

Anomaly Detection with a Spatio-Temporal Tracking of the Laser Spot

David ATIENZA ^{a,1}, Concha BIELZA ^a, Javier DIAZ ^{a,b} and Pedro LARRAÑAGA ^a

^a *Department of Artificial Intelligence, Technical University of Madrid, Spain*

^b *IKERGUNE A.I.E, Elgoibar, Spain*

Abstract. Anomaly detection is an important problem with many applications in industry. This paper introduces a new methodology for detecting anomalies in a real laser heating surface process recorded with a high-speed thermal camera (1000 fps, 32×32 pixels). The system is trained with non-anomalous data only (32 videos with 21500 frames). The proposed method is built upon kernel density estimation and is capable of detecting anomalies in time-series data. The classification should be completed in-process, that is, within the cycle time of the workpiece.

Keywords. Kernel density estimation, Anomaly detection, Time-series, Laser surface heating process

1. Introduction

Anomaly detection is the task of finding anomalous events that differ a great deal from the normal behavior of a system. Anomaly detection usually involves the computation of an abnormality score. If the abnormality score is higher than a given threshold, the instance is considered anomalous. Anomalies are usually associated with a negative impact on the process under study.

In this paper, a novel anomaly detection method is developed to a laser surface heat treatment process. A laser surface heat treatment process aims to modify the mechanical properties of the steel by increasing the surface temperature. An anomaly in laser surface heat treatment could mean that the surface is not heated to a high enough temperature or the heated zone is out of tolerance. Anomaly detection [1] has been applied to other laser applications. For example, [2] used continuous hidden Markov models to detect anomalies in laser welding. They also solved the same problem in [3] by tracking the sputters during the laser welding process using a high-speed thermal camera and Kalman filters [4]. Another interesting approach to anomaly detection is based on D -Markov machines [5]. D -Markov machines can take into account spatio-temporal information using overlapping sliding windows of size D and with a space discretization. Kernel density estimation (KDE) has also been used in many anomaly detection problems [6,7].

¹Corresponding Author: David Atienza, Computational Intelligence Group, Departamento de Inteligencia Artificial, Escuela Técnica Superior de Ingenieros Informáticos, Universidad Politécnica de Madrid, Boadilla del Monte, 28660 Madrid, Spain; E-mail: d.atienzag@alumnos.upm.es.

2. Data

2.1. Data Description

We have used data from a real laser surface heat treatment process applied to 32 cylindrical workpieces made of steel. A thermal camera (NIT Taychon 1024 μ Core@1000 fps, 32×32 pixels) recorded a video for each workpiece. Each pixel value ranges from 0 to 1024 and is equivalent to its temperature. Each video contains 21,500 frames (cycle time of 21.5 seconds for each workpiece). However, the laser spot is visible from the camera during approximately 20,700 frames (different videos can have different number of frames where the laser spot is visible). There is a constraint on the classification process: the workpiece should be classified within its cycle time (in-process classification). During the laser surface heating process, the laser spot moves with a fixed pattern (100Hz frequency) as shown in Figure 1(a). The recorded surface of the workpiece is 10×20 mm. Nevertheless, the pattern is modified during around 3800 frames (from frame 2200 to frame 6000) to avoid an obstacle in the cylindrical piece that cannot be heated by the laser. This change in the pattern should be taken into account because normal behaviour will be defined differently in specific time-frames of the process.

2.2. Data Preprocessing

Before the laser spot movement can be processed, we have to obtain the laser spot position in each frame. We then extract the differences between one frame and the next for the entire video. The result is a *subtraction* video that shows the variation in the surface temperature of the piece. As the laser spot is moving, it applies energy to different zones of the surface. The regions that are heated by the laser spot exhibit higher pixel values in the subtraction video. Finally, we compute the centroid for the regions of the surface with higher pixel values in the subtraction video. This gives us the coordinates of the position of the laser spot on each frame of the video, see Figure 1(b). The set of positions (one for each frame of the video) will be used as a summary of the video to detect anomalies in the laser spot movement.

3. Methodology

The methodology used to detect anomalies takes into account both the spatial and temporal characteristics of the laser spot positions by generating multiple KDE models to characterize the movements of the laser. Many techniques have been proposed to select a good bandwidth for the KDE, but we use Scott's rule [8] in all experiments.

