Preprint

OBJECT RECOGNITION USING SHAPE GROWTH PATTERN

Abbas Cheddad, Huseyin Kusetogullari, Håkan Grahn

Department of Computer Science and Engineering, Blekinge Institute of Technology,

SE-37141 / Karlskrona, Sweden

Abstract—**This paper proposes a preprocessing stage to augment the bank of features that one can retrieve from binary images to help increase the accuracy of pattern recognition algorithms. To this end, by applying successive dilations to a given shape, we can capture a new dimension of its vital characteristics which we term hereafter: the shape growth pattern (SGP). This work investigates the feasibility of such a notion and also builds upon our prior work on structure preserving dilation using Delaunay triangulation. Experiments on two public data sets are conducted, including comparisons to existing algorithms. We deployed two renowned machine learning methods into the classification process (i.e., convolutional neural network -CNN- and random forests -RF-) since they perform well in pattern recognition tasks. The results show a clear improvement of the proposed approach's classification accuracy (especially for data sets with limited training samples) as well as robustness against noise when compared to existing methods.**

Keywords—*Binary image dilations, shape growth pattern, pattern recognition, convolutional neural network, machine learning.*

I. INTRODUCTION

A. Motivation and Literature Review

Pattern recognition is a long standing problem that is challenging to tackle. Many algorithms have been proposed and developed to address it. Nevertheless, those algorithms may suffer in their performance under inconsistencies in intensity and/or chrominance tone or under irregularities in scales of the targeted region of interest (ROI) in grayscale or colour images. Binary images, on the other hand, are not prone to colour differences. Moreover, the algorithms for computing properties of binary images are easy to grasp, faster and computationally less expensive [1]. It is common, however, to see binary images being extensively used in various fields, ranging from Content-based Image Retrieval [2], [3], to industrial product inspection and robotics, to Optical Character Recognition and pattern recognition and to medical imaging. Binary image dilation, as a subfield of mathematical morphology, is a useful technique to, for example, increase thickness of handwritten text lines, eliminate noise, connect broken segments and to preprocess biometrics data, to name a few [4]- [9].

Before delving into the rest of the paper, we warrant herein a summary of our contributions:

i) **Shape growth pattern (SGP)**: We describe pattern recognition tasks through augmented feature space using SGP (a unique characteristic of shapes that has been overlooked in the literature). This can be of paramount importance to studies relying on data sets of binary patterns or to studies having small data sets.

ii) **Machine learning based on SGP**: CNN (adapted to shape characteristics) and RF are deployed in this work and we prove that the proposed approach has the best performance rate for classification on two independent data sets.

We previously proposed in [10], the Delaunay triangulation based binary image morphing (DTBIM) as an appealing alternative for morphological dilation that better preserves shape structures. Thus, this paper comes yet as another validation check. DTBIM is also a noise resilient method. This fact is highlighted in [10] and is tested independently in this paper.

B. Background

The SGP can be attained by using the common morphological dilation methods (i.e., dilations based on binary kernels, isotropic dilation) which are discussed briefly below together with the most commonly used machine learning methods.

1) Image dilation:

i) **Kernel based dilation**: In this paper we use a disk shaped structuring element or kernel defined as $S = \{(0,1), (1,0), (1,1), (1,2), (2,1)\}$, to perform the shift-invariant dilation which is equivalent to Minkowski addition. This type of dilation can be symbolically represented as follows: Let I and S be sets in the $2D$ space N^2, that correspond to the binary image and the structuring element, respectively. And let $i = i_1, i_2, ..., i_n$ and $s = s_1, s_2, ..., s_n$ be the elements in I and S, respectively. The dilation of I by S, denoted by $I \oplus S$, is defined as:

$$I \oplus S = \{d \in N^2 | d = i + s, i \in I, s \in S\} \quad (1)$$

In this paper, we use three dilation passes (i.e. $I \oplus S$, $(I \oplus S) \oplus S$, $((I \oplus S) \oplus S) \oplus S$).

