Selecting the best performing rotation invariant patterns in local binary/ternary patterns

Loris Nanni[1] Sheryl Brahnam[2] Alessandra Lumini[1]
[1] DEIS, IEIIT—CNR, Università di Bologna, Viale Risorgimento 2, 40136 Bologna, Italy
loris.nanni@unibo.it; Alessandra.Lumini@unibo.it
[2] Computer Information Systems, Missouri State University, 901 S. National, Springfield, MO 65804, USA
sbrahnam@missouristate.edu

Abstract *- This paper purposes a new method for selecting the most discriminant rotation invariant patterns in local binary patterns and local ternary patterns. Our experiments show that a selection based on variance performs better than the recently proposed method of using dominant local binary patterns (DLBP). Our method uses a random subspace of patterns with higher variance. Features are transformed using Neighborhood Preserving Embedding (NPE) and then used to train a support vector machine. Moreover, we extend DLBP with local ternary patterns (DLTP) and examine methods for building a supervised random subspace of classifiers where each bin of the histogram has a probability of belonging to a given subspace according to its occurrence frequencies. We compare several texture descriptors and show that the random subspace ensemble based on NPE features outperforms other recent state-of-the-art approaches. This conclusion is based on extensive experiments conducted in several domains using five benchmark databases.*

Keywords: texture descriptors; local binary patterns; local ternary patterns; non-uniform patterns; support vector machines.

1 Introduction

Rapid advances in imaging technology are leading to the creation of large stockpiles of images in numerous fields. This is giving rise to a host of new classification problems and a pressing need to develop new machine learning techniques. Nowhere is this more apparent than in medicine. If new methods for automatically extracting relevant medical images of multiple patients that share salient features could be found, this would greatly improve diagnosis and treatment. An example of new research in medicine that these databases are fostering is the potential use of machine face classification applied to medical diagnosis [1, 2]. Facial abnormalities often reflect disease, and recent research has demonstrated the effectiveness of detecting these conditions using face recognition techniques, see for example, [3] and [4]. Since a host of environmental and genetic syndromes leave signs in the face, medical face recognition systems will become an exciting new area of future medical research [5].

Many state-of-the-art machine learning approaches optimized for image classification make use of texture descriptors. Local Binary Patterns (LBP), first proposed by [6], are currently one of the most popular texture descriptors. The popularity of LBP as a feature descriptor is due in part to the fact that LBP is very resistant to lighting changes. This makes LBP a good choice for coding fine details. LBP is also powerful and low in computational complexity. Since the 1990s, LBP texture descriptors have been the focus of considerable research, especially by Prof. Pietikainen's group [6], as well as [7, 8].

In the field of medicine, LBP has been used to find relevant slices in brain MR (magnetic resonance) volumes [9], to classify true mammographic masses from normal parenchyma [10], and as textural features extracted from thyroid slices [11]. A number of researchers have recently investigated automated cell phenotype image classification using LPB (see, for example, [12]. A wide collection of papers that explore LBP in medical applications is available at
http://www.ee.oulu.fi/mvg/page/lbp_bibliography#biomedical. LPB has also been used in recent research in face recognition [7, 13], as well as in many other interesting areas of research, including work on smart guns [8] and fingerprint identification [14].

Until recently, LBP descriptors have utilized only the uniform patterns. Recent work has attempted to augment LBP by using non-uniform patterns. In [15] combining uniform patterns with a few non-uniform patterns was shown to improve performance. In [16] rotation invariant patterns are selected, instead of the uniform patterns. They propose choosing patterns that represents 80% of the patterns in the training data. Several other variants have recently been proposed, see, e.g., [17].

In this work we extend the work of [16] by focusing on different methods for selecting the best performing bins. Our goal is to enhance performance by selecting a set of rotation invariant patterns for LBP and LTP. We do this by comparing the simple variance selection, i.e., by selecting the histogram bins with higher variance using the training data, with the dominant pattern selection method proposed in [16]. Furthermore, we propose using a "supervised" random subspace of classifiers, where each bin of the

histogram has a probability of belonging to a given subspace according to its occurrence frequencies in the training data. In our experiments for both LBP and LTP, the simple variance selection process we propose outperforms the dominant patterns proposed in [16]. The best performance is obtained with a random subspace ensemble, where a set of a high variance features are projected by NPE before being fed into a support vector machine classifier.

