INPAINTING-BASED CAMERA ANONYMIZATION

Sara Mandelli, Luca Bondi, Silvia Lameri, Vincenzo Lipari, Paolo Bestagini, Stefano Tubaro

Dipartimento di Elettronica, Informazione e Bioingegneria
Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milano, Italy

ABSTRACT

Over the years, the forensic community has developed a series of very accurate camera attribution algorithms enabling to detect which device has been used to acquire an image with outstanding results. Many of these methods are based on photo response non uniformity (PRNU) that allows tracing back a picture to the camera used to shoot it. However, when privacy is required, it would be desirable to anonymize photos, unlinking them from their specific device. This paper investigates a new and alternative approach to image anonymization task. The proposed method leverages image inpainting described as an inverse regularized problem, and does not need any priors about the PRNU to remove. Specifically, we show how PRNU pattern can be strongly attenuated by reconstructing each pixel of an image from its neighbors, only slightly affecting visual quality. Results confirm this approach as a viable alternative solution for image anonymization.

Index Terms— PRNU removal, image anonymization, inpainting, inverse problem regularization

1. INTRODUCTION

Image source attribution has been deeply studied by the multimedia forensic community [1, 2, 3]. This problem consists in detecting which device has been used to acquire a specific image, thus tracing back an image to its owner [4, 5]. The most promising approaches exploit photo response non uniformity (PRNU) [6, 7, 8]. This is a multiplicative noise pattern characterizing each camera sensor, which is inevitably injected into every acquired picture. By estimating a noise fingerprint from an image under analysis, it is possible to compare it with the PRNUs of known camera instances, thus detecting which device shot the picture.

When privacy is a concern, being able to link a picture to its owner is clearly undesirable. As an example, photoreporters carrying out legit investigations may prefer to anonymize their shots in order to avoid being threated. For this reason, counter-forensic methods that enable deleting or reducing PRNU traces from images have been proposed in the literature. Among the developed techniques, some require the knowledge of the PRNU pattern to be deleted. As an example, the authors of [9] and [10] propose different iterative solutions to delete a known PRNU from a given picture. Other methods work by blindly modifying pixel values and scrambling their positions in order to make the underlying PRNU unrecognizable. For instance, [11] shows that multiple image denoising steps attenuate the PRNU. Alternatively, [12] proposes to anonymize images by applying seam-carving to change pixel locations. More recently, [13] compares patch-based methods to shuffle small image blocks.

In this paper we investigate parallel and fast inpainting techniques as methods for image anonymization. Inpainting is a well-known topic, which refers to the application of simple or sophisticated algorithms for reconstructing lost or corrupted portions of an image (e.g., by solving partial differential equations, with the application of sparse domain transformations or the regularization of inverse problems, etc.). The rationale behind our approach is that, by deleting and reconstructing each pixel from its neighbors, PRNU effect can be strongly attenuated. The proposed method mainly works in two steps: (i) each image pixel is substituted by its inpainted value, in order to corrupt the PRNU; (ii) pixels around edges are replaced by denoised versions of the original ones in order to mask possible visual artifacts around sharp discontinuities.

Even though the literature is wide and provides many advanced solutions, we investigate the use of simple yet effective inpainting schemes that keep computational complexity at bay. Specifically, we only consider solutions that reconstruct pixels by solving inverse regularized problems. In particular, it is important to point out that a regularization which perfectly reconstructs the original image would lead to a high visual quality result, but would be totally useless for what concerns the anonymization task, as the traces of PRNU would remain almost unchanged. Therefore, in order to achieve our goal, the algorithm proposes a trade off between quality and anonymization of the outputs.

The investigated architecture is validated against 600 color images of the well known Dresden Image Database [14], cropped to a common size of 512×512 pixels. Results show that the selected inpainting techniques actually hinders camera source attribution detectors. Moreover, the edge processing step improves visual quality of the anonymized images. This makes the proposed technique a viable solution for image anonymization task compared to recently introduced methods.

