

**Desenvolvimento de Sistema de Visão para Monitoramento Industrial**

**Resumo**

O monitoramento de ambientes em tempo real por meio de câmeras de videomonitoramento fornece informações que permitem detectar padrões de movimento específicos prevenindo acidentes ou ainda determinar se um processo poderá sofrer perda de qualidade. Através deste projeto, pretende-se desenvolver um sistema de visão para detecção de objetos e rastreamento de pessoas em uma planta de processos industriais, possibilitando a melhoria da segurança para as pessoas no ambiente.

**Palavras-chave:** monitoramento, visão computacional, segurança do trabalho.

**1. Introdução e Justificativa**

Em ambiente industrial, há uma permanente preocupação com a segurança e a saúde do operador de máquinas. Além disso há o fator econômico em prevenir a interrupção do processo de fabricação em consequência da presença humana, que gera atrasos e custos para a produção. Isso tem motivado o aparecimento de um campo de pesquisa recente, o qual trata justamente da coexistência segura de homens e máquinas móveis em zonas classificadas (ORIOLO, 2015).

A assim chamada, quarta revolução industrial, introduz no ambiente industrial o uso massivo de tecnologias digitais, dentre elas: a inteligência computacional e a robótica colaborativa. Para viabilizar a robótica colaborativa, são utilizados diversos tipos de sensores, com o objetivo de detectar a presença de obstáculos e informar ao robô para que ele calcule uma nova trajetória (BATISTA, da SILVA, THE, 2018). O sensor PMD 3D200 (IFM) é uma câmera óptica de medição de distâncias utilizada em ambientes industriais. Os pixels medem a distância entre a câmera e a superfície mais próxima. A unidade ilumina a cena usando sua própria fonte de luz interna e processa a luz refletida da superfície. Esse tipo de equipamento é muito preciso, porém tem custo muito elevado. Além de necessitar de uma interface de software com os sistemas de controle do robô. Por outro lado, é uma câmera que pode ser utilizada também para aplicações de segurança do trabalho, como em aplicações de detecção de movimentos em situações de risco que podem levar a acidentes de trabalho.

Câmeras como a PMD 3D200 não mais comuns nas indústrias de manufatura. Porém, câmeras de videomonitoramento estão presentes na maioria dos ambientes industriais. Essas câmeras de vídeo usam tecnologia CMOS e possuem um custo muito menor. Por isso, um mesmo ambiente costuma conter mais de uma câmera, disposta em diferentes ângulos e com diferentes enquadramentos do ambiente monitorado. Em geral, essas câmeras servem ao propósito da segurança patrimonial. Porém, as imagens capturadas também podem servir a outros fins.

A inteligência artificial, embutida nos sistemas de visão computacional, teve uma forte evolução a partir das redes neurais de aprendizagem profunda (*Deep Learning*), particularmente as redes neurais convolucionais (RNC). O maior problema com o uso de redes neurais desse tipo é que eles são bastante lentos para aplicações em vídeo. Um método recente, conhecido por YOLO, venceu essa barreira, tornando possível detectar e identificar objetos de uma forma que viabilizou a aplicação em vídeo (REDMON et al., 2016). Entretanto, o ganho em velocidade muitas vezes custa uma diminuição da acurácia (GUO et al., 2016), mesmo que o desempenho geral ainda seja bastante elevado (AGGARWAL, 2018).

Neste projeto, pretende-se desenvolver um sistema de visão computacional para o monitoramento de um ambiente industrial usando câmeras de videomonitoramento, com o objetivo de prevenir acidentes de trabalho. Para isso, as imagens das câmeras serão capturadas e processadas por um

computador, que fará a detecção dos objetos, o rastreamento das pessoas no ambiente e a indicação de riscos de acidentes nos casos em que membros dessas pessoas possam entrar em contato com máquinas rotativa ou eletrificadas, por exemplo. Dessa forma, pode-se melhorar a segurança no ambiente de trabalho, prevenindo acidentes sem, contudo, ser necessário grandes investimentos em sistemas de captura de imagem. O protótipo resultante deste projeto será desenvolvido posteriormente para a aplicação em robótica colaborativa, já que o ambiente onde ele será validado também possui um robô industrial, auxiliando no desenvolvimento de projetos de controle de trajetória de robôs que estão em curso.

## 2. Objetivos

Este projeto visa desenvolver um sistema de visão computacional para o monitoramento de um ambiente de trabalho de características industriais. O sistema deverá detectar objetos e rastrear pessoas e seus movimentos para informar a um sistema de automação industrial possíveis situações de risco aos ocupantes do ambiente. Para desenvolver as atividades são solicitadas **2 (duas) bolsas**. Os **objetivos específicos** deste projeto são:

- Desenvolver um sistema de detecção de objetos em um ambiente industrial controlado;
- Desenvolver um sistema de rastreamento de pessoas em um ambiente industrial controlado;
- Integrar o sistema de detecção de objetos e o sistema de rastreamento em um sistema de prevenção de acidentes.

