Random Forest Based Diagnosis Approach for Rail Fault Inspection in Railways

Yunus Santur[1], Mehmet Karaköse[1], and Erhan Akın[1]

[1]Firat University, Computer Engineering, Elazığ, Turkey
{ysantur, mkarakose, eakin}@firat.edu.tr

Abstract

Railway systems are one of the most preferred transport means worldwide. Some faults may occur on railway tracks due to several reasons and such faults may cause accidents. For this reason, railway tracks should be periodically inspected. In this study, a computer vision based approach was proposed for inspecting the faults in railway tracks. It was aimed to inspect the faults which may occur on rail surfaces such as scouring, breaking, and deficient fasteners such as bolts and sleepers with the experimental study presented. In this study, feature extraction was performed on a video image containing especially a healthy railway track. Then, feature extraction was again performed on the image containing healthy railway track by generating virtual faults, and these two data sets were labelled as faulty and healthy, and trained. Algorithm was applied on a video image including faulty and healthy frames, operating time and accuracy performance was measured and a decision making mechanism was established during test phase.

1. Introduction

Computer vision is commonly used in areas such as robotics, industrial, geographical information systems, medical and remote sensing. It can be said that a computer vision system consists of an image sensor, an application algorithm and a database [1]. Computer vision is commonly used in industrial applications for the determination of deficiencies in the product due to the advantages like time and cost saving [2]. Detection of deficiencies in agricultural products such as vegetables and fruits and assessment of their quality [3], real-time determination of deficiencies in textile products [4] and real-time determination of deficiencies in electronic circuits [5] can be given as examples in this area.

Railway is one of the most important transport means in Turkey and the world since it is safe, fast, easily managed, less harmful to the environment and has a high carrying capacity. High-speed trains can reach to 350 km/h due to widespread of railways. Deficiencies occur in railways may result in accidents and material damages. Therefore, rails should be periodically monitored, their faults should be detected and maintenance should be carried out.

The simplest inspection method is manual and visual inspection with human labor but such type of inspection is slow, unsafe and most of all it remains limited with the knowledge of the competent person [6]. As it is shown in Figure-1, in this method, the expert performs inspection via rail profile measurement devices named as Robel-A, Robel-B and SKM along the rail by walking or driving [7].

Fig. 1. Manual rail measurement [7]

A similar method is the inspection of the rails via mechanical devices. This method diagnoses faults by using graphs obtained from the friction between mechanical device and rail, it is fast and provides accurate results but it may damage the rails or increase the existing damage on the rails due to the requirement of contacting with the rail [8]. Due to the damages caused by contact-based methods performed under the supervision of an expert or with mechanical devices, rail analysis has been conducted with high accuracy and quickly with computer vision based technologies using camera or laser recently [9-19]. In Figure-2, a computer vision based rail profile study using computer, light source, encoder and camera is given.

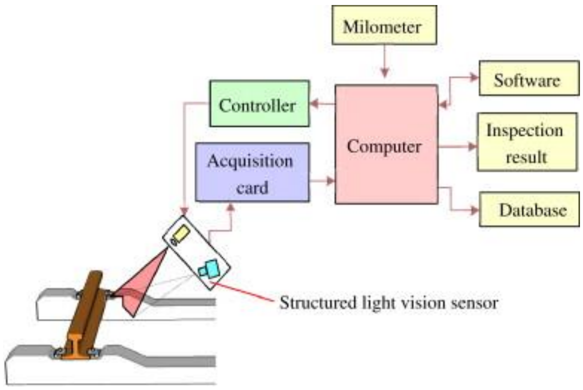

Fig. 2. Computer vision based rail analysis system [7]

There are some disadvantages of computer vision based fault diagnosis method; railway tracks do not have a uniform geometrical characteristics and texture, and the light is insufficient. As it is seen in study No:[10], an external light source is used due to insufficient light in computer vision based rail analysis applications. In real-time rail analysis applications, detection of the location of the fault precisely is as important as the detection of the fault. For this purpose, encoders with precision are used in the system. High resolution camera and fgpa/computer units on which the application algorithm operates constitute the other components of the system.