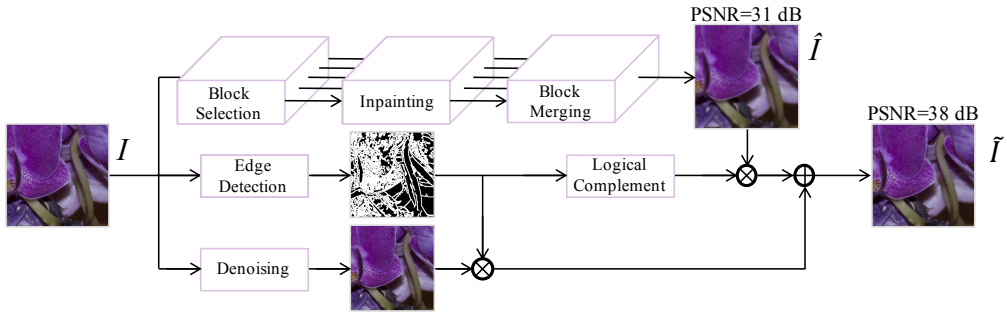

Fig. 1: Pipeline of the proposed anonymization strategy: an original image I undergoes block selection, and every modified version is inpainted in parallel. Results are merged to obtain $\hat{I}$, which is further processed around edges, thus obtaining the anonymized image $\tilde{I}$.

2. PROBLEM

The PRNU is a characteristic trace of each camera sensor introduced in acquired images as a multiplicative zero-mean noise pattern [4, 7]. Given an image I, it is possible to detect whether it has been acquired with a given camera by comparing a noise fingerprint extracted from the image and the camera PRNU K. Formally, the noise fingerprint can be estimated as $W = I - \dot{I}$, where $\dot{I}$ is a denoised version of I. In a nutshell, if the normalized cross-correlation (NCC) $\rho(W, I\,K)$ between W and a camera PRNU K pixel-wise scaled by I is higher than a confidence threshold Γ, the image is attributed to that camera [4, 7, 15].

In this paper we focus on the problem of image anonymization: given an image I, attenuate its PRNU traces so that it is impossible to attribute the image to the camera that shot it. In other words, if I is acquired with a camera whose estimated PRNU is K, it is reasonable that $\rho(W, I\,K) > \Gamma$. Our goal is to modify I through editing techniques in order to obtain an anonymized version $\tilde{I}$ with the following properties: (i) NCC between the noise extracted from the output image $\tilde{I}$ (i.e., $\tilde{W}$) and the camera PRNU gets to reasonably low values (i.e., $\rho(\tilde{W}, \tilde{I}\,K) < \Gamma$); (ii) image $\tilde{I}$ is not visually corrupted by the applied editing operations (i.e., peak-signal-to-noise ratio between I and $\tilde{I}$ assumes high values).

3. IMAGE ANONYMIZATION ALGORITHM

The proposed pipeline, shown in Fig. 1, is the following: (i) select and delete different blocks of I; (ii) blocks of I are processed in parallel, reconstructing the erased pixels from neighboring ones as in a classical inpainting problem; (iii) image $\hat{I}$ is reconstructed by splicing together all the inpainted blocks; (iv) the final anonymized image $\tilde{I}$ is obtained by further processing pixels around edges of $\hat{I}$, in order to mitigate inpainting artifacts near sharp discontinuities and increase visual quality. The algorithm is applied separately on each color plane. In the following, we describe each step considering focusing on a single color plane, without loss of generality. Implementation is available[1].

[1] https://bitbucket.org/polimi-ispl/

Block selection. The first step consists in selecting and deleting different image blocks as shown in Fig. 2. The selection of pixels is performed through N pixel selectors S, $s \in [1, N]$, applied in an element-wise fashion to I. Specifically, each pixel selector S_s is a matrix with the same size of the image, set to 0 if the pixel must be removed and to 1 elsewhere. This matrix, if multiplied pixel-wise with image I, selects and erases areas of $B \times B$ pixels, interleaved in both horizontal and vertical directions by a fixed gap of P pixels left at their original values. In order to gradually select and cancel all image pixels, each selector S_s is orthogonal to the others. This means that each selector deletes a different portion of the image, but the whole image is covered considering the effect of all selectors eventually. Fig. 2 shows an example in which 4 pixel selectors are multiplied by the original image in order to cancel some specific regions. We define each modified version of I as $I_s = S_s I$, with $s \in [1, N]$.

Image inpainting. We investigate the family of inpainting methods based on the solution of simple ℓ_1 and ℓ_2 regularized inverse problems. Our choice is motivated by the fact that this class of problems has been broadly implemented through standard and efficient iterative methods [16].

