

Using Derived kernel as a new Method for Recognition a Similarity Learning.

Ramadhan A. M. Alsaidi, Ayed R.A. Alanzi, Saleh R. A. Alenazi, Madallah Alruwaili

Abstract: A new technique for feature withdrawal by neural response is going to be familiarized in this research work by merging an entropy measure with Squared Pearson correlation Coefficient (SPCC) method. The process of choosing effective models on the basis of entropy measures was proposed further to enhance the ability to select templates. For more accurate similarity measure we used the statistical significant relationship between functions. The research illustrate that the proposed method is proficiently compared with the state-of-the-art methods.

Keywords : Feature Extraction; Hierarchical Learning; Entropy Measures; Pearson Correlation Coefficient; Pooling Operation; Sample.

I. INTRODUCTION

Since a long time, feature extraction task has been at a great attraction for the researcher in computer vision, machine to learn, and to recognize the pattern. In the field of image classification, significant progress has been observed recently [1–9]. Also a random related clutter and occlusion in input images generate the exact classification of interested items at engaging end [10]. Similarity measure is one of the most notable and important issues in applications such as image classification. These are important factors and challenging tasks in the image organization. Consequently, the main thing is the selection of more effective measure is to make the error rate of classification method lesser [10–12]. So many different studies have attracted them to the similarity measure among features in the last few decades. These studies comprise three approaches such as supervised [13], semi-supervised [14, 15] and unsupervised [16, 17]. In semi-supervised approach, Smale et al. in [18] anticipated an idea of derived kernel based neural response (NR) feature to present image naturally as a hierarchically built in light of neuroscience of visual cortex, and he defined the derived kernel as an internal product by the neural response [18–20]. It has the probability to further advanced and improved the several tasks of pattern recognition [21, 22]. Researcher has been attracted by Derived kernel based neural response (NR), and it has verified to be a powerful solution to a wide range of application fields,

specifically, in image processing, machine learning, and computer vision. In [11], Hong Li et al. suggested a pattern selection technique for the similarity measure capability of the derived kernel based neural response, when [23] presenting a hierarchical feature extraction method called local neural response (LNR) to recognize. Another technique is suggested in [24, 25] to a feature extraction scheme using neural response by linking the hierarchical architectures with the sparse coding. An image classification algorithm as a semi-supervised sparse based neural response was examined in [26]. Hierarchical sparse feature based on neural response introduced in [27] to choose sufficient templates. Certainly, the common method of the neural response is approving its dimensionality only on the number of patterns at the layer which is considered by neural response and the magnitude of the input image is dismissed. According to the neural response, the resemblance among the images of different sizes can directly calculate same like the derived kernel measures the similarity between images without subdivision. A training set of templates is recognized to be a main semantic module in the derived kernel to obtain the templates for an image set where randomly extracting image patches [18]. During the template modules are resampled, many variable resemblances are resulting at the randomly sampling, specifically when the quantity of templates is lesser, as well as the large number of templates as randomly chose. It does not always provide a better similarity measure and furthermore, it leads to rise in the running time. The aim of this research work is to solve the above issues and address it by improving the ability to distinguish in picking effective models based on entropy measures. Where Shannon entropy[28] plays a considerable role by computing several templates. Therefore leading to provide an efficient way to evaluate the insight properties of the neural response. Moreover, statistically major connection between the nominated templates to advance the ability of the similarity measure for derived kernel is performed by the Pearson correlation coefficient method. The paper is organized as following; introducing the necessary background for the derived kernel in Section II and Section III gives the introduction and analyses some basic characteristics and terminologies used in this study. By explaining both, the statistically significant similarity measure and selecting effective templates method, a planned technique in Section IV is described.

Revised Manuscript Received on February 10, 2020.

* Correspondence Author

Ramadhan A. M. Alsaidi, department of Mathematics, Jouf University, Gurayat, Saudi Arabia. Email: rsaidi@ju.edu.sa

Ayed R. A. Alanzi*, department of Mathematics, Majmaah University, Majmaah 11952, Saudi Arabia. Email: a.alanzi@mu.edu.sa

Saleh R. A. Alenazi, Computer Technology department, Tabuk College of Technical, Tabuk, Saudi Arabia. Email: sala207@hotmail.com

Madallah Alruwaili, College of Computer and Information Sciences, Jouf University, Skaka, Aljouf, Saudi Arabia. Email: madallah@ju.edu.sa

In Section V, We have confirmed the efficiency of the method that we assume with widespread experiments on recognized a set of data and then matched it by algorithms. Thus, we conclude and give some future study by the end of this work.

