Uniformed two Local Binary Pattern Combined with Neighboring Support Vector Classifier for Classification

Mohammed Ngadi[1] , **Aouatif Amine**[2], **Bouchra Nassih**[3] ,
Hanaa Hachimi[4], **Adnane El-Attar**[5]

Systems Engineering Laboratory
National School of Applied Sciences
Ibn Tofail University of Kenitra - Morocco
[1]Ngadi.mohammed@univ-ibntofail.ac.ma

[2]amine_aouatif@univ-ibntofail.ac.ma

[3]Nassih.bouchra@univ-ibntofail.ac.ma

[4]hanaa.hachimi@univ-ibntofail.ac.ma

[5]adnane.elattar@usmba.ac.ma

ABSTRACT

Due to its numerous real life applications, face classification has been one of the main research topics in computer vision and machine learning in recent years. In this paper, we develop an efficient and practical method for face classification. Our approch (LBP u2-NSVC) belongs to hybrid methods; it's based on the combination of Uniform Local Binary Patterns (LBPu2) and a the classifier Neighboring Support Vector Classifier (NSVC). To be very simple, given a face dataset, we start by computing the LBP features first. Later on, these features are used to train a classifier using the new NSVC algorithm. Our experiments show that by using this approach and tuning well the NSVC parameters, we can get better performances than state-of-the-art similar algorithms. To confirm this, we use the famous MIT-CBCL face dataset for the different tests.

Keywords: Face classification, Feature extraction, Neighboring Support Vector Classifier, Local Binary Pattern, Discrete Wavelet Transform.

Mathematics Subject Classification Number : 47N30
ACM Computing Classification System : I.5.2

1 Introduction

The problem of variables selection in classification is generally faced when the number of variables that can be used to explain the class of an individual, is very high. The needs have evolved in recent years with the handling of a large number of variables in such areas as genetic data or image processing (Skrjanc, Blazic and Agamennoni, 2005) and (Precup, Preitl and Korondi, 2007) and (Solos, Tassopoulos and Beligiannis, 2016) . Nevertheless, if one should treat the data described by a large number of variables, the classical methods of analysis, learning and data mining may prove to be ineffective or can lead to ambiguous results. In this article, we propose innovative methods to reduce the initial size of the data and to select sets of relevant variables for supervised classification.

The first contribution focuses on the selection of the primary function for features extraction (Amine, Ghouzali, Rziza and Aboutajdine, 2008) and (Majid, Khan and Mirza, 2003) and face detection, using the LBP descriptor (Ojala, Pietikainen and Maenpaa, 2002), (Ojala, Pietikäinen and Mäenpää, 2001), (Hadid, Pietikainen and Ahonen, 2004) and (Ojala, Pietikainen and Harwood, 1996). Indeed, using the LBP descriptor, central pixel is compared to its neighbors, this comparison gives rise to binary codes, which are then kept or rejected based on simple criteria (Ojala et al., 1996) . The motivation that pushed to use this operator is that a face can be seen as an assemblage of micro-patterns whose description by LBP is both good, robust against gray variations and fast to generate.

The second contribution concerns the selection of the second function within the classification. We use a new method based on two different families (unsupervised: Fuzzy C-Means and supervised: SVM).

The basic idea of the Neighboring Support Vector classifier (NSVC) (Ngadi, Amine, Hachimi and El-Attar, 2016) is to build new vicinal kernel functions, obtained by supervised clustering in the feature space. These vicinal kernel functions are then used for learning.

Finally, a comparison was made between our approach, and that based on the combination of Discrete Wavelet Transform (DWT) (Unser, 1995), (Singh, Tiwari and Shukla, 2012) , (Nixon and Aguado, 2012) and NSVC.

The rest of the paper is organized as follows. A brief description of proposed method and LBP is given in section 2 and 3. Section 4 introduces the NSVC based on supervised partitioning of features space. Experimental results are presented in section 5, while section 6 concludes the article.

2 Overview of the Proposed Method:

Any automatic face classification process must take into account several factors that contribute to the complexity of its task, because the face is a dynamic entity that is constantly changing under the influence of several factors. Accordingly, the following general approach should be adopted to design such a system:

- Analysis: (also called features extraction), it must retrieve image information that will be saved in memory to be used later in the Decision phase. The choice of these useful information rely on establishing a model for the face, they must be discriminating and not redundant.
- Learning: consists on memorizing the representations calculated in the analysis phase.
- Decision: Measurement of similarity.

