Anti-Forensics of Lossy Predictive Image Compression

Yuanman Li, *Student Member, IEEE* and Jiantao Zhou, *Member, IEEE*

***Abstract*—Image compression evidence has been utilized as an important forensic feature to justify image authenticity. However, some recent studies showed that the compression evidence of block transform-based image coding, e.g., JPEG and JPEG2000, can be effectively erased by adding designed dither noise in the transform domain. In this paper, we demonstrate that it is also feasible to hide the compression evidence of lossy predictive image coding, a class of compression paradigm widely employed in critical scenarios. To tackle the challenging issue of error propagation inherent to predictive coding, we design a prediction-direction preserving strategy, allowing us to add dither noise in the prediction error (PE) domain, while minimizing the incurred distortion. Extensive experimental results are provided to verify the effectiveness of the proposed anti-forensic algorithm for lossy predictive image coding.**

***Index Terms*—Anti-forensics, digital forensics, lossy predictive image compression**

I. Introduction

Digital images have been widely used in many applications, e.g., entertainment, medicine, sciences, space exploration, precision engineering. However, its digital nature also enables easy manipulations via the popular image editing software such as Photoshop. Therefore, a great deal of concern has been raised regarding the image legitimacy and authenticity.

Various image forensic approaches were proposed to assess the credibility of image [1]–[3]. Generally, image forensic methods can be classified into two categories: *extrinsic* solutions [1], in which the validation is based on the additional information embedded into the original image; and *intrinsic* solutions [1]–[3], in which no extra information needs to be embedded. The latter approach has received increasing attention, since very often the received image is the only information we have. Its success is due to the following observations: most image processing operations such as acquisition, compression, channel coding, etc. inevitably leave unique artifacts, which are detectable. Such *intrinsic fingerprints* are naturally and inherently generated in the processing chain, offering reliable evidence to help identify the origin and even detect the alternations of image.

Among the many types of intrinsic fingerprints, the one introduced by compression is of particular forensic significance [1]–[3]. This is because most digital images are subject to compression either by the on-camera engines or for more efficient storage. If the compression evidence can be detected and the associated parameters can be estimated accurately, then we can have better understanding of the image processing history, and further justify the image credibility. Along this line, many forensic techniques were suggested to extract the compression evidence of JPEG [2], JPEG 2000 [3], and lossy predictive image coding [3].

Nevertheless, recent studies argued that the forensic techniques may not be trustworthy if they do not preclude the existence of anti-forensic algorithms capable of hiding the image manipulation fingerprints [4]–[8]. In [4], Stamm *et al.* showed that anti-forensically modified image with smooth and continuous distribution of DCT/DWT coefficients can be produced to fool a variety of JPEG/JPEG2000 forensic algorithms, e.g., [2]. This is achieved by adding designed dither noise in the DCT/DWT domain. To improve the quality of the resulting image, Valenzise *et al.* employed an adaptive dither addition by solving a minimum-cost bipartite graph matching problem [5]. Recently, Fan *et al.* demonstrated that better tradeoff between forensic undetectability and image distortion can be achieved through a two-round TV-based de-blocking and histogram smoothing [6]. However, all the existing anti-forensic algorithms only addressed the block transform-based codecs, while leaving the lossy predictive coding unexplored. The lossy predictively coded images are mostly used in critical scenarios, e.g., medicine, remote sensing, and military, due to its superior performance at medium/high rates and the ability to better preserve fine structures [9]. These targeting applications make them easier to be threatened, compared with the ones coded by JPEG/JPEG2000. It is therefore crucial to know the feasibility and strength of image anti-forensic techniques tailored to lossy predictive image coding.

In this paper, we show that it is indeed possible to erase the compression evidence of lossy predictive image coding. To tackle the challenging issue of error propagation inherent to predictive coding, we design a prediction-direction preserving strategy, allowing us to add dither noise in the prediction error (PE) domain, while minimizing the incurred distortion. Note that the proposed method is not a trivial extension of the anti-forensic algorithms for block transform-based image codecs because: 1) direct addition of dither noise in the PE domain generates severely distorted image, due to the high sensitivity of PE sequence against disturbances; 2) in predictive coding, the future pixel reconstruction depends on all the previously coded pixels, while different blocks are processed independently in block transform-based coding. Such dependence makes the task of adding dither much more challenging than that in block transform-based coding.