The remainder of this paper is organized as follows. In section 2 we provide an overview of LBP as a descriptor. In section 3 we detail our proposed approach using dominant patterns. In section 4, we briefly describe the five datasets used in our experiments. In section 5, we report experimental results. Finally, in section 6, we provide a few concluding remarks and directions for future research.

2. The LBP operator

The LBP operator is calculated by evaluating the binary differences between the gray value of a pixel **x** and the gray values of P neighboring pixels on a circle of radius R around **x**. By selecting the smallest value of P-1 bitwise shift operations on the binary pattern, LBP is made rotation invariant. A pattern is *uniform* if the number of transactions in the sequence between 0 and 1 is less than or equal to two. The number of possible uniform patterns is $P+1$. The feature vector extracted from each cell is the LBP histogram of dimension P+2 (a single bin for non-uniform patterns).

Unfortunately, there is a loss of valuable anisotropic structural information using the circularly symmetric neighborhood to solve the rotation invariant problem. An elliptical neighborhood definition that preserves anisotropic structural information is proposed in [18].

Conventional LBP is also sensitive to noise in the near-uniform image regions. Local Ternary Patterns (LTP), as proposed by [19], overcome this problem. In LTP the difference between a pixel **x** and its neighbor **u** is encoded by 3 values according to a threshold τ: 1 if $\mathbf{u} \geq \mathbf{x} + \tau$; -1 if $\mathbf{u} \leq \mathbf{x} - \tau$; else 0. The ternary pattern is split into two binary patterns according to its positive and negative components, as illustrated in figure 1. The histograms that are computed from the binary patterns are then concatenated to form the feature vector.

An interesting variant of a 3-valued coding scheme that makes LBP more noise resistant is proposed in [20] [21]. In this work a fuzzy threshold function is applied. Another method for making LBP more robust in terms of noise is to use median binary patterns as proposed, for example, in [22], where the intensity space is mapped to LBP by thresholding a given pixel against the median value of its neighborhood.

The extracted histogram in conventional LBP is large. Using center-symmetric local binary patterns [23], it is possible to reduce the LBP histogram dimension. This is accomplished by comparing a given pixel with center-symmetric pairs of pixels. Given 8 neighbors, center-symmetric LBP produces only 2^4 binary patterns in contrast to the 2^8 different binary patterns produced by conventional LBP.

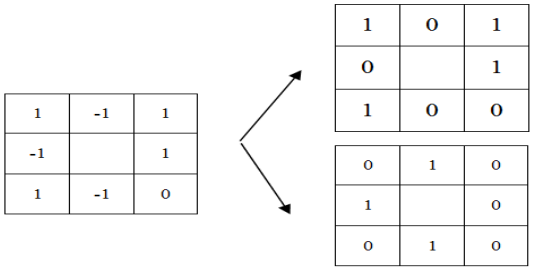

Figure 1. Example ternary code divided into positive and negative LBP codes.

A method for improving classification combines LBP descriptors with various preprocessing methods, see [14, 24]. In [24] Gabor wavelets are combined with the LBP operator to represent face images. This method of representing faces is high in dimensionality, however, since multiple Gabor transformations are performed. Zhang et al., [24] apply dimensionality reduction techniques to the output of the LBP operators to offset this problem.

3. Proposed approach

Our goal is to enhance classifier performance by selecting a set of rotation invariant patterns for LBP and LTP. Below is a step by step outline of our approach:

Step 1: The 250 bins with highest variance are extracted.

Step 2: A random subset of 125 features are then selected.

Step 3: PCA followed by NPE is used to reduce this set of features to 30.[1]

Step 4: A support vector machine is trained and tested using the features.

Step 5: Steps 2-4 are performed 50 times.

Step 6: The 50 classifier results are then combined using the sum rule.

[1] This is the default value in the NPE toolbox available at http://www.cs.uiuc.edu/homes/dengcai2/Data/ data.html. The number neighbors used for building the projection matrix is also the default value of 5.

