



HAL
open science

Image Classification Applied to the Problem of Conformity Check in Industry

Nour Islam Mokhtari, Igor Jovančević, Hamdi Ben Abdallah, Jean-José Orteu

► **To cite this version:**

Nour Islam Mokhtari, Igor Jovančević, Hamdi Ben Abdallah, Jean-José Orteu. Image Classification Applied to the Problem of Conformity Check in Industry. SOCO 2022 - 17th International Conference on Soft Computing Models in Industrial and Environmental Applications, Sep 2022, Salamanque, Spain. p. 340-349, 10.1007/978-3-031-18050-7_33 . hal-03813066

HAL Id: hal-03813066

<https://imt-mines-albi.hal.science/hal-03813066>

Submitted on 14 Oct 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Image Classification Applied To The Problem Of Conformity Check In Industry

Nour Islam Mokhtari, Igor Jovančević, Hamdi Ben Abdallah and Jean-José Orteu

Abstract This paper shows the application of several learning-based image classification techniques to conformity check, which is a common problem in industrial visual inspection. The approaches are based on processing 2D images. First, a classification pipeline has been developed. An effort has been invested into choosing an appropriate classifier. First experiment was performed with *HoG* features (Histogram Of Gradient) and Support Vector Machine (*SVM*). Further, to improve accuracy, we employed a *bag of visual words* (*BoVW*) and *ORB* detector for extracting features that we further use to build our *dictionary of visual words*. The final solution uses features extracted by passing an image through a pre-trained deep convolutional neural network *Inception*. Using these features a *SVM* classifier was trained and high accuracy was obtained. To augment our image data set, different transformations such as zoom and shearing were applied. Promising results were obtained which shows that state-of-the-art deep learning classification techniques can be successfully employed in the visual industrial inspection field.

Nour Islam Mokhtari
DIOTASOFT, 201 Pierre and Marie Curie Street, 31670 Labège, France

Igor Jovančević
DIOTASOFT, 201 Pierre and Marie Curie Street, 31670 Labège, France
Faculty of Natural Sciences and Mathematics, University of Montenegro, Cetinjska 2, 81000
Podgorica, Montenegro
e-mail: igorjovan@gmail.com

Hamdi Ben Abdallah and Jean-José Orteu
Institut Clément Ader (ICA) ; Université de Toulouse ; CNRS, IMT Mines Albi, INSA, UPS, ISAE
; Campus Jarlard, 81013 Albi, France

1 Introduction

Visual inspection of mechanical assemblies is an essential phase in the production process. The objective of this paper is to automate the inspection procedures by developing computer vision algorithms able to perform this task. The developed algorithms are exploiting 2D images and 3D point clouds and are running on a robotic platform or on a hand-held tablet. The Computer-Aided Design (CAD) model of the inspected mechanical assemblies is utilized as an input for in-house developed CAD-based 2D/3D localization module. This module provides a close enough estimate of the relative pose between the camera and the assembly being controlled. The pose estimate is a requirement imposed by the algorithms to properly observe an element of interest.

Treated inspection problem belongs to a large group of problems. The challenge is to verify the presence of certain elements at their pre-specified locations in an assembly. To validate the approach, four real use cases are solved. They concern metallic parts in the aerospace manufacturing domain and verifying their presence at predefined locations.

2 Related work

In the previous papers, the authors are dealing with visual inspection challenges in the aerospace industry, by employing conventional image processing [1, 2, 3] and 3D point cloud processing techniques [4, 5] or recent deep learning architectures on 3D point clouds [6]. In this work, a way has been found to provide enough 2D images in order to evaluate some known machine learning approaches and some deep learning architectures.

Inspection based on 2D image analysis has been in the focus of many works. Generally, the use of 2D image analysis for defect detection is preferred over 3D point clouds, because of its lower computation time. Leiva *et al.* [7] tackled a visual detection and verification of exterior aircraft elements. Authors in [8] developed a fully automated vision system to inspect the surface of crankshafts.