II. MODEL STRUCTURE

The first assumption in hierarchical learning algorithms are as every input image is possibly divided to a hierarchy of segments growing the size with semantic complication respectively. Indeed, the hierarchy is suitable to lessen the model complexity of the issue. Shown here is the similarity of the smallest segments, the ordered architecture repeatedly develops at each patch a similarity estimate with the following greater parts by the definition of the EPCC and combining processes. Naturally, the combining process induces similarity in the architecture whereas the SPCC step brings discrimination. Consequently, the interchanging presentations of the SPCC and pooling process produces a complex resemblance ratio of the contribution data with major differences and similar characteristics.

A. Knowledge of Categorization

The following is a brief description of the calculated framework, shown in [24, 25, 27], using the hierarchical model of the construction, every input here is consisted of segments of growing size.

1. Categorized Architecture

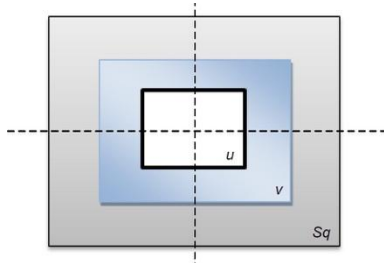


Fig. 1. Nested layers' domain

Here categorized architecture is comprises of three strata of spots in planned structure u , v and S_q in \mathbb{R}^2 , through $u \subset v \subset S_q$, that the study will be four-sided, centered and axis stretched (see Figure 1). We used a function space at each object nominated by $Im(S_q)$, $Im(v)$ then $Im(u)$ applied on the shapes S_q, v then u .

2. Conversion

A set of H_i Conversion that that provides a relation among the affiliated layers is undertaken in this study. Furthermore, it studies the model of H_i conversion which are the shapes from the minimum objects to the following maximum objects $h : u \rightarrow v$, and consequently H_i . Transformation $h \in H_i$ otherwise $h \in H_i$ simply involves restriction should be noted. such as, if $f \in T_v$ is an shape patch of magnitude v then

$h \in H_i$, and $f \circ h$ is a portion of the picture spot of size u (see Figure 2).

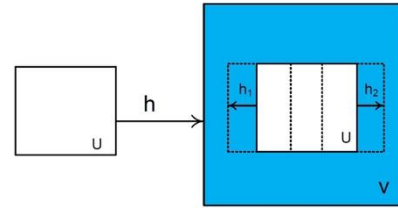


Fig. 2. Translation starting u to v , when h_1 and h_2 parallel to two diverse translations $h : u \rightarrow S_q$.

3. Patterns

The hierarchical construction is having a basic semantic module; the patterns which are mostly learned from information are an exercising set of patterns. Through the proposed model the templates are going to play the role of the similarity. Furthermore, it connects the architecture with the fundamental division of entry data. Another new technique to produce pattern is to model of the possibility circulation on the function space arbitrarily. Unfixed sampling will take it to various similarities when the patterns models are reshaped, specifically when the number of templates are minimum. Whereas, a maximum number of randomly chosen shapes do not always take it to better resembling calculation and the running time also grows. the important template choosing method will develop the similarity ratio ability of the resulting kernel. Thus, a template choice method is anticipated in this paper rely on entropy measures. The anticipated method makes use of the discriminative data in the image place and takes it to a small number of templates. The theoretical examination and tests performed on real data describe that the proposed algorithm is flexible to object recognition task and can increase recognition accuracy and competence. At last, the paper describes the pattern sets as $T_u \subset Im(u)$ and $T_v \subset Im(v)$, which are used to fix, separate, and capable with the uniform possibility measure (see Fig. 3). Basically, the output for the sparse depiction features relies mainly upon a good choice of experiment set. Investigational result illustrates that for some sets of signals, learned experiment sets can be profited from pattern sets, which eventually lead to a same or better recognition result as compare to other conventional procedures.