The rest of this paper is organized as follows. Section II

This work was supported in part by the Macau Science and Technology Development Fund under grants FDCT/009/ 2013/A1 and FDCT/046/2014/A1, in part by the Research Committee at University of Macau under grants MYRG2014-00031-FST and MYRG2015-00056-FST, and in part by the National Science Foundation of China under grant 61402547.

Yuanman Li and Jiantao Zhou are with the Department of Computer and Information Science, Faculty of Science and Technology, University of Macau. (Corresponding author: Jiantao Zhou, email: jtzhou@umac.mo).

briefly introduces the predictive image coding. Section III describes the proposed anti-forensic technique. The experimental results are given in Section IV, and Section V concludes.

II. Lossy predictive image coding and the compression evidence

Predictive image coding is a class of simple compression paradigm achieving good coding performance without significant computational overhead [10]–[12]. Compared with the state-of-the-art JPEG 2000, CALIC [10], [11], a classical predictive image codec, can obtain higher PSNR values and tighter error bound on every single pixel simultaneously, when the bit rate is above 1 bpp [9]. The superior performance at medium/high rate region and the ability to better preserve fine details make predictive coding an excellent compression choice for critical applications such as medical imaging, satellite imaging, etc. The demanding requirements in these scenarios make it crucial to investigate the strength of the forensic tools and the anti-forensic strategies designed specifically for lossy predictive image coding.

Without loss of generality, we here use the lossy version of CALIC as an representative predictive image codec to present our findings. The extension to the other predictive image codecs shall be straightforward. Under this compression framework, each pixel of the input image $\mathbf{I}$ is processed sequentially in a raster-scan order. To encode a pixel I_i, we first make a prediction $\tilde{I}_i$ according to the causal, *reconstructed* (not original) neighboring pixels. In CALIC, seven causal, neighboring pixels in the `west (w)`, `west west (ww)`, `north (n)`, `north west (nw)`, `north east (ns)`, `north north (nn)`, `north north east (nne)` directions are employed to drive the gradient adaptive predictor (GAP), as shown in (1). The relative positions to pixel i are illustrated in Fig. 1. The associated PE e_i can then be calculated by

$$e_i = I_i - \tilde{I}_i \tag{2}$$

which is further quantized into $\breve{e}_i$. Most lossy predictive image codecs utilize the following simple uniform scalar quantization

$$\breve{e}_i = \begin{cases} (2\tau+1)\lfloor (e_i+\tau)/(2\tau+1) \rfloor & \text{if } e_i \geq 0 \\ (2\tau+1)\lfloor (e_i-\tau)/(2\tau+1) \rfloor & \text{if } e_i < 0 \end{cases} \tag{3}$$

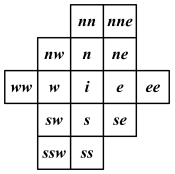

Fig. 1. Positions relative to pixel i.

where the non-negative integer τ is the quantization parameter balancing the distortion and the bit rate. Eventually, the corresponding quantization bin index is entropy encoded to produce the compressed bit stream.

The reconstructed pixel value can be obtained by

$$\hat{I}_i = \tilde{I}_i + \breve{e}_i \tag{4}$$

which is also maintained at the encoder side to calculate the prediction according to (1). Otherwise, the encoder and decoder cannot be synchronized.

Due to the quantization applied in the PE domain, the predictively coded image can be easily distinguished from the original, uncompressed one. As can be observed clearly in Fig. 3 and Fig. 4, the distribution of the prediction errors of the predictively coded image represents obvious comb-like structure. This phenomenon arises because the quantization and de-quantization operations force the prediction errors to take values in a discrete set. Lin *et al.* showed that such unique features can be easily detected by a forensic detector [3].

III. Proposed anti-forensic framework for lossy predictive image coding

Given the work of [4] and [6], a natural question arising is whether we can extend it to erase the compression evidence of lossy predictive image coding. A seemingly straightforward approach is to generate the dither noise sequence in a way similar to [4] and [6], add it with the PE sequence, and eventually transform the modified PE sequence back into the pixel domain. Unfortunately, this naive extension does not work, because of the high sensitivity of PE sequence against even tiny disturbances. Though the distortion in the PE domain is small, the incurred distortion in the pixel domain could be quite large. For instance, when $\tau = 5$, the reconstructed Lena from the PE sequence with straightforwardly added dither noise is only of 13.99 dB (see Fig. 2), which is too low to satisfy the requirement of anti-forensics.