In vision-based inspection, machine learning models that perform classification and object detection can be very useful bricks. Park *et al.* [9] present a convolutional neural network (CNN) to detect dirties, scratches, burrs, and wears. Miranda *et al.* [10] presented a CNN-based method to detect and inspect screws on aircraft fuselage images acquired by a UAV. Jong-Chih *et al.* [11] developed a machine-learning method to classify 4 types of visible surface defects on semiconductor wafers.

In most cases, there are not enough images to train the models because assemblies are rarely available for data acquisition. This paper will present an approach to employ classification while having little data.

3 Methodology and results

Following sections will demonstrate the experiments with multiple solutions and gradual development of the image classification pipeline to verify the presence of objects mounted on a mechanical assembly. It will also show how a small dataset has been augmented.

3.1 Problem statement - verifying presence by classification

The aim is to verify that a screw is present (as in Fig. 1a) in a predefined location on a mechanical assembly. The localization module enables cropping an input image and extracting a region of interest around the predefined location of the screw. Fig. 1 shows 3 samples that the classification algorithm should be able to differentiate. The methodology and the results of three different evaluated approaches will be outlined. The gradual improvement in the results will be demonstrated. The approaches have been applied to six different tasks but here only four of them will be presented.

3.2 SVM with HoG

Gaussian Support Vector Machine (SVM) with Histogram of Gradients (HoG) descriptor has been a popular solution to similar tasks [12, 13]. So this approach has been employed first. C and γ are parameters that can control the SVM decision boundary. To choose the best parameters C and γ for the classifier, *cross-validation* has been used. Namely, our dataset has been split into 2 subsets, 80% for training and 20% for testing. Then the classifier has been trained on the larger subset with the chosen parameters and then tested on the smaller set. This way, the parameters C and γ that maximize both the training accuracy and the testing accuracy have been chosen. Also, this way the over-fitting problem has been addressed, since the trained classifier has been tested on a subset that was not seen before. This will subsequently give an idea on how well the classifier will perform on new data in the future. The tests are listed in Tables 1 and 2. It can be noted that the results are not satisfactory.

3.3 SVM with BoVW

Further on, HoG was replaced by *bag of visual words* (BoVW) which is a feature representation technique that uses a partly unsupervised machine learning scheme based on *k-means* algorithm. First step is to construct a *visual dictionary* of BoVW descriptors in the points that were previously extracted by the ORB detector. The cross-validation approach has been used, as in section 3.2, to choose the optimal

SVM params	Accuracy (for train set and test set)	
$C = 0.1$	train set = 48.62%	test set = 56.52%
$C = 0.5$	train set = 50.00%	test set = 50.00%
$C = 1.0$	train set = 100.0%	test set = 50.00%
$C=2.0$	train set = 100.0%	test set = 60.00%
$C = 4.0$	train set = 100.0%	test set = 43.48%
$C = 8.0$	train set = 100.0%	test set = 45.65%
$C = 16$	train set = 100.0%	test set = 50.00%

Table 1: Different tests using SVM with HoG (C is variable, γ is fixed)

SVM params	Accuracy (for train set and test set)	
$\gamma = 0.1$	train set = 100%	test set = 58.69%
$\gamma = 0.5$	train set = 100%	test set = 56.52%
$\gamma = 1.0$	train set = 100%	test set = 50.00%
$\gamma = 2.0$	train set = 100%	test set = 52.17%
$\gamma = 4.0$	train set = 100%	test set = 41.30%
$\gamma = 8.0$	train set = 100%	test set = 32.61%
$\gamma = 16$	train set = 100%	test set = 54.35%

Table 2: Different tests using SVM with HoG (γ is variable, C is fixed)

parameters. Main focus was put on two parameters that have the most influence on the results: C for SVM and K for the size of the visual words dictionary. In tables 3 and 4 a significant improvement in accuracy can be observed, compared to SVM+HoG approach.

