STEEL STRIP SURFACE DEFECT IDENTIFICATION BASED ON BINARIZED STATISTICAL FEATURES

Zoheir MENTOURI[1 2], Abdelkrim MOUSSAOUI[3], Djalil BOUDJEHEM[1], Hakim DOGHMANE[4]

In the steel hot rolling process, flat products that are shaped by a gradual reduction of the thickness and the increasing of the length may exhibit different surface defects, which should be identified. The solution, widely adopted, and still considered as a challenge is the automatic inspection. It is assumed, allowing an immediate detection with accurate identification of the defect that starts appearing during production. However, for a perfect labeling of the occurring defects, inspection system should be provided with reliable algorithms. In this paper, tools are combined to provide a high-efficiency solution. The suggested method is based on the recent Binarized Statistical Image Feature extractor used, to date, in biometrics. Combined with a relevant reduction-data method and the K nearest neighbors classifier, this solution showed improved recognition rates of the strip surface defects in the hot rolling process, outperforming, the reported results in previous works.

Keywords: Computer vision, statistical features; classification, strip surface defects, hot rolling process.

1. Introduction

Surface defects may seriously impact the quality of steel products and lead to their rejection, causing needless additional expenses. To efficiently recognize them and enable an immediate decision-making, the online automatic surface inspection is the solution widely adopted and still considered as a challenge.

Then, for many categories of steel products, different techniques based on the frequency domain as the FFT or spatial-frequency analysis used with some adaptive learning methods as SVM or NN-BP reached rates around 90% in strips, thick plates and slabs defect categorization [1][2][3]. Whereas some other techniques as background difference, region growing or co-occurrence matrix,

[1] Laboratory of Advanced Control-LABCAV. Université 8 Mai 1945 Guelma, BP 401, Guelma 24000, Algeria. e-mail: mentouri.zoheir@univ-guelma.dz.
[2] Research Center in Industrial Technologies-CRTI, BP64, Chéraga 16014, Algiers, Algeria.
[3] Laboratory of Electrical Engineering-LGEG. Université 8 Mai 1945 Guelma, BP 401, Guelma 24000, Algeria
[4] Laboratory of Inverse Problems-PI: MIS. Université 8 Mai 1945 Guelma, BP 401, Guelma 24000, Algeria.

were used in detection of different types of hot and cold rolled strips and gave interesting results [4][5][6][7].

What should be noted, is that due to the lack of defect standards [1], there is no ideal method combination which could be suitable for all surface defects. In practice, their large number makes of their labeling a complex multi-class issue, of which, the solution always depends on application needs [8][9][10][11][12].

On the other hand, a key step to an efficient classification is the prior defect description task. As reported, it can be based on human experience in defining heuristics that view, for instance, scratch defects as sharp edges with almost white pixels and dents as areas with extreme gray values [13], or on some transformations as PCA-SOM to detect complex shape defects as oxidation, exfoliation and waveform and Hough transform for well-defined shapes as welding, clamps, and holes [14]. Moreover, the defect description can use the statistical properties as defect geometry, grayscale and shape, to build a feature space, wherein the elements have not only to satisfy the invariance to defect size and orientation changes, but should have a certain stability against noise too [15] [16].

In this paper, we present an efficient tool combination, to categorize the surface defects of hot rolled steel products. We introduce, in this field, a recent local image descriptor, namely the Binarized Statistical Image Features (BSIF) used, to date, in some face recognition applications [17][18]. Thus, an overview about the used filters is given in the second section, as well as, a BSIF operation. In the next section the suggested approach is presented with the general scheme. The section four is concerned with the experimental study, where we present the used defects database, we describe the applied image processing tasks and we present and discuss all results. The paper ends with a conclusion.

2. Overview on Filters determination and Binarized Statistical Image Features (BSIF)

Conversely to some known methods, where filters are predefined manually, those of BSIF descriptor, provided in [17], are based on a statistical information learning with 13 natural images from Hyvarinen [19].

