CW-SSIM Kernel based Random Forest for Image Classification

Guangzhe Fan[a], Zhou Wang[b], and Jiheng Wang[a],

[a]Dept. of Statistics and Actuarial Science, Univ. of Waterloo, Waterloo, ON, Canada

[b]Dept. of Electrical and Computer Engineering, Univ. of Waterloo, Waterloo, ON, Canada

ABSTRACT

Complex wavelet structural similarity (CW-SSIM) index has been proposed as a powerful image similarity metric that is robust to translation, scaling and rotation of images, but how to employ it in image classification applications has not been deeply investigated. In this paper, we incorporate CW-SSIM as a kernel function into a random forest learning algorithm. This leads to a novel image classification approach that does not require a feature extraction or dimension reduction stage at the front end. We use hand-written digit recognition as an example to demonstrate our algorithm. We compare the performance of the proposed approach with random forest learning based on other kernels, including the widely adopted Gaussian and the inner product kernels. Empirical evidences show that the proposed method is superior in its classification power. We also compared our proposed approach with the direct random forest method without kernel and the popular kernel-learning method support vector machine. Our test results based on both simulated and real-world data suggest that the proposed approach works superior to traditional methods without the feature selection procedure.

Keywords: image similarity measure, image classification, CW-SSIM, random forest, kernel, recursive partitioning, tree classification, hand-written digit recognition

1. INTRODUCTION

Due to the high dimensionality nature of digital images, image classification algorithms typically require a feature extraction process (such as corner detection) or an appearance-based dimension reduction stage (such as principle component analysis) before the application of statistical learning and classification algorithms. Meanwhile, there has been some interesting recent progress on defining similarity metrics between two images that are in their original 2D functional form. These include the structural similarity (SSIM) index[1] and its extension – complex wavelet SSIM (CW-SSIM) index[2,3]. Conceptually, these similarity metrics have the potentials to be used in image classification problems, but there has not been sufficient study on how this should be performed in real-world scenarios. In this paper, we propose a novel but efficient way to incorporate CW-SSIM in a tree-based learning algorithm – the random forest[4,5], which is a powerful ensemble of many randomly perturbed classification trees. A traditional random forest is difficult to be implemented directly for image classification because it relies on a categorical feature space to construct trees. Here we are interested in using the concept of "kernel-induced classification tree and random forest"[6], which uses kernel functions to construct splitting rules in a tree-based models and thus the random forest. This overcomes the limitation of traditional direct random forest methods in functional types of classification problems.

Image similarity indices play a crucial role in the development, assessment and optimization of a large number of image processing and pattern recognition systems. An image can be viewed as a 2-D function of intensity. Perhaps the simplest way to compare the similarity of two images is to compute the mean squared error between these two 2D functions. Unfortunately, such a point-wise similarity measure does not take into account the correlation between neighboring image pixels and has been shown to be problematic in many ways[7]. Recently, a substantially different approach called the SSIM index[1] was proposed, where the structural information of an image is defined as those attributes that represent the structures of the objects in the visual scene, apart from the mean intensity and contrast. Thus, the SSIM index separates the comparison of local structural patterns from local mean intensity and contrast comparisons. The SSIM index has shown somewhat surprising success in predicting perceptual image quality when compared with more sophisticated methods based on psychological models of the human visual system[7]. A common drawback of both MSE and SSIM metrics is their high sensitivity to small geometric distortions such as translation, rotation and scaling. The CW-SSIM measure overcomes this problem by transforming SSIM to the complex wavelet transform domain[2,3]. The key

idea behind CW-SSIM is that small geometric image distortions lead to consistent phase changes in local wavelet coefficients, and that a consistent phase shift of the coefficients does not change the structural content of the image. The potential of CW-SSIM has been demonstrated with a series of applications, including image quality assessment[2], digit recognition[2], line-drawing comparison[3], segmentation comparison[3], range-based face recognition[8] and palmprint recognition[9].

