Moving-camera Video Surveillance in Cluttered Environments using Deep Features

Bruno M. Afonso Lucas P. Cinelli Lucas A. Thomaz Allan F. da Silva Eduardo A. B. da Silva Sergio L. Netto

Signals, Multimedia and Telecommunications
COPPE/Poli-UFRJ

COPPE UFRJ

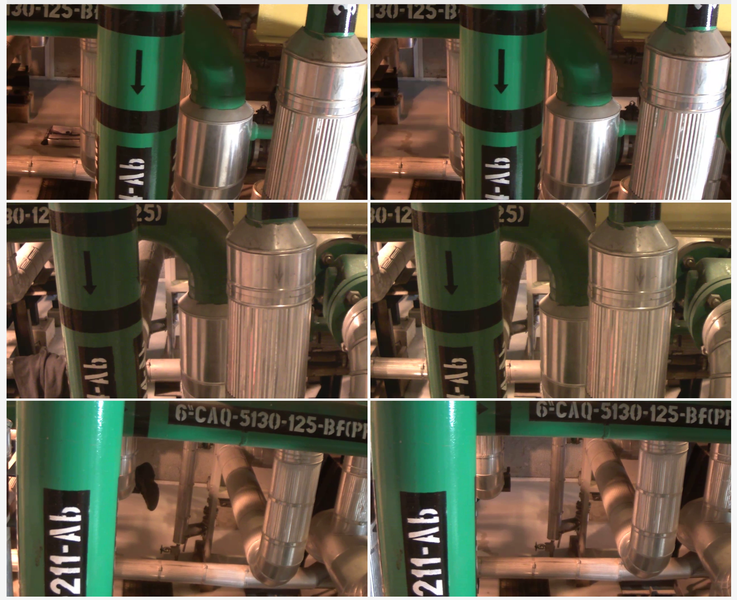
UFRJ

PROPOSED PIPELINE

- Anomaly detection in cluttered environments using videos acquired with a moving camera is a major unsolved issue
- New simple method of comparing a reference and a target video that uses ConvNets to leverage the robustness of the extracted features
- We investigate detection performance across layers as well as the influence of different classifiers

DATASET

- The Video Database of Abandoned Objects (VDAO) [1,2] consists of multiple videos recorded in a **cluttered industrial** environment with **camera moving** back and forth along a rail.
  - High resolution 1280x720 pixels
  - Different lighting conditions
  - Bounding-box annotation



ACKNOWLEDGMENTS

Thanks to CAPES, CNPq and FAPERJ for funding. We also acknowledge funding from CAPES PRO-DEFESA: Nº 23038.009094/2013-83 and NVIDIA for providing the hardware used in this work.

