Online Video-based Sequence Synchronization for Moving Camera Object Detection

Allan F. da Silva, Lucas A. Thomaz, Sergio L. Netto and Eduardo A. B. da Silva
PEE, COPPE, Universidade Federal do Rio de Janeiro, Brazil
Email: {*allan.freitas, lucas.thomaz, sergioln, eduardo*}*@smt.ufrj.br*

***Abstract*—This work presents a video-based sequence synchronization algorithm to be used in real-time video surveillance applications. The signals are aligned based on an online dynamic time warping approach that uses only video content information. The algorithm was tested in the alignment of reference and target videos acquired in a cluttered industrial environment with a moving camera. During each recording, the camera follows approximately the same trajectory but its speed along the trajectory may differ among recordings, thus generating time warping between the videos. The results shows that, when considering both alignment error and computational complexity, the use of the mean squared error as similarity metric leads to the best performance among metrics such as the SSIM, absolute error or even sophisticated matching techniques such as histograms of oriented gradients. The proposed method was also tested in a real-time application. Results show that the instantaneous alignment computation of a target frame can be performed successfully, at the cost of a three-fold increase in the error when compared to the case when latency is allowed.**

I. Introduction

Since it is easy to acquire signals from several portable sensors, surveillance systems should ideally be able to analyse multiple information from a large number of sources. In most such systems, several audio and video signals acquired from the same location are used to detect any anomaly or unusual behavior. This creates a demand for temporal alignment techniques, so that the contents of each signal can be identified and compared.

The most relevant technique for aligning two time series is dynamic time warping (DTW) [1], [2]. DTW considers that one sequence can be shrinked or stretched along the time axis to match the other. The problem of finding the best mapping can be described as the search for the path that minimizes the matching error between sequences. This technique has been initially used in the context of automatic speech recognition, but it has already been successfully applied in a wide range of applications, such as biomedicine [3], entertainment [4] and data-mining [5].

The standard DTW has a computational complexity that is quadratic with the number of elements of the sequences. Some constraints can be applied to save computation in the cost matrix. Sakoe-Chiba [2] bounds the path to lie inside a region around the diagonal of the cost matrix, while Itakura [1] restricts the path to be inside a parallelogram, requiring that one sequence will never be more than a certain number of times faster than the other sequence. Other approaches, such as the one in [6], prune the DTW matrix if the cost reaches a given threshold. A DTW algorithm with linear computational complexity was developed by Salvador et. al [7]. It employs a multi-resolution approach to spare computation. A drawback that is shared by these methods is the fact that the cost matrix must be computed beforehand. This creates difficulties for online processing.

A method that presents an online DTW with linear complexity was developed by Dixon [8]. This online DTW performs incremental alignment between two signals when one of them is being received in real time, so only a subsequence of one of the signals is currently known. This algorithm is applied in the alignment of audio signals to provide live analysis of musical performance.

A temporal alignment technique is frequently one of the steps of a video anomaly detection algorithm. In [9] a technique for detection of objects along a road was proposed. It applies a DTW algorithm to align present and past camera images using as similarity metric the position of the epipole between images. After aligning the frames, the algorithm applies a road registration and objects are detected by an image subtraction. A similar approach was developed by Mukojima et. al [10] in the context of railroad objects detection. This algorithm performs a time-alignment between reference and target videos by computing frame-by-frame correspondences and then applies a DTW algorithm using the variance of the angles between corresponding keypoints as similarity metric. After synchronizing both videos, the method performs spatial alignment between frames and computes two image subtraction metrics to detect object regions. Another method for detecting objects along a road was proposed in [11]. This approach applies a rough video alignment that uses only a GPS signal, that is followed by a geometric registration between frames, with objects in the road being detected by the computation of a correlation metric.

A surveillance system that uses a moving robotic platform along with signal processing algorithms is proposed in [12]–[14]. This system presents a generic framework able to perform several tasks such as: detection of audio anomalies [15], identification of gas leakage, detection of video anomalies [16], [17] and diagnosis of rotating machines [18]. Its video anomaly detection algorithm improves the ideas of [11] by adapting it to the context of real-time object detection in a cluttered industrial environment. The algorithm identifies changes between target and reference frames using a correlation metric dispensing with the use of any clues such

978-1-5090-3649-3/17/$31.00

as trail or road detection. During an alignment step, the algorithm computes horizontal displacements between consecutive frames and aligns different videos with a model-based maximum likelihood estimation, which makes this method not compatible with non-rectilinear camera movements.