3.1. Training

The system was trained using information about the normal processes. First, we had to split each video in space and time. We divided the frames into W non-overlapping consecutive temporal windows. Next, we partitioned the space in each temporal window. To do this, we used uniform partitioning [9] is used on the vertical (V partitions) and horizontal (H partitions) axes. This partition created a matrix of $W \times V \times H$ spatio-temporal regions.

After completing the spatial and temporal partitioning, we proceeded as follows to train a KDE model for each spatio-temporal region:

1. Find the set of laser spot positions in a spatio-temporal region d , i.e. $\text{Origin} = \{\mathbf{x}_t \mid \mathbf{x}_t \in [0, 32] \times [0, 32]\}$ where stands for the laser spot position found in the t -th frame of the video, represented by its vertical and horizontal components.
2. Train a KDE model for d with the successive positions of the set Origin , i.e. with the positions $\text{Destination} = \{\mathbf{x}_{t+1} \mid \mathbf{x}_{t+1} \in [0, 32] \times [0, 32]\}$. Destination laser spot positions could be outside region d .

The KDE model for a spatio-temporal region d gives an estimate of where the laser spot placed within region d in the current frame should be in the next frame. In a normal video, if two positions of the laser spot, \mathbf{x}_t and \mathbf{x}_k , are very close to each other, then the next positions at frame $t + 1$ and $k + 1$, \mathbf{x}_{t+1} and \mathbf{x}_{k+1} , are expected to be very close too. Uniform partitioning is better suited for our model because all the positions of a region (Origin set) are very close to each other. As we train the model with non-anomalous videos, the Destination set contains positions very close to each other. That implies that region-wise KDE models will be more accurate because the higher probabilities are confined to a small area on the surface of the workpiece (where the Destination positions lie). The result of the training process is a matrix of KDE models of size $W \times V \times H$. The estimated probability density of each bivariate individual KDE model is:

$$\hat{f}(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n \frac{1}{2\pi\sqrt{|\hat{\Sigma}|}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{x}^i)^T \hat{\Sigma}^{-1}(\mathbf{x} - \mathbf{x}^i)\right) \quad (1)$$

where n is equal to the number of training positions (positions in the Destination set), \mathbf{x}^i represents training positions and $\hat{\Sigma}$ the covariance matrix of each individual kernel.

3.2. Classification Phase

We define a movement as the transition from a laser spot position at frame t (origin) to a laser spot position at frame $t + 1$ (destination). Therefore, the number of movements in a video is equal to the “number of laser spot positions” – 1. The implemented anomaly detection uses the matrix of KDE models to evaluate the likelihood of each individual movement in a test video. First, the laser spot positions are extracted as described in Section 2.2. Then, the destination position probability density of each movement is computed using the KDE model corresponding to the spatio-temporal region of the origin position. In other words, if a movement goes from \mathbf{x}_t in region d to \mathbf{x}_{t+1} , \mathbf{x}_{t+1} probability density is computed (see Equation 1) with the KDE model of the region d . The result is a vector of likelihoods. Each position of the likelihood vector represents a movement likelihood. The likelihood vector can be computed on line: the likelihood of a movement is calculated when the camera sends a new frame. This is important for speeding up the classification process. The likelihood vector can be used to detect anomalies with two different approaches: log-likelihood evaluation and log-likelihood distribution.

In log-likelihood evaluation, we compute the negative log-likelihood of the complete sequence of movements using the following expression: $NLL = -\sum_{i=1}^n \log p_i$, where p_i represents the likelihood of the i -th movement and n stands for the number of movements

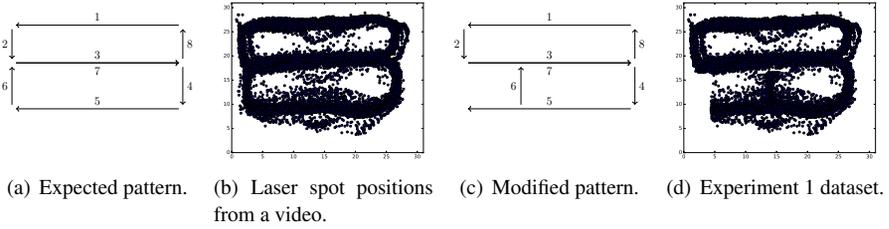


Figure 1. (a) Normal and (c) anomalous patterns. (b) Normal and (d) anomalous laser spot positions.

in the video. In order to evaluate the presence of anomalies, the negative log-likelihood is tested against a threshold T . The heat treatment is considered as anomalous if $NLL > T$ and as normal otherwise.