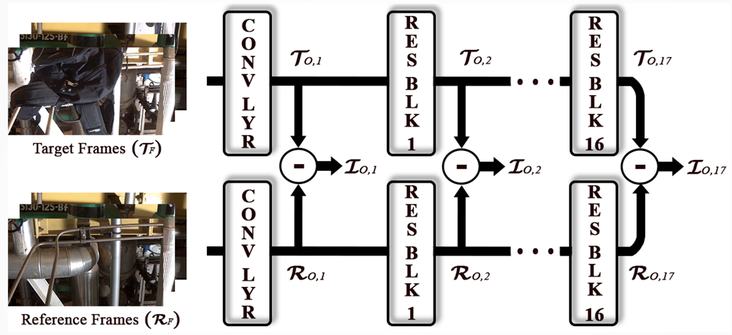
PROPOSED PIPELINE

Feature extraction

- Pre-trained ResNet50 [3]
- Siamese setup
- Feature map subtraction

Two different classifiers

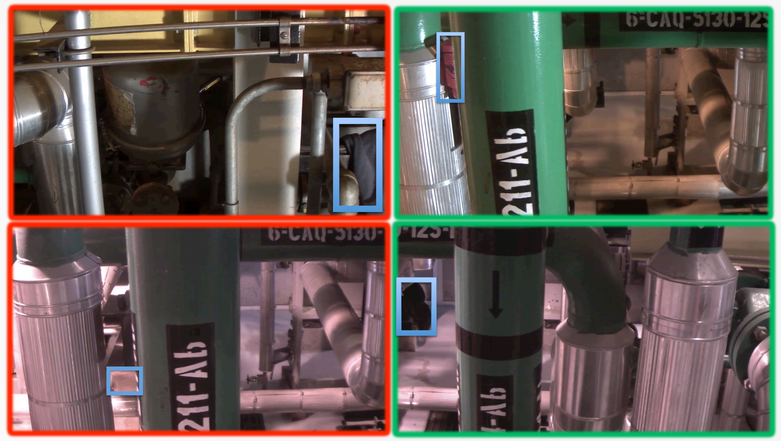
- MLP with 2 hidden layers
- Random forest with a maximum of 100 trees



RESULTS

	CNN+RF			CNN+MLP		
Layer	FP	TP	DIS	FP	TP	DIS
1	0.32	0.68	0.45	0.39	0.70	0.49
2	0.30	0.73	0.40	0.36	0.70	0.47
3	0.24	0.69	0.39	0.34	0.71	0.44
4	0.25	0.74	0.36	0.28	0.66	0.44
5	0.27	0.68	0.42	0.40	0.76	0.47
6	0.27	0.68	0.42	0.39	0.73	0.47

METHOD	FP	TP	DIS
STC-mc [4]	0.48	0.41	0.66
DAOMC [5]	0.89	0.46	0.47
MCBS [6]	0.99	0.98	0.98
mcRoSuRe-A [7]	0.95	0.37	0.37
CNN + MLP	0.66	0.28	0.44
CNN + RF	0.74	0.25	0.36



CONCLUSIONS

- State-of-the-art results
- Simple temporal consistency median filter reduces **DIS** by 4.4%
- The closest result to the ideal scenario on average for all test videos
- Shallower layers with considerably better results
- RF classifier gives better performance than MLP even though the latter is far more popular

REFERENCES

[1] A. F. da Silva, L. A. Thomaz, G. Carvalho, M. T. Nakahata, and E. Jardim, “An annotated video database for abandonedobject detection in a cluttered environment,” in International Telecommunications Symposium, Sao Paulo, Brazil, Aug. 2014, pp. 1–5.

[2] VDAO - Video Database of Abandoned Objects in a Cluttered Industrial Environment, Available: http://www02.smt.ufrj.br/~tvdigital/database/objects/.

[3] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, June 2016, pp. 770–778.

[4] M. T. Nakahta, L. A. Thomaz, A. F. da Silva, E. A. B. da Silva, and S. L. Netto, “Anomaly detection with a moving camera using spatio-temporal codebooks,” Multidimensional Systems and Signal Processing, pp. 1–30, Mar. 2017.

[5] H. Kong, J.-Y. Audibert, and J. Ponce, “Detecting abandoned objects with a moving camera,” IEEE Transactions on Image Processing, vol. 19, no. 8, pp. 2201–2210, Aug. 2010.

[6] H. Mukojima, D. Deguchi, Y. Kawanishi, I. Ide, H. Murase, M. Ukai, N. Nagamine, and R. Nakasone, “Moving camera background-subtraction for obstacle detection on railway tracks,” in IEEE International Conference on Image Processing, Phoenix, USA, Sept. 2016, pp. 3967–3971.

[7] L. A. Thomaz, E. Jardim, A. F. da Silva, E. A. B. da Silva, S. L. Netto, and H. Krim, “Anomaly detection in moving-camera video sequences using principal subspace analysis,” IEEE Transactions on Circuits and Systems I: Regular Papers, vol. 65, pp. 1003–1015, March 2018.