k-means	Accuracy (80% train; 20% test)	
$K = 10$	train set=83.88%	test set=82.97%
$K=30$	train set=86.71%	test set= 90.33%
$K = 50$	train set=88.56%	test set=87.81%
$K = 60$	train set=86.16%	test set=85.43%
$K = 80$	train set=88.55%	test set=86.81%
$K = 100$	train set=88.41%	test set=87.36%

Table 3: Accuracy when changing K and fixing SVM parameters $\gamma = 1$ and $C = 1$

SVM	Accuracy (80% train; 20% test)	
$C = 1$	train set=88.56%	test set=87.81%
$C = 2$	train set=90.34%	test set=90.91%
$C = 4$	train set=91.91%	test set=89.83%
$C = 10$	train set=92.19%	test set=92.01%
$C=20$	train set=92.67%	test set= 92.54%
$C = 30$	train set=4.03%	test set=90.93%

Table 4: Accuracy when changing C and fixing $K = 50$, $\max_{\text{iter}} = 2000$, $\epsilon = 0.01$ and $\gamma = 1$

The 2 parameters \max_{iter} and ϵ work as a stopping criteria for the SVM algorithm. The SVM optimization will stop if one of the 2 criteria is reached: either the loss function value became lower than ϵ or a maximum number of iterations (\max_{iter}) was reached.

In the experiments with BoVW, the reached accuracy was around 92.54% on the augmented data set. This means that the system is still making false predictions 7.5% of the time, which is unacceptable, by industry standards. This led us to using a more informative descriptor such as deep learning architecture which will be presented in the next section.

3.4 Linear SVM with CNN features

The final solution is based on deep learning feature extraction and SVM classification. For this approach, more data are required, hence they needed to be generated. In industry, there are usually two large problems: the lack of data and the classes being imbalanced. To tackle these two challenges, a data augmentation scheme was included, followed by some manual data cleaning. Starting from a dataset of 381 images, a dataset of 1826 images was obtained. This augmented data set was distributed by class as follows: 672 images that contain a screw (Fig. 1a), 673 images that do not contain a screw (Fig. 1b) and 517 images where it is not possible to decide (Fig. 1c). The affine transformations were applied to increase the dataset. Those transformations were mainly a combination of zooming (to account for scale), skewing (to account for different camera poses), and cropping. Three examples of augmented data are shown in Fig. 2.

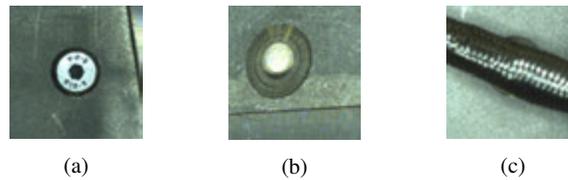


Fig. 1: Data before augmentation: (a) screw is present, (b) screw is absent, (c) unknown if screw is present

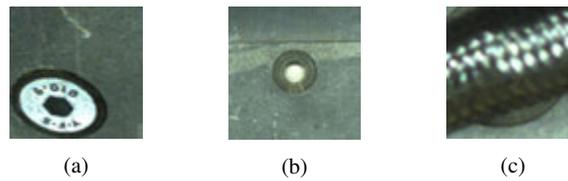


Fig. 2: Data after augmentation: (a) screw is present (skew, zoom, crop), (b) screw is absent (rotation, skew), (c) unknown if screw is present (zoom)

In paper [14], authors reported that, by using pre-trained networks such as VGG and Inception to extract features from images and then applying some traditional classifiers such as SVM, they were able to achieve the state of the art results in both classification and object detection. This approach is appropriate for the case treated in this paper, especially because not a lot of data is available for training a CNN from scratch, even with data augmentation. The working model of the proposed approach is demonstrated in Fig. 3.