Tree-structured classification models are popular alternatives to classical methods such as logistic regression and discriminant analysis. The Classification and Regression Trees (CART) methodology[10] has been a very popular and standardized approach in the statistics literature. CART and most other recursive partitioning algorithms grow a binary decision tree in a sequential fashion by using splitting rules based on predictor variables to split the data in such a way as to reduce variation in a response variable. In CART, specifically, the trees are grown as large as possible to avoid early stopping, and then pruned backward using a cross-validated cost-complexity criterion to avoid overfitting of the training data, so as to obtain the "right-sized" tree. When the data space is complicated, a single CART model is unstable and may not be able to fulfill a good job. Breiman[4,5] proposed the random forest procedure, which is a tree-based ensemble classifier to improve CART's predictive ability and stability. Basically, the data space is perturbed by either bootstrapped samples or random selections of predictor variables or both. Many different CART models are then grown in parallel on the perturbed data spaces. These different trees are combined to vote for the classification of new cases. Generally, random forest is much more accurate than a single CART model in classification.

2. METHODOLOGY

2.1 Complex Wavelet Structural Similarity Index (CW-SSIM)

The SSIM index was originally proposed to predict perceived image quality[1,7]. The fundamental principle is that the human visual system is highly adapted to extract structural information from the visual scene, and therefore, a measurement of structural similarity should provide a good approximation of perceptual image quality. In particular, SSIM attempts to discount those distortions that do not affect the structures (or local intensity patterns) of the image. In the spatial domain, the SSIM index between two image patches $x = \{x_i \mid i = 1, 2, …, M\}$ and $y = \{y_i \mid i = 1, …, M\}$ is defined as

$$S(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)},\tag{1}$$

where μ and σ are the sample mean, standard deviation and covariance terms of x, y and xy, respectively, and C_1 and C_2 are two small positive constants to avoid instabilities. The maximum value 1 is achieved if and only if x and y are identical.

The major drawback of the spatial domain SSIM algorithm is that it is highly sensitive to translation, scaling, and rotation of images. The CW-SSIM index is an extension of the SSIM method to the complex wavelet domain. The goal is to design a measurement that is insensitive to "non-structural" geometric distortions that are typically caused by nuisance factors, such as changes in lighting conditions and the relative movement of the image acquisition device, rather than the actual changes in the structures of the objects. The CW-SSIM index is also inspired by the impressive pattern recognition capabilities of the human visual system1. In the last three decades, scientists have found that neurons in the primary visual cortex can be well-modeled using localized multi-scale bandpass oriented filters that decompose natural image signals into multiple visual channels. Interestingly, some psychophysical evidence suggests that the same set of visual channels may also be used in image pattern recognition tasks[11]. Furthermore, phase contains more structural information than magnitude in typical natural images, and rigid translation of image structures leads to consistent phase shift. The CW-SSIM index is defined as

$$\tilde{S}(c_x, c_y) = \frac{2 | \sum_{i=1}^{N} c_{x,i} c_{y,i}^* | + K}{\sum_{i=1}^{N} | c_{x,i} |^2 + \sum_{i=1}^{N} | c_{y,i} |^2 + K}.\tag{2}$$

Here c_x and c_y are the sets of local coefficients (in the neighboring spatial locations of the same wavelet subband) extracted from the complex wavelet transforms (e.g. the complex version of the steerable pyramid decomposition[12]) of

the two images being compared, respectively, $c*$ denotes the complex conjugate of c, and K is a small positive constant. The purpose of K is mainly to improve the robustness of the CW-SSIM measure when the local signal-to-noise ratios are low. We consider CW-SSIM as a useful measure of image structural similarity based on the beliefs that 1) the structural information of local image features is mainly contained in the relative phase patterns of wavelet coefficients, and 2) constant phase shift of all coefficients does not change the structure of the local image feature.

2.2 Kernel-induced Classification Tree and Random Forest

A *kernel* is a function K, such that for all $\vec{x}_i$ and $\vec{x}_j \in X^p$, $i, j = 1, 2, \ldots, n$

$$K(\vec{x}_i, \vec{x}_j) = <\varphi(\vec{x}_i), \varphi(\vec{x}_j)>,$$ (3)

where Φ is a (non-linear) mapping from the input space to an (inner product) feature space. If the observation i is fixed in the training sample, and observation j is a new input, then the kernel function above can be treated as a new feature defined by observation i, denoted as $K(\vec{x}_i, \cdot)$. Some popular kernels are inner product kernel, polynomial kernel and Gaussian (radial basis) kernel.