## 3. Revisão Bibliográfica

O método YOLO (You Look Only Once) (REDMON et al., 2016) é o estado da arte em sistemas de reconhecimento de objetos em tempo real, aliando velocidade e acurácia. A YOLO é uma RNC de interesse particular para este projeto por ser de código aberto e livre de licenças de uso. Essa rede neural pode ser aplicada para a detecção de objetos em imagens de vídeo. Contudo, a rede foi originalmente treinada para objetos comuns do dia-a-dia e necessita ser ajustada ou refinada.

Existem basicamente duas estratégias de treinamento das RNCs. *Transfer Learning* (transferência de aprendizado) é uma estratégia comumente empregada para treinar Redes Neurais Convolucionais (RNC) que consiste em aproveitar os pesos obtidos durante o treinamento da rede em uma tarefa similar (CHEN et al., 2020). Particularmente, essa abordagem permite que as RNCs extrapolem o conhecimento adquirido em um problema para outro semelhante, contribuindo para operacionalização desses modelos em cenários onde a quantidade de dados rotulados disponíveis é reduzida. Essa técnica é geralmente complementada por uma ajuste fino (*fine-tuning*), no qual os pesos, importados do outro problema, são refinados com dados do problema em questão a fim de alcançar melhor convergência (TAJBAKHSI et al., 2016). Com o *fine-tuning*, geralmente existe uma necessidade de dados de treinamento bem inferior ao necessário se o treinamento do zero, com pesos iniciais nulos ou aleatórios, fosse empregado. As abordagens comuns para o *fine-tuning* são: reajustar todos os pesos ou fazer o ajuste apenas para algumas camadas. Outra possibilidade, é adicionar novas camadas a rede, mantendo as previamente existentes com os pesos importados do outro problema e treinando do zero as novas camadas. Transferir aprendizado em conjunto com técnicas de refinamento permitem às RNCs alcançarem o estado-da-arte em diversas aplicações, como desenvolvimento de carros de autônomos e sistemas CAD (*Computer Aided Diagnosis*).

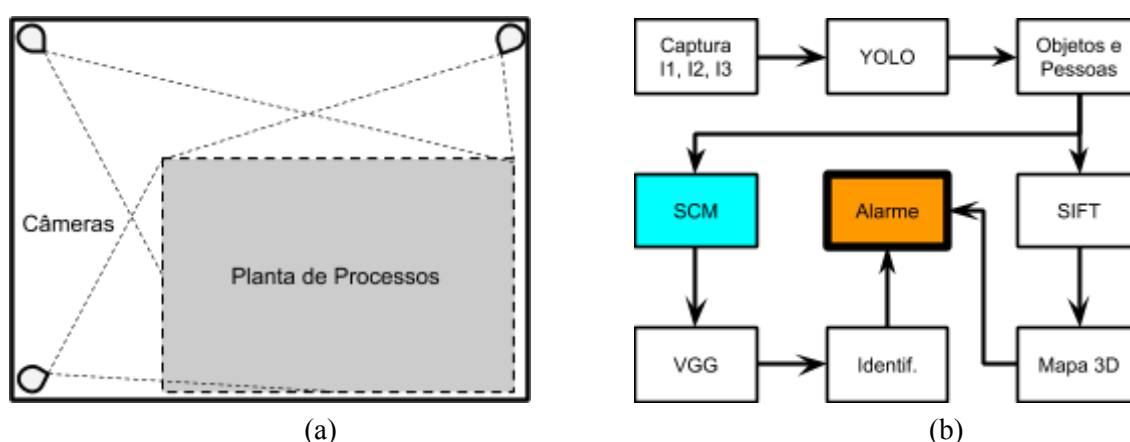
Rebouças Filho et al. (2017) compararam o desempenho de classificação e tempo de extração dos métodos GLCM, Hu moments, Statistical moments, Zernike's moments, Elliptic Fourier features, Tamura's features, Structural Co-occurrence Matrix (SCM) e Analysis of Human Tissue Densities

(AHTD) na análise de imagens do pulmão e cérebro. Devido ao seu bom desempenho como extrator rápido e eficaz, trabalhos recentes têm empregado a SCM (RAMALHO *et al.*, 2016) na extração de características de imagens visando a classificação para a detecção de doenças (PEIXOTO *et al.*, 2018; SOUZA *et al.*, 2018; FILHO *et al.*, 2018). Mais recentemente, a SCM tem sido empregada na ajuda ao refinamento de RCNs, conforme sugerido por FERREIRA *et al.* (2019).