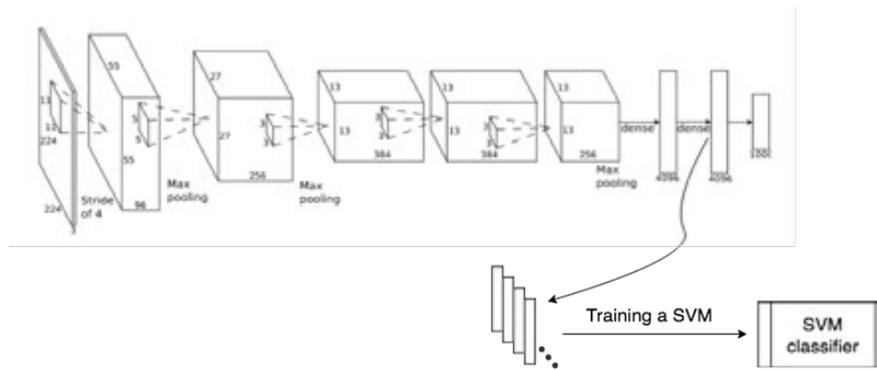


Fig. 3: Our scheme at training time

First, the images are passed through a pre-trained network such as Inception or VGG. Further, for each image, the output of the bottleneck layer is extracted, also called the transfer layer (referring to transfer learning). It has the size of 2048 in the Inception network. Finally, feature vectors from all the images are collected and used for training an SVM classifier.

Performed operations at test time are shown in Fig. 4. First, the test image is passed through the network, then the transfer layer output is extracted and passed through the trained SVM classifier. Finally the corresponding prediction is obtained.

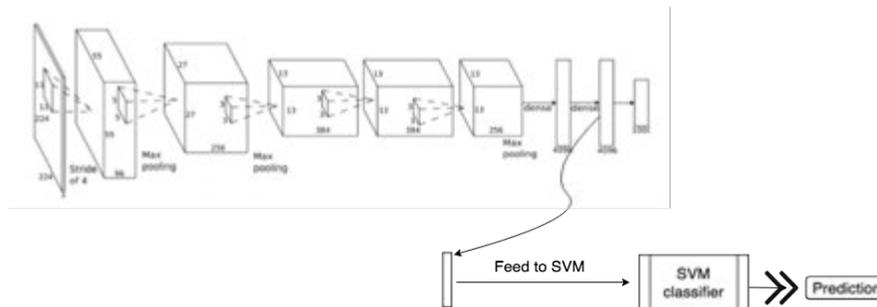


Fig. 4: Scheme at test time

To test this approach, a linear SVM with the extracted features is used. The results are shown in Table 5. It can be noted that the classification accuracies are very high. This is the best classifier used and the efficiency of this classifier comes first and foremost from the pre-trained CNN *Inception v1*. The reason why the extracted features are very discriminative is because the network has been trained on millions of images from ImageNet data set [15] and it has learned to extract the most relevant parts from the image; most relevant in the sense of distinguishing between different

SVM	Accuracy
$C = 1$	training=100%, testing=99.73%
$C = 5$	training=100%, testing=100%

Table 5: Inception v1 and linear SVM

classes. Developed pipeline was evaluated on five other and even more challenging classification tasks and some very good results were obtained.

In Fig. 5 another inspection use case with 5 different classes can be observed. In this task, the objective is to build a system that can distinguish between 5 different types of screws that can be found in some airframes.

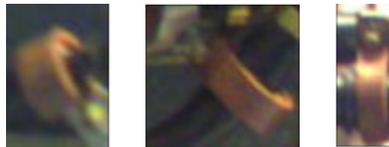


Fig. 5: Small support screws samples

Dataset contains 1227 images: 137 in class1, 116 in class2, 395 in class3, 310 in class4 and 269 in class5. The dataset has been split in an 80%/20% manner. Obtained accuracies are very high: 100% on the train set and 94.5% on the test set.