A classification tree model is a recursive partitioning procedure in the feature space. Starting from the root node, at each step, a greedy exhaustive search is implemented to find the best splitting rule such as "$X_i < c$" for numerical features. If the answer is yes, then the observation will move to the left child node, and otherwise to the right child node. The procedure is implemented recursively until a very large binary tree is constructed. A large tree usually overfits the training sample. Next, a cross-validation process is employed to prune the tree back to its proper size. A single classification tree constructed using the procedure described above is highly interpretable but quite unstable and weak in prediction. Figure 1 gives an example. A random forest algorithm is simply a replication of the classification tree procedure while introducing a random vector in the construction space. This may be implemented by limiting the number of features to be searched at every step growing the tree and/or bootstrapping the data set. The trees in a random forest are usually very large and need no pruning. Due to the instability of classification trees, they are quite diversified due to the random vector introduced in the process. Each classification tree generally is a low-bias but high-variance model. When they are combined to vote for a decision, the variance is reduced and the classification power becomes very strong. Another attractive property is that including more trees in the random forest will not overfit the training samples. However, the random forest approach described above usually works in a feature space. For functional types of data such as images, they are difficult to be used directly.
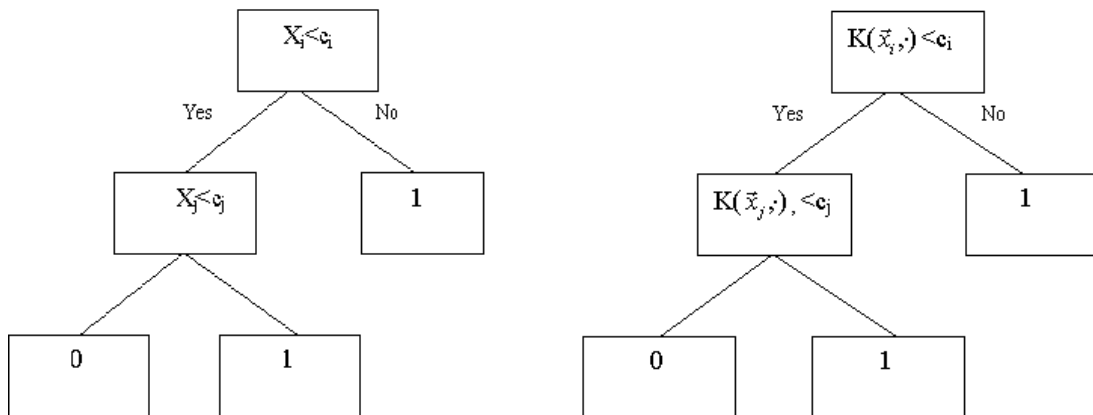

Figure 1. Left: an example of a classification tree with 3 terminal nodes. Note that X_i and X_j are features in the data space. Right: an example of a kernel-induced classification tree with 3 terminal nodes. Note that X_i and X_j are observation vectors in the data space and the $K(,)$'s are kernel functions defined on these observations

Here we use the concept of kernel-induced classification tree to overcome this limitation. Instead of using the raw signal features in the data space to construct splitting rules, kernel functions based on observations are employed. Since the definition of kernel is very flexible to handle various types of data, the potential of random forest is greatly enhanced and extended. Not only the feature space is enlarged, but also some complicated and non-linear hidden patterns between observations can be captured by the random forest learning procedure. An example of a kernel-induced classification tree is illustrated in the right figure of Figure 2. A kernel-induced random forest is simply a replication of many such trees with a random vector introduced. In this paper, the types of kernel we consider are Gaussian, inner product and CW-SSIM kernels.

3. RESULT

3.1 Experiment on Simulated data

To demonstrate the proposed algorithm, here we consider the classification problem of hand-written digit images as an example. The experiment is first carried out on simulated data. A training/testing digit image database of 500 images was created by shifting, scaling, rotating, and blurring ten hand-written template digit images. Figure 2 shows a random subset of examples in our image database.

Figure 2. A random sample of simulated hand-written digit images.

The random forest-based classification model with three different kernels (CW-SSIM, Gaussian and inner-product) was trained using 450 images randomly selected from the database, and then tested using the rest 50 images. This process is replicated 100 times, each with a different random division between training and testing sets. The misclassification rate and its standard deviation for each kernel are then computed. Since the number of trees in the random forest also has impact on the overall classification performance, we tested the performance of the algorithms with different numbers of trees. The test results are summarized in Figure 3, where the error bars represent ±1 standard deviations from the average test error rates. It can be observed that CW-SSIM kernel based random forest algorithm performs the best in all situations, while the Gaussian kernel method results in the worst performance. This verifies our hypothesis that CW-SSIM is a better kernel in the construction of image classification algorithms. Moreover, increasing the number of trees does not seem to improve Gaussian kernel based algorithm, but has positive impact on the other two methods. When the number of trees is reasonably large, the test error rate is likely to converge. We also observe that including more trees in the forest does not lead to overfitting.