Our approach can be summarized in the following diagram:

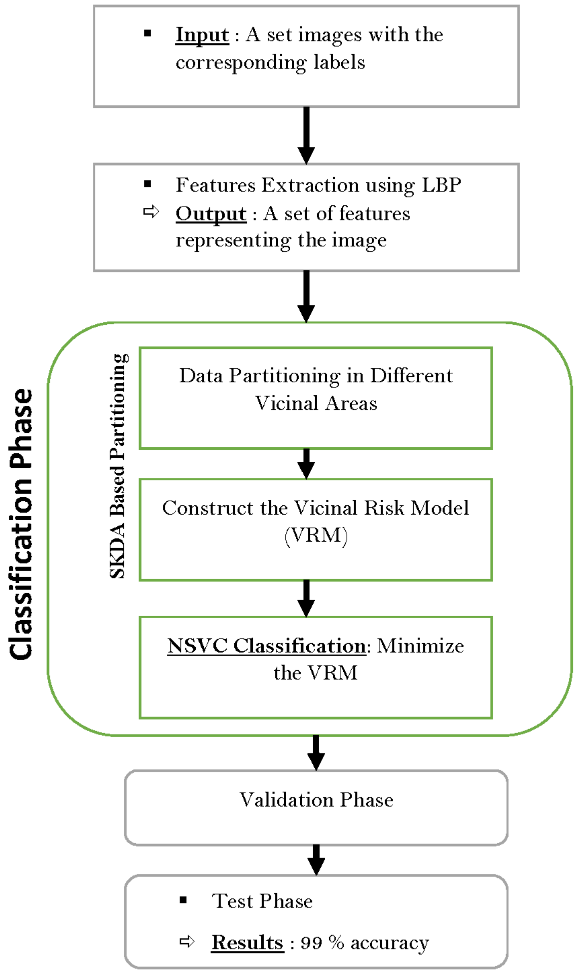

Figure 1: Feature extraction with classification process

3 Local Binary Patterns (LBP)

The LBP allows you to encode the local distribution of the image intensities by comparing each pixel to its vicinity (Nanni, Lumini and Brahnam, 2012) . It is an invariant measure of texture derived from analysis of the nearest neighbors. This technique allows to characterize regular patterns representing areas of these textures in an image. This operator is derived from the techniques of textures analysis known as Gabor filters (Grigorescu, Petkov and Kruizinga, 2002), co-occurrence matrices (Latif-Amet, A Ertüzün and Erçil, 2000), and so on (Mäenpää and Pietikäinen, 2005).

The first version of this descriptor consisted on extracting areas of high contrast in an image by analyzing the eight nearest neighbors of a pixel. Later multi scales versions have been proposed. According to the scale of the neighborhood used, certain areas of interest as the corners or edges can be detected by this descriptor (Fig.2).

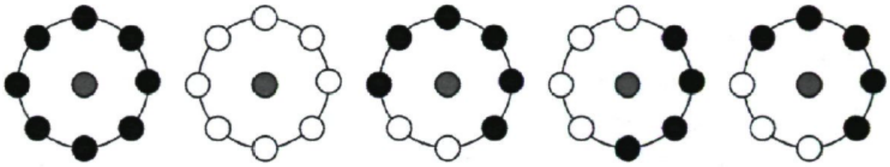

Figure 2: Example of features textures detected by the LBP operator.

The LBP operator can be defined on a symmetrical circular neighborhood having a radius R ($R > 0$) around the central pixel (xc, yc) and P ($P > 0$) is the number of pixels in the neighborhood (Fig.3). The coordinates of the P points are calculated by:

$$P_i = (x_c + Rcos(2\Pi p/P), y_c - Rsin(2\Pi p/P)) \tag{3.1}$$

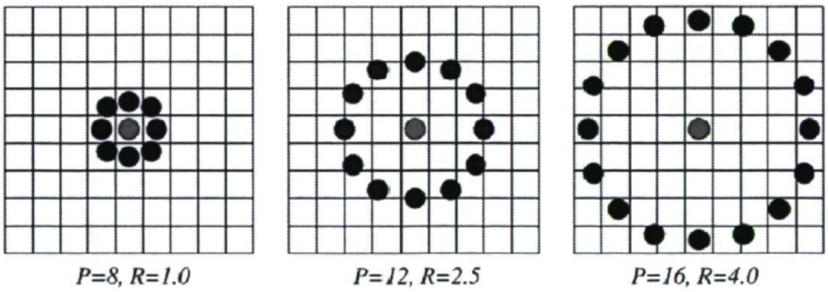

Figure 3: Sample neighborhoods used to calculate a LBP at multiple scales.