This work proposes a DTW-based video alignment algorithm that can be used in video detection systems. The online DTW is adapted and optimized in the context of real-time video alignment for anomaly detection algorithms. One concern when using the DTW in this application is that one of the videos can have regions with visual information that do not exist in the other video. The video anomalies in this case are objects removed, moved or placed during a new recording. In order to develop a method able to deal with this issue, in this work we considered and tested several image distance metrics according to their efficiency in the video temporal alignment and their computational complexity. The performance of the video alignment using data acquired by the several sensors in the surveillance system is also presented for comparison purposes.

The remaining of this paper is organized as follows. Section II presents a brief description of the standard DTW algorithm while Section III shows the online DTW approach. In Section IV an overview of the surveillance system is presented, and the proposed algorithm is described in Section V. Section VI shows the experimental results of the proposed method. Section VII presents the conclusions and final discussions.

II. Standard Dynamic Time Warping

Given two time series $\mathbf{X} = [x_1, x_2, \cdots, x_N]$ and $\mathbf{Y} = [y_1, y_2, \cdots, y_M]$, the DTW aims to find the minimum-cost path $W = [w_1, w_2, \cdots, w_L]$ with $w_k = (i_k, j_k) \in [1, 2, \cdots, N] \times [1, 2, \cdots, M]$ such that x_{i_k} and y_{j_k} are aligned. This path W usually should satisfy some constraints:

- Boundary: $w_1 = (1, 1)$ and $w_L = (N, M)$;
- Monotonicity: $i_1 < i_2 < \cdots < i_L$, $j_1 < j_2 < \cdots < j_L$;
- Continuity: $w_{k+1} - w_k \in \{(1, 0), (0, 1), (1, 1)\}$;

To find the optimal warping path that aligns the time series $\mathbf{X}$ and $\mathbf{Y}$, one can create a cost matrix $\mathbf{d}$ of size $N \times M$ where each element $d(i, j)$ represents a similarity measurement between the samples x_i and y_j that is also the cost of their misalignment. The optimal warping path is the one that minimizes the sum of the costs along the path:

$$\mathrm{DTW}(\mathbf{X}, \mathbf{Y}) = \min \sum_{(i,j) \in W} d(i, j). \tag{1}$$

The problem given by Eq. (1) can be easily solved by dynamic programming, creating an accumulated-cost matrix $\mathbf{D}$ using the following recursive formulation:

$$D(i, j) = d(i, j) + \min \left\{ \begin{matrix} D(i-1, j) \\ D(i, j-1) \\ D(i-1, j-1) \end{matrix} \right\}. \tag{2}$$

The path is obtained starting at the element $D(N, M)$ and testing each previous element $D(N-1, M)$, $D(N, M-1)$ and $D(N-1, M-1)$ in the recursion. Whichever has the smallest value is added to the path and the recursion continues from it until the element $D(1, 1)$ is reached.

III. Online Dynamic Time Warping

The standard DTW requires that all samples from both sequences are known at the start of the execution of the algorithm, since it aligns the initial and final samples from each sequence beforehand. One of its drawbacks is that when one of the sequences is only partially known the boundary conditions cannot be satisfied. The online DTW seeks the best alignment of a partially unknown target sequence and a subsequence of the reference, restricting the search to a window so that the algorithm has linear complexity.