In the study conducted by Li et.al [9], a real-time system based on image processing for the detection of faults such as punctures on rail surfaces was proposed. In the study, wear and puncture type rail surface faults were divided into two classes according to the extent of the fault, and the system was tested in accordance with operating speed and accuracy performance criteria.

With the high accuracy and fast function of the method, dust and grease residues on rail tracks may be detected as fault. Same situation is also the common disadvantage of computer based fault diagnosis methods [9, 10].

For this reason, usage of laser cameras for fault detection in railway tracks has been expanding recently [11, 12]. In the study conducted by Aytekin et.al [11], a system using laser camera for deficient bolt detection on railway tracks was proposed. The system was trained by using totally seven dimension reduction feature extraction and classification method and then performance assessment was carried out based on test images. The highest accuracy values among applied methods were obtained from principal component analysis and random forests classification algorithm.

In a study conducted by Leszek et.al [12], a method for real-time detection of carbon pantograph wears via fpga and laser cameras for electrical rail devices was proposed. In this proposed method, pantographs which were detected and measured via laser camera were classified as newly worn and highly worn by using the center of gravity of bright areas as feature extraction. A similar study were presented in [13] for arc detection in pantograph systems.

Real-time systems operating with high accuracy are needed in the analysis of railway tracks. Image processing methods such as template matching [11] remain insufficient in terms of high accuracy and real time operation requirements for rail analysis. Therefore, feature extraction methods seem to be used commonly in railway analysis in the literature. General block diagram of a system which can be used for this purpose is given in Figure 3.

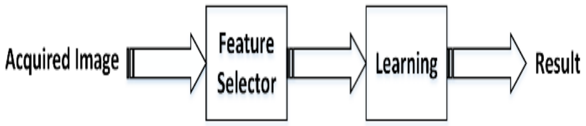

Fig. 3. Block diagram of an inspection system

In this research, an experimental study was conducted by using feature extraction methods for the inspection of fault types which may occur on railway tracks on the basis of the researches in literature. In these experimental studies, a decision making mechanism was established by measuring operation times and accuracy performances of feature extraction methods on faulty and healthy frames.

2. Railway Analysis

The section on which railway vehicles move via wheels and which constitutes the most important component of the railway analysis in terms of transportation safety is named as *railway track*.

Some sections constituting the railway track are given in Figure 4 [14]. Structural material transmitting the load to the ballast on the railway track is named as *sleeper*. Sleepers which are mostly wooden may be made of concrete, iron or steel. Track structural material consists of gravels, distributing the load coming from sleeper homogeneously and supporting the sleeper is named as *ballast*.

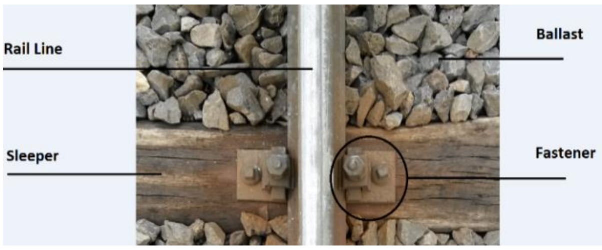

Fig. 4. Railway track components

Finally, elements in different geometric types such as hexagonal, circular, etc. which are at both sides of the railway track and provide connection between railway track and sleeper are called as *fastener* [14, 15].

Studies in the literature on railway analysis generally focus on the detection and/or classification of deficiencies in such fasteners and faults with high accuracy by using computer vision.

Fault types which may occur on railway upper surfaces in the form of scouring, wear, cracks and etc. are given in Figure 5 [16]. In this study conducted by R. Huber-Mörk et.al, training was performed by using feature extraction and statistical based methods and aforementioned rail surface faults were classified according to their types.