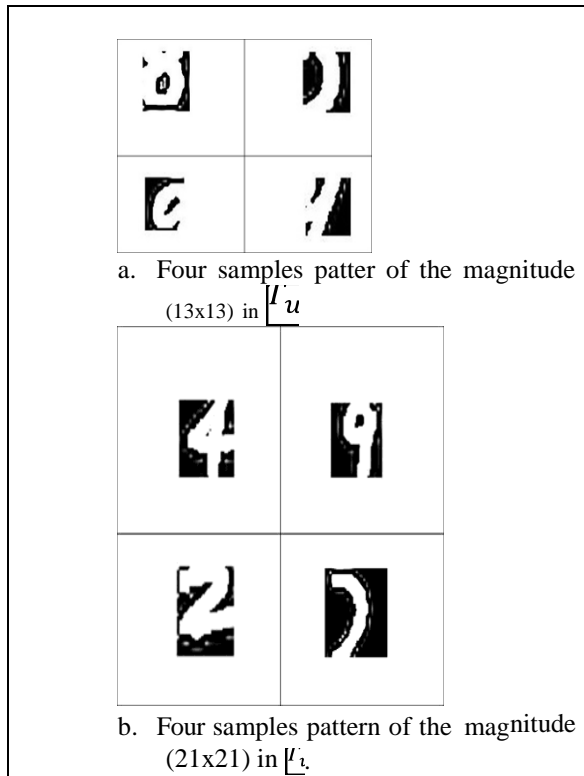


Fig. 3. Eight samples from the group of patterns (MNIST)

III. ESSENTIAL ELEMENTS FOR PROPOSED MODEL

A. Entropy

In object reshaping, object registration and the object division Entropy has been shown at the large range of presentation fields. Actually, around 1865 the experimental study in entropy concept was started, when Rudolf Clausius invented the entropy for the calculation of the quantity of energy in a thermodynamic method. Claude Shannon anticipated “A Mathematical Theory of Communication” In 1948. This theory was highly appreciated at the beginning and to model information sources this theory was used by him. The Shannon’s entropy of distinct arbitrary flexible X has a probability distribution $p_x = (p_1, p_2, \dots, p_n)$ then expressed as:

$$H(X) = H(p_x) \triangleq \sum_i^n p_i \log \frac{1}{p_i} \quad (1)$$

IV. ENTROPY OF AN IMAGE

Entropy model increases in the learning of physics in the heat engines which calculates the amount of system condition. To recognize the spread of states which is used to approve a system is another technique to describe entropy. If in a shape, these conditions resemble to gray stages then definite pixels can be accepted. For instance, in the 8-bit pixel 256 same states are there. The variety of states is larger if all these spots are occupied similarly, as in the case of flawlessly histogram shown as in the pattern in the entropy of the shapes. Whereas, if the image is adjacent, then only two sides are covered and entropy remains

less. Certainly, the shape is consisted of pixels, then every pixel takes a worth. The entropy is nothing if the worth of the pixels is alike. Figure 3 describes the original shape and the diagram. The given histogram is calculation of the fundamental possibility of pixel strengths, such as: $p(i) = h_{ii}(i)/N$, where $h_{ii}(i)$ describes the histogram access of power value i in shape U then N is whole number of pixels U . for this pattern expansion, the measurement of the entropy of shape is like.

$$H(U) = \sum_i h_{ii} \log N / h_{ii}(i) \quad (2)$$

A. Assortment of explanatory pattern

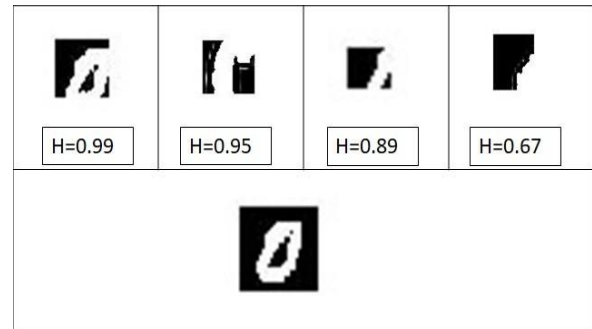


Fig. 4. Four descriptive examples of templates displaying the growing quantity of data exposed about the shape.

The main aim of view choice is to routinely propose best or supreme shapes which exploit the quantity of data gathered in the experimental set of the provided shapes dataset. A fine templates reveal important information regarding existing information. Hence, displaying that sooner to spectators can increase equally the speed and the effectiveness of data accepting. For instance, in Fig. 5, three descriptive scenes of patterns with various amounts of data. Visibly, the 1st one on the right side parallels to the finest view that displays the extreme quantity of data by showing the object in the less unclear way. Also projected patterns choice method is defined as: Let’s organize first template set is arbitrarily skimming image layers from the object set of size u and v . remember the size of u remains comparatively smaller because it has only the fundamental components of the shapes so the outcome of the pattern of size u is represented as

$$T_u = \{t_1^u, t_2^u, \dots, t_{q^u}^u\}$$

(here q^u be there the cardinality of T_u then $T_u \in I_u$).

whereas, the elementary pattern of size v and (here size of v is enough to bring different model) shown in

$$T_v = \{t_1^v, t_2^v, \dots, t_{q^v}^v\}$$

(here q^v be there the cardinality of T_v and $T_v \in I_v$)(To

generate the terms of pattern u then v vibrant, the sample is shown in Fig. 3). No doubt, T_u then T_v take high variance as the pattern are resamples, specially, where q^u and q^v are smaller. In contrast, a large amount of q^u and q^v are the reason to grow at the run time not always take to better pixels.