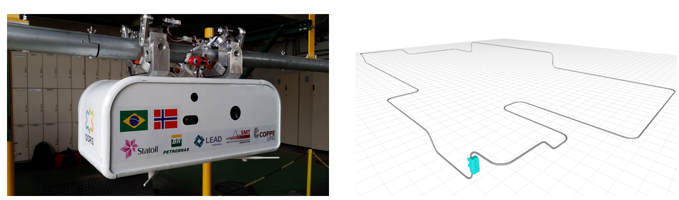
Starting with reference and target subsequences having the size of the search window c, the algorithm applies the standard DTW to find an initial warping path, inserting a weight 2 for diagonal steps in the accumulated cost matrix of Eq. (2), so that there is no bias for diagonal movements:

$$D(i, j) = \min \left\{ \begin{matrix} D(i-1, j) + d(i, j) \\ D(i, j-1) + d(i, j) \\ D(i-1, j-1) + 2d(i, j) \end{matrix} \right\}. \tag{3}$$

For each new iteration, the algorithm uses the warping path found in a previous iteration to decide whether to increase the size of the reference or the target subsequences, and computes similarity measurements between the new sample from one sequence and the last c samples from the other sequence, including them in the cost matrix $\mathbf{d}$. The accumulated-cost matrix $\mathbf{D}$ is also updated by applying Eq. (3) and a new warping path between the reference and target subsequences is found.

IV. Surveillance System Description

DORIS - Monitoring Robots for Offshore Facilities is a surveillance system designed for remote supervision, diagnosis, and data acquisition on offshore facilities. The system is composed of a rail-guided robot that moves inside a cluttered industrial environment. The robotic platform can carry different interchangeable sensors that provide measurement and analysis of several properties of the robot and the environment. Fig. 1 shows an image of the robotic platform and a model of the rail installed in a cluttered environment. Specific details on the robotic platform and the control mechanisms can be found in [12]–[14].



(a) (b)

Fig. 1. DORIS System. (a) Robotic platform in a cluttered environment. (b) 3D model of the rail (gray) and the robotic platform (blue).

In the DORIS system, RGB and thermal cameras are employed in the detection of video anomalies such as abandoned objects, gas leakage and fire. An anomaly detection algorithm [16], [17] designed for this system compares frames from a new video that may contain an anomaly with the most similar frames from a reference video (a previous recording which was validated by an operator as having no anomalies). In order to find the similar frame in the reference video, the algorithm computes displacements between consecutive frames from each video and aligns the videos using a template matching. Fig. 2 shows a frame from a video with an abandoned object and a frame with the same view from the reference video.

(a) (b)

Fig. 2. Example of a reference frame and an aligned target frame with video anomalies to be detected. (a) Reference frame. (b) Target frame from the same scene with a few different objects marked by a red square.

The system also has several sensors measuring orientation, speed, acceleration and power consumption. This information can be used, for instance, to estimate the current position and provide temporal and spatial alignment. The system has a native algorithm that uses only the power consumption information to provide a rough estimate of the linear displacement between the current instant and the starting position, which can also be used to align any video with a reference video. In [19], a correlation metric is used to find the delay that best aligns the outputs of sensor set corresponding to two different recordings. This delay is used to estimate the optimal alignment between other signals of interest such as audio or video signals.

V. Online Video Alignment for Moving Camera Object Detection

The framework of surveillance systems with moving camera object detection imposes several constraints that must be satisfied by the alignment algorithm. In this application, one of the sequences, the reference signal, is fully known and the other sequence, the target sequence, is being received in real-time and must be aligned and processed on-the-fly, which makes the online DTW seen in Section III a suitable approach. However, in order to be used in this framework, some innovations had to be made to the online DTW algorithm.

Since the original online DTW was developed for a music application, a new cost function must be applied in order to align video frames. Furthermore, when dealing with videos acquired in a surveillance operation, one can often deal with frames recorded in the same position that have regions with different information. As can be seen in Fig. 2, the frames from the target video may have regions with video anomalies. In this case there may be objects that either did not exist or were in a different position during the reference recording. Thereby, the alignment algorithm must be able to align frames even when one of them has small regions that do not match the ones in the other. In this work, we propose the use of a simple metric, the mean square error (MSE) between subsampled frames (L_2 norm) and show that it produces the best compromise between error rate and processing time. We also show that it has some of the best results with low computational complexity, which is an important requirement in real-time applications.