ii) **Isotropic Dilation**: When it comes to binary image dilation, our core topic, a thresholded form of the distance transformation is sometimes referred to as the isotropic dilation. The connection between the two, namely the distance transformation and dilation, is discussed in [11]. The distance transformation map which we use is the one proposed by Maurer et al. [12]. It can be computed using several distance metrics, however, in this paper we resort to using the Euclidean distance as it is the

most adopted metric. Isotropic dilation is naturally more flexible than the previous method that performs the shift-based dilation in the sense that no kernel construction is required and the computation is done just once after which multiple dilations can be accomplished by merely performing thresholding at multiple levels (i.e., 2, 3, 4).

2) Machine Learning CNN and RF: Convolutional Neural Networks (CNNs) are large networks that have recently become an important and effective approach for computer vision problems such as image classification, object recognition, face recognition and human detection and tracking [15], [18], [19]. Besides this, in order to obtain high accurate results using CNNs, it is necessary to train the models effectively with large data sets [20] a condition which may not always be attainable. Data sets with limited number of samples will deteriorate the success recognition rate in the computer vision applications. For data sets with limited and binary training samples, we therefore propose SGP to effectively train the CNN models.

We also want to compare the performance of the conventional use of CNN, when it estimates and extracts features from a 2D image on its own, to the CNN performance when assisted by the SGP.

The second classifier we use in this paper is the random forests which fall under a larger family of classifiers known as ensemble learning methods that generate many sub-classifiers and aggregate their results. The two well-known methods are boosting and bagging of classification trees [21]. Breiman [22] proposed random forests (RFs), an ensemble learning non-parametric statistical method for classification and regression, which adds an additional layer of randomness to bagging. Breiman's approach to construct the RF classifier is the one we adopt in this paper.

II. METHODOLOGY

A. Shape Growth Pattern (SGP)

When examining the previous published studies on pattern recognition, it is clear that the majority are either extracting features from the examined images, from their multiscale versions (e.g., wavelet decomposition, pyramid representation, etc.), or from their transformations (e.g., Hessian, Radon, Gabor, Gaussian, etc.). To the best of our knowledge, there is no prior work that dealt with capturing one of the very valuable characteristics of shapes, namely, the growth pattern that a given shape exhibits when performing binary morphological dilation. The notion which we coined here is termed shape growth pattern (SGP) and its usefulness becomes even more apparent when having small data sets that are insufficient to characterise the different shape instances. SGP can augment the feature space of each instance, therefore, helps in classification using different machine learning methods. The SGP can be attained by using the common morphological dilation methods (i.e., dilations based on binary kernels, isotropic dilation). Moreover, this paper further validates our recently published dilation method that preserves the binary shape structure. Unlike existing methods, it allows for geometric variations during the dilation operation to probe an image. Hence, we will contrast the performance of SGP based on the existing methods, described in section I, and also on our own dilation method which we termed **D**elaunay triangulation based
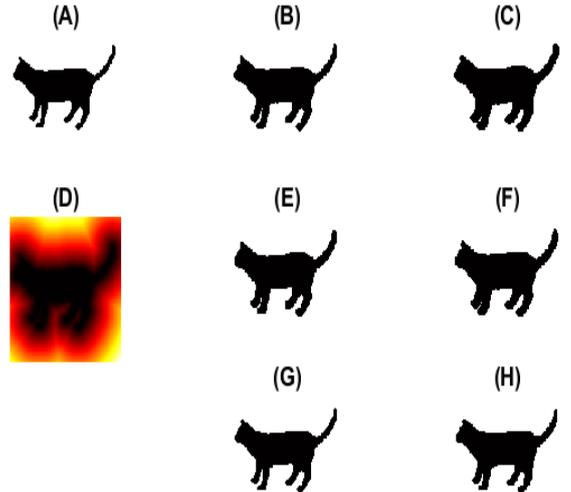

Fig. 1: Dilation examples using the methods reported in this paper. (A) Original binary image from the Data set II database, (B & C) consecutive dilations using a disk structuring element of size 3×3, (D) the distance transformation image of (A) –image enhanced for display–, (E & F) consecutive isotropic dilations thresholded at levels 2 and 3 in the Euclidean space, respectively and (G & H) consecutive dilations using the DTBIM [10].

binary **i**mage **m**orphing (DTBIM). More details on DTBIM are provided in [10].