Building of compact pattern sets by superior discriminative capability is required. Our aim was to choose the patterns models \overline{P}_u and \overline{P}_v from the basic pattern models \overline{T}_u and \overline{T}_v such as:

$$\overline{P}_u \subset \overline{T}_u \subset I_u \text{ and } \overline{P}_v \subset \overline{T}_v \subset I_v$$

1. Stages to choose \overline{P}_u from \overline{T}_u

- Calculate the rate of entropy of separate candidate pattern $\overline{T}_i^u (i = 1, 2, \dots, q^u)$.
- Sort \overline{T}_u in soaring sequence rendering to their entropy.
- Choose the final \overline{T}_u candidate pattern in order to build \overline{P}_u .

Similarly, we may choose \overline{P}_v after \overline{T}_v .

Algorithm 1 Building of \overline{P}_u

Contribution: Full $\overline{T}_u = \{t_1^u, t_2^u, \dots, t_{q^u}^u\}$ then the number of pattern set d .

Output: \overline{P}_u

- (1) count=1
- (2) $\overline{P}_u = \emptyset$
- (3) $t_k = \text{argmax}_{t_k \in \overline{T}_u} H(t_k)$
- (4) $\overline{P}_u = \overline{P}_u \cup t_k$
- (5) $\overline{T}_u = \overline{T}_u \setminus t_k$ and
count = count + 1
- (6) if count < d go to 3
- (7) besides
- (8) $\overline{P}_u = \overline{P}_u \cup \emptyset$
- (9) finish if

Consequently, it is notable that only at the phase of selecting pattern, there is a variation then the remaining steps remain same as [24, 25].

B. Pearson Correlation Coefficient (PCC)

In the statistical analysis, pattern identification and image processing the Pearson's method is commonly used. Applications on the latter comprise equating two images for the purpose of object registration, difference of measurement, etc [29]. It is illustrated in

$$r = \frac{\sum(x_i - x_m)(y_i - y_m)}{\sqrt{(\sum(x_i - x_m)^2) \sum(y_i - y_m)^2}} \quad (3)$$

here \overline{x}_i and \overline{y}_i are the strength of the i^{th} pixel in image 1 and image 2, respectively, \overline{x}_m and \overline{y}_m are the mean intensity of image 1 and image 2, respectively. In this study we deal with the squared Pearson correlation coefficient (SPCC) r^2 . The SPCC indicates the strength of the linear relationship between the two subimages \overline{x} and \overline{y} . If $r^2(x, y) = 0$, then \overline{x} and \overline{y} are said to be dissimilar. The nearer the value of $r^2(x, y) = 1$, the stronger likeness between the two subimages. On the other hand, SPCC is known to be an example of reproducing kernels.

V. THE PROPOSED METHOD

About the image recognition, the recommended outline consists of three patches: Formation Patch, Initial Operation patch then Matching patch. This method describes an indirect mapping after the below patch to the \overline{p} above patch. This mapping method has remained as an indirect mapping purpose \overline{f} for instance:

$$\overline{N}^v = \overline{f}(x^v, T^u)$$

Here, the major nonlinear mapping function f comprises of a series of methods such as: $\overline{f} = p \circ q$, here \overline{p} shows combining process (Combining process can be considered as mean, maximum, and energy, though, in this specific research work, we are dealing only with maximum combining.) then \overline{q} describes PCC function that can be used as a resemblance measure:

$$\overline{r} = \overline{g}(x, \hat{x})$$

where \overline{x}, \hat{x} are the column shows $\in I_u$, and \overline{r} is the Pearson correlation coefficient (PCC). Furthermore, it will be more appropriate to work with the squared Pearson correlation coefficient (SPCC) \overline{r}^2 . As we claimed that the SPCC provided an indication on the strength of the linear relationship between the two sub-images \overline{x} and \hat{x} . If $\overline{r}^2(x, \hat{x}) = 0$, then \overline{x} and \hat{x} are considered separate. The nearly matching the value of $\overline{r}^2(x, \hat{x}) = 1$, the stronger the correlation between the two sub-images. The importance of this connection rely on the fact where function \overline{f} is capable of mapping a set of feature paths from the lower patches to a solitary feature path at the upper patches. Reader is recommended to consult [24, 25, 28] for further detail treatment.