The original algorithm proposed by Dixon [8] performs an optimal warping between two videos which can include repetitions of any of the frames of the videos. However, in an object detection application, the real concern is finding a frame in the reference video that is equivalent to each new frame in the target video. Therefore, the proposed algorithm computes the optimal warping path and, for each target frame, finds the aligned reference frame with the minimum cost.

In addition, the online DTW algorithm computes the path in the forward direction, incrementally computing the optimal warping for each new sample. In the proposed system, a latency in the warping path computation is introduced, by computing the alignment for a given target frame only after the k subsequent frames were received. This approach was discussed in [8] and was deemed unnecessary in the context of music alignment. However, since this work deals with a different application, this approach with latency is also considered.

VI. Results

The system described in Section IV was used in the acquisition of six videos: two reference videos and four target videos that contain anomalous objects. For three target videos the robotic platform was programmed to vary its speed along the trajectory, with speed varying between $0.2\,m/s$ and $0.4\,m/s$, thus generating videos with a warping in time when compared to the reference video. The fourth target video was recorded with a constant speed of $0.1\,m/s$, but it contains regions with bigger anomalous objects than the other three target videos, as the one seen in Fig. 2. The reference videos were recorded with the robotic platform moving with two constant speeds: a recording with speed of $0.2\,m/s$ that is a reference to the three target videos with varying speeds, and a recording with speed of $0.1\,m/s$ that is a reference to the constant-speed target video. All videos have a spatial resolution of 800×450 pixels and have a frame rate of around 2.5 fps. Tab. I summarizes some of the properties of the videos. The proposed algorithm was implemented in C++ and tested with several target video and reference video alignment configurations. The tests were made using a computer with an Intel Core i7-3630 QM processor with 2.40 GHz clock and 16 GB of RAM.

A. Cost Function

To test the robustness of the DTW algorithm in this application, several cost functions were considered. In [4], a DTW is developed which uses a subsampled version of the frame as the frame descriptor, and uses as cost function the L_1-norm

TABLE I
PROPERTIES OF THE VIDEOS USED IN THE TESTS.

	Duration (s)	Total frames	Camera speed (m/s)
Target 1	560	1400	0.2 to 0.3
Target 2	486	1215	0.2 to 0.4
Target 3	488	1218	0.2 to 0.4
Target 4	1329	3176	0.1
Reference 1	649	1622	0.2
Reference 2	1346	3050	0.1

between frame descriptors. The moving-camera background-subtraction algorithm proposed in [10] employs the normalized vector distance (NVD) [20] to compare frames and detect anomalies, which can also be applied as a cost function in a DTW algorithm. Other common metrics for comparing frames also tested include the structural similarity (SSIM) [21] and distance between histogram of oriented gradients (HoG) descriptors [22].

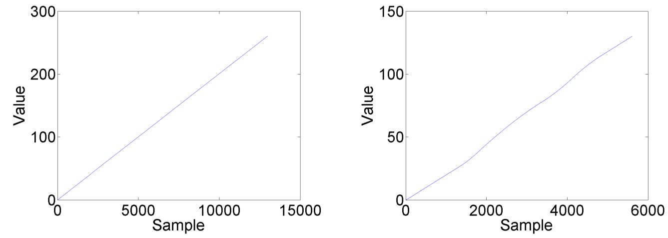
In the tests, cost functions based on the L_1 and L_2 norms were considered. We subsample the reference and target frames to the size 16×9, stack the lines from each frame in a descriptor vector and compare the descriptors by computing the L_1 and L_2 norms of the difference. The HoG cost function uses a descriptor based on the implementation given by [23]. For the NVD cost function, the original frames are subsampled to the size 80×45 and each frame is divided into 25 image patches. The SSIM is applied in the comparison of the frames after downsampling to the size 32×18.

The standard DTW was tested in the videos for comparison purposes. This version of the algorithm computes the similarities between all frames from both videos, having a computational complexity that is quadratic on the length of the videos. Due to the amount of processing time this method takes to align each pair of videos, it was only tested with the cost function that computes the L_2-norm of the error between subsampled frames.