Trabalhos recentes têm demonstrado aplicações de construção de um mapa 3D de um ambiente usando o SIFT para extração de pontos característicos das imagens de cenas capturadas em vários ângulos (WANG, BAO, LIN, WANG, 2019; PRADEEP, MAHAJAN, BHARTI, PAL SINGH, JOSYULA, KUMAR, 2018).

#### 4. Materiais e Métodos

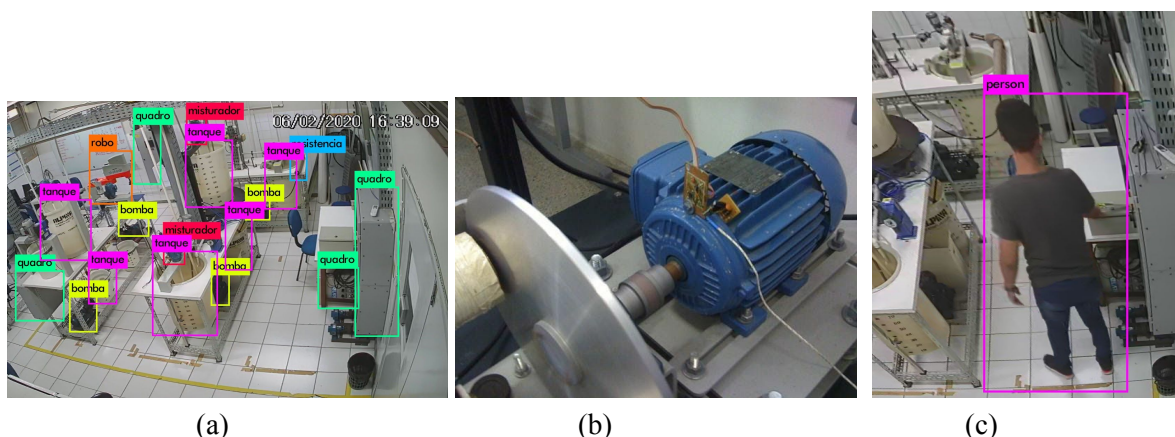
O método proposto utiliza imagens de 3 câmeras de videomonitoramento de um laboratório que abriga uma planta industrial de processos. O ambiente tem acesso controlado, permitindo estudar e testar diversos algoritmos sem grandes interferências. A disposição das câmeras é indicada na Figura 1(a) e um modelo do método a ser desenvolvido é mostrado na Figura 1(b). As três câmeras possuem campos de visão estratégicos para diminuir as oclusões dos equipamentos. O método inicia com a captura das imagens das 3 câmeras. Em seguida é aplicado o método YOLO para detectar objetos nas imagens. Com esses objetos, utiliza-se o método SIFT para correlacioná-los nos 3 ângulos obtidos, possibilitando criar um mapa tridimensional do ambiente com a posição estimada de cada objeto. As pessoas detectadas podem então ser rastreadas no mapa 3D e caso seus movimentos sejam detectados nas proximidades de equipamentos energizados ou em movimento, um alarme será emitido para o sistema de automação executar medidas preventivas. Apesar de também classificar os objetos, para a aplicação proposta neste projeto, é necessário realizar um refinamento da rede YOLO (ou outra RNC) para as imagens específicas. O método de extração de características SCM (RAMALHO *et al.*, 2016) é utilizado para auxiliar no treinamento da RNC (refinamento da YOLO) na identificação dos objetos. O processo de refinamento está indicado genericamente como VGG na Figura 1 (b), representando o refinamento da camada de saída da RNC, ou seja, atuando na classificação dos objetos detectados pela YOLO.



**Figura 1.** (a) Disposição das câmeras do ambiente monitorado. (b) Diagrama de blocos do método.

A Figura 2 ilustra a detecção dos objetos e das pessoas que circulam no ambiente usando a rede YOLO como algoritmo inicial de *deep learning*. Essa é a etapa inicial descrita pelo modelo da Figura 1 (b). Todos a infraestrutura necessária para o projeto estão disponíveis por meio dos laboratórios

LARI (aquisição de imagens e integração com o sistema de automação) e LAPISCO (infraestrutura para processamento das imagens e treinamento das redes neurais em GPUs) do campus Fortaleza.



**Figura 2.** (a) Ambiente monitorado. (b) Detalhe de equipamento detectado. (c) Pessoa detectada no ambiente monitorado.

## 5. Cronograma de atividades

A execução deste projeto está dividida em 12 meses. As atividades principais foram planejadas conforme o cronograma individual dos bolsistas detalhados nas Tabelas 1 e 2.

