. IEEE, 2017. p. 121-129.
- ¹⁴ HUANG, Shao et al. Egocentric Temporal Action Proposals. **IEEE Transactions on Image Processing**, v. 27, n. 2, p. 764-777, 2018.