For the sake of comparison, the system's native displacement estimation algorithm was also employed in the computation of a video alignment between reference and target videos and its results were used as ground-truth. This algorithm uses only the measurement of the power consumption to estimate the robot's linear position during the recording, that gives the linear position at which each frame of a video was acquired. Using this information, one can achieve video alignment by finding the frames in the two videos that were recorded in closest positions. In Fig 3, one can see an example of the linear position estimate for a reference video recorded at a constant speed and a target video recorded with speed variation.

We also investigate the use of the signals acquired by all the other sensors available in the DORIS system in the alignment problem. These sensors have a timestamp that enables a synchronism with a video signal acquired simultaneously, so the alignment obtained from the sensor data for the reference and target recordings can be used to estimate a time-warping between the two videos. For this purpose, be $\mathbf{r}_i$ a vector containing the set of sensors samples acquired at a time i during the reference recording and $\mathbf{t}_j$ a vector containing the set of sensors samples acquired at a time j during the target recording. We applied the DTW with cost function:

$$d(i,j) = \|\mathbf{r}_i - \mathbf{t}_j\|^2 . \qquad (4)$$



(a) (b)

Fig. 3. Example of the linear position estimate provided by the DORIS system. The trail has a length of approximately 130 m. (a) Reference video with constant speed (2 laps). (b) Target video with speed variation.

For these tests, we consider only the search window provided by the online DTW algorithm, and compute the final warping path using the whole reference and target sequences. A first experiment analyses the impact of the several cost functions and the reduced cost matrix in the alignment between the target and reference sequences. In a second experiment we perform an analysis of the quality of the incremental path estimated on-the-fly by the online approach.

Tab. II presents the alignment error between the several cost functions tested in the DTW algorithm and the ground truth. Tab. III presents the average processing time for each cost function. As can be seen from the results, the cost function based on the L_2-norm (which represents, up to scale, the MSE between subsampled frames) is only outperformed, in terms of alignment error, by the cost functions based on NVD and SSIM. However, it is at least 2 times faster than both of them. Given that it is advantageous that the alignment step be as simple as possible due to the very complex nature of the anomaly detection step, this indicates that the MSE is the best cost function to be used in a real-time application. One should also notice that, despite computing only a reduced number of frame similarities, which are restricted to a search window, the online approach produces the same result as the standard DTW, showing that the algorithm correctly estimates a region of interest around the best warping.

B. Real-time warping

In a real-time anomaly-detection application, a frame from the target video must be synchronized to a frame from the reference video in order to produce an output without prior knowledge of any posterior target frames. If the system can allow a fixed latency by producing a detection output for the target frame N only after K new target frames have been received, the optimal alignment for frame N can be estimated with a fixed look into the future. In this case, the optimal warping path used to align target frame N can be computed

TABLE II

ALIGNMENT ERROR (IN FRAMES) FOR SEVERAL COST FUNCTIONS USED IN THE DTW ALGORITHM. THE ERROR IS COMPUTED AS THE AVERAGE DIFFERENCE BETWEEN THE ALIGNMENT GIVEN BY THE DTW METHOD AND THE ALIGNMENT PROVIDED BY THE SYSTEM'S NATIVE DISPLACEMENT ESTIMATE. THE BEST 3 RESULTS ARE MARKED IN BLUE.

		Average error (frames)			
		Video 1	Video 2	Video 3	Video 4
Online DTW	L_1	0.95	0.92	1.45	0.44
	L_2	0.58	**0.48**	**0.80**	**0.37**
	SSIM	**0.41**	**0.39**	**0.66**	**0.36**
	NVD	**0.48**	**0.60**	**0.78**	**0.32**
	HoG	**0.45**	0.61	0.95	0.38
	Sensor	4.88	6.08	5.86	4.84
Standard DTW	L_2	0.58	0.48	0.80	0.37

TABLE III

PROCESSING TIME FOR SEVERAL COST FUNCTIONS USED IN THE DTW ALGORITHM. THE BEST 3 RESULTS ARE MARKED IN BLUE.

		Processing time (s)			
		Video 1	Video 2	Video 3	Video 4
Online DTW	L_1	**76**	82	89	**283**
	L_2	**83**	**80**	**87**	308
	SSIM	189	187	189	516
	NVD	243	233	236	671
	HoG	89	**79**	**83**	**280**
	Sensor	**19**	**18**	**20**	**76**
Standard DTW	L_2	41311	52954	40567	132571

between a subset of the reference sequence and the target sequence up to the frame $N + K$.