The proposed algorithm, SGP, leverages the performance of machine learning in binary shape classification by capturing a new dimension of growth pattern which is unique for each shape. Such pattern is achieved by successive dilation passes. In each pass, features that characterise a shape are extracted. The aggregated feature vectors' stack forms a relationship between each other which is highly descriptive of the sought-after shape growth pattern.

Shape descriptors have been very much reliant on the robustness in capturing the unique characteristics of a specific shape for pattern matching. Whether that is carried out in the spatial domain or in the frequency domain, a descriptor's algorithm departs only from the available shape images. The SGP can be achieved based on structuring elements (a.k.a. kernels) or based on spatial distance transformation (isotropic dilation).

B. SGP using DTBIM

In this section, we delve into the new method whose shape structure preservation property teases it apart from the aforementioned methods. Shih [16] stated in his book that "traditional mathematical morphology uses a designed structuring element that does not allow geometric variations during the operation to probe an image." (Ch. 11, p.341). The method, DTBIM, implements point geometry-aware dilation algorithm that exploits the versatile structure of Delaunay triangulations. One of the favourable properties of DTBIM is its small incremental expansion that is unreachable by the smallest SE used in the common methods. Therefore, binary shape characteristics may survive longer chained dilations before the object is totally deformed. Such a feature can leverage the performance of several algorithms for further processing.
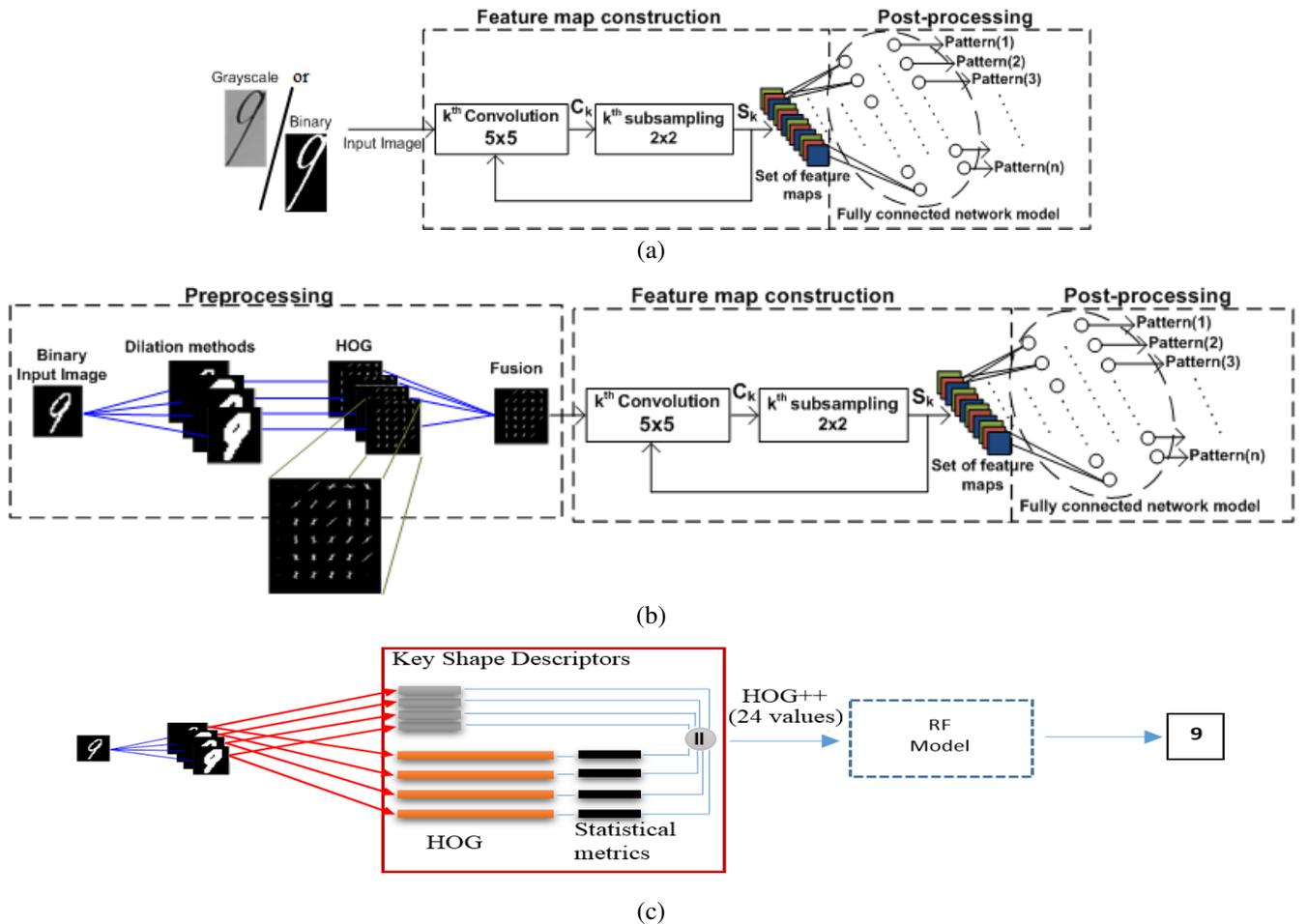
(a)