The convolution of a Filter with an image provides pixels description based on their neighborhood. Hence, for an original image patch X of a size m x m, and a filter F_i of equal size, the corresponding value of the filtered patch is determined by the product of their respective vectors:

$$h_i = f^T.x \tag{1}$$

Then, as formulated in equation 2, the use of n appropriate filters, allows getting a vector of n filtered outputs that are statistically independent.

$$h = F.x \tag{2}$$

Where, F is the filter matrix of dimension n x m^2.

This filter matrix is assumed to be the product of two appropriate matrices defined as follows:

$$h = M.P.x = M.y \tag{3}$$

Where, M is an estimated orthogonal matrix and P is a canonical preprocessing matrix, computed respectively by Independent Component Analysis and Principal Component Analysis Algorithms.

In practice, for the use of Principal Component Analysis, image patches are randomly chosen and centered around zero and the covariance matrix is computed then, eigen-decomposed by:

$$C = U.D.U^T \tag{4}$$

Where D: the diagonal matrix of the eigenvalues, organized in a decreasing order.

The canonical preprocessing matrix is defined by:

$$P = D^{-1/2}.U^T \tag{5}$$

Of which, the elements are the whitened principle components, computed with the first (most significant) eigenvalues of D.

The next step concerns the determination of the orthogonal estimated matrix M. Given the Equations 3 and 5, ICA algorithm is used to determine M by:

$$y = M^{-1}.h \tag{6}$$

Then, the product of the two computed matrices P and M gives the filter matrix F. An example of a filter set responses is presented in Fig. 1, below.

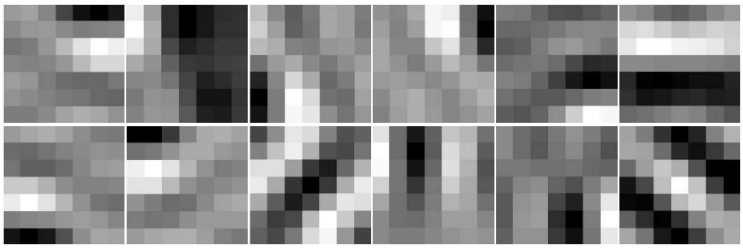

Fig. 1: Set of 7x7 sized pre-defined filters (number=12)

For the determination of a local image description, a BSIF procedure is applied as shown in Fig. 2.

It consists in applying a filter set matrix of nxm^2 of a size to an image patch X of m x m of size, then in attributing to the filtered-response R, a binary code, based on a zero thresholding as in equation 7.

$$\begin{cases} B = 1 \;\; if \;\; f^T.x \succ 0 \\ B = 0 \;\; Otherwise \end{cases} \tag{7}$$

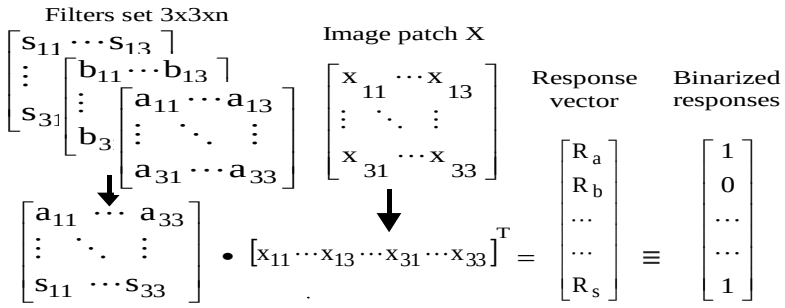

Fig. 2: BSIF encoding of a patch central-pixel

3. Application of the suggested approach

For defects identification, the implementation scheme is based on three main stages, as shown in Fig. 3. Defect images are firstly filtered by applying BSIF filters to provide new-scaled images. This transformation makes data more relevant for the subsequent steps that are features extraction and classification.