A. Fundamentals Stage \overline{u}

The formation layer mostly deals with the building of two models (First and Second layers) at two sub-layers which are the first and the second layer Formation respectively. To start, the primary stages deals with description of patterns models \overline{t}_i of mass \overline{u} by arbitrarily anchoring \overline{n} , then we use entropy scheme to choose the small size of sub-image layers of dimension \overline{u} from every gray-scale contribution pictures. The second stage dealing with the uniting of sub-images to an individual matrix \overline{T}_u , i.e.:

$$\overline{T}_u = [t_1, t_2, \dots, t_M] = \begin{pmatrix} t_{1,1} & \dots & t_{M,1} \\ \vdots & \ddots & \vdots \\ t_{1,N} & \dots & t_{M,N} \end{pmatrix} \in \mathbb{R}^{N \times M}$$

Here $\overline{t}_i (i = 1, 2, \dots, M)$ remains the i^{th} member then $\overline{T}_u \subset Im(u)$. Now \overline{N} stands the measurement of every feature path \overline{t}_i (remember these patterns \overline{t}_i in \overline{T}_u would be altered into paths) then \overline{M} is the whole number of each pattern set from whole categories. whereas, \overline{t}_i can be thought as denoting feature vector of class \overline{i} .) all template set from all categories. \overline{T}_u is supposed to be an over-complete training set when $\overline{N} \ll \overline{M}$.



Similarly, second patch patterns can be built \overline{T}_v with size variation \overline{v} such as:

$$\overline{T}_v = [t_1, t_2, \dots, t_L] = \begin{pmatrix} t_{1,1} & \dots & t_{L,1} \\ \vdots & \ddots & \vdots \\ t_{1,K} & \dots & t_{L,K} \end{pmatrix} \in \mathfrak{R}^{K \times L}$$

For instance: \overline{L} indications of quantity \overline{K} . In the meantime, the finest measurement of the patterns is defined experimentally. In contrast, the training set is marked as:

$$\overline{x} = [x_1, x_2, \dots, x_q]$$

here x_j ($j = 1, 2, \dots, q$) stands the j^{th} training picture then $\overline{x} \subset Im(S_q)$

B. Basic Process Patch (\overline{v})

This patch comprises of two processes for example. Basic SPCC and local pooling

1. Basic SPCC

The stages are revealed here:

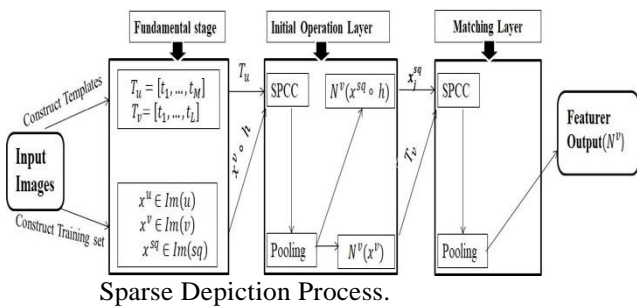
- Having a training picture $x^v \in Im(v)$.
- Creation of picture bits of dimension \overline{u} using subsequent connection:

$$\overline{x}_j^v = \{x^v \circ h_j, j = 1, 2, \dots, |H_u|\}$$

It is necessary to note that every such alteration h_j requires a means to divide a sub-image into a group of smaller patches. In contrast, we don't need $|H_u|$ comprising all possible alterations $h: u \rightarrow v$.

- Assess the connecting factors of two pictures layers x_i^v then x_j^v using SPCC.

Fig. 5. Patched Proposed Categorized building of



Feature path \overline{x}_j^v can be described as:

$$\overline{S}^v = [s_1, s_2, \dots, s_{|H_u|}]$$

2. Local Pooling Stages

After applying the meagre coding, local pooling is needed to execute on SPCC \overline{S}^v (like. the pooling process, process row-wise on the medium of SPCC \overline{S}^v) consistent to the similar atomic indication in \overline{T}_v as:

$$\overline{N}^v(x^v)(t) = \max\{S^v : x^v \in Im(v), t \in T_v\}$$

Therefore, the basic patch neural reply of training picture x_i by a provided candidate pattern $t_i \in T_v$ stands by $\overline{N}^v(x_j^v)(t_i)$.

C. Similar Patch (\overline{S}_q)

Once more, this patch function similarly to the above basic process patch.