(b)

(c)

Fig. 2: Graphical representations of the experimental setups. (a) Depicts the common use of the CNN to recognise handwritten digits. (b) A pictorial representation of the CNN architecture using the proposed SGP. (c) A graphical representation of the proposed approach using SGP to classify a given shape using RF (shown is RF in the testing phase).

In order to fathom how Delaunay triangulations are constructed, the reader is referred to the numerous books discussing 2D geometry and mesh surface generations such as in [13]. Due to the limited space, we just note here that we discuss in details the DTBIM and the notion of constrained Delaunay triangles to achieve morphological dilation in our paper [10]. However, we warrant here-after an illustration of dilation using the different methods, see Fig. 1. Worth noting, DTBIM shares a common property with the isotropic dilation as both do not need internally any structuring elements, while additionally DTBIM does not require the user to specify any thresholding level. The arrangement of the ON pixels constituent a binary shape dictates the DTBIM's dilation behaviour; a property that other methods lack and which contributes to the overall property of structure preservation. Finally, as in [10] we also show DTBIM's robustness against noise in pattern recognition, see Fig. 3.

C. The HOG Feature Vector

Typically, dilated shapes are post-processed to extract feature descriptors. These features are then used as inputs for pattern recognition using machine learning algorithms; the CNN and the RF in our case. There exist myriad features which can be extracted from a given shape. As our aim is not to come up with a new feature descriptor, but rather to improve the performance of the existing descriptors, we resort to using existing shape descriptors.

Of the many techniques currently in vogue for shape feature extraction, the histogram of oriented gradients (HOG) descriptor comes as one of the most useful techniques. The HOG descriptor has been receiving much interest as it can be used to train machine learning models to detect or recognize different shapes [17]. Note that, in this paper, the cell size of HOG is set to 5×5. This cell size's setting encodes the adequate amount of spatial information in order to recognise an object while restricting the dimension of the HOG feature vector, this eventually helps to speed up training. For instance, a cell size of 8×8 encodes only a limited amount of shape information as compared to a cell size of 2×2 which encodes a more information but at the expense of the HOG feature vector dimensionality which increases significantly. Thus we found that a cell size of 5×5 is a good compromise.