To have a more diverse range of comparisons, we also included support vector machine (SVM) and direct random forest methods (without kernel) in our experiments. To perform a fair comparison with the proposed kernel-based approach, no

feature extraction and dimension reduction process is applied in our implementations of the SVM and direct random forest methods. For SVM, we adopt the commonly used Gaussian kernel and tune its parameter to achieve the best test set result. For random forest, we use the default settings which are similar to that of Kernel-induced random forest. Again, we used 450 images for training and 50 images for testing, and the process is replicated 100 times. The mean misclassification rate and its standard deviation for each method for all the algorithms under comparison are given in Table 1. It can be observed that the proposed CW-SSIM kernel random forest method significantly outperforms SVM, direct random forest alone, as well as other kernel-based random forest algorithms.
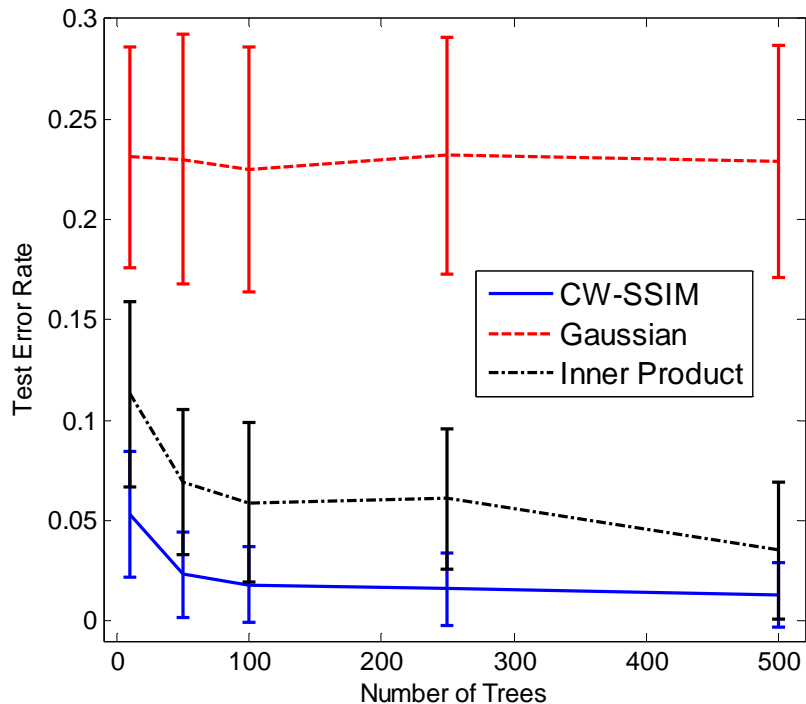

Figure 3. Test error rate plot of three different kernels and different number of trees in the kernel-induced random forest.

Table 1. The results of CW-SSIM induced random forest, original random forest and SVM on the simulated image data.

	Mean error rate (%)	Std.ev (%)
Inner product	5.8	3.4
Gaussian	22.9	5.8
CW-SSIM	1.3	1.6
random forest	7.7	4.2
SVM	15.1	4.9

3.2 Experiment on MINST database

We tested our method on the well-known MNIST benchmark of handwritten digit images[13]. The MNIST database is composed of 60000 training and 10000 test examples, where the data were collected among Census Bureau employees

and high school students. The original images have a normalized size of 28×28 and contain gray levels for the purpose of anti-aliasing. Some sample images are shown in Figure 4. In our preliminary test, we extracted a subset from the database for training and testing.