The values of points which do not fall on a real pixel are used to get the value of the pixel by bilinear interpolation (in the literature we find various neighbourhood topologies that can be different from the diagram in Fig.4) (Liao, Zhu, Lei, Zhang and Li, 2007).

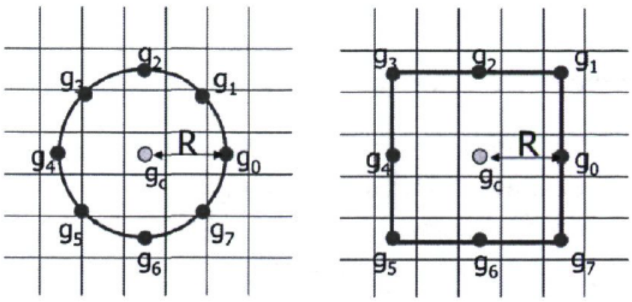

Figure 4: Example of neighborhood obtained by interpolation. The topology of the neighborhood may vary from that shown in this figure.

The calculation of the LBP is based on a binary code describing the reason for the local texture obtained by comparison of the central pixel with its neighbors. This descriptor is constructed by thresholding the local neighborhood by its central pixel grey-level value. The N neighboring are marked using a binary code 0, 1 obtained by comparing their values to the central pixel value. If the tested pixel has a gray level below the level of the central pixel gray value, it is marked with the value 0; otherwise this pixel is marked with the value 1:

$$P_i' = \begin{cases} 1 & \text{if } P_i \geq P_0 \\ 0 & \text{otherwise} \end{cases} \tag{3.2}$$

P_i' is the resultant binary code, P_i is the gray level value of the original pixel at position i and P_0 the central pixel value.
With this technique, and in its original version, there are $256(2^8)$ possible models (or texture neighborhood 3×3 units). The resulting binary marking value is then multiplied by the weight given to each pixel in the vicinity. The weight is given by the value 2^{i-1} (with i the position given in the vicinity: $1, ..., 8$).
The sum of the values obtained gives the LBP measure for the selected pixel (Fig.5):

$$LBP = \sum_{i=1}^{8} P_i' 2^{i-1} \tag{3.3}$$

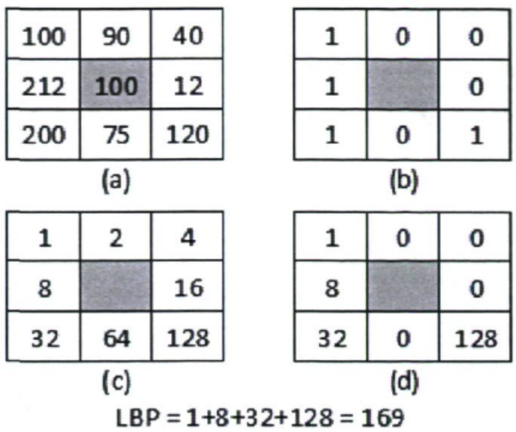

Figure 5: Example of the LBP descriptor calculation on a 3 x 3 neighborhood: (a) gray-level in a 3 x 3 area; (b) binary marking (binary mask obtained); (c) mask with the weights that we used; (d) result of the multiplication of the binary mask with weights. The calculated value of the LBP descriptor is equal to 169.

The central pixel is replaced with the value obtained by the calculation of the LBP descriptor (169 in the example in Figure 5 replaces the original value that was equal to 100). By repeating this procedure on all pixels in the image, one gets the image in the LBP texture space. The descriptor can be easily extended to include a larger area of neighborhood, leading to different results.

The NSVC is a classifier adaptive to different datasets. It is based, on one hand, on a non-supervised approach such as K-means or FCM and, on the other hand, on a supervised approach: SVM.

