

Image splicing detection using 2-D phase congruency and statistical moments of characteristic function

Wen Chen^a, Yun Q. Shi^a, Wei Su^b

^aDepartment of Electrical and Computer Engineering
New Jersey Institute of Technology, Newark, NJ 07102, USA
{wc47@njit.edu, shi@njit.edu}

^bU.S. Intelligence & Information Warfare Directorate, Fort Monmouth, NJ 07703
{wei.su@us.army.mil}

ABSTRACT

A new approach to efficient blind image splicing detection is proposed in this paper. Image splicing is the process of making a composite picture by cutting and joining two or more photographs. The spliced image may introduce a number of sharp transitions such as lines, edges and corners. Phase congruency has been known as a sensitive measure of these sharp transitions and hence been proposed as features for splicing detection. In addition to the phase information, the magnitude information is also used for splicing detection. Specifically, statistical moments of characteristic functions of wavelet subbands have been examined to catch the difference between the authentic images and spliced images. Consequently, the proposed scheme extracts image features from moments of wavelet characteristic functions and 2-D phase congruency for image splicing detection. The experiments have demonstrated that the proposed approach can achieve a higher detection rate as compared with the state-of-the-art.

Keywords: Image splicing, phase congruency, statistical moments, characteristic functions, wavelet decomposition

1. INTRODUCTION

The manipulation of photographs has a long history since the invention of photography in the mid 19th century. In Victorian era, the combined photography (photomontage) was made by darkroom masking. A photomontage is a pictorial composite made by joining several pictorial elements together. Photomontage as an art form was fully developed after W. W. I. Today, with the development of digital technology, the creation of digital photomontage is an easy task by using software such as Photoshop. The photomontage may be used for different purposes. On the one hand, people can discover amusement from the photomontage. On the other hand, photomontage may be used for malicious purpose. For example, one can stick the head of a celebrity on a naked body and then broadcast this composite picture in the internet.

In the creation of photomontage, the most fundamental and essential operation is the image splicing. Image splicing is the process of combining image fragments from the same or different images without further post-processing such as smoothing of boundaries among different fragments. Even without the post-processing, the artifacts introduced by the image splicing may be almost imperceptible. In such situation it is hard for human observers to perceive the splicing involved. For example, the spliced images in Fig. 1 [1] look just like the authentic images.



Fig. 1. The examples of spliced images

The goal of the blind detection is to discriminate automatically spliced images from non-spliced (authentic) images. The blind detection of the image splicing is a challenging task. In the past, several blind detection approaches have been developed to tackle this specific and basic type of image forgery. In [2] a statistical model for “natural” image is applied to design a system to detect image manipulation which includes the image splicing. Ng *et al.* have proposed a detection method by extending an effective technique designed for detection of human speech splicing based on bicoherence to image splicing [3]. By detecting the inconsistency in lighting with respect to different parts in an image, Johnson and Farid have developed a technique of image splicing detection [4]. Hsu and Chang used geometry invariants and camera characteristic consistency to detect spliced image [5].

In this paper we introduce a novel approach to image splicing detection by exploiting the magnitude and phase information of a given test image. In our previous work [6], moments of wavelet characteristic functions have been successfully used in the steganalysis to distinguish the stego images from non-stego images. Inspired by this successful image model used in the steganalysis, we propose to use the moments of wavelet characteristic function as one part of the image features to detect the spliced images. It has been pointed out that Fourier phase conveys important information about image structure and features [7]. Image features such as step edges and lines have identical arrival phase of all Fourier harmonics [8, 9]. Phase congruency provides an absolute measure of the significance of such image features [8, 10]. The image splicing introduces some “unnatural” characteristics such as the lighting inconsistencies [4], camera characteristic inconsistency [5] and sharp image features. The sharp image features (edges, lines, corners, etc) give rise to high phase congruency. We propose to extract another part of features from the local phase congruency. The experiment showed that the combination of moments and phase congruency gave an improved classification performance as compared with the state-of-the-art [3].

The rest of this paper is organized as follows. Section 2 discusses the proposed approach. The characteristic function and concept of phase congruency are introduced. In Section 3, the experiments are presented. The proposed detection scheme with 120 image features is used. Furthermore, the discussion of the experimental results is presented. Finally conclusions are drawn in Section 4.