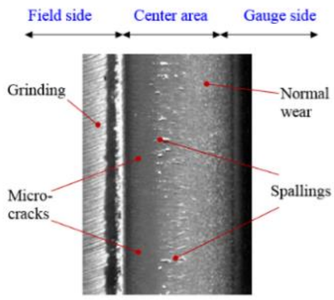

Fig. 5. Fault types which may occur on rail surfaces [16]

2.1. Rail Faults

Fault types which may occur on rail surfaces can be classified in four groups; *Rail Breakages*, *Rail Scouring*, *Corrugation Faults* and *Headcheck Faults*. In following section, information including the definition and potential causes of these fault types will be given [15].

Rail Breakages: The gaps on the rail surface which are longer than 50 mm and deeper than 10 mm are named as breakage. Rail breakages can be formed as a result of manufacturing defects, welding defects, operation of trains at a speed exceeding the specified limits, sleeper defects and excess load.

Rail Scouring: Fault types on the rail surfaces which are shorter than 60 mm and have a depth lower than 1 mm are named as scouring. Locomotive skidding, starting and stopping of the train can cause scouring type faults or can increase existing scours.

Corrugation Faults: Corrugated wears in different sizes occur on rail surfaces are named as corrugation. Deficiencies and defects in rail sleeper connections, excess or unbalanced loads, presence of rail depressions can cause such type of faults.

Headcheck Faults: Hairtrack cracks occur on rail surfaces are named as headcheck. Such type of faults may occur during starting and breaking of the railway vehicles.

3. Proposed Method

In this study, a computer vision based method for the inspection of faults which may occur on railway tracks was proposed. It was aimed to detect scouring type faults and deficiencies of fasteners such as bolts and sleepers with the method proposed. General block diagram of the system is given in Figure 6.

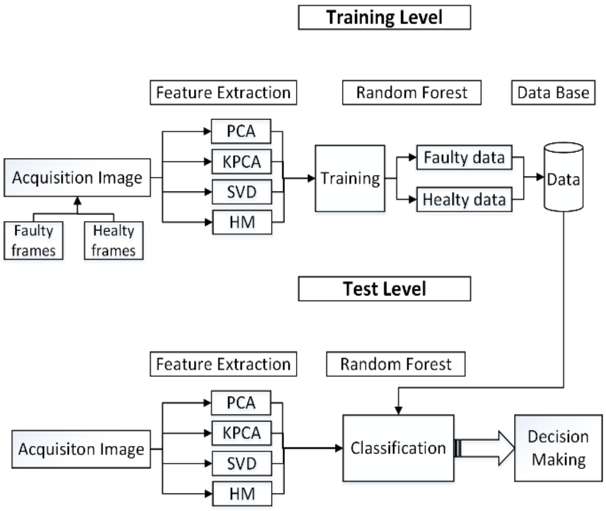

Fig. 6. Block diagram of the proposed system

A video image containing healthy railway track was used in the study. Application consists of two phases; training and testing.

Eigenvalues were obtained by using the method of feature extraction from each frame coming from video image containing healthy railway track image during training phase [17]. Obtained data set were labelled as healthy and recorded in database in order to be used in training algorithm. Then, virtual fault was created on the same image and eigenvalues were obtained by feature extracting from each frame from the video flow containing fault, and this data set was labelled as faulty and recorded in database in order to be used in training algorithm.

Obtained data from these two classes were subjected to training with random forest (RF) algorithm [20] and decision forest was created. High accuracy and rapid functioning of RF method on large data is the biggest reason of selecting this method.

A video image not previously used in the system and containing frames with no fault was used during test phase. Eigenvalue obtaining procedure is repeated on the frames obtained from video flow during test phase; obtained value is compared within the decision forest; the closest class is determined and it is found whether a fault exists within the frame.