4 Neighboring Support Vector Classifier (NSVC)

Support vector machines, first introduced by Vapnik and colleagues for the problems of classification and regression, can be seen as a new training technique based on traditional polynomial and radial basis function (RBF) . As discussed before, SVMs have attracted considerable attention because of their high generalization ability and higher classification performance relative to other pattern recognition algorithms.
However, the assumption that the training data are identically generated from unknown probability distributions may limit the application of SVM to the problems of everyday life (Muller, Mika, Ratsch, Tsuda and Scholkopf, 2001).
To relax the assumption of identical distribution, the NSVC (Yang, Cao, Song, Schaefer and Su, 2014) , (Cao, Song, Yang, Liu and Guo, 2004) uses a set of vicinal cores functions built based on supervised clustering in the feature space induced by the kernel. The basic idea of the NSVC is to build new vicinal core functions obtained by supervised clustering in the feature space. These vicinal core functions are then used to SVM training.
This approach consists of two steps:

- Supervised clustering step based on SKDA algorithm (for Supervised Kernel-based Deterministic Annealing) used to partition the training data in different vicinal areas.
- A training step where the SVM technique is used to minimize the Vicinal Risk function (VRM) under the constraints defined in clustering step based on SKDA.

Consider the following input output data together:

$$(x_i, y_i)_{i=1}^{l}, x_i \in R^n, y_i \in \{-1, 1\} \tag{4.1}$$

Where l is the number of input data points, and n is the dimension of the input space.
The vicinity functions $v(x_i)$ of the x_i data points are built if test data points satisfy two assumptions:

- The unknown density function is smooth in the neighborhood of each point x_i.
- The function minimizing the functional risk is also smooth and symmetric in the neighborhood of each point x_i.

The optimization problem based on the principle of VRM named vicinal linear SVM (Vapnik, 2000), (Chapelle, Weston, Bottou and Vapnik, 2000), can then be formulated as:

$$\begin{aligned} minimize : \phi(w) = \frac{1}{2} w^T w + C \sum_{i=1}^{l} \xi_i \\ subject\,to : y_i \int_{V(x_i)} ([< x, w > + b] p(x|V(x_i)))dx \geq 1 - \xi_i \\ \xi_i \geq 0, i = 1, ..., l \end{aligned} \tag{4.2}$$

Where w is a weight, C is a punishment constant for ξ_i, b is the offset, $v(x_i)$ is the vicinity associated with the test point x_i , and $p(x|V(x_i))$ is the conditional probability of the respective

vicinity in the input space.
The following theorem for the vicinal SVM solution is true (see (Vapnik, 2000) for a proof):

$$f(x) = \sum_{i=1}^{l} y_i \beta_i L(x, x_i) + b \tag{4.3}$$

Where to define the coefficients β_i one has to maximize:

$$\begin{aligned} W(\beta) = \sum_{i=1}^{l} \beta_i - \frac{1}{2} \sum_{i,j=1}^{l} \beta_i \beta_j y_i y_j M(x_i, x_j) \\ subject\, to \sum_{i=1}^{l} \beta_i y_i = 0 \\ \beta_i \geq 0, \end{aligned} \tag{4.4}$$

Where $L(x, x_i)$ is called the mono-vicinal kernel and $M(x_i, x_j)$ is the bi-vicinal kernel of the vicinal SVM (Vapnik, 2000).

4.1 Supervised kernel-based deterministic annealing for NSVC

The clustering of training data in the feature space is a well-documented subject (Camastra and Verri, 2005), (Leski, 2004) . It consists on non-linearly mapping the observed data of an input low-dimensional space to a high dimensional feature space using a kernel function, which facilitates the separation of linear data. Denoting a non-linear transformation of the input space X to a high-dimensional space using a kernel function as:
$\Phi : \Re^n \to F$
$x_i \to \Phi(x_i), j = 1, ..., l$
Where $\Phi(x_i)$ is the transformed point x_i.
All training data points are distributed in c vicinities / clusters in the feature space, where $\phi_k(z)$ is the center of mass of the k-th vicinity residing in F. This is a similar representation to clustering based on the characteristic space of k-means:

$$\phi_k = \sum_{i=1}^{l} \alpha_{k_i} z_i \quad , k = 1, 2, ..., c \tag{4.5}$$

Where c is the number of clusters, α_{ki} are the parameters to be defined by the clustering technique (SKDA) and $z_i = y_i \phi(x_i)$ denotes the data points labeled in the feature space.
The classification problem is usually defined mathematically by a cost function to be minimized, for NSVC case, this function is the distortion function . Similar to the notation used in (Rose, 1988), we let $p(\phi_k|z_i)$ denote the probability of association of points z_i mapped to the cluster center ϕ_k. Using the square distance $D_k(z_i)$ (Yang et al., 2014) between the center ϕ_k and the training vector z_i, the distortion function in the function space becomes:

$$J_\phi = \sum_{i=1}^{l} \sum_{k=1}^{c} p(z_i) p(\phi_k|z_i) D_k(z_i) \tag{4.6}$$

Since no a priori knowledge of the distribution of data is assumed, over all possible distributions which give a given value of J_ϕ we choose the one that maximizes the conditional Shannon entropy in the feature space :

$$H_\phi = - \sum_{i=1}^{l} \sum_{k=1}^{c} p(z_i) p(\phi_k|z_i) log p(\phi_k|z_i) \tag{4.7}$$

The optimization problem can be reformulated as the minimization of the Lagrangian:

$$F_\phi = J_\phi - T H_\phi \tag{4.8}$$

Where T is the Lagrange multiplier.