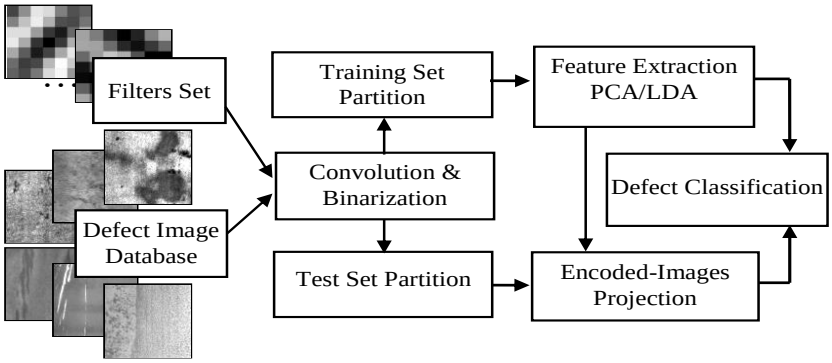

Fig. 3: Application Scheme for Strip Defects Identification

In the application, one BSIF operation consists in convolving the whole z x z sized image with a filter F_l of m x m size, binarizing the response, affecting a weight to the obtained binary code and repeating the operation n times, according to the number of filters in the considered set, to obtain a pixel code word. Final pixels new-values are computed by summing all the $2^{(n-l)}$ - weighted response-codes as formulated in equation 8.

$$\sum_{l=1}^{n}[Code(Conv(F_l, X)).2^{(n-l)}] = \sum_{l-1}^{n}(B_l.2^{(n-l)}) \qquad (8)$$

Where, n and l are respectively the number of filters and the rank of a filter in the concerned filters set.

The described procedure is applied to all images in the database and an example of their transformed aspect is shown in Fig. 4.

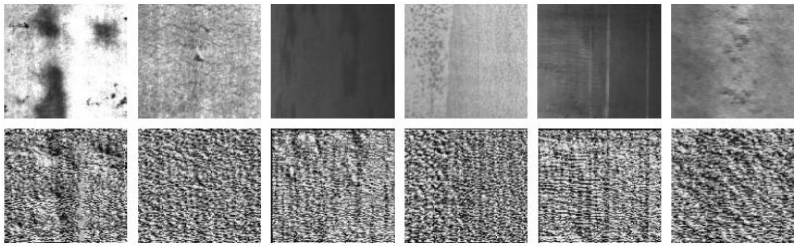

Fig. 4: Example of six image-defect samples. The 1st row, from left to right: Patch, Crazing, Inclusion, Pitted surface, Scratch and rolled-in-scale. The 2nd row : the corresponding binarized images using Filters set of size 7x7x12

Further, the invariance of the histogram to rotation and translation as well as, its low computational cost, makes of this tool and efficient means to represent, by a single vector, image statistical properties. Thus, the pixel values of new encoded image, are stacked into 2^n -long and [0 1]-normalized proper histograms H_i, which are concatenated in a single global matrix H, defined by:

$$H = \lfloor H_1, H_2, .., H_s \rfloor \tag{9}$$

Where, s is the number of the used images of all classes.

4. Experimental Study

4.1. Description of Strip Steel Surface Dataset

Typically, surface defects in rolling process are quite diverse. What may explain the fact that in many studies it was dealt, each time, with only some of them, depending on the application needs. Therefore, in this study, we use a Northeastern University surface defect database [20], shown in Fig. 5. It includes six types of common defects of the hot rolled strips, with 300 variants of each one and totaling 1800 defect grayscale-images. The different samples present many variabilities in defect orientation, size and grayscale level, which might be subject to the illumination effect in industrial environment.

The defects consist in Patches: surface with oxide not completely removed by a faulty pickling process; Crazing: a type of network of fine cracks; Inclusions: non-metallic particles that show through at the surface of the steel; Pitted surface: sharp depressions in the surface related to chemical attack; Scratches: sharp indentation in the surface caused by a machine and Rolled-in scale: a scale partially rolled into the surface of the steel sheet.

As shown in the figure below, the defects may be, localized and of a compact appearance, with relatively clear edges such as patches and scratches or scattered while affecting the whole surface such as pitted surface or crazing. Owing to this pattern heterogeneity, we consider, in this application, the entire image as an area of interest. There is no need to, more, reduce its size or remove

any possible useless regions. The medium size of 200x200 pixels presents a good compromise between the amount of information in the images and time processing considerations. It, also, eases the comparison with the state of art.