2. PROPOSED APPROACH

The splicing detection can be considered as a two-class pattern recognition problem. The input images are categorized into two classes: spliced image and non-spliced (authentic) image. The meaning of spliced image has been given in Section 1. The authentic image is defined as the output of an image acquisition device with no splicing involved. To accurately separate spliced images from authentic images, the selection of features with distinguishing ability is critical. In this section, we will describe how to extract the proposed features based on statistical moments of wavelet characteristic function and 2D phase congruency.

2.1 Moments of characteristic function

Image histogram has been widely used in image analysis. It is well-known that the histogram of a digital image or its wavelet subbands is essentially the probability mass function (pmf), if the image grayscale values or the wavelet coefficient values are treated as a random variable. That is, pmf is the normalized version of histogram. Any pmf p_x may be expressed as a probability density function (pdf) f_x by using the relation

$$f_x(x_0) = \sum_a p_x(a) \delta(x_0 - a) \quad (1)$$

Therefore we have the probability density function if each component of the histogram is multiplied by a corresponding shifted unit impulse. According to probability and random process theory [11, pp.145-148], the pdf and the characteristic function (CF) forms a Fourier transform pair (with the sign in the exponential reversed). Hence the Fourier transform of the histogram is essentially the characteristic function.

Denote histogram of an image (or its wavelet subbands) and its CF by $h(f_i)$ and $H(f_i)$, respectively, we propose to use the statistical moments of the CFs of both a test image and its wavelet subbands as image features. The n th moment of the CF is defined as follows

$$M_n = \sum_{j=1}^{N/2} f_j^n |H(f_j)| / \sum_{j=1}^{N/2} |H(f_j)| \quad (2)$$

where $H(f_j)$ is the CF component at frequency f_j , N is the total number of points in the horizontal axis of the histogram. Let's call these moments as features generated from the test image for short.

The splicing process often introduces sharp transition in the image 2-D array in terms of edges, lines and corners which are characterized by high frequency components in the Fourier transform domain. In order to enhance the content containing the high frequency components, we propose to predict each pixel grayscale value in the original test image by using its neighboring pixels' grayscale values, and obtain a prediction-error image by subtracting the predicted image from the test image. It is expected that this prediction-error image removes low frequency information and keeps high frequency information, thus making the splicing detection more efficient. Or, we may state that by using prediction-error image, we reduce the within-class distance while increasing the between-class distance. In addition to features from the test image, we also extract features in the same manner from the prediction-error image. The prediction-error image is the difference between the test image and its predicted version. The prediction algorithm is given by [12].

$$\hat{x} = \begin{cases} \max(a, b) & c \leq \min(a, b) \\ \min(a, b) & c \geq \max(a, b) \\ a + b - c & \text{otherwise} \end{cases} \quad (3)$$

where a, b, c are the context of the pixel x under considerations, \hat{x} is the prediction value of x . The locations of a, b, c are illustrated as in Fig. 2.

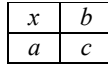


Fig.2: Prediction context

2.2 2-D Phase congruency

The image splicing leaves traces of image manipulation especially at locations where sharp image transition is introduced. The Local Energy Model developed by Morrone et al. [13] postulates that the sharp features are perceived at points of maximum phase congruency in an image. Phase congruency (PC) was first defined by Morrone and Owens [9] in terms of the Fourier series expansion of a signal at some location x as

$$PC_1(x) = \max_{\bar{\phi}(x) \in [0, 2\pi]} \frac{\sum_n A_n \cos(\phi_n(x) - \bar{\phi}(x))}{\sum_n A_n} \quad (4)$$

where A_n is the amplitude of the n th Fourier component, $\phi_n(x)$ is the local phase of the n th Fourier component at position x , and $\bar{\phi}(x)$ is the amplitude weighted mean local phase angle at position x . If PC equals to a maximal value of 1, all frequency components are in phase, as in the case of step edge in Fig.3 [14], otherwise, PC takes on values between 0 and 1.