In this subsection, we analyze how the DTW algorithm behaves when providing a frame alignment in a real-time application. Using as cost function the MSE between subsampled frames, we vary the allowed latency in the system. The results can be compared to the ones shown in Tab. II, which represents the alignment error with the maximum amount of latency, when all information from the target video is available in the computation of the video alignment.

Fig. 4 shows the alignment error obtained when using several values of latency in the warping computation. The results show that, contrary to what is stated in [8], the use of latency can reduce the alignment error up to a third of the one obtained when using only the current target frame to estimate the warping. The graphs also show that with a latency of 50 frames, which corresponds to approximately 20 s , the error becomes close to the minimum, reaching a value similar to the one in Tab. II. For all cases, a good trade-off is found when using a latency of around 15 frames, which represents a delay of approximately 6 s and is not prohibitive for the considered application.

VII. CONCLUSIONS

This work presented a video-based temporal alignment algorithm for surveillance systems based on a dynamic time-warping approach that is able to compute the time-warping between sequences in real-time. The algorithm was tested in a real application that consists of a platform that moves inside a cluttered environment and records a video signal which is used in the detection of video anomalies. The results

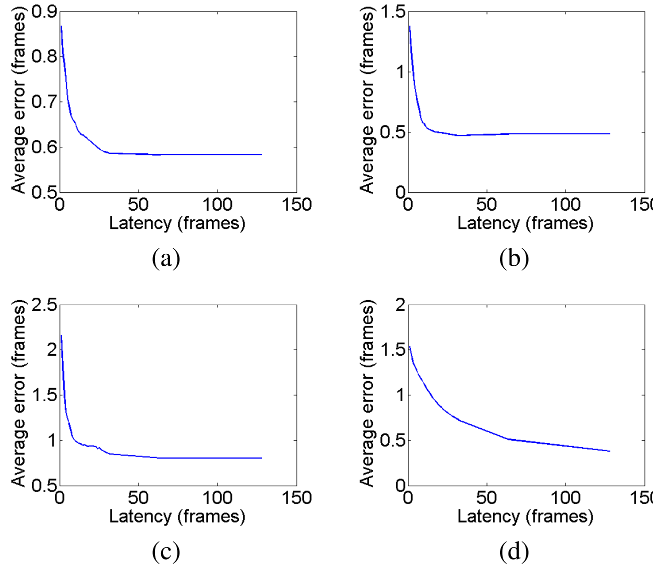


Fig. 4. Average alignment error (in frames) using as cost function the MSE between subsampled frames. (a) Video 1. (b) Video 2. (c) Video 3. (d) Video 4.

show that the mean square error between frames provides the best compromise between alignment error and processing time when used as a cost function for the DTW algorithm. The results also show that the real-time computation of the warping produces a three-fold penalty in the alignment error when compared to the warping computed using all samples from the sequences. If some latency in the computation is allowed, a good trade-off between the alignment error and the delay introduced is obtained using a latency of 6 s.

REFERENCES

[1] F. Itakura, "Minimum prediction residual principle applied to speech recognition," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 23, no. 1, pp. 67–72, Feb. 1975.

[2] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 26, no. 1, pp. 43–49, Feb. 1978.

[3] Y. Yuan, Y.-P. P. Chen, S. Ni, A. G. Xu, et al., "Development and application of a modified dynamic time warping algorithm (dtw-s) to analyses of primate brain expression time series," *BMC Bioinformatics*, vol. 12, no. 1, 2011.

[4] A. Kappeler, H. Iliadis, M.and Wang, and A. K. Katsaggelos, "Block based video alignment with linear time and space complexity," in *IEEE International Conference on Image Processing (ICIP)*, Sept. 2016, pp. 3324–3328.

[5] T. Rakthanmanon, B. Campana, A. Mueen, G. Batista, et al., "Searching and mining trillions of time series subsequences under dynamic time warping," in *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Aug. 2012, pp. 262–270.