Although we have selected the HOG descriptor [14] as a candidate to test for the performance of the dilation methods,

any other binary shape feature descriptor (e.g., Haar features, SIFT/SURF descriptors, AP/BAP features [15] etc.) would be equivalently appropriate to use.

D. Handwritten Digits Recognition: SGP as an input to CNN

According to [20], using eight layers in CNN gives promising results for the recognition problems so eight layers have been used to recognize handwritten characters and different shape objects in this paper. The model contains eight learned layers, namely, five convolutional and three fully-connected [20]. The architecture of the overall CNN model is shown in Fig. 2(b). Unlike the existing use of CNNs in the literature (see, Fig. 2 (a)), in this paper, feature vectors of binary images are first extracted from the different multi-pass dilations using the DTBIM [10], kernel based dilation [17] and the isotropic dilation [17]. Namely, the HOG [14] descriptors are retrieved from each dilated image. As a result, there will be more than one feature map ($n \times 24 \times 24$). Then feature maps are zero padded to match the original image size.

By referring to Fig. 2 (b), the preprocessing phase can be formulated as follows: Let X be the binary input image and B_i be the dilated four binary images where $i = 1, 2, 3$ and 4. The feature maps F^i of the dilated binary images are obtained using HOG and each feature map is of size 28×28. Subsequently, one $2D$ weight matrix w (sets to 0.25) is applied to each obtained feature map which are then fused as follows:

$$h = \sum_{i=1}^{4} \sum_{x=1}^{H} \sum_{y=1}^{W} (F^i(x,y) * w(x,y)) \qquad (2)$$

Where x and y denote the coordinates of F^i, H and W are the height and width of F^i. Thus, four feature maps are fused in equation 2. The resulting fused feature map after zero-padding is of the same size as the original (28×28) and it is used as an input for the CNN method. In the convolution process, the kernel filter strides with one step over the fused map h to estimate the kernel's central values until it reaches the end of h. After that, the pooling layer is applied with the size of 2×2 to down-sample the convolved h spatially in both directions. In this work, the convolution process has been set to $k = 5$ times.

The obtained convolved features are used as input for the fully connected neural network model. The network model is based on the backpropagation which allows updating the weights between the nodes in the network model by decreasing the error rate. Thus, optimal weights which interconnect the nodes will be automatically formulated for the recognition problem.

E. Binary Shape Recognition: SGP as an input to CNN & RF

We applied each of the dilation methods to eventually yield multi-level feature vectors. Then, in addition to using the CNN classifier, we also trained the RF classifier on these sets of features [22]. The aggregation process in RF is carried out to mitigate the effects of over-fitting and to improve generalization. The number of trees was set to 500 in the constructed model. Based on the findings of Leo Breiman, the number of trees to grow in each iteration was recommended to be 500 (in which the out-of-bag (OOB) error rate was the lowest). Besides any inferior performance of RF is more likely to be a result of the data characteristics rather than the number

of trees used, additionally random forests do not overfit as more trees are added, see [21]. The extracted features, denoted by HOG++ in Table II, are: the standard deviation, the entropy, the ratio of the median over the standard deviation of the HOG descriptor, in addition to the shape properties (eccentricity, area, and solidity). In total, we represent each shape pattern in the data set by 24 values (i.e. 6 statistical values from the original shape and subsequently from each of the 3 dilation passes, see Fig. 2 (c)).

III. EXPERIMENTAL ENVIRONMENT

A. Experimental Setup

Fig. 2 depicts the three distinct architectures which we followed to execute the methodology. As mentioned earlier in sections I and II, we are using CNN and RF as test beds for the effectiveness of the proposed approach in enhancing the recognition accuracy rate of the different classifiers. Two dilation techniques, described in section I, are chosen for comparison with our proposed method in terms of classification accuracy. In the experiments, we also incorporate the DTBIM algorithm [10]. Unlike the other dilation algorithms, the DTBIM algorithm dilates shapes without severely deforming their structure. It is important to note that Fig. 2 (a) illustrates the state-of-the-art usage of CNN in conjunction with image-based pattern recognition. The classifiers shown in (Fig. 2 (a), (b) and (c)) are contrasted in Tables I and II.