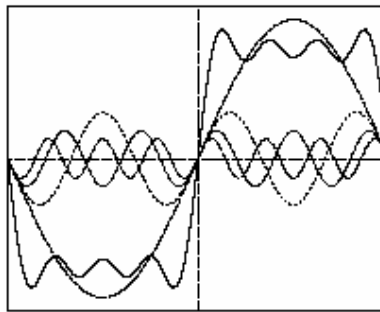


Fig.3: All the Fourier components for a square wave are all in phase at step point

Kovesi [14] addressed the problems encountered in the calculation of PC according to Equation (4) and developed a new and more sensitive measure of phase congruency. Kovesi also extended the 1-D PC to allow the calculation of 2-D PC of image by applying 1-D analysis over several orientations and combining the results in some way. To calculate 2-D PC of a given image, the image is first convolved with a bank of log-Gabor filters. Let the image denoted by $I(x, y)$, the even-symmetric filter and odd-symmetric filter at scale n and orientation o denoted by M_{no}^e and M_{no}^o , respectively. The responses of each quadrature pair of filters are a vector:

$$[e_{no}(x, y), o_{no}(x, y)] = [I(x, y) * M_{no}^e, I(x, y) * M_{no}^o] \quad (5)$$

where $*$ is the convolution operator. From Equation (5), the amplitude of this response is given by

$$A_{no}(x, y) = \sqrt{e_{no}^2(x, y) + o_{no}^2(x, y)} \quad (6)$$

and phase is given by

$$\phi_{no} = \text{atan}(e_{no}(x, y), o_{no}(x, y)) \quad (7)$$

The 2-D phase congruency is then calculated by

$$PC_2(x, y) = \frac{\sum_o \sum_n W_o(x, y) [A_{no}(x, y) \Delta\Phi_{no}(x, y) - T_o]}{\sum_o \sum_n A_{no}(x, y) + \varepsilon} \quad (8)$$

where $[\]$ denotes that the enclosed quantity is equal to itself if it is positive, and equal to zero otherwise; $W_o(x, y)$ is a measure of significance of frequency spread; ε is a small positive constant used to prevent division of zero; T_o is a quantity introduced to compensate image noise; and $\Delta\Phi_{no}(x, y)$ is a sensitive phase deviation function defined as

$$\Delta\Phi_{no}(x, y) = \cos(\phi_{no}(x, y) - \bar{\phi}_o(x, y)) - \left| \sin(\phi_{no}(x, y) - \bar{\phi}_o(x, y)) \right| \quad (9)$$

The sharp image features such as edge and line will be perceived and highlighted in the 2-D phase congruency array. We construct part of the proposed features from the 2-D phase congruency. MATLAB code for computing the 2-D phase congruency can be found in [15].

2.3 Feature extraction procedure

As demonstrated in Fig.4, the first 78 dimensional features are collected from the test image I and its prediction-error image \hat{I} . These features are moments of characteristic functions. The three-level Daubechies wavelet (we use db2 wavelet in MATLAB 6.5) decomposition of the test image is first carried out. At each level i , $i = 1, 2, 3$, there are four wavelet subbands (approximation, horizontal, vertical and diagonal). Totally there are 13 subbands if the image itself is considered a subband at level 0. For each wavelet subband, the first three moments are derived according to Equation (2), resulting in 39 dimensional features. Similarly, another 39 dimensional features are extracted from prediction-error image.

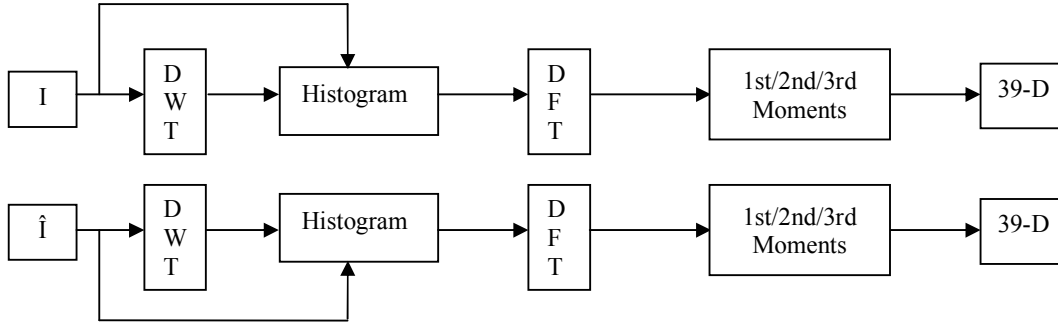


Fig.4: A block diagram of 78-D feature extracted from the test image and its prediction-error image

The additional 42 dimensional features are collected from three reconstructed images generated from the test image and three reconstructed images generated from the prediction-error image. To generate the reconstructed image I_{0i} ($i = 1, 2, 3$) from the test image, we first perform discrete Daubechies wavelet transform to the test image, erase the information (force the wavelet coefficients to be zero) contained in the approximation subband LL_i ($i = 1, 2, 3$), and then perform inverse Daubechies wavelet transform. In other words, I_{01} , I_{02} , and I_{03} are generated by erasure of approximation subband LL_1 , LL_2 and LL_3 , respectively. From each reconstructed image I_{0i} ($i = 1, 2, 3$), we extract seven image features: first three moments are calculated according to Equation (2), and four statistics – mean, variance, skewness and kurtosis – are computed based on 2-D array of phase congruency of the reconstructed image. The same procedure is conducted for each reconstructed image \hat{I}_{0i} ($i=1, 2, 3$) from the prediction-error image to collect another group of seven features. The procedure is illustrated in Fig.5.