We tested the performance of random forest algorithms with three different kernels for different numbers of trees. The test results are summarized in Table 2, in which Columns 1 to 5 show 5 trials of tests where 4000 images randomly selected from 5000 images were used for training and the remaining 1000 images from the same set were employed for testing. The results in Column 6 used all the 5000 images for training and a separate set of 2000 images for testing. Clearly, CW-SSIM kernel based random forest algorithm performs significantly better than the other methods in all test situations, while the Gaussian kernel method results in the worst performance. The performance of the proposed method as a function of the number of trees in random forest is shown in Figure 5. It can be observed that the test error rate reduces to a low level with around 500 trees. Further increasing the number of trees does not improve the performance but leads to small fluctuations. In our experiment, SVM method without any preprocessing and feature selection process does not give reasonable result, so we do not report its performance in Table 2.
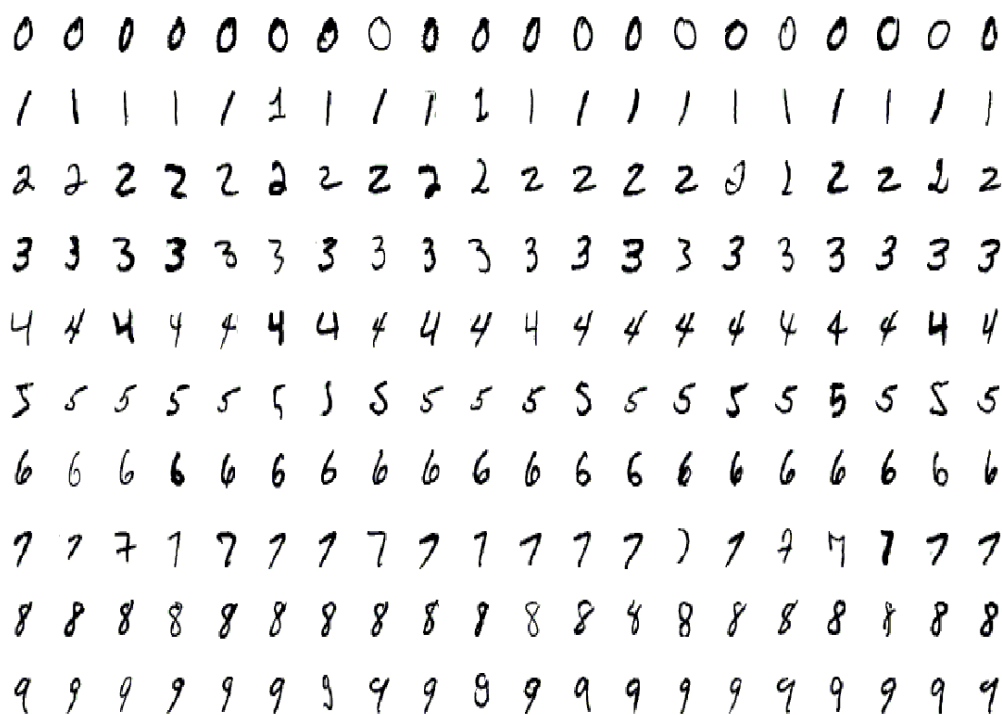

Figure 4. A random sample of the original hand-written digit images from MINST data

Table 2. Comparison of misclassification rates of direct and three kernel-based random forest methods on MNIST database

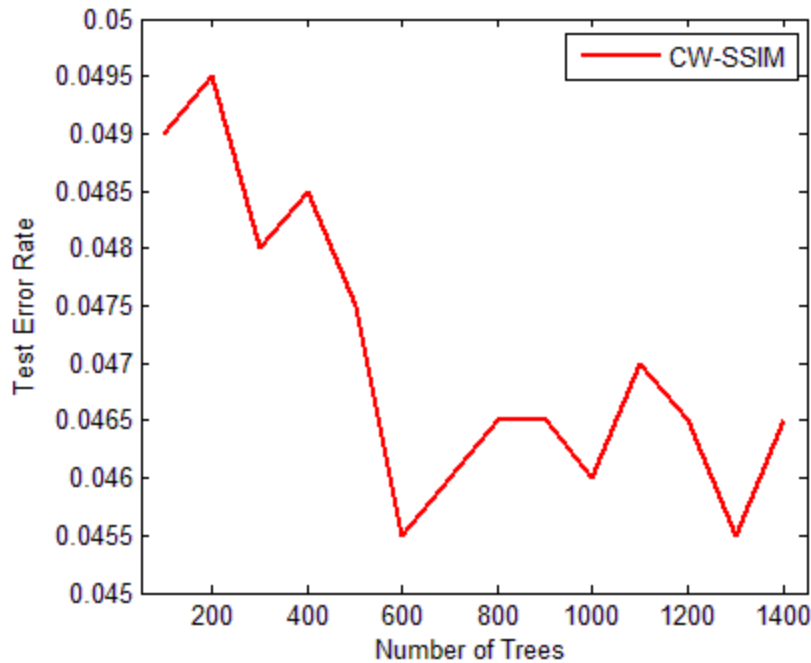|	Test 1000	Test 1000	Test 1000	Test 1000	Test 1000	Test 2000
cwssim	7.5	5.4	5.5	6.3	5.5	4.8
Gaussian	8.3	8.4	8.3	8.3	8.2	8.4
Inner Prod.	12.7	10.9	10.9	12.5	10.5	10.5
random forest	8.1	6.1	6.5	6.7	6.3	6.1