B. Data Sets

In order to evaluate the performance of the recognition algorithms, we have used two different data sets which are:

- **Data set I**: The MNIST[1] (comprising a training set of 60,000 grayscale images, and a test set of 10,000 grayscale images).

- **Data set II**: The Shape Kimia-216[2] (comprising a training set of 300, and a test set of 100 images).

The first data set has one channel grayscale images with the size of 32×32 and the latter one has binary images with an average size of 141×141. Note that, for the Data set I, we used different sets of images to train the CNN model. Namely, in **case I**: 1,000 images, in **case II**: 5,000 images and in **case III**: 10,000 images have been used. In each of the three cases (I, II, and III), the corresponding number of image samples were selected from each digit. For instance, in **case I**, the first 100 images from each label of the ten digits were selected. Thus, there will be 1000 training images in total for **case I**. By using the CNN for all cases, we can analyse the behaviour and performance of the constructed CNN model trained on fewer samples.

IV. RESULTS AND DISCUSSION

A. Handwritten Digits Recognition (Data Set I)

As shown in Table I, the best accuracy results of recognition are obtained using the SGP: DTBIM Dilation in all of the three cases (see the description of cases in section III) and the

---

[1] http://yann.lecun.com/exdb/mnist/

[2] http://vision.lems.brown.edu/content/available-software-and-databases

lowest results are obtained using the CNN with direct input of grayscale images. According to the results shown in Table I, using DTBIM dilation in the preprocessing part increases the accuracy rate of recognition of digits in the MNIST data set to 78.8%, 86.6% and 90.6% for case I, II and III, respectively. By observing the increase in accuracy as a function of the trained samples' size (Table I), we can easily infer that DTBIM Dilation has achieved the accuracy of grayscale images (in Case III - 10,000 samples) with merely half of the sample size (Case II).

TABLE I: Accuracy results of the methods using CNN on the Data set I (MNIST). Case I: 1,000 images, case II: 5,000 images and case III: 10,000 images.

Method	Accuracy % Case: I	Accuracy % Case: II	Accuracy % Case: III
Input: 2D Image			
Single Grayscale image	71.4	80.1	86.2
Single Binary image	76.1	84.5	85.6
Input: SGP (features from multi-dilations)			
SGP: Disk Dilation	75.8	83.7	86.2
SGP: Isotropic Dilation	73.4	85.2	87.4
SGP: DTBIM Dilation	78.8	86.6	90.6

B. Binary Shape Recognition (Data Set II)

Since the data set II has a limited number of samples, it becomes a suitable platform to exercise the augmented feature space using SGP which we advocate for. Table II presents the accuracy results of using different dilation methods with the CNN and RF to recognize shape objects in the data set II. Based on the results, it is apparent that DTBIM dilation still outperforms the other binary dilation methods when using either RF or CNN, on one hand, and the CNN when images are inputted directly, on the other hand. The main reason is that the performance of machine learning methods can be boosted when shapes or digits are uniquely characterised by a set of features; in our case it is the feature set that captures the SGP. The CNN's internal feature extraction mechanism is unable to reach this hybrid dimension of SGP on its own.

TABLE II: Accuracy results of the methods using CNN and RF on the shape Data set II (Shape Kimia-216).