IF $(d_v - d_h > 80)$ {sharp horizontal edge} $\tilde{I}_i \leftarrow \hat{I}_w$ ($\hat{I}_x$ is the *reconstructed* pixel value at direction x relative to i)
ELSE IF $(d_v - d_h < -80)$ {sharp vertical edge} $\tilde{I}_i \leftarrow \hat{I}_n$
ELSE{ {no obvious edge} $\tilde{I}_i \leftarrow (\hat{I}_w + \hat{I}_n)/2 + (\hat{I}_{ne} - \hat{I}_{nw})/4$
  IF $(d_v - d_h > 32)$ {horizontal edge} $\tilde{I}_i \leftarrow (\tilde{I}_i + \hat{I}_w)/2$
  ELSE IF $(d_v - d_h > 8)$ {weak horizontal edge} $\tilde{I}_i \leftarrow (3\tilde{I}_i + \hat{I}_w)/4$
  ELSE IF $(d_v - d_h < -32)$ {vertical edge} $\tilde{I}_i \leftarrow (\tilde{I}_i + \hat{I}_n)/2$
  ELSE IF $(d_v - d_h < -8)$ {weak vertical edge} $\tilde{I}_i \leftarrow (3\tilde{I}_i + \hat{I}_n)/4$}
where $d_h = |\hat{I}_w - \hat{I}_{ww}| + |\hat{I}_n - \hat{I}_{nw}| + |\hat{I}_n - \hat{I}_{ne}|, d_v = |\hat{I}_w - \hat{I}_{nw}| + |\hat{I}_n - \hat{I}_{nn}| + |\hat{I}_{ne} - \hat{I}_{nne}|$. (1)

Fig. 2. Reconstructed image (13.99 dB) by directly adding dither in the prediction error domain.

In this work, our contribution is to specifically address the challenging issue of suppressing the error propagation effect in the process of adding dither, allowing us to design a feasible anti-forensic scheme to hide the compression evidence of lossy predictive image coding. Our design goal is two-fold: 1) make the distribution of the PE sequence of the modified image smooth, continuous, and indistinguishable from the original one; 2) maximally preserve the quality of the resulting anti-forensically modified image.

We can easily know the quantization parameter τ and get all the quantized prediction errors $\breve{e}$ from the available, predictively coded image. We then anti-forensically modify each quantized PE by *strategically* adding designed dither noise in the following form

$$\breve{e}' = \breve{e} + d \tag{5}$$

where d is the anti-forensic dither and $\breve{e}'$ is the modified PE. The resulting image can be obtained by converting the sequence of $\breve{e}'$ back to the pixel domain. There are two key problems that need to be solved: 1) how to generate the dither sequence? and 2) how to add it with the quantized PE? Our contribution primarily lies in solving the latter challenge, as the first one has been largely resolved, as discussed in [4] and [6].

A. Dither Generation

The dither sequence is designed in such a way that the distribution of $\breve{e}'$ corresponds to a model distribution of the PE sequence obtained from an original image. For uncompressed image, the PE can be satisfactorily modeled as a zero-mean Laplacian distribution [9]. Because the AC components of DCT transform can also be modeled by the same distribution, many results of [4] and [6] can be directly applied here. Following the framework of [4], we can derive the dither distribution as

$$f_D(D=d|\breve{E}=\breve{e}) = \begin{cases} \frac{\hat{\lambda}}{2(1-y)}\exp\left(-\hat{\lambda}|d|\right), & \text{if } \breve{e}=0 \text{ and } d\in[-\tau,\tau]\bigcap\mathcal{Z} \\ \frac{\hat{\lambda}}{y^{-1}-y}\exp\left\{-\text{sgn}(\breve{e})\hat{\lambda}d\right\}, & \text{if } \breve{e}=n\Delta \text{ and } d\in[-\tau,\tau]\bigcap\mathcal{Z} \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

where $\mathcal{Z}$ is the set of integers, $y = \exp(-\hat{\lambda}\Delta/2)$, $\hat{\lambda} = -\frac{2}{\Delta}\ln(\gamma)$, $\Delta = 2\tau+1$,

$$\gamma = \frac{-N_0\Delta}{2N\Delta+4S} + \frac{\sqrt{N_0^2\Delta^2-(2N_1\Delta-4S)(2N\Delta+4S)}}{(2N\Delta+4S)}, \tag{7}$$