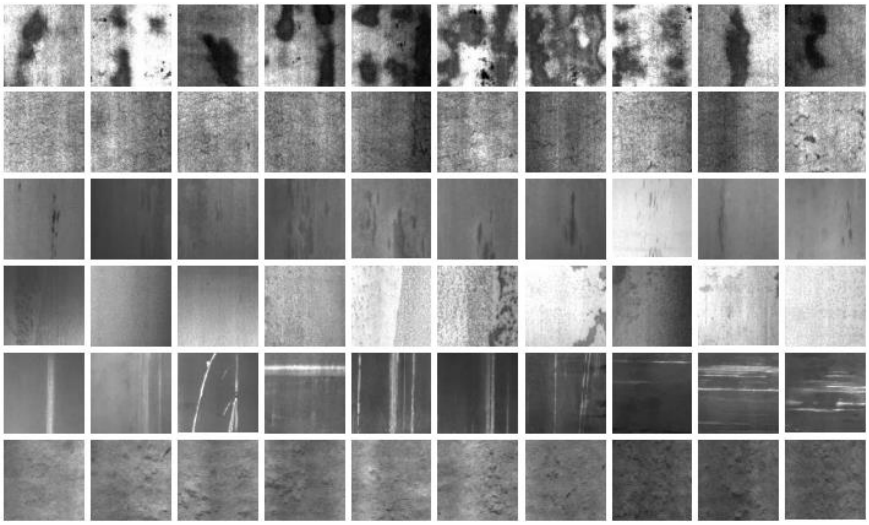

Fig. 5: Samples of defect images of NEU Database. One defect type per row: Patches, Crazing, Inclusion, Pitted surface, Scratches and Rolled-in-scale

4.2. Data Projection and similarities computing

In data dimensionality-reduction and projection, the unsupervised Principal Component Analysis method (PCA), known for its class variations description, is widely used as a means of data representation, even though, without, necessarily, being efficient in class distinction [21].

Given a number A of vectors randomly selected from the matrix H, and representing all classes, they are organized in a training matrix Tr [2^n, A].

The covariance matrix (total scatter matrix) of Tr is computed by equation 10 and eigen-decomposed to find the optimal projection matrix W_{pca}, given by equation 11, and where the retained best vectors N<A, correspond to the N largest eigenvalues.

$$C = \varphi.\varphi^T \tag{10}$$

$$W_{pca} = \arg\max\left|W^T C W\right| = \left[W_1 W_2 \ldots W_N\right] \tag{11}$$

Where φ is the normalized and centered matrix of Tr and the w_i (i=1, 2, .., N) are the N kept eigenvectors.

The other method, which is, rather, class discrimination-oriented, is the Linear Discriminant Analysis. This latter is concerned with searching the best

discriminant vectors and is based on a maximization of a ratio of the between-class scatter matrix of equation 12:

$$S_b = \sum_{i=1}^{c} (\varphi_{c_i} - \varphi)(\varphi_{c_i} - \varphi)^T \tag{12}$$

and the with-in class scatter matrix of equation 13:

$$S_w = \sum_{i=1}^{c} \sum_{k \in c_i}^{q_i} (\gamma_k - \varphi_{c_i})(\gamma_k - \varphi_{c_i})^T \tag{13}$$

Where γ_k is the k^{th} sample in class c_i, φ_{ci} is the mean vector in class c_i , φ is the overall mean of the data-classes, c is the number of classes and q_i is the vector number in the class c_i .

Then, The scheme followed, to reduce the dimensionality of the BSIF histograms and project data into a smaller subspace, is the one that takes benefits from the two presented methods and addresses their shortcomings [22][23]. The flowchart of Fig. 6 gives an overview of the applied model computing procedure.