To determine the α_{ki} parameter, we minimize the free energy function $F\ w.r.t$ the likelihood of association (Rose, 1988), which is related to the Gibbs distribution as:

$$p(\phi_k|z_i) = \frac{p(\phi_k) e^{-\frac{D_k(z_i)}{T}}}{\sum_{m=1}^{c} p(\phi_m) e^{-\frac{D_m(z_i)}{T}}} \tag{4.9}$$

Where $p(\phi_k)$ is the mass probability for k-th cluster:

$$p(\phi_k) = \sum_{i=1}^{l} p(z_i) p(\phi_k|z_i) \tag{4.10}$$

And so the energy function is:

$$F^*_\phi = min_{p(\phi_k|z_i)}(J_\phi - T H_\phi) = -T \sum_{i=1}^{l} p(z_i) log \sum_{k=1}^{c} p(\phi_k) e^{-\frac{D_k(z_i)}{T}} \tag{4.11}$$

The partial derivative of F $w.r.t\ \phi_k$:

$$\frac{\partial(F^*_\phi)}{\partial(\phi_k)} \ = \ 0, \tag{4.12}$$

Accordingly:

$$\sum_{i=1}^{l} p(z_i) p(\phi_k) e^{-\frac{D_k(z_i)}{T}} [z_i - \phi_k] = 0 \tag{4.13}$$

By dividing by the normalization factor:

$$Z_{z_i} = \sum_{m=1}^{c} p(\phi_m) e^{-\frac{D_m(z_i)}{T}} \tag{4.14}$$

And so,

$$\sum_{i=1}^{l} \frac{p(z_i) p(\phi_k) e^{-\frac{D_k(z_i)}{T}}}{Z_{z_i}} z_i = \sum_{i=1}^{l} \frac{p(z_i) p(\phi_k) e^{-\frac{D_k(z_i)}{T}}}{Z_{z_i}} \phi_k \tag{4.15}$$

Using Eq.(4.9) leads to :

$$\sum_{i=1}^{l} p(z_i) p(\phi_k|z_i) z_i = \sum_{i=1}^{l} p(z_i) p(\phi_k|z_i) \phi_k \tag{4.16}$$

$$\phi_k = \sum_{i=1}^{l} \frac{p(z_i) p(\phi_k|z_i)}{\sum_{i=1}^{l} p(z_i) p(\phi_k|z_i)} z_i = \sum_{i=1}^{l} \alpha_{ki} z_i \tag{4.17}$$

Finally, we obtain the expression of α_{ki} that will be used to construct the vicinal kernel for NSVC functions:

$$\alpha_{ki} = \frac{p(z_i) p(\phi_k|z_i)}{\sum_{j=1}^{l} p(z_j) p(\phi_k|z_j)} \tag{4.18}$$

4.2 NSVC with the feature space partitioning

The optimization problem based on feature space partitioning is formulated as follows (Vapnik, 2000):

$$\begin{aligned} minimise: \phi(w) = \frac{1}{2}w^T w + C\sum_{k=1}^{K}\xi_k \\ subject\,to: y_k\int_{V(\phi_k)}[< z, w > +b]p(z|\phi_k)dz \geq 1-\xi_k, i = 1, .., l \\ \xi_k \geq 0, k = 1, ..., K \end{aligned} \tag{4.19}$$

Where $v(\phi_k)$ represents the k^{th} vicinity associated with the mass center ϕ_k in the feature space, and $p(z/\phi_k)$ is the conditional probability of respective vicinity in the feature space. According to Bayes theorem, we have:

$$p(z_i|\phi_k) = \frac{p(z_i)p(\phi_k|z_i)}{p(\phi_k)} = \frac{p(z_i)p(\phi_k|z_i)}{\sum_{j=1}^{l} p(z_j)p(\phi_k|z_j)} \tag{4.20}$$