Another way to detect anomalies using the likelihood vector is to analyze the log-likelihood distribution. For each movement, we compute the negative log-likelihood of getting the negative log-likelihood vector: $\mathbf{NLLV} = (-\log p_1, \dots, -\log p_i, \dots)$. A probability distribution of the negative log-likelihood vector can be calculated using KDE over \mathbf{NLLV} . This probability distribution is calculated for each video. First, a reference distribution is calculated by combining multiple (in our work, 32) non-anomalous negative log-likelihood probability distributions. Multiple negative log-likelihood distributions can be combined in several manners. In this paper, we obtain the average distribution of all non-anomalous distributions. Then, we compare the reference distribution with the test negative log-likelihood distribution using the Kullback-Leibler (KL) divergence [10]. In order to evaluate the presence of anomalies, the KL divergence is tested against a threshold T as in log-likelihood evaluation.

4. Results

4.1. Simulating Anomalous Data

Because of the reliability of the process, there are no anomalous data. As we have not seen any real anomaly, we do not know exactly how it behaves. Therefore, the test was performed using simulated anomalous data obtained by modifying the normal datasets, using two types of geometrical anomalies:

- **Experiment 1:** The positions in segment 6 of Figure 1(a) have been moved to the right. The other positions were left unmodified. The objective was to simulate an incorrect figure. The result is shown in Figure 1(d). Around 1300 laser spot positions (6.2% of the total) were moved in each normal dataset.
- **Experiment 2:** Gaussian noise was introduced to each laser spot position found in the normal dataset. Different variances were used to test our method: 0.2, 0.1 and 0.05. The objective was to simulate anomalies in the laser positioning system.

4.2. Results and Discussion

All tests were conducted using $W = 413$ (around 50 frames on each temporal window), $V = 2$, $H = 2$. Table 1 shows the results of the log-likelihood evaluation using

Table 1. Results for the log-likelihood evaluation. A lower NLL is better.

	AVG _{Norm}	AVG _{Anom}	MAX _{Norm}	MIN _{Anom}	Gap	RATIO
Experiment 1	24531.63	29829.66	27580.59	27157.82	-422.77	0.98
Variance 0.2	24531.63	42764.28	27580.59	41428.01	13847.43	1.5
Variance 0.1	24531.63	36373.23	27580.59	34378.32	6797.74	1.25
Variance 0.05	24531.63	31453.18	27580.59	29093.37	1512.78	1.05

leaving-one-out cross-validation. For each experiment, we computed the average negative log-likelihood for both normal (AVG_{Norm}) and anomalous (AVG_{Anom}) datasets. In addition, we show the maximum negative log-likelihood (MAX_{Norm}) value obtained by a normal dataset and the minimum negative log-likelihood value of an anomalous dataset (MIN_{Anom}). The difference between MIN_{Anom} and MAX_{Norm} is called gap. All datasets can be correctly classified if the gap, using a threshold T such that $MAX_{Norm} \leq T < MIN_{Anom}$, is positive. The RATIO (MIN_{Anom}/MAX_{Norm}) indicates how many times greater MIN_{Anom} is than MAX_{Norm}. The RATIO is insensitive to the scale. Therefore, it is included to compare the log-likelihood evaluation and the log-likelihood distribution (see Table 2). RATIO > 1 means that all datasets can be correctly classified.

Using the log-likelihood evaluation criterion, we were able to correctly classify all the anomalous datasets in Experiment 2. We got the same result even with relatively low variance. For example, using the variance 0.05, we are introducing a variation lower than 0.67 in 99.7% of the positions. The gap in Experiment 1 is negative, so we could not classify all anomalous datasets correctly. Nevertheless, the average score is around 21% higher for the anomalous datasets although only one subset of the laser spot positions was changed.