As can be seen in step 1, we use a variance selection process. We select the histogram bins with the highest variance in the training data instead of the dominant pattern selection method proposed in [16]. So our methods can easily be reproduced, we modify the original LBP code found at http://www.ee.oulu.fi/mvg/page/lbp_matlab.[2]

After selecting a random subset of features in step 2, we use Neighborhood preserving embedding (NPE) in step 3, as a bin selector. NPE [25] is a feature transform technique that differs from principal component analysis (PCA) in that it preserves the local neighborhood structure on the data manifold and is thus less sensitive to outliers than PCA. However, we use PCA to reduce the size of the features. Since we retain 99.999% of the variance[3] in the data, little information is lost in our application of PCA as a feature reduction method. This greatly reduces computational complexity.

In step 4, we feed the features in a support vector machine (SVM) [26]. We repeat this process 50 times and in step 4 combine the results of the 50 classifiers using the sum rule.

In addition to the above process, we run experiments using a *supervised* random subspace of classifiers coupled with standard DLBP and DLTP. In the tested supervised random subspace method, each feature is not randomly selected but has a probability of being chosen based on the occurrence frequencies in the training data. This is accomplished as follows: given $\mathbf{x}_i$, the sum of the occurrence frequency of the i-th bin of the histogram in the training data, the probability of the i-th bin to be chosen is given by $\mathbf{x}_i/\sum_i \mathbf{x}_i$.

4. Datasets

Below we describe each of the datasets used in our experiments, along with the evaluation protocols. In general, if the database contains two classes, we use a base SVM. If the database contains more than two classes, we use the standard method of handling multiclass problems with SVM, viz., one versus one SVM.

[2] we have extracted the uniform patterns with *getmapping011*(16,'riu2') while the non-uniform patterns (i.e. all the rotation invariant patterns) are extracted by *getmapping011*(16,'ri')

[3]
```
w=pca(dataset(TR),0.99999);%function of the PRTools 3.1.7
TR=+(w*dataset(TR));
TE=+(w*dataset(TE));
clear W
options = [];
options.k = 5;
options.NeighborMode = 'Supervised';
options.gnd = yTR;
[eigvector, eigvalue] = NPE(options, TR);
```

The performance measure adopted in the experiments reported in this paper is the area under the ROC-curve [27] in the 2-classes problems and the accuracy in the multi-class problems. The area under the ROC-curve is a two-dimensional measure of classification performance that plots the probability of classifying correctly the genuine examples against the rate of incorrectly classifying impostor examples.

Infant COPE database and evaluation protocols

First described in [1], the Infant COPE (Classification Of Pain Expressions) database contains 204 facial images of 26 neonates experiencing a variety of nonpain stressors and one pain stimulus (to the puncture on the heel of a lance followed by repeated squeezing of the heel as blood samples were taken for a state mandatory blood exam). Of the 204 images,144 fall in the category of nonpain and 60 in the category of pain. For complete details of the experimental design, see [4]. We use the following evaluation protocol in the infant COPE experiments. First we divide the images by subject. The images of a given subject, *s*, form the testing set while the remaining subjects form the training set. This procedure is repeated for each subject. The images are resized to 100×120 pixels. Then 64 overlapping cells of dimension 25×25 are created at steps of 11 pixels for each image. A different classifier is trained on each of these cells, and the 64 classifier decisions are then combined.

2D HeLa dataset

The 2D HeLa dataset contains 862 single-cell images [28].[4] Each image is a 16 bit greyscale image of size 512 by 382 pixels. The dataset can be classified into ten classes: ActinFilaments, Endsome, ER, Golgi Giantin, Golgi GPP130, Lysosome, Microtubules, Mitochondria, Nucleolus, and Nucleus. The protocol used in our experiments is a 5-fold cross validation technique, with each dataset randomly divided into $4/5^{ths}$ for training and $1/5^{th}$ for testing. When using DLBP/DLTP, the bin with the higher occurrence is discarded because it represents the black background.

Pap smear dataset

The pap smear database contains 917 samples collected at the Herlev University Hospital using a digital camera and microscope [29]. Two skilled cyto-technicians classified each cell into the two classes of normal versus abnormal. Each cell was examined by two cyto-technicians. A medical doctor examined cells that were difficult for the cyto-technicians to classify. To calculate the area under the ROC-curve, a 5-fold cross validation technique is

[4] HeLa dataset is available at http://murphylab.web.cmu.edu/.

employed, where each dataset is randomly divided into $4/5^{ths}$ for training and $1/5^{th}$ for testing.