By comparing Eq.(4.17) and Eq.(4.20), we get:

$$\phi_k = \sum_{i=1}^{l} p(z_i|\phi_k)z_i \tag{4.21}$$

And the optimization constraint becomes:

$$\begin{aligned} & y_k\int_{V(\phi_k)}[< z, w > +b]p(z|\phi_k)dz \\ = \; & y_k[< \int_{V(\phi_k)} p(z|\phi_k)zdz, w > + \int_{V(\phi_k)} bp(z|\phi_k)dz] \\ = \; & y_k[< \sum_{i=1}^{l} p(z_i|\phi_k)z_i, w > + \sum_{i=1}^{l} bp(z_i|\phi_k)] \\ = \; & y_k[< \phi_k, w > +b] \end{aligned} \tag{4.22}$$

Let define the mono and bi-vicinal kernels as:

$$L_k(x) = \sum_{i=1}^{l} y_i\alpha_{ki}K(x, x_i), k = 1, 2, ..., K \tag{4.23}$$

$$M_{km}(x) = \sum_{i=1}^{l}\sum_{j=1}^{l} y_i y_j \alpha_{ki}\alpha_{mj}K(x_i, x_j), k, m = 1, 2, ..., K \tag{4.24}$$

Where the α_{ki} parameters are obtained from the SKDA clustering step. The decision boundary is:

$$f(x) = \sum_{k=1}^{c} \beta_k y_k L_k(x) + b \tag{4.25}$$

Where β_k is the coefficient that maximize the dual function:

$$\begin{aligned} maximize\; W(\beta) = \sum_{k=1}^{c}\beta_k - \frac{1}{2}\sum_{k,m=1}^{c}\beta_k\beta_m y_k y_m M_{km}(x) \\ subject\,to \sum_{k=1}^{c}\beta_k y_k = 0, \quad \beta_k \geq 0 \end{aligned} \tag{4.26}$$

In order to obtain a sparse solution at the cost of the extra clustering procedure, a good selection of the number of clusters is required.

5 Results and discussion

After deriving the mathematics behind the proposed approach as illustrated in figure 1, we will now test it and compare it to other similar state-of-the-art methods. To do so, we use the very famous MIT-CBCL face database [1].

5.1 Database

We perform face classification using an extended version of the MIT-CBCL face database. The original database has 6,977 images with 2,429 faces and 4,548 non-faces, Fig.6.

Figure 6: A subset of MIT CBCL face database used for classification.

5.2 Results of NSVC:

The basic of NSVC is to build new neighboring kernel functions, obtained by supervised clustering in feature space. These neighboring kernel functions are then used in SVM based learning.
When using polynomial kernels, we have used cross-validation in order to compute optimal learning parameters for both kernels.
We evaluate the accuracy of each feature extraction method with NSVC. The results obtained are shown in Table 1.

Table 1: Accuracy of LBP-NSVC with polynomial kernel.

	LBP u2	LBP ri	LBP ri-u2
Accuracy %	**99.99**	99.94	99.94
Parameter of Kernel	2	3	3

[1] http://faculty.ucmerced.edu/mhyang/face-detection-survey.html

Where: 'u2' is uniform LBP, 'ri' is rotation-invariant LBP, and 'ri-u2' is uniform rotation-invariant LBP.
To validate our approach we used another method as DWT-NSVC. The results obtained are shown in Table 2.

Table 2: Accuracy of DWT-NSVC with polynomial kernel.

Method	Parameter of Kernel	Accuracy %
DWT-NSVC	3	99.97

An accurate and robust face classification system was developed and tested. This system exploits the feature extraction capabilities of the LBP for NSVC based classification to increase its robustness to variations conditions. The proposed technique performed much better compared to the other technique of feature extraction like DWT combined NSVC.
The following sections shows a comparison of the accuracy achieved with our experiences and other classifiers.
We found that the LBP-u2 method is particularly interesting when used as a preprocessing step with other classifiers. Figure 7 shows the results of LBP-u2 combined with Adaboost [95%]. The method of Boosting is interesting because we can choose the number of weak classifiers in order to achieve the desired error rates on samples examples. Moreover, we observe that the error rate decreases exponentially with the number of used weak classifiers.