In this scenario, the inpainted image $\hat{I}_s$ is estimated by solving the minimization problem:

$$\hat{I}_s = \arg\min_{\bar{I}_s} \; \|S_s\,\bar{I}_s - I_s\|_{\mathrm{F}}^2 + \mu\,\|R(\bar{I}_s)\|_p^p\,. \qquad (1)$$

The pixel selector S_s selects the pixels from $\hat{I}_s$, $\|\cdot\|_{\mathrm{F}}$ represents the Frobenius norm, μ is the penalty weight associ-

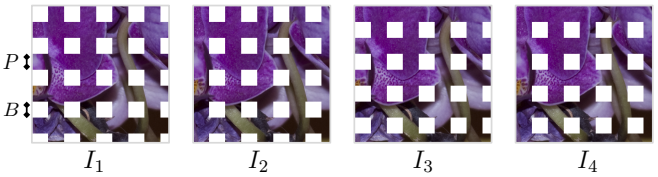

Fig. 2: Results of applying 4 pixel selectors to image I: each selector S_s selects and deletes different blocks of the original image. White areas represent deleted pixels.

ated to the regularizer operator $R(\cdot)$, and $\|\cdot\|_p$ is the entry-wise ℓ_p norm. The first term of the objective function (i.e., $\|S_s\,\bar{I}_s - I_s\|_F^2$) is a fitting condition. It basically imposes that the inversion result approximately honors the known samples of the image. The second term (i.e., $\mu\,\|R(\bar{I}_s)\|_p^p$) is the regularizer, which provides the condition to be respected by the inpainted pixels. Different regularizing operators and norms lead to different inpainting results.

A reasonable regularizer condition consists in imposing some smoothness constraints (i.e., $R(\cdot)$ should be a roughening operator). This ensures inpainted values to be not too dissimilar from neighboring pixels, which is desirable especially in flat regions of natural images (i.e., far from edges). Among the many regularization operators available in the literature, we make use of $R(\cdot) = \sqrt{(D_x(\cdot))^2 + (D_y(\cdot))^2}$, where operations are applied element-wise and $D_x(\cdot)$ and $D_y(\cdot)$ are the horizontal and vertical derivative operators, respectively. For instance, the result of applying $D_x(\cdot)$ to an image $\hat{I}_s$ is defined as

$$\bar{I}_{s_x}(i,j) = \begin{cases} \bar{I}_s(i,j) - \bar{I}_s(i,j+\Delta) & j \in [1, H-\Delta] \\ 0 & j \in [H-\Delta+1, H] \end{cases}, \tag{2}$$

where $\Delta > 0$ is the desired pixel gap for the derivative calculation, H represents the image height and width, as we are dealing with square images, and (i,j) represent the pixel locations within the image. By setting different gaps Δ, various definitions of derivative can be used.

For what concerns the norm applied to the regularization term, we consider two strategies: (i) a ℓ_2 norm leading to the well-known Tikhonov regularization [17], which is the most widespread technique for inverting ill-conditioned problems; (ii) a ℓ_1 norm strategy, known as Total Variation (TV) regularization [18], which exhibits edge preserving properties. While the ℓ_2 strategy is very simple from an implementation point of view, it is known to introduce an overall smoothing effect that may be undesirable around sharp edges. For this reason we also investigate the ℓ_1 norm applied to the previously defined operator, which is widely used for preserving edges and discontinuities in the final inpainting solution.

Block Merging. After obtaining the inpainted versions $\hat{I}_s$ of each image I_s, we construct the image $\hat{I}$ by merging the inpainted blocks:

$$\hat{I} = \sum_{s=1}^{N} (1 - S_s)\,\hat{I}_s. \tag{3}$$

Note that each pixel of $\hat{I}$ is an inpainted pixel, and no original pixels of I survive this operation. Moreover, due to S_s definition, each pixel of $\hat{I}$ is reconstructed from a single $\hat{I}_s$.