DaimlerChrysler Pedestrian dataset

The DaimlerChrysler pedestrian dataset [30][5] contains a set of images with pedestrians and a set without pedestrians. This dataset is difficult to classify because the non-pedestrian samples include a number of images where a shape-based pedestrian detector resulted in low confidence matches. In our experiments we extract a set of 4,900 images from the original dataset. We use a 5-fold cross validation protocol, where each dataset is randomly split into $4/5^{ths}$ for training and $1/5^{th}$ for testing.

LOCATE mouse protein sub-cellular localization endogenous database

The LOCATE mouse protein sub-cellular localization endogenous database contains approximately 50 images. Each image contains somewhere between 1 and 13 cells per class [31].[6] The dataset is divided into eleven classes: Actin-Cytoskeleton, Cytoplasm, Endosomes, ER, Golgi, Lysosomes, Microtubule, Mitochondria, Nucleus, Peroxisomes, PM. We use a 5-fold cross validation protocol, where each dataset is randomly split into $4/5^{ths}$ for training and $1/5^{th}$ for testing. When using DLBP/DLTP, the bin with the higher occurrence is discarded because it represents the black background.

5. Experimental results

In table 1, we report our experimental results using the system architecture and datasets with performance indicators described in sections 3 and 4. In the first set of experiments, we used LBP with P=16 and R=2. *LBP-riu* reports experiments using the standard rotation invariant uniform LBP descriptor. *Dom K=X%* reports experiments using DLBP with $K=X$. VR(X) reports experiments using the X bins with highest variance in the training data. R-Dom K reports experiments obtained using our *supervised* random subspace method coupled with *Dom K=90%*. R-VR(X) reports experiments using a random subspace of SVM obtained starting from the features selected by VR(X). R-(VR(250)+ NP) reports experiments using the ensemble method, detailed in section 2, based on the combination of random subspace, VR(250), and NPE. Finally, R-RIU reports the performance obtained by a random subspace of SVM obtained starting from the features RIU.

A similar set of experimental results is reported in Table 2. In these experiments LTP has P=16 and R=2. The threshold τ in LTP is 3.

Examining the two tables below, we can make the following conclusions:

- Both dominant LBP and variance selection for LBP outperform standard LBP;
- None of the methods work well with LBT. Both Dom K=80% ad Dom K=90%, when coupled with LPT, does not improve LTP. Also VR(250) does not improve LTP. Only VR(100) results in a performance that is similar to LTP;
- NP does not perform well;
- It is interesting to note that both for LBP and LTP, R-VR(250) outperforms VR(250), and R-(VR(250)+ NP) outperforms R-VR(250);
- A random subspace of SVMs does not perform remarkably better when coupled with RIU (see the row R-RIU in the tables above).
- Our idea of coupling random subspace, variance selection, and NPE, and the supervised random subspace for dominant LBP/LTP obtains the best performance. In our opinion this is due to the correlation among the selected bins.

We chose to couple random subspace with VR(250) because it contains many more features than VR(100). Using 250 features is a problem because of the possible correlation between the features and the high dimensionality of the vector (and the resulting curse of dimensionality). In the literature is well known that both these problems can partially be solved using random subspace ensemble.

We also ran some experiments with R-VAR(100). The results are as follows:

- LOCATE dataset: 0.843 (LBP) – 0.892 (LTP);
- 2D-Hela dataset: 0.883 (LBP) – 0.915 (LTP);
- PAP dataset: 0.822 (LBP) – 0.846 (LTP).

Looking at the these results with R-VAR(100), it is clear that there is no one clear winner in all the datasets (among the methods for selecting a set of bins). More experiments need to be performed to select the best way for selecting the most important bins in LBP/LTP.

6. Conclusion

In this paper, we perform a set of empirical experiments on several benchmark databases to determine the best method for selecting bins from the rotation invariant LBP/LTP. We compare several methods and reach the following conclusions: 1) both dominant and variance

[5] The DaimlerChrysler dataset is available at http://www.science.uva.nl/research/isla/dc-ped-class-benchmark.html

[6] The LOCATE mouse protein sub-cellular localization endogenous database is available at http://locate.imb.uq.edu.au/

Table 4. Performance obtained by LBP with *P*=16 and *R*=2.