The same approach that consists of a linear SVM with InceptionV1 extractor was used in two other tasks that are described below.

Task 1: The goal is to develop a system that can recognize one of three possible scenarios : a red clump exists (Fig. 6), no red clump exists (Fig. 7) and unknown state (the clump might be hidden by another part) (Fig. 8).

Fig. 6: Class 1: presence of red clumps (*valid* state)

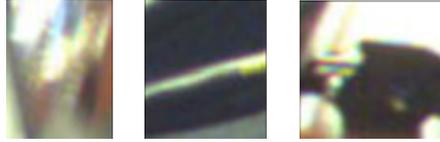


Fig. 7: Class2: absence of clumps (*invalid state*)



Fig. 8: Class 3: *Unknown state* (decision can not be made, often acquisition or localization problem)

Dataset contains 712 images distributed among 3 classes: 234 *valid state*, 164 *invalid state*, 37 *unknown state*). The set has been split as previous sets: 80%/20%. Obtained accuracy on training set was 100% and on testing set: 98.03%.

Task 2: the goal is to be able to distinguish between three different states on a mechanical engine. The first state: the metallic box is mounted on the engine (Fig. 9). The second state: the box is absent (Fig. 10). The third state: it is not possible even for a human to decide whether the box is mounted or not, for example when it is hidden by other parts of the engine (Fig. 11).

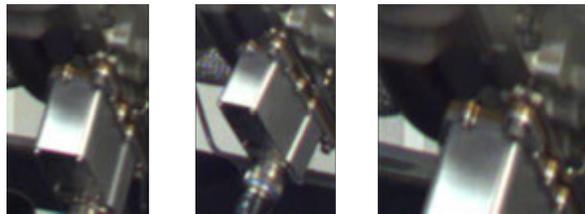


Fig. 9: Class 1: presence of the metallic box (*valid state*)

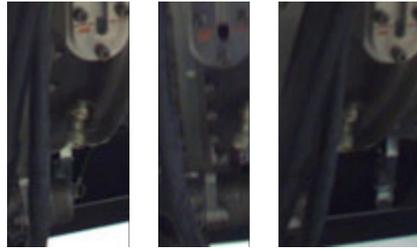


Fig. 10: Class 2: absence of the metallic box (invalid state)

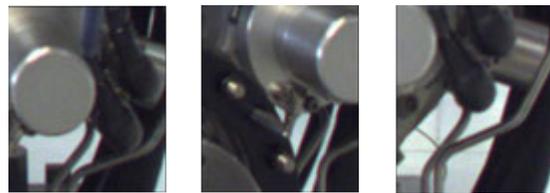


Fig. 11: Class 3: unknown state (decision can not be made, often acquisition or localization problem)

Dataset consisted of 1022 images, distributed among 3 classes: 633 images where the box is present, 201 images with no box and 188 images with unknown state. After splitting the dataset 80%/20%, the following accuracies were reached: 100% on training set, and 98.03% on testing set.

4 Conclusion

This work applied state-of-the-art machine learning to a known industrial visual inspection problem: conformity check. The main tool is classification for 2D images. Developed classification pipeline consists of a feature extractor and a supervised machine learning scheme, i.e. SVM classifier. It is designed to solve the problem of verifying the presence of different mechanical parts. Different feature extractors were evaluated and it has been found that a CNN architecture outperformed traditional feature extractors: HoG and BoVW. Namely, a pre-trained neural network *Inception v1* was employed to extract feature vectors from images. Further, these feature vectors were used to train a linear SVM. This approach generated very high accuracies. This work shows that machine learning can be employed for industrial applications even with small size datasets. The experiments that were done can easily be extended to other industries such as manufacturing or automotive industry.

Acknowledgements This research has been carried out within the joint research laboratory "Inspection 4.0" between IMT Mines Albi/ICA and DIOTA. This laboratory is co-funded by the French "Région Occitanie". We would like to thank the Région Occitanie for its financial support.