Edge Processing. Inpainting applied with the aforementioned solutions may still introduce some undesirable visual artifacts around edges (the effect can be noticed in Fig. 1, where image $\hat{I}$ has low peak signal to noise ratio). Therefore, to increase image visual quality, a further edge processing operation is applied. More specifically, we substitute pixels of $\hat{I}$ around image edges with pixels coming from a denoised version of I, in order to avoid the reintroduction of too much PRNU information. Indeed, [11] shows that denoising can attenuate PRNU, thus motivating our approach.

Formally, the edge reconstruction pipeline is as follows: (i) the original image I is denoised using two successive steps of BM3D algorithm (with variance parameter $\sigma = 7$) [19], thus obtaining $\dot{I}$; (ii) I edges are extracted using Canny edge detector (with its default parameters in MATLAB®), obtaining a binary edge mask E (the same size of I), which is 1 only at edge locations and 0 elsewhere; (iii) edge-mask E is dilated by means of a disk structural element (with radius of 3 pixels); (iv) the final anonymized image is defined as $\tilde{I} = \hat{I}\,(1 - E) + \dot{I}\,E$, where each operation is applied pixel-wise. Fig. 1 shows the effect of applying edge processing to $\hat{I}$. Note that peak signal to noise ratio (PSNR) of $\tilde{I}$ is increased by 7 dB. Moreover, the computational complexity of this step is mainly due to denoising operation, which is almost negligible with respect to the rest of the inpainting pipeline.

4. RESULTS

In this section we report all the details about the conducted experimental campaign. To this purpose, we first introduce the used experimental methodology. Then, we report results obtained with different inpainting strategies.

Methodology. Our experimental setup mimics that of Entrieri *et al.* [13], one of the most recent image anonymization approaches. For this reason, as image dataset, we randomly selected 600 never-compressed Adobe Lightroom images from the Dresden Image Database [14] coming from six camera instances (Nikon D70, Nikon D70s, Nikon D200, two devices each). All images were synchronized to landscape orientation and cropped to the central portion of size 512×512 pixels. The estimation of the clean sensor fingerprint K for each camera was obtained from 25 homogeneously lit flatfield images as typically suggested [4, 15]. Noise residuals W were computed with the Wavelet-based filter [20] commonly used in PRNU extraction [4].

We show results in terms of receiver operating characteristic (ROC) curves of a camera identification PRNU-based detector. Specifically, fixing the PRNU of a given camera, NCCs obtained from anonymized images taken with that camera define the set of positive samples, whereas the set of negative ones includes NCC values from all images not taken with that camera. Our goal is to reduce as much as possible the area under the curve (AUC), thus making the PRNU-based detector not working. Moreover, to evaluate image quality level, we report the relationship between PSNR and true positive rate (TPR) calculated at a fixed false positive rate (FPR) of 1%, which we denote as $\text{TPR}_{0.01}$. The goal is to reach a high PSNR with low values of TPR and AUC.