In the system, principal component analysis (PCA), kernel principal component analysis (KPCA), Singular Value Decomposition (SVD) and Histogram Match (HM) feature extraction methods were used [17].

Video images containing 250x480 frames were used in the method proposed. Four different scenarios to be in different sections of the image and to have different sizes were realized during the creation of the anomaly. Methods were compared by measuring the accuracy performance and real time operation times on these scenarios.

3.1. Feature Extraction

Representing a large data in a smaller sized space containing eigenvalues is named as dimension reduction or feature extraction [17]. According to equation 1, obtaining D eigenvalues at X mxn size for input data with W transformation matrix can be named as dimension reduction or feature extraction. In general, the first element within the eigenvalue vector represents the most significant eigenvalue, and significance level within the vector decreases.

$$D = WX \tag{1}$$

Processing the raw data constitutes disadvantage in terms of both speed and memory consumption. Therefore, eigenvalue obtaining method by using dimension reduction methods and feature extraction are commonly used in computer vision applications requiring real-time operation [2-12]. Feature extraction methods used in the study are described below.

PCA: The method of expressing p pieces of data having correlation in terms of k pieces of data (p>k) having no correlation is known as PCA. The first eigenvalue obtained after PCA transformation is the value that defines the data set best. It was selected in this study due to its high accuracy and rapid function. PCA method is commonly used in applications such as face recognition, object recognition and fastener detection in railway track [11, 21].

KPCA: It is very similar to PCA method. A kernel function is used instead of a transformation matrix during transformation in KPCA method. It has been used in fault diagnosis with PCA method in the literature [22].

SVD: It is a tracker algebra-based method which is used as a basic step in some dimension reduction methods like PCA. It was selected during the study since it functions faster than PCA method. It is especially used in web page ranking in the literature [due to its rapid function 23].

HM: In this method, histogram mean and standard deviation value are selected as feature. In the study conducted by Aytekin et.al [11], Histogram comparison method provided faster and high accuracy results for the determination of deficient fasteners on the right and left sides of the railway track.

3.2. Random Forest

RF is a classification method using multiple decision trees with multiple classifiers. These decision trees are created by randomly selecting variables from data set and then optimum tree is tried to be selected by ranking.

It is one of the most suitable classification algorithms to be used in fault diagnosis since it functions rapidly even in large data sets, provides good results on lost data and generates high accuracy results [20]. There is no need to prune the trees while creating random forest and it is resistant to overfitting problem. High memory consumption depending on the size of the data set is the biggest disadvantage of this method.

During the creation of random forest, selecting the number of trees, variables having no correlation and to be used in training and the number of variables are very important and this operation is generally performed experimentally. In this study, we use downscaled and vectorized rgb images as input features and the number of trees for RF was selected as 10.

4. Experimental Results

Experimental study consists of two phases such as training and testing. Variables obtained by feature extraction from faulty and healthy frames coming from the video flow were used in RF method training algorithm and classified during training phase. In testing phase, frames coming from the video flow were scanned in decision environment which was created with RF and they are diagnosed for being faulty. At the end of the study, accuracy performances and operating times of algorithms were compared by creating different scenarios.

Fig. 7. Frames obtained from the video images which were used during training phase

Frames obtained from the video image can be seen in Figure 7. Faulty frames in training algorithm were obtained by crating virtual fault on the frames within the video image. Size of the fault created in the first scenario was selected to be 0.2% of the total size of the frame having a size of 250x480 pixels. Accuracy performances of selected methods were obtained with class labels.

The frames constituting the first 50% of the video image contain anomaly and remaining 50% do not contain anomaly in the experimental studies. Class label "1" was selected for correct classification and "0" for incorrect classification were selected for the results obtained from PCA, KPCA, SVD and HM. Feature values and class labels obtained with PCA, KPCA, SVD and HM for a sample test frame are given in Table 1. 3 values for all methods except HM and 2 values for HM method were used.