[6] D. F. Silva and G. E. A. P. A. Batista, "Speeding up all-pairwise dynamic time warping matrix calculation," in *SIAM International Conference on Data Mining*, Miami, USA, May 2016, pp. 837–845.

[7] S. Salvador and P. Chan, "Toward accurate dynamic time warping in linear time and space," *Intelligent Data Analysis*, vol. 11, no. 5, pp. 561–580, Oct 2007.

[8] S. Dixon, "Live tracking of musical performances using on–line time warping," in *8th International Conference on Digital Audio Effects (DAFx)*, 2005.

[9] H. Kyutoku, D. Deguchi, T. Takahashi, Y. Mekada, et al., “Subtraction-based forward obstacle detection using illumination insensitive feature for driving-support,” in *IEEE International Conference on Computer Vision (ICCV)*, Berlin, Germany, 2012, pp. 515–525.

[10] H. Mukojima, D. Deguchi, Y. Kawanishi, I. Ide, et al., “Moving camera background-subtraction for obstacle detection on railway tracks,” in *IEEE International Conference on Image Processing (ICIP)*, Sept. 2016, pp. 3967–3971.

[11] 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.

[12] G. Carvalho, G. Freitas, R. Costa, G. Carvalho, et al., “DORIS - Monitoring Robot for Offshore Facilities,” in *Offshore Technology Conference*, Rio de Janeiro, Brazil, Oct. 2013.

[13] M. Galassi, A. Røyrøy, G. Carvalho, G. Freitas, et al., “DORIS - A Mobile Robot for Inspection and Monitoring of Offshore Facilities,” in *Anais do XX Congresso Brasileiro de Automática*, Belo Horizonte, Brazil, Sept. 2014.

[14] R. Freitas, M. Xaud, I. Marcovistz, A. Neves, et al., “The embedded electronics and software of doris offshore robot,” in *IFAC Workshop on Automatic Control in Offshore Oil and Gas Production*, Florianópolis, Brazil, May 2015.

[15] T. M. Prego, A. A. de Lima, S. L. Netto, and E. A. B. da Silva, “Audio anomaly detection on rotating machinery using image signal processing,” in *IEEE 7th Latin American Symposium on Circuits Systems (LASCAS)*, Feb. 2016, pp. 207–210.

[16] G. Carvalho, J. F. L. de Oliveira, E. A. B. da Silva, S. L. Netto, et al., “Um sistema de monitoramento para detecção de objetos em tempo real empregando câmera em movimento,” in *XXXI Simpósio Brasileiro de Telecomunicações*, Fortaleza, Brazil, Sept. 2013.

[17] G. H. F. de Carvalho, *Automatic detection of abandoned objects with a moving camera using multiscale video analysis*, Ph.D. thesis, COPPE - Universidade Federal do Rio de Janeiro, Rio de Janeiro - Brasil, 2015.

[18] A. A. de Lima, T. M. Prego, S. L. Netto, E. A. B. da Silva, et al., “On fault classification in rotating machines using fourier domain features and neural networks,” in *IEEE 4th Latin American Symposium on Circuits and Systems (LASCAS)*, Feb. 2013, pp. 1–4.

[19] A. F. da Silva, L. A. Thomaz, E. A. B. da Silva, S. L. Netto, et al., “Alinhamento de sinais obtidos em trajetórias fechadas utilizando um conjunto genérico de sensores,” in *XXXI Simpósio Brasileiro de Telecomunicações*, Santarém, Brazil, Sept. 2016, pp. 329–333.

[20] T. Matsuyama, T. Ohya, and H. Habe, “Background subtraction for non-stationary scene,” in *Asian Conference on Computer Vision (ACCV)*, Taipei, Taiwan, Jan. 2000, pp. 622–667.

[21] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: from error visibility to structural similarity,” *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, April 2004.

[22] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, Washington, USA, June 2005, pp. 886–893.

[23] O. L. Junior, D. Delgado, V. Goncalves, and U. Nunes, “Trainable classifier-fusion schemes: An application to pedestrian detection,” in *12th International IEEE Conference on Intelligent Transportation Systems*, Oct. 2009, pp. 1–6.