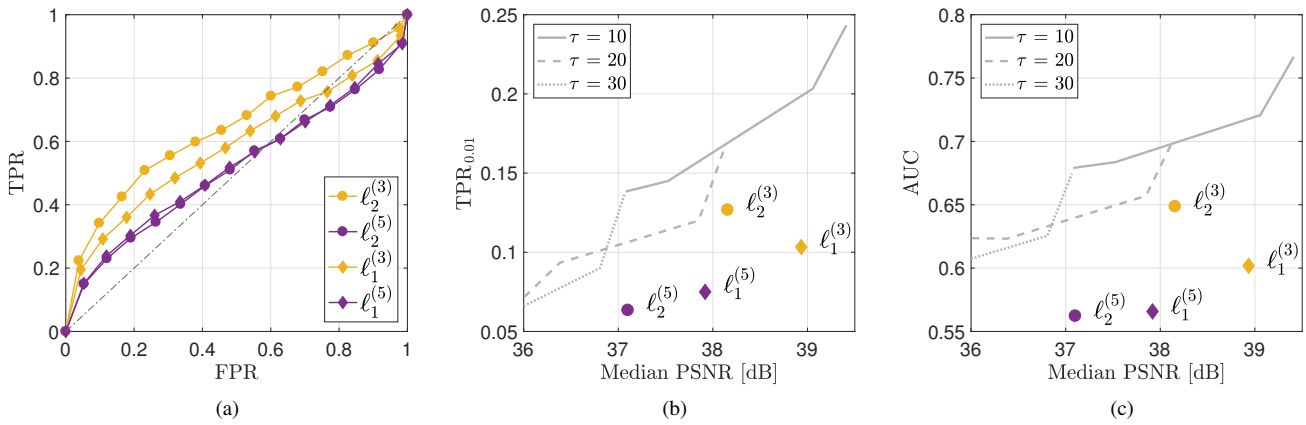

Fig. 3: (a) ROC curves after camera anonymization process (each color represents a different inpainting strategy). (b) Median PSNR vs. $\text{TPR}_{0.01}$ for different inpainting strategies, in comparison with PatchMatch-based (gray lines) patch replacement [13]. (c) Median PSNR vs. AUC for different inpainting strategies, in comparison with PatchMatch-based (gray lines) patch replacement [13].

Results are always averaged on all six devices.

To distinguish between the proposed inpainting strategies, we use the following notation: each technique is identified by the label $\ell_p^{(B)}$, where ℓ_p represents the selected norm for the cost function regularization (i.e., $p \in \{1, 2\}$), while the superscript defines the size of the $B \times B$ regions deleted by the pixel selectors. For what concerns the derivative implementation, we noticed that $\Delta = 3$ provides a good trade off between high reconstruction quality and low PRNU correlation. Indeed, smaller Δ tend to inpaint the image from nearer (and more correlated) pixels, consequently enhancing the correlation with the camera fingerprint. Conversely, larger Δ result in low PSNR. The penalty weight associated to the regularizer is $\mu = 10^{-20}$ for both ℓ_2 and ℓ_1 norms, since higher values oversmooth the image, thus reducing the quality.

Testing. The first experiment aims at showing that the proposed pipeline is actually able to fool PRNU-based camera attribution detectors. To this purpose, we tested both ℓ_2 and ℓ_1 inpainting by considering different block sizes $B \in \{3, 5\}$. Fig. 3(a) reports promising ROC curves, as their slopes are approximately 45^{o} degrees (i.e., perfect anonymization). Every strategy seems to provide satisfying performances, even though, the bigger the block dimensions ($B = 5$ instead of 3), the better the anonymization results.

For what concerns the inpainting effect on image quality, Figs. 3(b) and 3(c) show encouraging results, both in terms of $\text{TPR}_{0.01}$ and AUC as functions of the median PSNR of inpainted images. As the goal of image anonymization is to reduce TPR or approaching AUC $\simeq 0.5$ still granting high PSNR, the best solutions for Figs. 3(b) and 3(c) are those in the lower right quadrant of the graph. In particular, notice that with the ℓ_1 strategy we are able to reduce $\text{TPR}_{0.01}$ under 11% and to obtain an AUC of 0.6 by still achieving a median PSNR around $39\,\text{dB}$.

As state-of-the-art comparison, Figs. 3(b) and 3(c) also show results of the PatchMatch-based image anonymization proposed in [13], adapted to work on color images. This algorithm depends on two parameters: (i) τ is an error threshold; (ii) σ is a smoothing factor. Each gray line in Figs. 3(b) and 3(c) represents results obtained for a given τ and changing $\sigma \in \{0.1, 0.75, 2, 4\}$. This comparison shows that the proposed methodology is an effective alternative to PatchMatch-based anonymization, both in terms of $\text{TPR}_{0.01}$ and AUC. For instance, the ℓ_1 solution for both $B \in \{3, 5\}$ is able to match state-of-the-art performances in terms of median PSNR, while gaining about 0.1 in terms of both $\text{TPR}_{0.01}$ and AUC.

5. CONCLUSIONS

In this paper we proposed a pipeline for image anonymization against PRNU-based detectors. This approach is based on image inpainting to reconstruct image pixels, and edge processing to increase image visual quality. We tested different inpainting strategies, showing that it is possible to attenuate PRNU traces even exploiting simple inpainting solutions.

Considering that results are competitive with up to date state-of-the-art blind PRNU removal solutions [13], the investigated pipeline proves an interesting alternative method for image anonymization. Moreover, the proposed framework is computationally efficient because of its high parallelization potential. Indeed, inpainting is computed in parallel on different versions of the image, and results are merged at the end.

Given the intrinsic flexibility of the proposed pipeline to any inpainting algorithm, future work will be devoted to further increase the anonymization capability, investigating more sophisticated inpainting solutions. Moreover, the effect of working with JPEG compressed images will be studied.