Table 1. Eigenvalues and class labels obtained from frames (for 1 frame)

	PCA	KPCA	SVD	HM
Feature Values	0.07 0.01 -0.0087	-0.4 -0.07 -0.1	0.001 -0.001 -0.005	468 357
RF Label	1	1	0	1

Table 2. Study conducted by selecting the fault size as 0.2% and with a test data of 1 second

		PCA	KPCA	HM	SVD
For 1 frame (sn)	Time of feature extraction	0.02	0.03	0.03	0.03
	Time of training	0.003	0.003	0.01	0.003
	Time of test	0.14	0.14	0.12	0.04
Accuracy (%)		80	60	60	60

In test results given in Table 2, the highest accuracy were obtained with PCA method and the fastest operating time was achieved with SVD. While HM provided good results for frames with deficient sleeper, it did not provide good results throughout the application.

In all methods, the number of trees for RF was selected as 10. Again, in all methods, the number of features extracted from frames was selected as 3 and 2 for HM method. In Figure 8, test results obtained for 12 frames coming from the video flow were given one under the other for the four methods applied. It was seen that in results obtained from PCA frames No. 6 and 9 were incorrectly classified and the remaining 10 frames were correctly classified.

While KPCA method gave lower accuracy than PCA, the frame No. 6 which was incorrectly classified in PCA method was correctly classified here. As it is seen in Table 3, SVD and HM methods have the lowest accuracy values; they correctly classify most part of the frames with anomaly.

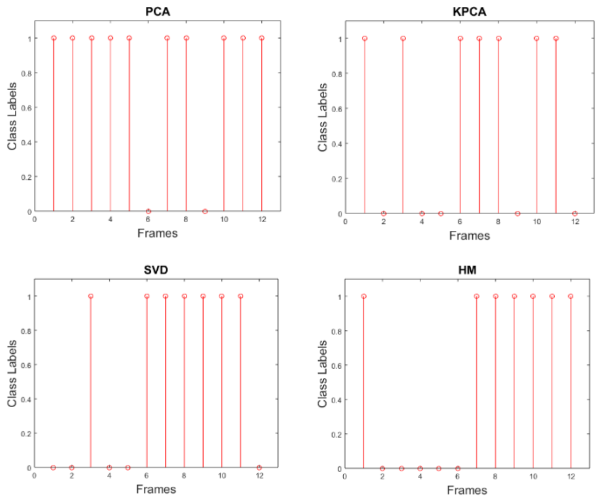

Fig.8. Accuracy comparison of the methods (for 12 frames)

Table 3. Test results accuracy table and correct classification of the methods (for 12 frames)

Frame No	1	2	3	4	5	6	7	8	9	10	11	12
Faulty ?	n	n	n	n	n	n	y	y	y	y	y	y
PCA	+	+	+	+	+	-	+	+	-	+	+	+
KPCA	+	-	+	-	-	+	+	+	-	+	+	-
SVD	-	-	+	-	-	+	+	+	+	+	+	-
HM	+	-	-	-	-	-	+	+	+	+	+	+

The study was repeated in 2 scenarios realized by setting the anomaly size as 0.2% on the same video and by changing the location of the anomaly. On the frames, depending on the frame background and the feature of the area with anomaly and opacity of the color and the frame, it was observed that accuracy performance in PCA and KPCA methods may slightly change while it does not change in HM and SVD methods.

The video test time was increased up to 4 seconds in 3rd scenario realized. Accuracy ratio in feature extraction operation decreased due to physical shape and location differences of fasteners like sleepers and bolts from frame to frame along with the extension of video time.

Accuracy performance of the methods for a test time of 4 seconds is given in Table 4. Classification graphs obtained for 48 frames can be seen in Figure 9. HM method classified all faulty frames correctly.