References

1. Jovancevic, I., Larnier, S., Orteu, J.-J., and Sentenac, T., "Automated exterior inspection of an aircraft with a pan-tilt-zoom camera mounted on a mobile robot," *Journal of Electronic Imaging* **24**(6), 1–15 (2015).
2. Ben Abdallah, H., Jovancevic, I., Orteu, J.-J., and Brèthes, L., "Automatic Inspection of Aeronautical Mechanical Assemblies by Matching the 3D CAD Model and Real 2D Images," *Journal of Imaging* **5**, 81–108 (10 2019).
3. Viana, I., Orteu, J.-J., Cornille, N., and Bugarin, F., "Inspection of aeronautical mechanical parts with a pan-tilt-zoom camera: An approach guided by the computer-aided design model," *Journal of Electronic Imaging* **24**, 061118 (12 2015).
4. Jovancevic, I., Pham, H.-H., Orteu, J.-J., Gilblas, R., Harvent, J., Maurice, X., and Brèthes, L., "3D point cloud analysis for detection and characterization of defects on airplane exterior surface," *Journal of Non Destructive Evaluation* **36**, 74 (12 2017).
5. Ben Abdallah, H., Orteu, J.-J., Jovancevic, I., Brèthes, L., and Dolives, B., "3D point cloud analysis for automatic inspection of complex aeronautical mechanical assemblies," *Journal of Electronic Imaging* **29**(04), 1–22 (2020).
6. Mikhailov, I., Jovancevic, I., Mokhtari, N., and Orteu, J.-J., "Classification using a three-dimensional sensor in a structured industrial environment," *Journal of Electronic Imaging* **29**(4), 1–14 (2020).
7. Leiva, J., Villemot, T., Danguomeau, G., Bauda, M.-A., and Larnier, S., "Automatic visual detection and verification of exterior aircraft elements," in [2017 IEEE International Workshop of Electronics, Control, Measurement, Signals and their Application to Mechatronics (ECMSM)], 1–5, IEEE (05 2017).
8. Tout, K., Bouabdellah, M., Cudel, C., and Urban, J.-P., "Automated vision system for crankshaft inspection using deep learning approaches," in [Fourteenth International Conference on Quality Control by Artificial Vision], 78, SPIE (07 2019).
9. Park, J.-K., Kwon, B.-K., Park, J.-H., and Kang, D.-J., "Machine learning-based imaging system for surface defect inspection," *International Journal of Precision Engineering and Manufacturing-Green Technology* **3**, 303–310 (July 2016).
10. Miranda, J., Veith, J., Larnier, S., Herbulot, A., and Devy, M., "Machine learning approaches for defect classification on aircraft fuselage images acquired by an uav," in [Fourteenth International Conference on Quality Control by Artificial Vision], 10, SPIE (07 2019).
11. Chien, J.-C., Wu, M.-T., and Lee, J.-D., "Inspection and classification of semiconductor wafer surface defects using cnn deep learning networks," *Applied Sciences* **10**(15) (2020).
12. Navneet, D. and Bill, T., "Histograms of Oriented Gradients for Human Detection," in [International Conference on Computer Vision & Pattern Recognition (CVPR '05)], Schmid, C., Soatto, S., and Tomasi, C., eds., **1**, 886–893, IEEE Computer Society, San Diego, United States (June 2005).
13. Xiao, J., Hays, J., Ehinger, K., Oliva, A., and Torralba, A., "Sun database: Large-scale scene recognition from abbey to zoo," in [2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition], 3485–3492, IEEE (06 2010).
14. Razavian, A., Azizpour, H., Sullivan, J., and Carlsson, S., "Cnn features off-the-shelf: An astounding baseline for recognition," in [2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops], **1403**, 512–519, IEEE (03 2014).
15. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A., and Fei-Fei, L., "Imagenet large scale visual recognition challenge," *International Journal of Computer Vision* **115** (09 2014).