Method	Accuracy % CNN	Accuracy % RF (HOG++)
Input: 2D Image		
Single Binary image	77.13	78.14
Input: SGP (features from multi-dilations)		
SGP: Disk Dilation	79.88	80.97
SGP: Isotropic Dilation	80.26	78.54
SGP: DTBIM Dilation	83.89	81.78

C. Resilience to Noise

Another advantage of using DTBIM, is its quasi-immunity to noise as compared to the other common dilation methods. In order to verify that, we exposed the test images to additive noise between each dilation pass with various density values. Subsequently, we extracted the HOG features and tested the classifier performance using the model which had been trained without exposure to noise. The added salt & pepper white noise's density values where set to: $d = (1, 2, 3, 4, 5) * 10^{-3}$.
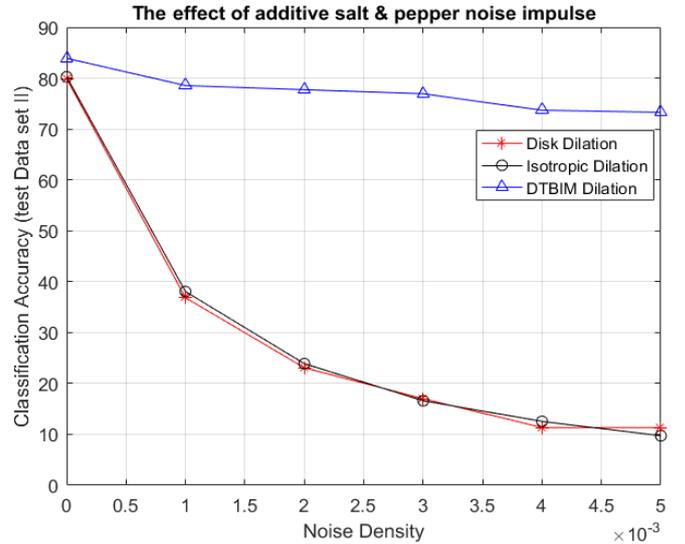

Fig. 3: Another intrinsic property of DTBIM; noise resilience. Shown are the accuracy results of the different dilation methods after exposure to additive noise. All methods were trained on the RF.

The results shown in Fig. 3 deduce that DTBIM is less prone to the additive noise which eventually helps in the task of pattern recognition.

It is clearly seen that the highest accuracy rate of recognition is estimated when the DTBIM dilation method is used in the recognition task. As a result, the DTBIM-based SGP is a more stable method against noise as compared to using other dilation methods to accurately recognize binary objects.

V. CONCLUSION

In this paper, different successive dilation methods with the HOG descriptor have been applied to binary images in order to improve the performance of pattern recognition algorithms. Two different machine learning techniques, namely, the CNN and the RF, have been used to classify binary shapes and handwritten digits. The successive dilation is meant to capture what we termed herein the shape growth pattern (SGP).

The results show that using SGP is deemed important in the pattern recognition field. Furthermore, we can see that DTBIM has outperformed other common dilation methods in capturing the very essence of SGP. For example, when trained using CNN (Table I), SGP with DTBIM's accuracy reaches to 90.6% as compared to disk-based SGP (86.2%) and isotropic-based SGP (87.4%) on the handwritten digits -Case III-. Our approach also shows better performance on the shape data set (Table II) where it reaches to 83.89% accuracy as compared to the former two methods which produce 80.97% and 80.26% accuracy, respectively. Moreover, DTBIM-based SGP also improves upon the CNN and RF classifiers when they are trained in the conventional way (direct 2D image input to the CNN and HOG features input to the RF). The CNN with 2D grayscale image input reaches an accuracy of 86.2% (data set I) while the RF with HOG features reaches to an accuracy of 78.14%. The SGP augments the feature bank and thus becomes a useful approach especially when training on small data sets.

Finally, it is also observed that the proposed method is less prone to additive salt & pepper noise unlike the other dilation methods whose performance suffer dramatically in the presence of additive impulse noise.

REFERENCES

[1] R. Jain, R. Kasturi, B. G. Schunck, "Machine Vision," *McGraw-Hill 1st edition*, 1995.

[2] M. Z. Pwint, T. T. Zin, M. Yokota, M. M. Tin, "Shape descriptor for binary image retrieval," *In Proc. IEEE 5th Global Conference on Consumer Electronics*, pp. 1-14, 2016.