1. SPCC Application

The stages comprise are specified as:

- (1) Substitute of the training template \overline{T}_u by $\overline{N}^v(T_v)$ (such as, since we have been studying the sequence we utilized the result of the higher patch).
- (2) Having a gray-worth input picture $x^{S_q} \in Im(S_q)$.
- (3) Dispersal of picture x^{S_q} into bits of picture layer of mass \overline{v} by using the relation below:
$$\overline{x}^{S_q} = N^v(x^{S_q} \circ h_j), j = 1, 2, \dots, |H_v|$$
- (4) measurement related to a SPCC for each bit $x_i^{S_q}$ by applying:

$$s_j = g(x_j^{S_q}, N^v(T_v)) = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_L \end{pmatrix}$$

Repeating by the finale of SPCC we found:

$$\overline{S}^{S_q} = [s_1, s_2, \dots, s_{|H_v|}]$$

2. Universal Pooling Stage

The last stage is about the calculation of a universal extreme general measures of \overline{S}^{S_q} like:

$$\overline{N}^{S_q}(x^{S_q})(t) = \max\{S^{S_q} : x^{S_q} \in Im(S_q), t \in T_v\}$$

When, \overline{N}^{S_q} denoted a fresh feature. Practice these methods to both training then experiment, a new model of features and recognition is found in the fresh feature space. This practiced method is concluded in Figure 6.

VI. RESEARCH AND ARGUMENTS

In this research work, we practiced proposed technique over the two fields images and discourse. In the field of picture, we assessed categorized sparse on the recognized M-NIST digit acknowledgment benchmark. Additionally, we castoff the images like feature extraction stage aimed at an organization algorithm i.e. Support Vector Machines (SVM) then 1-Nearest Neighbors (1NN).

A. Image Recognition

1. Handwritten digit recognition

First of all, we confirmed the developed techniques with item recognition experiments with MNIST hand-written numeral recognition benchmark, and we found $\overline{70,000}$ of data samples, and everyone is $\overline{28 \times 28}$ grayscale



Picture. Also, during the research, we took some images arbitrarily chosen after the MNIST data samples. After that, we took eight modules of pictures: $\lfloor 2 \rfloor$ to $\lfloor 9 \rfloor$ (see Figure 7). The numbers in this data have a minor amount of ordinary conversion, alternation, measuring, distributing and other twists like one may be hopping to find in a corpus covering the handwriting of mortal themes. We have used marked images set having 5 examples per class, on the other hand the non-sample test groups have 30 samples per class.



Fig. 6. Eight samples from the set of experiment examples. Images here have been ftatic at $\lfloor 28 \times 28 \rfloor$ pixels.

Here $\lfloor T_n \rfloor$ and $\lfloor T_n \rfloor$ stand pattern sets created by arbitrarily withdrawn 500 pictures layers (of size $\lfloor u \rfloor$ then/or $\lfloor v \rfloor$) through pictures, that are neither used in the experiment else experiment sets (In the test we use $\lfloor D = T_n \rfloor$). In the numeral’s dataset, patterns magnitude $\lfloor 10 \times 10 \rfloor$ pixels are bigger enough to enhance semi-circles then separate stroke joints, whereas larger patterns, near $\lfloor 20 \times 20 \rfloor$, can be seen to add closely complete, numbers were in more discriminative construction is existing. To investigate, we use the first patch pattern $\lfloor u = 11 \times 11 \rfloor$ pixels then the 2nd patch pattern $\lfloor v = 19 \times 19 \rfloor$ pixels.

Table 1. The investigational results with MNIST dataset.

Fea. Ex. Meth.	SVM	SVM
Smale[18]	0.6793	0.7867
Sparse[24]	0.8277	0.8288
Developed method	0.86	0.86

When the features are extracted we can achieve the classification exactness by the application of a $\lfloor k - NN \rfloor$ through $\lfloor k = 1 \rfloor$ then SVM remember that the classifier are average above 50 random experiment pattern, having the training and pattern fixed sets. The research is accomplished beneath the same Smale’s environment; assessment consequences describes that the established model constantly outclasses the Smale approaches as displayed in Table 1.

VII. CONCLUSION

This research work built a new neural response algorithm established on SPCC similarity measure and Smale’s structure to study pictures and discourse. After the experiments it is proved that the planned algorithm is applicable to various object recognition tasks. It helps us to disperse tasks and divide into subtasks. This technique provides enough correctness for image recognition task (containing MNIST handwritten dataset); it guides to more fixed, discriminative image and speech representation. Furthermore, in future adaptive filtering methods can be used to eliminate the noise from image so accuracy can be enhanced. In addition to this, we would also encourage the idea to advance this work to the mission of other language’s sound recognition to find out how much gain in the classification can be altered to phonetic recognition.