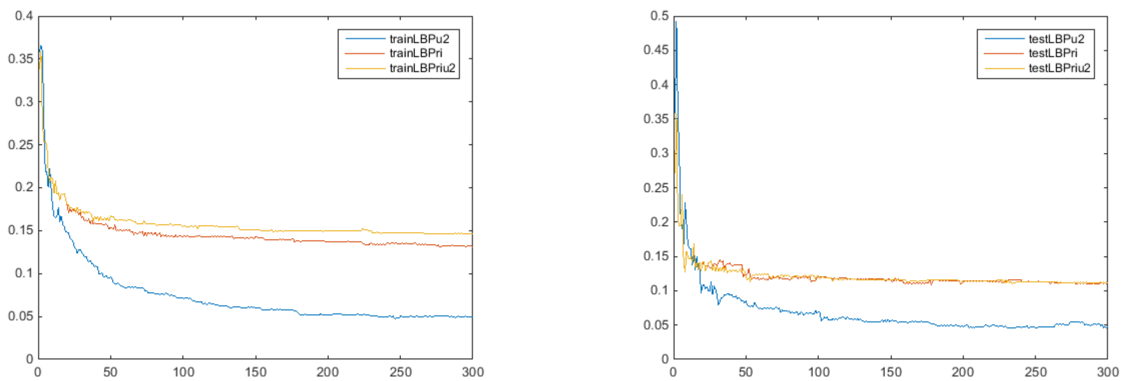

Figure 7: Classification error with respect to the number of weak classifiers.

5.3 Discussion:

From figure 7, its clear that the LBP-u2 gives a better error-rate than other LBP version when used with adaboost. Finally, to show the robustness of our proposed method, we compare it to other state of-the-art classification methods (Figure 8), some of them uses different classification approach than the NSVC such as : Naive Bayesn [95%], Decision Tree [88%], K Nearest Neighbors (KNN) [**98%**], linear Support Vector Machines (linear SVM) [96%], and Random

Forest [93%].

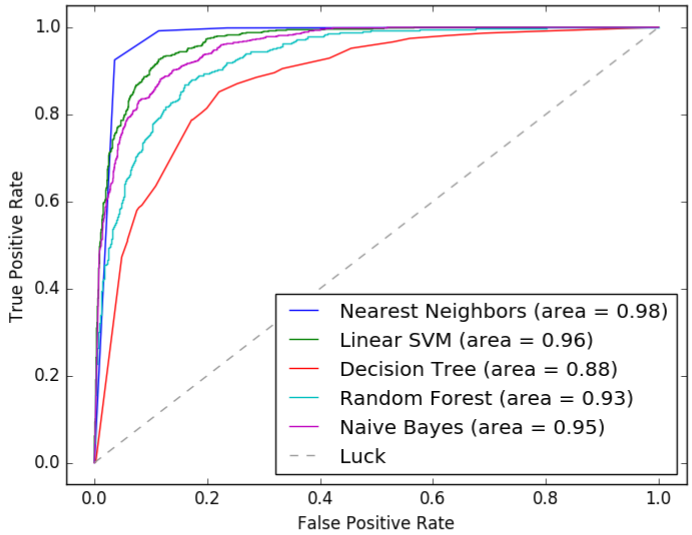

Figure 8: Comparison results of different classifiers methods on LBP u_2

It is clear that the approach proposed LBPu2-NSVC produced the best or equal classification accuracy (99.99%) compared to other methods (98%).

In addition to its high performance, the NSVC is a new theoretical method of classification which combines two methods of classification belonging to two different classification families (unsupervised: Fuzzy C-Means and supervised: SVM).

6 Conclusion

In this article, we presented a new interesting approach LBP-NSVC, which is based on Local Binary Patterns combined with the Neighboring Support Vector Classifier. The experimental results on MIT-CBCL show that the proposed method gives very good results.

The comparison with DWT-NSVC demonstrates that our approach is better. The best precision obtained by LBP-NSVC exceeded 99 %.

Our goal in the near future is to continue the study of LBPu2-NSVC to test it on different datasets from other research areas and try to find the best compromise between precision and execution time.

References

Amine, A., Ghouzali, S., Rziza, M. and Aboutajdine, D. 2008. Investigation of feature dimension reduction based dct/svm for face recognition, *IEEE Symposium on Computers and Communications.Marrakech, Morocco.* pp. 188–193.

Camastra, F. and Verri, A. 2005. A novel kernel method for clustering, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **27**(5): 801–805.

Cao, A., Song, Q., Yang, X., Liu, S. and Guo, C. 2004. Mammographic mass detection by neighboring support vector machine, *International Joint Conference on Neural Networks, IEEE, Budapest, Hungary* **3**: 1953–1958.