6. REFERENCES

[1] M. Kharrazi, H. T. Sencar, and N. Memon, "Blind source camera identification," in *IEEE International Conference on Image Processing (ICIP)*, 2004.

[2] S. Bayram, H. Sencar, N. Memon, and I. Avcibas, "Source camera identification based on CFA interpolation," in *IEEE International Conference on Image Processing (ICIP)*, 2005.

[3] L. Bondi, L. Baroffio, D. Güera, P. Bestagini, E. J. Delp, and S. Tubaro, "First steps toward camera model identification with convolutional neural networks," *IEEE Signal Processing Letters (SPL)*, vol. 24, no. 3, pp. 259–263, March 2017.

[4] J. Lukáš, J. Fridrich, and M. Goljan, "Digital camera identification from sensor pattern noise," *IEEE Transactions on Information Forensics and Security (TIFS)*, vol. 1, pp. 205–214, 2006.

[5] M. Kirchner and T. Gloe, "Forensic camera model identification," in *Handbook of Digital Forensics of Multimedia Data and Devices*. John Wiley & Sons, Ltd, 2015.

[6] M. Chen, J. Fridrich, M. Goljan, and J. Lukas, "Source digital camcorder identification using sensor photoresponse nonuniformity," in *SPIE Electronic Imaging (EI)*, 2007.

[7] M. Chen, J. Fridrich, M. Goljan, and J. Lukas, "Determining image origin and integrity using sensor noise," *IEEE Transactions on Information Forensics and Security (TIFS)*, vol. 3, pp. 74–90, 2008.

[8] L. Gaborini, P. Bestagini, S. Milani, M. Tagliasacchi, and S. Tubaro, "Multi-clue image tampering localization," in *IEEE International Workshop on Information Forensics and Security (WIFS)*, 2014.

[9] A. Karaküçük and A. E. Dirik, "Adaptive photoresponse non-uniformity noise removal against image source attribution," *Journal of Digital Investigation*, vol. 12, pp. 66–76, 2015.

[10] H. Zeng, J. Chen, X. Kang, and W. Zeng, "Removing camera fingerprint to disguise photograph source," in *IEEE International Conference on Image Processing (ICIP)*, 2015.

[11] K. Rosenfeld and H. T. Sencar, "A study of the robustness of PRNU-based camera identification," in *IS&T/SPIE Electronic Imaging (EI)*. International Society for Optics and Photonics, 2009.

[12] A. E. Dirik, H. T. Sencar, and N. Memon, "Analysis of seam-carving-based anonymization of images against PRNU noise pattern-based source attribution," *IEEE Transactions on Information Forensics and Security (TIFS)*, vol. 9, pp. 2277–2290, 2014.

[13] J. Entrieri and M. Kirchner, "Patch-based desynchronization of digital camera sensor fingerprints," in *IS&T Electronic Imaging (EI)*, 2016.

[14] T. Gloe and R. Böhme, "The dresden image database for benchmarking digital image forensics," *Journal of Digital Forensic Practice*, vol. 3, pp. 150–159, 2010.

[15] G. Chierchia, S. Parrilli, G. Poggi, C. Sansone, and L. Verdoliva, "On the influence of denoising in PRNU based forgery detection," in *ACM Workshop on Multimedia in Forensics, Security and Intelligence (MiFor)*, 2010.

[16] K. Papafitsoros, C. B. Schoenlieb, and B. Sengul, "Combined first and second order total variation inpainting using split bregman," *Image Processing On Line (IPOL)*, vol. 3, pp. 112–136, 2013.

[17] A. Tikhonov and V. Arsenin, *Solutions of ill-posed problems*, Scripta series in mathematics. Winston, 1977.

[18] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, pp. 259268, 1992.

[19] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3D transform-domain collaborative filtering," *IEEE Transactions on Image Processing (TIP)*, vol. 16, pp. 2080–2095, 2007.

[20] M. K. Mihcak, I. Kozintsev, K. Ramchandran, and P. Moulin, "Low-complexity image denoising based on statistical modeling of wavelet coefficients," *IEEE Signal Processing Letters (SPL)*, vol. 6, pp. 300–303, 1999.