LBP	***DATASETS***				
	2D-Hela	**PAP**	**COPE**	**Pedestrian**	**LOCATE**
RIU	0.827	0.749	0.849	0.835	0.833
Dom K=80%	0.837	0.831	0.852	0.838	0.825
Dom K=90%	0.797	0.829	0.861	0.836	0.858
VR(100)	0.826	0.822	0.864	0.827	0.843
VR(250)	0.808	**0.851**	0.821	0.852	0.849
NP	0.738	0.801	0.801	0.755	0.827
R- Dom K	0.900	0.828	0.860	**0.892**	0.880
R-VR(250)	0.839	0.843	0.861	0.857	0.870
R-(VR(250)+ NP)	**0.896**	0.848	**0.867**	0.845	**0.871**
R-RIU	0.776	0.730	0.874	0.868	0.774

Table 5. Performance obtained by LTP with *P*=16 and *R*=2.

LTP	***DATASETS***				
	2D-Hela	**PAP**	**COPE**	**Pedestrian**	**LOCATE**
RIU	0.920	0.829	0.925	0.918	0.913
Dom K=80%	0.893	0.834	0.900	0.927	0.884
Dom K=90%	0.855	0.822	**0.928**	0.815	0.871
VR(100)	0.892	0.849	0.915	0.945	0.872
VR(250)	0.853	0.853	0.897	0.842	0.886
NP	0.579	0.823	0.852	0.653	0.841
R- Dom K	0.922	0.853	0.917	**0.952**	0.913
R-VR(250)	0.901	0.852	0.918	0.909	0.886
R-(VR(250)+ NP)	**0.932**	**0.868**	0.922	0.911	**0.929**
R-RIU	0.909	0.814	0.925	0.951	0.907

set. Moreover, the power of our approach is tested on a broad spectrum of datasets: the *Infant COPE* database of selection work well with LBP; 2) none of the methods work very well with LTP; and 3) our approach obtains the best performance in almost all the datasets. It is clear that random subspace enables a selection of a wider set of bins and that it handles the correlation problems of the selected neonatal facial images; the *2D HeLa dataset* and the *Locate endogenous dataset* of fluorescence microscope images; the *Pap smear* dataset of smear cells images; the *Pedestrian* dataset of pedestrian images.

In the future, we plan to study the performance of the proposed texture descriptors when the feature extraction is performed from images that have been pre-processed using the different methods (e.g., Gabor filters).

References

[1] Sheryl Brahnam, Chao-Fa Chuang, Frank Y. Shih *et al.*, "Svm classification of neonatal facial images of pain."

[2] Sheryl Brahnam, Chao-Fa Chuang, Randal Sexton *et al.*, "Machine assessment of neonatal facial expressions of acute pain," *Decision Support Systems,* vol. 43, pp. 1247-1254, 2007.

[3] Jiandong Fang, Shiaofen Fang, Jeffrey Huang *et al.*, "Digital geometry image analysis for medical diagnosis." pp. 217-221.

[4] Sheryl Brahnam, Loris Nanni, and Randall Sexton, "Neonatal Facial Pain Detection Using NNSOA and LSVM," in The 2008 International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV'08) Las Vegas, 2008.

[5] Paul Ekman, Thomas S. Huang, Terrence J. Sejnowski *et al.*, *Final report to nsf of the planning workshop on facial expression understanding,* Technical Report, Human Interaction Lab, University of California, San Francisco, CA, 1992.

[6] Timo Ojala, Matti Pietikainen, and Topi Maeenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *Ieee transactions on pattern analysis and machine intelligence,* vol. 24, no. 7, pp. 971-987, 2002.

[7] L. Nanni, and A. Lumini, "RegionBoost Learning for 2D+3D based Face Recognition," *Pattern Recognition Letters,* vol. 28, no. 15, pp. 2063-2070, 2007.

[8] X. Shang, and R. Veldhuis, "Local absolute binary patterns as image preprocessing for grip-pattern recognition in smart guns," in First IEEE International Conference on Biometrics: Theory, Applications, and Systems, 2007, pp. 1-6.

[9] D. Unay, and A. Ekin, "Intensity versus texture for medical image search and retrieval " in 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 2008, pp. 241-244.