[3] RX Hu, W. Jia, H. Ling, Y. Zhao, and J. Gui, "Angular pattern and binary angular pattern for shape retrieval," *IEEE Transactions on Image Processing*, vol. 23, no. 3, pp. 1118-27, 2014.

[4] D. Zelenika, J. Povh, B. Enko, "Text detection in document images by machine learning algorithms," *In Proc. of the 9th International Conference on Computer Recognition Systems*, vol. 403, pp. 169-179, 2015.

[5] A. A. Desai, "Support vector machine for identification of handwritten gujarati alphabets using hybrid feature space," *CSI Transactions on ICT*, vol. 2, no. 4, pp. 235-241, 2015.

[6] A. T. Jamal, N. Nobile, C. Y. Suen, "End-Shape recognition for Arabic handwritten text segmentation," *In Proc. 6th IAPR TC 3 Int. Workshop on Artificial Neural Networks in Pattern Recognition*, vol. 8774, pp. 228-239, 2014.

[7] M. Khayyat, L. Lam, C. Y. Suen, F. Yin, C. L. Liu, "Arabic handwritten text line extraction by applying an adaptive mask to morphological Dilation," *In Proc. 10th IAPR International Workshop on Document Analysis Systems (DAS)*, pp. 100-104, 2012.

[8] M. N. Abdi and M. Khemakhem, "A model-based approach to offline text-independent Arabic writer identification and verification," *Pattern Recognition*, vol. 48, no. 5, pp. 18901903, 2015.

[9] C. C. Han, H. Y. M. Liao, G. J. Yu, L. H. Chen, "Fast face detection via morphology-based pre-processing," *Pattern Recognition*, vol. 33, no. 10, pp. 1701-1712, 2000.

[10] A. Cheddad, "Structure preserving binary image morphing using Delaunay triangulation," *Pattern Recognition Letters*, vol. 85, pp. 8-14, 2017.

[11] L. F. Costa, R. M. Cesar, "Shape Analysis and Classification: Theory and Practice," CRC Press 2009.

[12] C. R. Maurer, R. Qi, and V. Raghavan, "A linear time algorithm for computing exact Euclidean distance transforms of binary images in arbitrary dimensions," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 25, no. 2, pp. 265-270, 2003.

[13] J. D. Boissonnat, J. P. Pons and M. Yvinec, "From segmented images to good auality meshes using Delaunay refinement," I*n: Emerging Trends in Visual Computing*, Berlin: Springer, pp. 1337, 2009.

[14] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," *In Proc. of the IEEE conf. on computer vision and pattern recognition*, pp. 886893, 2005.

[15] G. Hu, Y. Yang, D. Yi, J. Kittler, W. Christmas, S. Z. Li, and T. M. Hospedales, "When face recognition meets with deep learning: an evaluation of convolutional neural networks for face recognition," *In Proc. ICCV Workshops*, 2015.

[16] F. Y. Shih, "Image Processing and Mathematical Morphology: Fundamentals and Applications," CRC Press, 2009.

[17] M. S. Nixon and A. S. Aguado, "Feature Extraction and Image Processing for Computer Vision," *Elsevier Ltd.*, 3rd Ed, pp. 241, 2012.

[18] J. Fan, W. Xu, Y. Wu, and Y. Gong, "Human tracking using convolutional neural networks," *IEEE Transactions on on Neural Networks*, vo. 21, no. 10, pp. 1601-1623, 2010.

[19] M. Liang and X. Hu, "Recurrent convolutional neural network for object recognition," *In Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3367-3375, 2015.

[20] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *In Advances in Neural Information Processing Systems*, pp. 11061114, 2012.

[21] A. Liaw and M. Wiener, "Classification and Regression by Random Forest," *R News*, vol. 2, no. 3., pp. 18-22, 2002.

[22] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, pp. 532, 2001.