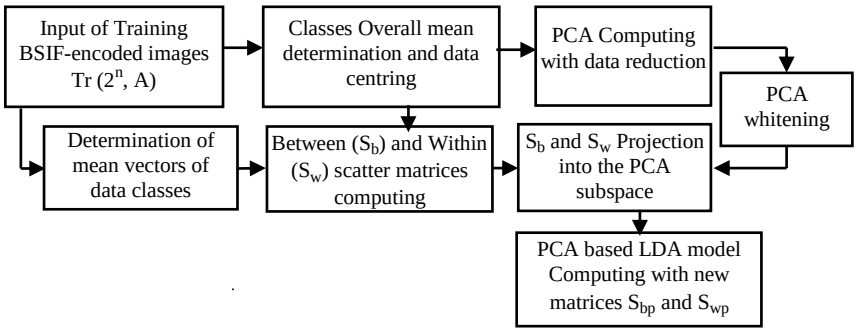

Fig. 6: The concept of PCA based LDA Data reduction

The PCA method is pre-applied to the training set Tr as well as, a whitening operation of the resulting data. The PCA subspace W_{pca} is then, created with an adopted reduction level around 80% (Recommended at 75% in [20])

In a second step, an LDA procedure is applied. However, instead of using actual scatter matrices, computed by equations 12 and 13, it uses their projections into the computed PCA subspace in searching W_{fld} , which is the optimal LDA projection matrix, defined by the Fisher criterion of equation 14:

$$W_{fld} = \frac{W^T W_{pca}^T S_b W_{pca} W}{W^T W_{pca}^T S_w W_{pca} W} \tag{14}$$

Where $W_{pca}^T S_b W_{pca}$ and $W_{pca}^T S_w W_{pca}$ are (S_{bp} and S_{wp}) respectively the projection of S_b and S_w matrices into the PCA subspace.

The obtained LDA subspace is used to project BSIF histograms of the test partition images, providing, thus, histogram features of the original defect images.

In the classification step, distance similarities are computed to label all defects of the test dataset. This task is performed with the two commonly used techniques: The non-parametric K-nearest neighbor classifier (with K=3), based on the Euclidean matching distance and the supervised learning machine algorithm: multi-class SVM, with a radial basis kernel function (Rbf).

4.3. Results and Compared methods

In order to implement the most suitable BSIF descriptor, all filter sets, provided in [18], are evaluated. Their application on the NEU Database, shown in Fig. 7, revealed that the image content is better captured by 7x7x12 filters set, which is the one we retained for the method evaluation.

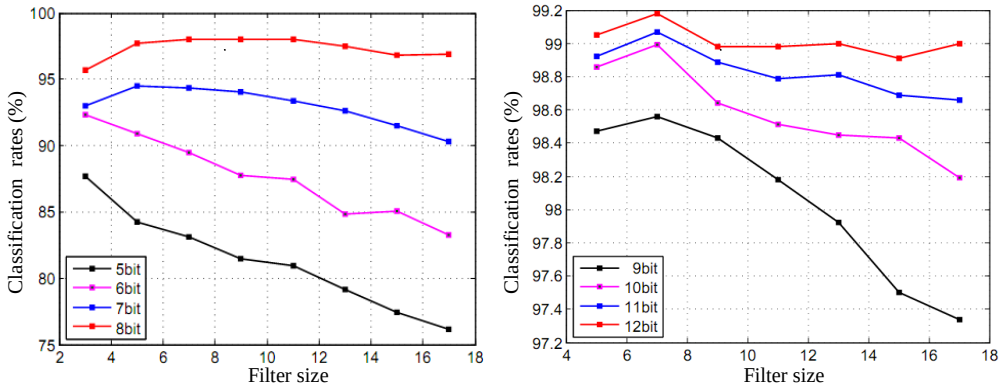

Fig. 7. Defect identification rates with different Filter Sets applied to NEU Database

Further, to efficiently assess the performance of the whole algorithm and its reliability, a high number of computing trials is chosen and where, for each new trial, the global histogram matrix H is randomly partitioned to select, a different pair of training and test sets. Then, each presented result, below, is, in fact, the average value of five hundred rates, obtained by as much execution of the algorithm.