Table 2 reports the results for the log-likelihood distribution. As opposed to the log-likelihood evaluation, all anomalous datasets can be correctly classified. The resulting ratios are considerably better than for log-likelihood evaluation. On the other hand, the log-likelihood distribution involves the construction of an additional KDE model and the computational cost of computing the KL divergence is higher. Variants of Experiment 1 were executed by moving only a subset of points in segment 6. The gap turns negative when the number of moved points is around 850 (almost 4% of the total). Therefore, the proposed approach can classify correctly every anomalous dataset with more than 850 positions moved.

As a comparison with state-of-the-art methods, two single multivariate KDE models were created (using all the positions of the whole region). One KDE model uses the positions of a normal video and the other uses the test positions. The KL divergence was computed using the train and test KDEs. The results (not included in the paper) show a similar performance in Experiment 1, but a poor behaviour in Experiment 2 (the gap is negative even with a variance of 1.05). That is because it does not take into account any temporal information. Furthermore, the computational cost of the single multivariate KDE models is much higher (it cannot be considered in-process).

Table 2. KL divergence using the log-likelihood distribution. A smaller KL divergence is better.

	AVG _{Norm}	AVG _{Anom}	MAX _{Norm}	MIN _{Anom}	Gap	RATIO
Experiment 1	3.05e-3	2.17e-2	9.05e-3	1.62e-2	7.15e-3	1.79
Variance 0.2	3.05e-3	1.80e-1	9.05e-3	1.49e-1	1.40e-1	16.44
Variance 0.1	3.05e-3	7.90e-2	9.05e-3	5.35e-2	4.45e-2	5.92
Variance 0.05	3.05e-3	3.04e-2	9.05e-3	1.24e-2	3.35e-3	1.37

5. Conclusions and Future Work

We developed a new approach to anomaly detection taking into account spatio-temporal information. The system uses multiple KDE models to output a score of abnormality using two different approaches: log-likelihood evaluation and log-likelihood distribution. Experimental tests using simulated anomalous data showed the sensitivity of the proposal even with sub-pixel variations in the laser spot positions. The classification can be completed in-process because part of the process can be executed on line. There are several potential future lines of research. First, new ways of setting the abnormality score might be investigated. The combination of a few anomalous with a large number of normal datasets with a view to building a more accurate model is another potentially interesting topic of research. Anomalous data should also be used for testing. The detection of the degradation of the expected pattern is another problem that could be explored by monitoring the evolution of the abnormality score.

References

- [1] V. Chandola, A. Banerjee, and V. Kumar. Anomaly detection: A survey. *ACM Computing Surveys*, 41(3):15:1–15:58, 2009.
- [2] M. Jäger, C. Knoll, and F. A. Hamprecht. Weakly supervised learning of a classifier for unusual event detection. *IEEE Transactions on Image Processing*, 17(9):1700–1708, 2008.
- [3] M. Jäger, S. Humbert, and F. A. Hamprecht. Sputter tracking for the automatic monitoring of industrial laser-welding processes. *IEEE Transactions on Industrial Electronics*, 55(5):2177–2184, 2008.
- [4] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing*, 50(2):174–188, 2002.
- [5] R. Chinmay, R. Asok, S. Soumik, and Y. Murat. Review and comparative evaluation of symbolic dynamic filtering for detection of anomaly patterns. *Signal, Image and Video Processing*, 3(2):101–114, 2009.
- [6] D. Yeung and C. Chow. Parzen-window network intrusion detectors. *Proceedings of the 16th International Conference on Pattern Recognition*, volume 4, 385–388, 2002.
- [7] L. Tarassenko, P. Hayton, N. Cerneaz, and M. Brady. Novelty detection for the identification of masses in mammograms. *Fourth International Conference on Artificial Neural Networks*, 442–447, 1995.
- [8] D. W. Scott. *Multivariate Density Estimation: Theory, Practice, and Visualization*. Wiley Series in Probability and Statistics, 2015.
- [9] Y. Yang and G. I. Webb. Discretization for naive-Bayes learning: Managing discretization bias and variance. *Machine Learning*, 74(1):39–74, 2008.
- [10] S. Kullback and R. A. Leibler. On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86, 1951.