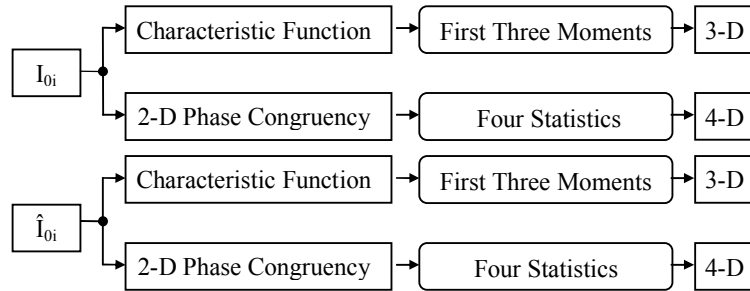


Fig.5: Feature extraction from the reconstructed images

The motivation for extracting image features based on phase congruency is that phase congruency provides an absolute measure of the significance of the features such as edges [8, 10]. The splicing process introduces edges or boundaries that may cause inconsistencies in the region. The phase congruency may sensitively capture the edge discontinuities due to splicing.

The motivation for using reconstructed images is that the splicing normally introduces the disturbance in the high frequency components in the spliced images, and the wavelet-based image sharpening by zeroing the approximation subband enhances the difference between the authentic and spliced images. This agrees with our experience in using prediction-error image in steganalysis [6]. Alternatively, we can state that this strategy is inspired by our successful experience in using prediction-error image in steganalysis.

3. EXPERIMENT

3.1 Image dataset

The image dataset is kindly provided by DVMM, Columbia University. It consists of 933 authentic and 912 spliced image blocks which all have the size of 128x128. Examples of authentic image are shown in Fig.6 below. Refer back to Fig.1 for spliced image examples. More details about the image sets can be found in [1].



Fig.6: Some examples of authentic image

3.2 Support Vector Machine classifier

Support Vector Machine (SVM) is used as the classifier in our experiment. At the training stage, given a set of training sample-label pairs (\bar{x}_i, y_i) , $i = 1, \dots, m$, where \bar{x}_i is the feature vector of the i th training sample and $y_i \in \{1, -1\}$ is the corresponding class label, SVM solves the following optimization problem [16]:

$$\begin{aligned} \min_{\bar{w}, b, \alpha_i} & \frac{1}{2} \bar{w}^T \bar{w} + C \sum_{i=1}^m \alpha_i \\ \text{subject to} & y_i (\bar{w}^T \phi(\bar{x}_i) + b) \geq 1 - \alpha_i, \alpha_i \geq 0 \end{aligned} \quad (10)$$

where C is the penalty parameter of the error term, and the function ϕ maps the training feature vector \bar{x}_i into a higher dimensional space. The kernel function $K(\bar{x}_i, \bar{x}_j) = \phi(\bar{x}_i)^T \phi(\bar{x}_j)$ is used to avoid the explicit computation of ϕ . In the experiment, we use one of the basic kernels - radial basis function (RBF):

$$K(\bar{x}_i, \bar{x}_j) = \exp\left(-\gamma \|\bar{x}_i - \bar{x}_j\|^2\right), \gamma > 0 \quad (11)$$

where γ is the kernel parameter. The penalty parameter C and kernel parameter γ must be carefully chosen so that the RBF kernel SVM can give accurate classification results. The “grid-search” method with cross-validation [17] was employed to find the optimal C and γ .

3.3 Experimental results and discussion

To evaluate the performance of the proposed classification system, we performed 20 runs of linear SVM and 20 runs of RBF kernel SVM with the best parameters C and γ . In each run, the training samples were randomly selected from the image dataset to train the classifier. We selected the training sample size to be 5/6 of images (775 authentic and 760 spliced), 1/2 of images (467 authentic and 456 spliced) and 1/3 of images (311 authentic and 304 spliced), respectively. The remaining images were used in testing.