The table 1 summarizes average rates of the identification methods, as well as their corresponding standard deviation (St), which characterizes results dispersion and allows getting, in a sense, an idea about the method robustness.

Table 1

Recognition rates of strip surface defects of NEU Database

F. Descriptor	Gabor_LDA	Gabor_LDA	BSIF_LDA	BSIF_LDA	BSIF_PCA
Classifier	KNN	SVM	KNN	SVM	KNN
Results (%)	86.63±1.11	88.36±1.01	**99.18±0.30**	84.64±1.14	94.92±0.75

The PCA method, tested in our evaluation, showed a relative low result (94.92%) what may justify our use of a two projections method in dimensionality reduction, allowing a good classes discrimination.

With the highest average (99.18%) and the lowest St (0.30), the BSIF descriptor with the two-projections PCA-LDA method and the KNN classifier constitute the combination (BSIF_LDA_KNN) that outperforms the compared methods and presents the best choice in term of computational cost.

What should be, further, reminded, is the efficiency of the Nearest Neighbor classifier. However, our choice the KNN classifier has been driven by the voting approach in this latter, which is assumed helping to minimize labeling errors and, really, in our application, gives an improved end-result

The lowest obtained average, around 84%, concerns the multiclass SVM classifier when used with BSIF. Even though, the method is known for its interesting results, as reported in many works, it has not been more investigated, here, since the classifier seems to take more computing time.

The proposed solution is compared to Gabor filtering too, which is the method frequently used for defect detection in steel industries (mainly for thick plates and slabs). The method has been tested with its parameters set to the optimal values of 8, 5 and 64 respectively, the numbers of filter bank orientations, scales and a down-sampling factor. Moderate classification rates with the SVM classifier are reached and slightly less with KNN, as shown in table 1.

The main limitation encountered with Gabor filtering is the amount of data. In spite of reducing image size before and after filtering, the dimension of relevant vectors that are kept, remains important comparing to BSIF vectors, what has been considered as a constraint with respect to memory and time savings.

To better appreciate, how powerful, is the suggested combination, performance curves of Fig. 8 are constructed. The results show that, whatever is the size of the selected training set, our suggested method, widely, outperforms all the others.

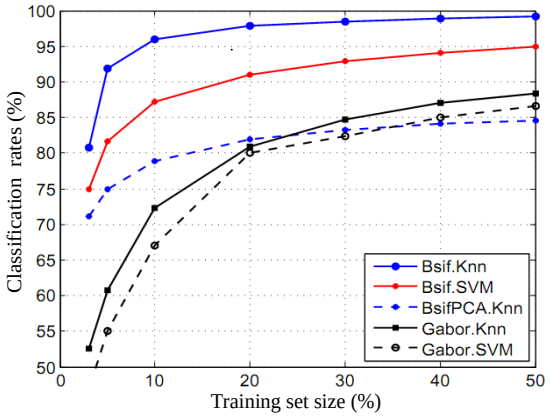

Fig. 8. Averages of identification rates of NEU Database Defects

The confusion matrix, in Table 2, is created in order to analyze the identification errors of the proposed method. The numbers of classified defects in this table, concern an identification rate of 99.22%, related to the last selected pair of sets (Train - test:50%-50%) among the five hundred trials of which the average rate is the highest one in the Table l.

Table 2

Confusion matrix of NEU Database defects classification

		Predicted Class					
		Cr	**In**	**PS**	**Pa**	**RS**	**Sc**
Actual Class	Cr	150	0	0	0	0	0
	In	3	146	1	0	0	0
	PS	0	2	148	0	0	0
	Pa	0	0	0	150	0	0
	RS	0	0	0	0	150	0
	SC	0	1	0	0	0	149

The matrix indications show that the highest rated classifications (at 100%) concern 3 of 6 defect types, which are visibly distinguishable from the other defect (Fig. 5). With 4 false identifications from the 150 predicted (correct at 97.33%), the inclusion defects present more than 50% of global error, whereas the pitted surface defect is identified with 2 errors (correct at 98.66%).