**Tabela 1 – Cronograma do Bolsista 1 - Desenvolvimento do sistema de detecção e rastreamento**

Atividade	0 1	0 2	0 3	0 4	0 5	0 6	0 7	0 8	0 9	1 0	1 1	1 2
Estudos e levantamento bibliográfico sobre algoritmos de aprimoramento do treinamento de redes convolucionais	■	■										
Especificação e implementação de melhorias no sistema de detecção de objetos e pessoas			■	■	■	■						
Realização de ensaios e validação dos sistemas no laboratório de automação						■	■	■	■	■	■	■
Elaboração de relatórios						■						■
Elaboração de artigos							■	■	■	■	■	■

**Tabela 2 – Cronograma do Bolsista 2 - Desenvolvimento do sistema de integração com a planta de processos**

Atividade	0 1	0 2	0 3	0 4	0 5	0 6	0 7	0 8	0 9	1 0	1 1	1 2
Estudos e levantamento bibliográfico sobre o mapeamento 3D de ambientes	■	■	■									
Especificação e implementação do sistema supervisor de monitoramento da planta e interface com o sistema de visão				■	■	■	■	■				
Realização de ensaios e validação da integração com o sistema de automação						■	■	■	■	■	■	■
Elaboração de relatórios						■						■
Elaboração de artigos ou patentes										■	■	■

A etapa de levantamento bibliográfico consiste na atividade de revisão bibliográfica e de antecedentes sobre o tema através de consulta a periódicos nacionais e internacionais, e a pesquisa em repositórios de patentes. Nessa etapa, os bolsistas realizarão um levantamento sobre as tecnologias que tornam viável a implementação dos sistemas. Também consiste em adquirir conhecimento sobre outros algoritmos que têm potencial de serem explorados. A atividade de especificação e implementação consiste em testar e avaliar diferentes algoritmos no ambiente de produção. Também consiste em modelar e implementar um sistema supervisor que oferece a retaguarda necessária para

a realização de ensaios e validação dos sistemas. A elaboração de relatórios, prevista no cronograma, envolve não apenas os relatórios semestrais exigidos pelo edital, mas também relatórios técnicos e outros documentos que serão produzidos para registro dos resultados da pesquisa. A partir do sexto mês, planeja-se ter dados suficientes para escrever artigos técnicos/científicos, com a participação direta dos bolsista, visando divulgação do projeto em eventos nacionais ou internacionais.

## Referências

- AGGARWAL, C. **Neural Networks and Deep Learning**. Springer. 2018. 512 p.
- BATISTA, J., da SILVA, J. L. N. e THÉ, G. A. P. **Can Artificial Potentials Suit for Collision Avoidance in Factory Floor? - A Case Study of Harmonic Machine-machine Coexistence**. In Proceedings of the 15th International Conference on Informatics in Control, Automation and Robotics (ICINCO 2018) - v. 2, p. 547-556, 2018.
- CHEN, Y., et al. Transfer learning with deep neural networks for model predictive control of HVAC and natural ventilation in smart buildings. **Journal of Cleaner Production**, v.254, p. 119866, 2020
- FERREIRA, D. S. et al. Saliency-driven system models for cell analysis with deep learning. **Computer Methods and Programs in Biomedicine**. v.182, p. 105053, dez, 2019.
- GUO, Y., LIU, Y., OERLEMANS, A., LAO, S., WU, S. e LEW, M. S. Deep learning for visual understanding: A review. **Neurocomputing**. v.18, p. 27-48, 2016.
- ORIOLO, G. **A case study of safe human/robot coexistence, Research Proposal**. DIAG Robotics Lab Sapienza University of Rome, 2015.
- PRADEEP, MAHAJAN, A., BHARTI, V., PAL SINGH, H., JOSYULA, L., KUMAR, P. Construction of a 3D Map of Indoor Environment. **Procedia Computer Science**. v.125, p. 124-131, 2018
- RAMALHO, G. L. B. V. Rotation-invariant feature extraction using a structural co-occurrence matrix, **Measurement**. v. 94, p. 406-415, nov. 2016.
- REBOUÇAS FILHO, P. P.; Reboucas, E. S.; MARINHO, L. B.; SARMENTO, R. M., TAVARES, J. M. R. e ALBUQUERQUE, V. H. C. Analysis of human tissue densities: A new approach to extract features from medical images. **Pattern Recognition Letters**, vol. 94, pp. 211-218, 2017.
- REDMON, J. et al. **You only look once: Unified, real-time object detection**. IEEE Conference on Computer Vision and Pattern Recognition, pp. 779-788, 2016.
- TAJBAKSHI, N. et al. **Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning?** IEEE Transactions on Medical Imaging, v. 35, n. 5, Mai, 2016.
- WANG, Z., BAO, W., LIN, D., WANG, Z. A Local Feature Descriptor Based on SIFT for 3D Pollen Image Recognition. **IEEE Access**. v.7, p. 152658 - 152666, 2019.