N_0 and N_1 are the number of observations satisfying $\breve{e}_i = 0$ and $\breve{e}_i \neq 0$, respectively. In addition, $N = N_0 + N_1$, and $S = \sum_{i=1}^{N}|\breve{e}_i|$. As suggested by [6], we use different $\hat{\lambda}_k$ for different quantization bin k, so as to further improve the undetectability-distortion tradeoff, where $\hat{\lambda}_k$ can be determined in a similar fashion as [6].

B. Prediction-direction preserving strategy of adding dither

Upon getting the dither sequence, the key task left is how to add it with the PE sequence. As mentioned earlier, completely random addition will seriously degrade the quality of the resulting image. To resolve this challenge, our contribution is to propose a prediction-direction preserving strategy when adding the dither, minimizing the distortion of the resulting image. Recall that the prediction in predictive image coding is carried out in an edge guided manner, averaging the pixels along the dominating edges. In this work, we define the prediction direction as the dominating edge direction along which the prediction is performed. Take GAP given in (1) for example. There are totally seven prediction directions, labeled by 'sharp horizontal edge', 'sharp vertical edge', etc., as shown in (1). Our dither addition strategy is based on the following key observation: if the prediction direction can be preserved upon dither addition, i.e., the dominating edge direction keeps unchanged, then the error propagation effect can be effectively suppressed to a large extent. Intuitively, keeping the prediction direction (dominating edge direction) helps preserve the local structures, which are mainly formed by edges. We also have the following *Proposition 1*, further explaining that, if the prediction direction can be preserved and the dither is bounded, then the error in the predicted value upon dither addition is bounded as well.

Proposition 1: Let $\tilde{I}_i$ be the predicted value from a set of pixels $\hat{I}_j$, where $j \in \Omega_i$, in the following form

$$\tilde{I}_i = \sum_{j\in\Omega_i} a_j \hat{I}_j \tag{8}$$

where $\sum_j |a_j| \leq \rho$ and ρ is a pre-determined bound on the magnitudes of the prediction coefficients [1]. Let also $\tilde{I}'_i$ be the predicted value from a set of disturbed pixels $\hat{I}'_j$, where $j \in \Omega_i$ and $|\hat{I}_j - \hat{I}'_j| \leq \epsilon$ holds for all j. Then we have

$$|\tilde{I}_i - \tilde{I}'_i| \leq \rho\cdot\epsilon \tag{9}$$

Proof:

$$\begin{aligned} |\tilde{I}_i - \tilde{I}'_i| &= \Big|\sum_{j\in\Omega_i} a_j(\hat{I}_j - \hat{I}'_j)\Big| \\ &\leq \sum_{j\in\Omega_i}|a_j||\hat{I}_j - \hat{I}'_j| \\ &\leq \sum_{j\in\Omega_i}|a_j|\epsilon \leq \rho\cdot\epsilon \end{aligned} \tag{10}$$

This completes the proof.

[1] This bound may vary for different prediction approaches.

This proposition implies that, when the prediction direction is preserved, namely, the set Ω_i and the prediction coefficients a_j's are unchanged, and the reconstruction errors of the causal pixels are bounded, then the error of the predicted value will be bounded as well, limiting the error propagation.

Therefore, our strategy is to keep the prediction direction as much as we can during the process of adding dither. We first define a set $\mathcal{M}_i$ to record the locations whose predictions depend on $\hat{I}_i$. For GAP [11], $\mathcal{M}_i = \{\mathtt{e}, \mathtt{ee}, \mathtt{sw}, \mathtt{s}, \mathtt{se}, \mathtt{ssw}, \mathtt{ss}\}$, as shown in Fig. 1. We also define a direction vector $\mathbf{v}_i$ to record the prediction direction of the locations in $\mathcal{M}_i$ before any anti-forensic operations are performed.

When adding the dither at location i using (5), we attempt to maximally keep $\mathbf{v}_i$ unchanged. To this end, we force d_i to be added within the following set

$$\mathcal{D}_i = \left\{ d_i \middle| \mathbf{v}_i' = \mathbf{v}_i, d_i \in [-\tau, \tau] \bigcap \mathcal{Z} \right\} \tag{11}$$

where $\mathbf{v}_i'$ is the updated direction vector upon adding d_i.