The average detection rate of the 20-run experiments is shown in Table 1, where TP (true positive) represents the detection rate of spliced images, TN (true negative) represents the detection rate of authentic images, and accuracy is the average detection rate. These experimental results show that non-linear classifier outperforms linear classifier. For example, when 5/6 of images are randomly selected as training samples, RBF SVM achieves average detection rate of 82.32% which is 4.53% higher than linear SVM. The results also showed that all the detection rates are higher than 70% which was reported in [3] based on the same image dataset.

Table. 1. Detection rates

Training size	5/6			1/2			1/3		
	TP	TN	Accuracy	TP	TN	Accuracy	TP	TN	Accuracy
Linear SVM	0.8112	0.7538	0.7819	0.8070	0.7394	0.7728	0.7992	0.7199	0.7591
RBF SVM	0.8289	0.8177	0.8232	0.8094	0.7941	0.8017	0.7987	0.7737	0.7861

Generally, the available sample size for classifier design is limited, thus the performance of the trained classifier will vary with the samples available for classifier design. We compared the performance of the classifier for three different training sizes: 5/6, 1/2 and 1/3. Fig. 7 shows that the performance improves when the training sample size increases. It was also found that the proposed features are very effective because the classification performance did not degrade much after the sample size significantly decreases. For example, if the training sample size is reduced from 5/6 to 1/3 of the total images (from 1535 to 615), the number of training samples is reduced by about 60%, but the area under ROC curve only decreases by 3.69% from 0.8980 to 0.8611, the detection rate only drops by 3.71% from 0.8232 to 0.7861 (Table 1). This implies that the proposed features perform well even for small training size.

Finally, we list the detection rates achieved by applying different feature sets alone in Table 2 and examine how effective the different combination of the features sets is in splicing detection. The results in Table 2 are obtained when 5/6 sample images are used for training while 1/6 sample images are used for testing. Note that 78D feature set is moments of characteristic functions derived from the test image and its prediction-error image, 18D feature set is moments of characteristic functions from the reconstructed images, and 24D feature set are features are collected from 2-D phase

congruency of reconstructed images. Table 2 shows that 78 dimensional moment features or 24 dimensional phase features alone can achieve detection rate of more than 70%. If these two sets are combined, the detection rate is even higher and exceeds 80%. The best detection rate of 82.32% is achieved by the 120 dimensional features.

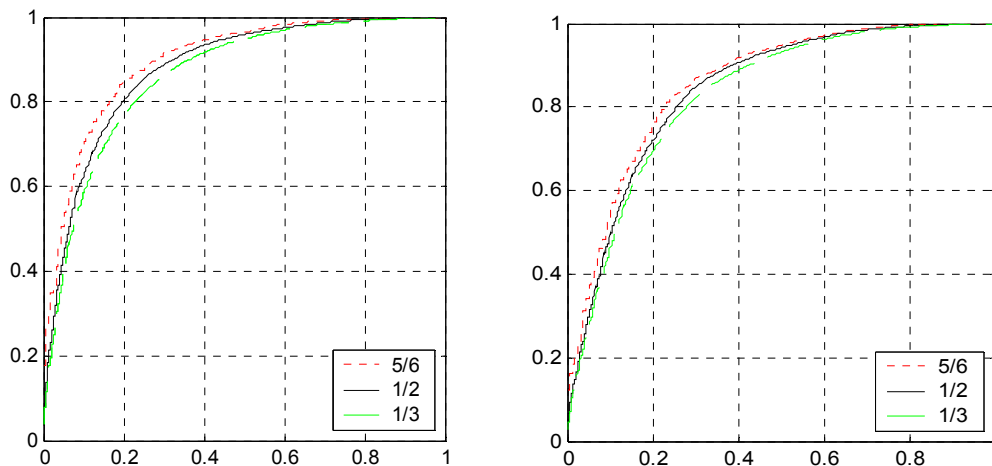


Fig.7: RBF kernel SVM (left), and Linear SVM (right)

Table 2: Detection rates for the combination of different feature sets

Feature Set	78D	18D	24D	78D+18D	18D+24D	24D+78D	78D+18D+24D
TP	0.7273	0.7227	0.7138	0.7724	0.8141	0.8289	0.8289
TN	0.7845	0.6737	0.7111	0.7883	0.7491	0.7959	0.8177
Accuracy	0.7565	0.6977	0.7124	0.7805	0.7810	0.8121	0.8232