ACKNOWLEDGMENT

ARA ALANZI WOULD LIKE TO THANK THE DEANSHIP OF SCIENTIFIC RESEARCH AT MAJMAAH UNIVERSITY FOR SUPPORT AND ENCOURAGEMENT.

REFERENCES

1. S. Bahrampour, N. M. Nasrabadi, A. Ray, and W. K. Jenkins, “Multimodal task-driven dictionary learning for image classification,” *IEEE transactions on Image Processing*, vol. 25, no. 1, pp. 24–38, 2015.
2. T. Liu and D. Tao, “Classificationliu2015classification with noisy labels by importance reweighting,” *IEEE Transactions on pattern analysis and machine intelligence*, vol. 38, no. 3, pp. 447–461, 2015.
3. X. Li, X. Zhao, Z. Zhang, F. Wu, Y. Zhuang, J. Wang, and X. Li, “Joint multilabel classification with community-aware label graph learning,” *IEEE Transactions on Image Processing*, vol. 25, no. 1, pp. 484–493, 2015.
4. Y. Li, X. Shi, C. Du, Y. Liu, and Y. Wen, “Manifold regularized multi-view feature selection for social image annotation,” *Neurocomputing*, vol. 204, pp. 135–141, 2016.
5. Y. Luo, D. Tao, K. Ramamohanarao, C. Xu, and Y. Wen, “Tensor canonical correlation analysis for multi-view dimension reduction,” *IEEE transactions on Knowledge and Data Engineering*, vol. 27, no. 11, pp. 3111–3124, 2015.
6. Y. Luo, Y. Wen, and D. Tao, “On combining side information and unlabeled data for heterogeneous multi-task metric learning,” in *IJCAI*, pp. 1809–1815, 2016.
7. F. Shen, C. Shen, X. Zhou, Y. Yang, and H. T. Shen, “Face image classification by pooling raw features,” *Pattern Recognition*, vol. 54, pp. 94–103, 2016.
8. C. Xu, D. Tao, and C. Xu, “Multi-view intact space learning,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 37, no. 12, pp. 2531–2544, 2015.
9. Y. Qu, L. Lin, F. Shen, C. Lu, Y. Wu, Y. Xie, and D. Tao, “Joint hierarchical category structure learning and large-scale image classification,” *IEEE Transactions on Image Processing*, vol. 26, no. 9, pp. 4331–4346, 2016.
10. E. López-Iñesta, F. Grimaldo, and M. Arevalillo-Herráez, “Classification similarity learning using feature-based and distance-based representations: A comparative study,” *Applied Artificial Intelligence*, vol. 29, no. 5, pp. 445–458, 2015.
11. H. Li, Y. Wei, L. Li, and Y. Yuan, “Similarity learning for object recognition based on derived kernel,” *Neurocomputing*, vol. 83, pp. 110–120, 2012.
12. T. M. Malas, S. S. Taifour, and G. A. Abandah, “Toward optimal arabic keyboard layout using genetic algorithm,” *Proc. 9th Int’l Middle Eastern Multiconf. on Simulation and Modeling (MESM 2008)*, pp. 50–54, 2008.
13. P. Mitra, C. Murthy, and S. K. Pal, “Unsupervised feature selection using feature similarity,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 24, no. 3, pp. 301–312, 2002.
14. M. Bilenko, S. Basu, and R. J. Mooney, “Integrating constraints and metric learning in semi-supervised clustering,” in *Proceedings of the twenty-first international conference on Machine learning*, p. 11, ACM, 2004.
15. C. Hoi, W. Liu, and S.-F. Chang, “Semi-supervised distance metric learning for collaborative image retrieval and clustering,” *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 6, no. 3, p. 18, 2010.