Chapelle, O., Weston, J., Bottou, L. and Vapnik, V. 2000. Neighboring risk minimization, in: Advances in neural information processing systems, *MIT Press, MA, USA* pp. 416–422.

Grigorescu, S. E., Petkov, N. and Kruizinga, P. 2002. Comparison of texture features based on gabor filters., *IEEE Transactions on Image Processing* **11**(10): 1160–1167.

Hadid, A., Pietikainen, M. and Ahonen, T. 2004. A discriminative feature space for detecting and recognizing faces., *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Washington, DC, USA* pp. 797–804.

Latif-Amet, A., A Ertüzün, A. and Erçil, A. 2000. An efficient method for texture defect detection: sub-band domain co-occurrence matrices ., *Image and Vision Computing* **18**(6-7): 543–553.

Leski, J. M. 2004. Fuzzy c-varieties/elliptotypes clustering in reproducing kernel hilbert space, *Fuzzy Sets and Systems* **141**(2): 259–280.

Liao, S., Zhu, X., Lei, Z., Zhang, L. and Li, S. 2007. Learning multi-scale block local binary patterns for face recognition., *Proceedings of the International Conference Advances in Biometrics* pp. 828–837.

Mäenpää, T. and Pietikäinen, M. 2005. Texture analysis with local binary patterns., *Handbook of Pattern Recognition and Computer Vision, 3rd ed, World Scientific* pp. 197–216.

Majid, A., Khan, A. and Mirza, A. 2003. Gender classification using discrete cosine transformation: a comparison of different classifiers., *7th International Multi Topic Conference INMIC* pp. 8–9.

Muller, K., Mika, S., Ratsch, G., Tsuda, K. and Scholkopf, B. 2001. An introduction to kernel based learning algorithms, *IEEE Transactions on Neural Networks* **12**: 181–201.

Nanni, L., Lumini, A. and Brahnam, S. 2012. Survey on lbp based texture descriptors for image classification. expert systems with applications., *Expert Systems with Applications* **39**(3): 3634–3641.

Ngadi, M., Amine, A., Hachimi, H. and El-Attar, A. 2016. A new optimal approach using nsvc for breast cancer diagnosis classification., *International Journal of Imaging and Robotics.* **16**(4): 24–36.

Nixon, M. and Aguado, A. S. 2012. Feature extraction and image processing for computer vision, *3rd Edition, Academic Press* .

Ojala, T., Pietikainen, M. and Harwood, D. 1996. A comparative study of texture measures with classification based on feature distributions., *Pattern Recognition* **29**(1): 51–59.

Ojala, T., Pietikäinen, M. and Mäenpää, T. 2001. A generalized local binary pattern operator for multiresolution gray scale and rotation invariant texture classification., *In Advances in Pattern Recognition - ICAPR* pp. 399–408.

Ojala, T., Pietikainen, M. and Maenpaa, T. 2002. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns., *IEEE Transactions on Pattern Analysis and Machine Intelligence* **24**: 971–987.

Precup, R. E., Preitl, S. and Korondi, P. 2007. Fuzzy controllers with maximum sensitivity for servosystems, *IEEE Transactions on Industrial Electronics* **54**(3): 1298–1310.

Rose, K. 1988. Deterministic annealing for clustering, compression, classification, regression, and related optimization problems, *Proceedings of the IEEE* **86**(11): 2210–2239.

Singh, A. K., Tiwari, S. and Shukla, V. P. 2012. Wavelet based multi class image classification using neural network., *International Journal of Computer Applications* **37**(4): 21–25.

Skrjanc, I., Blazic, S. and Agamennoni, O. 2005. Identification of dynamical systems with a robust interval fuzzy model, *Automatica* **41**(2): 327–332.

Solos, I. P., Tassopoulos, I. X. and Beligiannis, G. N. 2016. Optimizing shift scheduling for tank trucks using an effective stochastic variable neighbourhood approach, *International Journal of Artificial Intelligence* **14**(1): 1–26.

Unser, M. 1995. Texture classification and segmentation using wavelets frames., *IEEE Transactions on Image Processing* **4**(11): 1549–1560.

Vapnik, V. 2000. The nature of statistical learning theory, *2nd Edition, Springer Verlag, NY, USA* .

Yang, X., Cao, A., Song, Q., Schaefer, G. and Su, Y. 2014. Vicinal support vector classifier using supervised kernel-based clustering, *Artificial Intelligence In Medicine* **60**(3): 189–196.