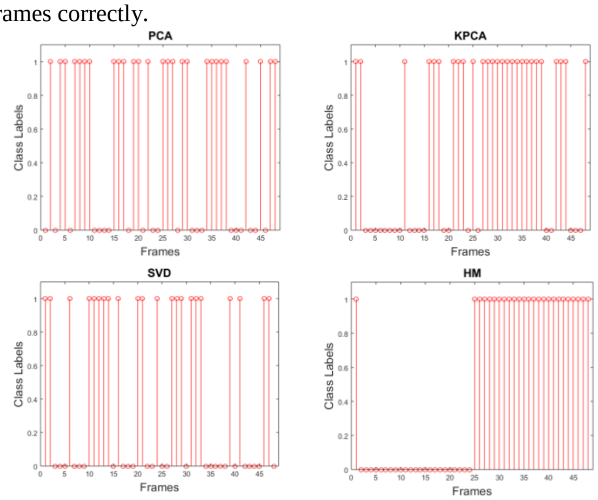

Fig. 9. Accuracy comparison of the methods (for 48 frames)

Table 4. Study conducted by selecting the fault size as 0.2% and with a test data of 4 seconds

Accuracy	Faulty	Healthy	Total
Frames (%)	50	50	100
PCA (%)	40	80	60
KPCA (%)	35	75	55
SVD (%)	38	54	46
HM (%)	98	4	52

In the scenario realized, virtual fault was created to set the number of anomalies in 4th scenario maximum 3 and total anomaly size as maximum 0.5%, and the results were compared with a 20-second test video. With the extension of the time of the video used for the test, accuracy performances in PCA and KPCA methods partially decreased. HM method classified all faulty frames with a 50% of general accuracy performance. Accuracy comparisons of the methods are given in Table 10 and performance comparisons can be found in Table 5.

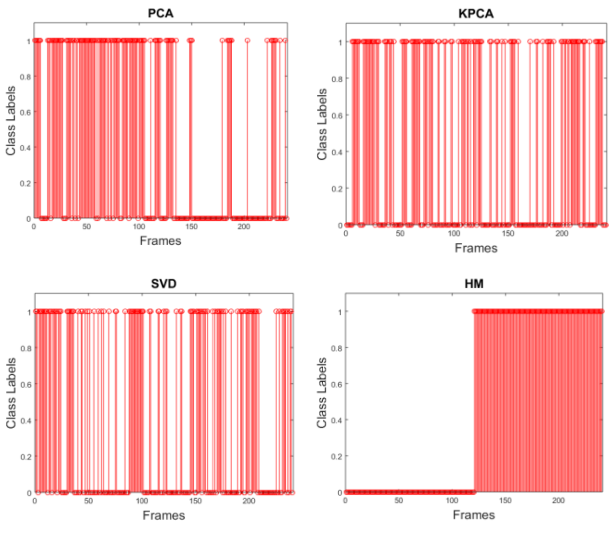

Fig. 10. Accuracy comparisons of the methods (for 240 frames)

Table 5. Study conducted by selecting the number of faults as 2, the fault size as 0.5% and with a test data of 20 seconds

Accuracy	Faulty	Healthy	Total
Frames (%)	50	50	100
PCA (%)	42	68	55
KPCA (%)	33	67	50
SVD (%)	38	54	46
HM (%)	98	2	50

The numbers of faulty and healthy frames coming from the video flow were selected same in experimental studies. In four different scenarios realized, PCA method gave the highest accuracy ratio in general and the fastest operating time was achieved with SVD method. When accuracy ratios were separately compared for faulty and healthy frames, PCA method has the greatest accuracy ratio in inspection of healthy frames while HM method has the highest accuracy ratio in the inspection of faulty frames. Equation 2 was used in order to increase the general accuracy rate. Results obtained from PCA and HM methods for the same frame in p, h order are shown in Equation 2, and the general accuracy ratio of the method was calculated as 85%. This situation is summarized in Table 6. Consequently, the block diagram of the method which is capable of functioning in real-time and has a high accuracy rate in the test phase is given in Figure 10.