These types of defects are of more complex shape and sparseness. What may explain the misclassification of some defect variants, adding to that the illumination parameter. In such a system, even if provided with a highly efficient feature extractor and classifier, the illumination should, normally, be above critical conditions.

As for the previous studies that dealt with the same dataset to assess some methods, the reported identification scores, although interesting, remain below to those of our suggested approach. The table 3 presents some of these results, where the training and test partition sizes, are as used in our application.

Table 3

Comparison of identification rates (%) of NEU surface defect Database

Work Ref.	**Features Descriptor**	**Classifier**	**Results (%)**
Kenchen S.[15]	SCN	SVM	98.60±0.59
Kenchen S.[16]	AECLBP	SVM	98.93±0.63
Kenchen S.[16]	CLBP	SVM	98.28±0.51
Li Yi [17]	CNN	CNN	99.05
Suggested	BSIF_LDA	KNN	**99.18±0.30**

5. Conclusion

The variety of the published works related to the identification of surface steel defects reported different performance level. The noted difficulty in such a task is, mainly, related to the complexity of these defects which may exhibit similarities between defects that belong to different classes and distinction between defects of the same class.

In this paper, we suggested a new combination of tools, where the recent BSIF descriptor, used, to date, in some biometric applications, is introduced and assessed in the description of the surface defect of hot rolled products. Further, data are processed by a combined reduction method to present an optimal features space that eases the classification task. This latter is performed with KNN classifier assuring an efficient defect identification with a low computational cost. Applied to a defect database which contains enough defects and defect variabilities, this tool combination outperforms the methods that dealt with similar defects [1], and even the one that used the same database as in our application. Indeed, the proposed solution shows high classification rates with an interesting dispersion level of results. Moreover, identification results are consistently higher while varying the size of the training image set. Finally, the obtained results demonstrate the applicability of the used descriptor in this field and show the efficiency and reliability of the proposed solution, what confirms its suitability for industrial application.

R E F R E N C E S

[1]. *N. Neogi, D. K Mohanta and P. K Dutta*, Review of vision-based steel surface inspection systems, EURASIP Journal on Image and Video Processing (2014). Available at: http://jivp.eurasipjournals.com/content /2014/1/50.

[2]. *X. Xie*, A Review of Recent Advances in Surface Defect Detection using Texture analysis Techniques. Electronic Letters on Computer Vision and Image Analysis 7(3):1-22, 2008, ELCVIA ISSN:1577-5097, Published by Computer Vision Center / Universitat Autonoma de Barcelona, Spain

[3]. *C. Tikhe, J. S. Chitode*, Metal Surface Inspection for Defect Detection and Classification using Gabor Filter, International Journal of Innovative Research in Science, Engineering and Technology (An ISO 3297: 2007 Certified Organization), 3 (6) June 2014.

[4]. *G Wu, H Kwak, S Jang, K Xu, J Xu*, Design of Online Surface Inspection System of Hot Rolled strips, Proceedings of the IEEE International Conference on Automation and Logistics Qingdao, China, 2008, 2291–2295

[5]. *JH. Cong, YH. Yan, HA. Zhang, J. Li,* Real-time surface defects inspection of steel strip based on difference image, international symposium on photoelectronic detection and imaging, related technologies and applications. Proc. of SPIE 6625,1–9, 2007

[6]. *P. Caleb, M. Steuer*, Classification of surface defects on hot rolled steel using adaptive learning methods, KES'2000. Fourth international conference on knowledge-based intelligent engineering systems and allied technologies. Proc. 1, 103–108, 2000

[7]. *Maoxiang CHU, Anna WANG, Rongfen GONG and Mo SHA*, Strip Steel Surface Defect

Recognition Based on Novel Feature Extraction and Enhanced Least Squares Twin Support Vector Machine, ISIJ International, **Vol. 54**, No. 7, 2014, 1638–1645

[8]. *Florent Dupont, Christophe Odet, Michel Carton*, Optimization of the recognition of defects in flat steel products with the cost matrices theory, NDT&E International, 1997, Elsevier Science Ltd , **Vol. 30**, No. 1, pp. 3-10, 1997