4. CONCLUSIONS

In this paper, a novel splicing detection scheme is proposed. To detect the spliced images, the distinguishing image features are extracted by exploiting both magnitude and phase information of a given image. The first part of image features is the statistical moments of characteristic functions of a test image, its prediction-error image, and their wavelet subbands. This image model has been shown to be able to catch changes caused by image splicing. Furthermore, the approximation (LL) subband at different levels is individually erased (i.e., wavelet coefficients forced to be zero), and inverse wavelet transform is applied in order to enhance high frequency components. From these reconstructed images with LL subbands erased, moments of wavelet characteristic functions form the second part of features. Finally, the statistics (mean, variance, skewness and kurtosis) of 2-D phase congruency array associated with the above-mentioned reconstructed images are the third part of features for splicing detection. There are totally 120 dimensional image features which include 96 moment features and 24 phase features.

The Support Vector Machine (SVM) classifier is used in the proposed classification system. Both linear SVM and non-linear SVM with RBF classifiers have been used in experimental investigation. The experimental results have demonstrated that the proposed method outperforms the state-of-the-art by about 10% higher detection rate. The dependence of classification performance on the training sample size is also investigated and it has been found that the proposed image features are very effective and not sensitive to the training size.

ACKNOWLEDGEMENTS

Appreciation is directed to Dongdong Fu who brought the concept of phase congruency into our group discussion.

REFERENCES

1. Data set of authentic and spliced image blocks, DVMM, Columbia Univ., <http://www.ee.columbia.edu/dvmm/researchProjects/AuthenticationWatermarking/BlindImageVideoForensic/>
2. H. Farid, "A Picture Tells a Thousand Lies", *New Scientist*, 179(2411), pp. 38-41, Sept. 6, 2003.
3. T.-T. Ng, S.-F. Chang and Q. Sun, "Blind Detection of Photomontage Using Higher Order Statistics", *IEEE ISCAS*, May 2004.
4. M. K Johnson and H. Farid, "Exposing digital forgeries by detecting inconsistencies in lighting", *ACM Multimedia and Security Workshop*, 2005.
5. Y.-F. Hsu and S.-F. Chang, "Detecting image splicing using geometry invariants and camera characteristics consistency", *IEEE ICME*, July 2006.
6. Y.Q. Shi, G. Xuan, D. Zou, J. Gao, C. Yang, Z. Zhang, P. Chai, W. Chen and C.H. Chen, "Steganalysis based on moments of characteristic functions using wavelet decomposition, prediction-error image and neural network", *IEEE ICME*, July 2005.
7. A. V. Oppenheim and J. S. Lim, "The importance of phase in signals", *Proc. Of the IEEE*, vol. 69, pp. 529-541, May 1981.
8. M.C. Morrone and R.A. Owens, "Feature detection from local energy", *Pattern Recognition Letters*, vol. 6, pp. 303-313, 1987.
9. M.C. Morrone and D. C. Burr, "Feature detection in human vision: A phase-dependent energy model", *Proc. R. Soc. Lond. B*, vol. 235, pp. 221-245, 1988.
10. P. Kovési, "Phase congruency: A low-level image invariant", *Psych. Research*, vol. 64, pp. 136-148, 2000
11. A. Leon-Garcia, *Probability and Random Processes for Electrical Engineering*, 2nd Edition, Reading, MA: Addison-Wesley Publishing Company, 1994.
12. M. Weinberger, G. Seroussi, and G. Sapiro, "LOCOI: A low complexity context-based lossless image compression algorithm", *Proc. of IEEE Data Compression Conf*, pp.140-149, 1996.
13. M. C. Morrone, J. R. Ross, D. G. Burr, and R. A. Owens, "Mach bands are phase dependent", *Nature*, 324(6094):250-253, November 1986.
14. P. Kovési, "Image features from phase congruency", *Journal of Computer Vision Research*, 1(3): 1-26, 1999
15. P. D. Kovési, MATLAB code for calculating phase congruency and phase symmetry/asymmetry, April 1996. <http://www.cs.uwa.edu.au/~pk/Research/research.html>.
16. B., I. Guyon, and V. Vapnik, "A training algorithm for optimal margin classifiers", *In Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, 1992.
17. C.-W. Hsu, C.-C. Chang, and C.-J Lin, "A practical guide to support vector classification", July 2003.