$$z = 1 - (1 - p)(1 - h) \tag{2}$$

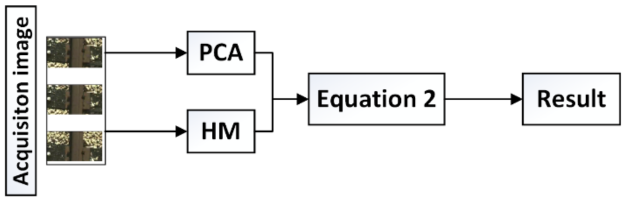

Fig. 10. Increasing the accuracy with PCA+HM

Table 6. Accuracy ratios of HM and PCA methods for faulty and healthy frames and combination of the two methods such in Equation 2

	Faulty frames	Healthy frames	Total
Frames (%)	50	50	100
PCA (%)	42	**68**	55
HM (%)	**98**	2	50
PCA+HM (%)		**85**	

5. Conclusions and Feature Work

In this research, a study was proposed for computer vision based inspection of the faults which may occur on railway tracks. In the study, virtual fault in the frames of the video image which is used for testing in the inspection of rail faults were created. The reliability of the study was tested by repeating the four different scenarios with different fault types and test times. In general, the highest accuracy ratio was obtained with PCA method while the highest operating time was achieved with SVD method.

7 frames per second can be processed in test phase with PCA method which has the highest accuracy ratio in general. The system proposed with this value shall be capable of operating with

55% of accuracy ratio on a test vehicle running at a speed of 25 km/h.

Highest accuracy ratio (98%) with HM method was achieved only for the inspection of faulty frames. 9 frames per second can be processed in test phase if HM method is used. The system proposed with this value shall be capable of operating with 50% of accuracy ratio on a test vehicle running at a speed of 30 km/h.

General accuracy rate of the system was increased to 85% with the combination of PCA and HM methods according to the results obtained from experimental studies. The system proposed with this value shall be capable of operating with 85% of accuracy ratio on a test vehicle running at a speed of 15 km/h. It is planned to conduct two different studies in the future associated with the current study.

In the application, accuracy ratio increased with the combination of two different methods but operating time extended correspondingly. For this reason, the first study planned will be building a parallel software and hardware architecture for the method proposed.

Since images acquired from rgb camera were used in the study, there is a possibility for objects like grease stains, dust residues and gravels to be detected as a fault. Therefore, it is planned to improve the algorithm in second study in order to obtain more reliable accuracy values and to eliminate the disadvantages due to the images.

6. Acknowledgment

This work was supported by the TUBITAK (The Scientific and Technological Research Council of Turkey) under Grant No: 114E202.

7. References

[1] R. D. Shachter, T. S. Levitt, L. N. Kanal, J. F Lemmer, "Utility-based control for computer vision", Uncertainty in Artificial Intelligence", 4, 407, 2014.

[2] E. N. Malamas, E. G. Petrakis, M. Zervakis, L. Petit, J. D. Legat, "A survey on industrial vision systems, applications and tools", Image and vision computing, pp.171-188, 2003.

[3] T. Brosnan, D. W. Sun, "Inspection and grading of agricultural and food products by computer vision systems a review", Computers and electronics in agriculture, 36(2), pp.193-213, 2002.

[4] A. Kumar, "Computer-vision-based fabric defect detection: a survey", Industrial Electronics, IEEE Transactions on, 55(1), 348-363, 2008.

[5] C. Benedek, O. Krammer, M. Janóczki, L. Jakab, "Solder paste scooping detection by multilevel visual inspection of printed circuit boards", Industrial Electronics, IEEE Transactions on, 60(6), 2318-2331, 2013.

[6] H. Guler, "Decision support system for railway track maintenance and renewal management", Journal of Computing in Civil Engineering, 27(3), 292-306, 2012.