[9]. *J. Masci, U. Meier, D. Ciresan, G. Fricout*, Steel Defect Classification with MaxPooling Convolutional Neural Networks, The 2012 International Joint Conference on Neural Networks (IJCNN),10-15 June 2012, Brisbane, QLD, Australia

[10]. *Maoxiang CHU, Rongfen GONG and Anna WANG*, Strip Steel Surface Defect Classification Method Based on Enhanced Twin Support Vector Machine, ISIJ International, **Vol. 54**, No. 1, pp. 119–124, 2014

[11]. *Shiyang Zhou, Youping Chen, Dailin Zhang, Jingming Xie, Yunfei Zhou*, Classification of Surface Defects on Steel Sheet Using Convolutional Neural Networks, MATERIALS AND TECHNOLOGY, available at: https://www.researchgate.net/publication/ 313896846, doi:10.17222/mit.2015.335, 2016

[12]. *Li Yi, Guangyao Li, and Mingming Jiang*, An End-to-End Steel Strip Surface Defects Recognition System Based on Convolutional Neural Networks, Steel research int. 87, 2016, No. 9999. DOI: 10.1002/srin.201600068, available at: www.steel-research.de.

[13]. *Smriti H. Bhandari, S. M. Deshpande, S. M. Deshpande*, A Simple Approach to Surface Defect Detection, 2008 IEEE Region 10, Colloquium and the Third International Conference on Industrial and Information Systems, Kharagpur, December 8-10, 2008, INDIA.

[14]. *Luiz A. O. Martins, Flavio L. C. Padua, Paulo E. M. Almeida*, Automatic Detection of Surface Defects on Rolled Steel Using Computer Vision and Artificial Neural Networks, Conference Paper (2010), DOI:10.1109/IECON.2010.5675519, Source:IEEE Xplore

[15]. *Huijun Hu, Yuanxiang Li, Maofu Liu and Wenhao Liang*, Classification of defects in steel strip surface based on multiclass support vector machine, Multimed Tools Appl. (2014) 69:199–216, Published online by Springer Science+Business Media, New York, 2012

[16]. *Kechen Song, Yunhui Yan*, A noise robust method based on completed local binary patterns for hot-rolled steel strip surface defects, Applied Surface Science 285P (2013), 858-864, available at : www.elsevier.com/locate/apsusc.

[17]. *J. Kannala and E. Rahtu*, Bsif: binarized statistical image features, Proceedings of 21st international conference on pattern recognition (ICPR 2012), Tsukuba, Japan, 1363–1366

[18]. *Ylioinas J., Kannala J., Hadid A., Pietikäinen M*, Face Recognition Using Smoothed High-Dimensional Representation. In: Paulsen R., Pedersen K. (eds) Image Analysis. SCIA 2015. Lecture Notes in Computer Science, **Vol. 9127**. Springer, Cham

[19]. *A. Hyvarinen, J. Hurri, P. O. Hoyer*, Natural Image Statistics. A probabilistic approach to early computational vision, February 27, 2009, Springer

[20]. *Kechen SONG, Shaopeng HU and Yunhui YAN*, Automatic Recognition of Surface Defects on Hot-rolled Steel Strip using Scattering Convolution Network, Journal of Computational Information Systems 10: 7, 3049–3055, 2014

[21]. *A. M. Martinez and A. C. Kak*, PCA versus LDA, IEEE Transactions on Pattern Analysis and Machine Intelligence. **Vol. 23**, N°.2, pp.228-233, 2001

[22]. *P. N. Belhumeur, J. P. Hespanha and D. J. Kriegman*, Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection, IEEE Transactions on Pattern Analysis and Machine Intelligence, **Vol. 19**, N°7, July 1997

[23]. *D. L. Swets, J. (Juyang) Weng*, Using Discriminant Eigenfeatures for Image Reteival, IEEE Transactions on Pattern Analysis and Machine Intelligence, **Vol. 18**, N°8, August 1996