[10] A. Oliver, X. Lladó, J. Freixenet *et al.*, "False positive reduction in mammographic mass detection using local binary patterns," *Medical Image Computing and Computer-Assisted Intervention (MICCAI), Lecture Notes in Computer Science 4791*, pp. 286-293, Brisbane, Australia: Springer, 2007.

[11] E. G. Keramidas, D. K. Iakovidis, D. Maroulis *et al.*, "Thyroid texture representation via noise resistant image features," in Twenty-First IEEE International Symposium on Computer-Based Medical Systems (CBMS 2008), 2008, pp. 560-565.

[12] L. Nanni, and A. Lumini, "A reliable method for cell phenotype image classification," *Artificial Intelligence in Medicine,* vol. 43, no. 2, pp. 87-97, 2008.

[13] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 2037- 2041, 2006.

[14] L. Nanni, and A. Lumini, "Local binary patterns for a hybrid fingerprint matcher," *Pattern Recognition,* vol. 11, pp. 3461-3466, 2008.

[15] H. Zhou, R. Wang, and C. Wang, "A novel extended local binary pattern operator for texture analysis," *Information Sciences* vol. 178, no. 22, pp. 4314-4325, 2008.

[16] S. Liao, M. W. K. Law, and A. C. S. Chung, "Dominant local binary patterns for texture classification," *IEEE Transactions on Image Processing,* vol. 18, no. 5, pp. 1107 – 1118, 2009.

[17] L Nanni, A. Lumini, and S. Brahnam, "Local binary patterns variants as texture descriptors for medical image analysis," *Artificial intelligence in medicine*, In Press.

[18] S. Liao, and A.C.S. Chung, "Face recognition by using elongated local binary patterns with average maximum distance gradient magnitude," *Computer Vision*, pp. 672-679, 2007.

[19] X. Tan, and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *Analysis and Modelling of Faces and Gestures,* vol. LNCS 4778, pp. 168-182, 2007.

[20] T. Ahonen, and M. Pietikäinen, "Soft histograms for local binary patterns," in Finnish Signal Processing Symposium (FINSIG 2007), Oulu, Finland, 2007.

[21] D. K Iakovidis, E. Keramidas, and D. Maroulis, "Fuzzy local binary patterns for ultrasound texture characterization," *Image Analysis and Recognition, 5th International Conference (ICIAR 2008), Lecture Notes in Computer Science 5112*, pp. 750-759: Springer, 2008.

[22] A Hafiane, G. Seetharaman, K. Palaniappan *et al.*, "Rotationally invariant hashing of median binary patterns for texture classification," in ICIAR 2008, 2008, pp. 619-629.

[23] M. Heikkilä, M. Matti Pietikäinen, and C. Schmid, "Description of interest regions with local binary patterns," *Pattern Recognition*, 2008.

[24] W. Zhang, H. Shan, X. Chen *et al.*, "Local gabor binary patterns based on mutual information for face recognition," *nternational Journal of Image and Graphics,* vol. 7, no. 4, pp. 777–793, 2007.

[25] He Xiaofei, Deng Cai, Shuicheng Yan *et al.*, "Neighborhood preserving embedding."

[26] N. Cristianini, and J. Shawe-Taylor, *An introduction to support vector machines and other kernel-based learning methods*, Cambridge, UK: Cambridge University Press, 2000.

[27] T Fawcett, *ROC graphs: Notes and practical considerations for researchers*, HP Laboratories, Palo Alto, USA, 2004.

[28] A Chebira, Y. Barbotin, C. Jackson *et al.*, "A multiresolution approach to automated classification of protein subcellular location images," *BMC Bioinformatics,* vol. 8, pp. 210, 2007.

[29] J. Jantzen, J. Norup, G. Dounias *et al.*, "Pap-smear benchmark data for pattern classification," in Nature inspired Smart Information Systems (NiSIS), Albufeira, Portugal, 2005, pp. 1–9.

[30] S. Munder, and D. M. Gavrila, "An experimental study on pedestrian classification," *IEEE transactions on pattern analysis and machine intelligence,* vol. 28, no. 11, pp. 1863-1868.

[31] J. L. Fink, R. N. Aturaliya, M. J. Davis *et al.*, "LOCATE: A protein subcellular localization database.," *Nucleic Acids Research,* vol. 34, 2006.