[7] Z. Liu, J. Sun, H. Wang, G. Zhang, "Simple and fast rail wear measurement method based on structured light", Opt. and Lasers in Eng. , 49(11), pp.1343-1351, 2011.

[8] J. Chen, C. Roberts, P. Weston, "Fault detection and diagnosis for railway track circuits using neuro-fuzzy systems", Control Eng. Practice,16(5), pp.585-596, 2008.

[9] Q. Li, S. Ren, "A real-time visual inspection system for discrete surface defects of rail heads", Instrumentation and Measurement, IEEE Trans, 61(8), pp.2189-2199, 2012.

[10] W. Jin, X. Zhan, B. Jiang, "Non-contact rail-wear inspecting system based on image understanding", In Mechatronics and Automation, 2007. ICMA 2007. International Conference on pp. 3854-3858, IEEE, 2007.

[11] C. Aytekin, Y. Rezaeitabar, S. Dogru, I. Ulusoy, "Railway Fastener Inspection by Real-Time Machine Vision". Systems, Man, and Cybernetics: Systems, IEEE Transactions on, 2015.

[12] L. Jarzebowicz, S. Judek, "3D machine vision system for inspection of contact strips in railway vehicle current collectors", In Applied Electronics (AE), 2014 International Conference on pp. 139-144, 2014.

[13] I. Aydin, M. Karakose, E. Akin, "Robust anomaly detection in pantograph-catenary system based on mean-shift tracking and foreground detection". In Systems, Man, and Cybernetics (SMC), 2013 IEEE International Conference on pp. 4444-4449, 2013.

[14] H. Feng, Z. Jiang, F. Xie, P. Yang, J. Shi, L. Chen, "Automatic fastener classification and defect detection in vision-based railway inspection systems". Instrumentation and Measurement, IEEE Trans, 63(4), pp.877-888, 2014.

[15] G. Karaduman, M. Karakose, E. Akin, "Experimental fuzzy diagnosis algorithm based on image processing for rail profile measurement", In Proceedings of 15th International Conference Mechatronika, 5-7 Aralık, Prague, 1-6, 2012.

[16] R. Huber-Mörk, M. Nölle, A. Oberhauser, E. Fischmeister, "Statistical Rail Surface Classification Based on 2D and 21/2D Image Analysis", In Advanced Concepts for Intelligent Vision Systems, pp.50-61, 2010.

[17] A. O. Kozal, M. Teke, H. A. Ilgin, "Comparative analysis of hyperspectral dimension reduction methods", In Signal Processing and Communications Applications Conference (SIU), 2013 21st pp. 1-4, 2013.

[18] O. Yaman, M. Karakose, E. Akin, I. Aydin, "Image processing based fault detection approach for rail surface", Signal Processing and Comm. Applications Conference (SIU), 23th. IEEE, pp.1118-1121, 2015.

[19] C. Tastimur, E. Akin, M. Karakose, I. Aydin, "Detection of rail faults using morphological feature extraction based image processing", Signal Processing and Comm. App. Conference (SIU), 23th.IEEE, pp.1244-1247, 2015.

[20] V. F. Rodriguez-Galiano, B. Ghimire, J. Rogan, M. Chica-Olmo, J. P. Rigol-Sanchez, "An assessment of the effectiveness of a random forest classifier for land-cover classification", ISPRS Journal of Photogrammetry and Remote Sensing, 67, pp.93-104, 2013.

[21] H. Abdi, L. J. Williams, "Principal component analysis". Wiley Interdisciplinary Reviews: Computational Statistics, 2(4), pp.433-459, 2010.

[22] J. M. Lee, C. Yoo, I. B. Lee, "Fault detection of batch processes using multiway kernel principal component analysis". Computers & chemical engineering, 28(9), pp.1837-1847, 2004.

[23] F. Kleibergen, R. Paap, "Generalized reduced rank tests using the singular value decomposition". Journal of econometrics, 133(1), pp.97-126, 2006.