16. I. W. Tsang, J. T. Kwok, C. Bay, and H. Kong, "Distance metric learning with kernels," in Proceedings of the International Conference on Artificial Neural Networks, pp. 126–129, Citeseer, 2003.
17. K. Q. Weinberger and L. K. Saul, "Distance metric learning for large margin nearest neighbor classification," Journal of Machine Learning Research, vol. 10, no. Feb, pp. 207–244, 2009.
18. S. Smale, L. Rosasco, J. Bouvrie, A. Caponnetto, and T. Poggio, "Mathematics of the neural response," Foundations of Computational Mathematics, vol. 10, no. 1, pp. 67–91, 2010.
19. M. Riesenhuber and T. Poggio, "Hierarchical models of object recognition in cortex,"
20. Nature neuroscience, vol. 2, no. 11, pp. 1019–1025, 1999.
21. T. Serre, M. Kouh, C. Cadieu, U. Knoblich, G. Kreiman, and T. Poggio, "A theory of object recognition: computations and circuits in the feedforward path of the ventral stream in primate visual cortex," tech. rep., DTIC Document, 2005.
22. J. Bouvrie, T. Poggio, L. Rosasco, S. Smale, and A. Wibisono, "Generalization and properties of the neural response," 2010.
23. A. Caponnetto, T. Poggio, and S. Smale, "On a model of visual cortex: learning invariance and selectivity from image sequences. cbcl paper 272," tech. rep., CSAIL technical report 2008-030, MIT, Cambridge, MA, 2008.
24. H. Li, Y. Wei, L. Li, and C. P. Chen, "Hierarchical feature extraction with local neural response for image recognition," Cybernetics, IEEE Transactions on, vol. 43, no. 2, pp. 412–424, 2013.
25. R. A. M. Alsaidi, H. Li, Y. Wei, R. Khaji, and Y. Y. Tang, "Hierarchical sparse method with applications in vision and speech recognition," International Journal of Wavelets, Multiresolution and Information Processing, vol. 11, no. 02, 2013.
26. R. Alsaidi, H. Li, Y. Wei, et al., "Improved hierarchical sparse method with application to offline handwritten arabic character recognition," International Journal of Scientific, Engineering Research, vol. 5, no. 03, 2014.
27. H. Li, H. Li, Y. Wei, Y. Tang, and Q. Wang, "Sparse-based neural response for image classification," Neurocomputing, vol. 144, pp. 198–207, 2014.
28. R. A. ALSAIDI, H. LI, H. LI, and R. KHAJI, "Selection of informative template in hierarchical sparse method," 2018.
29. T. M. Cover and J. A. Thomas, "Elements of information theory john wiley & sons," New York, vol. 68, pp. 69–73, 1991.
30. A. M. Neto, A. C. Victorino, I. Fantoni, D. E. Zampieri, J. V. Ferreira, and D. A. Lima, "Image processing using pearson's correlation coefficient: Applications on autonomous robotics," in 2013 13th International Conference on Autonomous Robot Systems, pp. 1–6, IEEE, 2013.



Saleh R. A. Alenazi, received the B.S. degree in Computer Science (Third Hons) from Multimedia University (MMU), Malaysia, in 2012, the M.S. degree in Computer Science from Northern University of Malaysia, Malaysia, in 2014, and the Ph.D degree in Computer Science from The National University of Malaysia, Malaysia, in 2019. He is currently a First Trainer A of

Computer Sciences at Technical and Vocational Training Corporation. His research interests include Data Analysis, Big Data, Programming.



Madallah Alruwaili, was born in Aljouf, KSA, in 1983. He received the bachelor degree (Hons) from Jouf University, KSA, in 2005, the M.S. degree from University of Science, Malaysia, in 2009, and the Ph.D degree from Southern Illinois University, Carbondale, Illinois, USA, in 2015. The PhD dissertation entitled "Enhancement and Restoration of Dust Images".

He is currently an Assistant Professor of Computer Engineering and Networks at Jouf University. He is now dean of College of Computer and Information Sciences. His research interests include image processing, image quality analysis, pattern recognition, computer vision, and biomedical imaging.

AUTHORS PROFILE



Dr. Ramadhan A. M. Alsaidi, was born in Ibb, Yemen in 1974. He received the B.S. in mathematics from the Sana'a University, in 1996, M.S. degrees in computation mathematics from USTC, China, in 2010 and the Ph.D. degree in applied mathematics (computer software and technology) from HUST, China, in 2014.

From 2007 to 2014, he was a Research Assistant with the computation and applied mathematics Laboratory. Since 2017, he has been an Assistant Professor with the Mathematics Department, Jouf University, science & art College in Gurayat. He is the author of more than 6 articles. His research interests include applied mathematics and image processing.



Dr. Ayed Rheal A. Alanzi, received the B. E. degree in Mathematics from Aljouf University, Kingdom of Saudi Arabia (KSA), in 2006, the M.S. degree in Statistics from Malaya University, Malaysia, in 2009, the M.S. degree in Mathematics from Southern Illinois University, Carbondale, Illinois, USA, in 2014, and the PhD degree in Statistics from The University of New Mexico, Albuquerque, New Mexico, USA, in 2017. He is

currently an Assistant Professor of Statistics at Majmaah University, Majmaah, KSA. His research interests include Bayesian statistics, data analysis, regression analysis, statistics in phylogenetics, biostatistics, and the analysis of stock market time series data.