Figure 5. Test error rate of CW-SSIM kernel based random forest on MNIST database as a function of the number of trees.

4. CONCLUSION AND FUTURE WORK

We have proposed a kernel-based random forest learning and classification algorithm, which appears to be an effective and reliable tool for image classification when an appropriate kernel such as CW-SSIM is provided. An interesting feature of our approach is that no feature extraction or dimension reduction process is involved. Instead, our algorithm relies on the powerful CW-SSIM index that provides useful similarity measure between two misaligned images even without a registration stage at the front end. The random forest procedure is then applied to this CW-SSIM based kernel information space in a novel way to produce superior classification performance. Other kernel-based learning algorithms, such as SVM, may not easily generalize themselves to incorporate the CW-SSIM kernel and take advantage of this useful similarity measure. Our proposed work is novel, original and efficient.

The present work is only an initial attempt along a new research direction. There are a number of ways to extend our work in our future studies. First, we need a better understanding of the roles of kernels for different types of image classification problems. Second, the random forest learning algorithm needs to be adjusted in order to accommodate large training datasets. Third, the CW-SSIM measure may be combined with certain feature extraction methods to produce more reliable kernels. Finally, the proposed approach may be applied and tested in more diverse types of image classification applications.

REFERENCES

1. Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. (2004)
2. Wang Z. & Simoncelli E. P., "Translation insensitive image similarity in complex wavelet domain," *IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, vol. II, pp. 573-576, Philadelphia, PA, March (2005)
3. M. P. Sampat, Z. Wang, S. Gupta, A. C. Bovik and M. K. Markey, "Complex wavelet structural similarity: A new image similarity index," *IEEE Transactions on Image Processing,* vol. 18, no. 11, pp. 2385-2401, Nov. (2009)
4. Breiman, L., "Bagging Predictors," *Machine Learning*, 24, 123-140 (1996)

5. Breiman, L., "Random forests," *Machine Learning*. 45, 5-32 (2001)

6. Fan, G., "Kernel-Induced Classification Tree and Random Forest," Technical Report, Dept. of Statistics and Actuarial Science, Univ. of Waterloo (2009). http://www.stats.uwaterloo.ca/stats_navigation/techreports/09techreports.shtml

7. Z. Wang and A. C. Bovik, "Mean squared error: love it or leave it? - A new look at signal fidelity measures," *IEEE Signal Processing Magazine*, vol. 26, no. 1, pp. 98-117, Jan. (2009)

8. S. Gupta, M. P. Sampat, Z. Wang, M. K. Markey and A. C. Bovik, "Facial range image matching using the complex wavelet structural similarity metric," *IEEE Workshop on Applications of Computer Vision*, Austin, TX, Feb. 21-22, (2007)

9. L. Zhang, Z. Guo, Z. Wang and D. Zhang, "Palmprint verification using complex wavelet transform," *IEEE International Conference on Image Processing*, San Antonio, TX, Sept. 16-19, (2007).

10. Breiman, L., Friedman, J., Olshen, R., and Stone, C., Classification and Regression Trees, Belmont, CA: Wadsworth, (1984).

11. J. A. Solomon and D. G. Pelli, "The visual filter mediating letter identification," *Nature*, vol. 369, pp. 395–397, (1994).

12. letter identification," *Nature*, vol. 369, pp. 395–397, 1994.J. Portilla and E. P. Simoncelli, "A parametric texture model based on joint statistics of complex wavelet coefficients," *International Journal of Computer Vision*, vol. 40, pp. 49–71, (2000).

13. Y. LeCun and C. Cortes, The MNIST Database, http://yann.lecun.com/exdb/mnist/.