More specifically, the steps for performing the dither addition are given as follows

Step 1: Initialize location index $i = 0$, and generate dither sequence $\mathbf{q}_k$ according to (6) for each quantization bin k.

Step 2: Determine the quantization bin index k associated with $\breve{e}_i$.

Step 3: Form a dither searching window $\mathcal{W}$ by including the first $S = 10$ elements from $\mathbf{q}_k$, where the window size is determined empirically.

Step 4: Calculate the intersection $\mathcal{C} = \mathcal{W} \bigcap \mathcal{D}_i$.

Step 5: If $\mathcal{C} \neq \emptyset$, then assign the first element of $\mathcal{C}$ to d_i, and generate $\breve{e}_i' = \breve{e}_i + d_i$. Update $\mathbf{q}_k$ by excluding the element that has been selected.

Step 6: If $\mathcal{C} = \emptyset$, choose d_i as a random integer in the range $[-\tau, \tau]$ generated by the distribution in (6), and calculate $\breve{e}_i'$ accordingly.

Step 7: Repeat ***Steps 2-6*** until all locations are processed.

Step 8: Convert the modified $\breve{e}'$ sequence back to the pixel domain to produce the image $\mathbf{I}'$.

IV. Experimental Results

To test the efficiency of the proposed method, we perform extensive experiments over the UCID-v2 corpus consisting of 1338 images. Due to the length limit, we in Figs. 3-4 only show the results for `Lena` and `Barbara` with different settings of τ. As can be observed, the distributions of the modified PE sequences of both images are continuous, smooth, and obey Laplacian shape, which are indistinguishable from the counterparts of the uncompressed versions. In addition, the anti-forensically modified images are still of high quality.

Further, we evaluate the performance of the forensics detector in [3] on the predictively coded images before and after the dither addition. For each uncompressed image in the UCID-v2, we created five compressed versions with $\tau = 1 \sim 5$ and form the test set consisting of 6690 images. The detection results are tabulated in Table I. It can be seen that the detector of [3] is very successful in determining which images are coded, with success rate (SR) consistently being 100% for all τ. However,

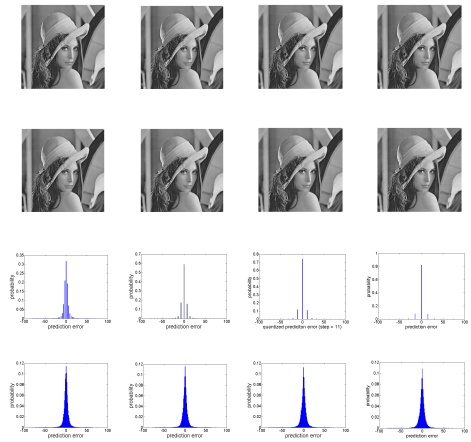

Fig. 3. From top to bottom: CALIC compressed image, anti-forensically modified image, histogram of quantized PE, histogram of the modified PE; From left to right: $\tau = 1, 3, 5, 7$. The PSNR values of the modified image w.r.t the compressed image are $\mathbf{50.80}$, $\mathbf{42.81}$, $\mathbf{38.80}$ and $\mathbf{36.51}$ dB, respectively.

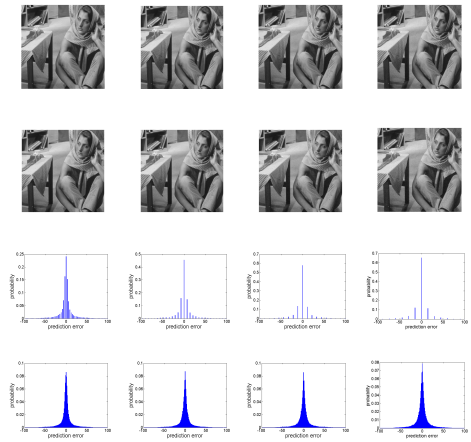

Fig. 4. From top to bottom: CALIC compressed image, anti-forensically modified image, histogram of quantized PE, histogram of the modified PE; From left to right: $\tau = 1, 3, 5, 7$. The PSNR values of the modified image w.r.t the compressed image are $\mathbf{50.81}$, $\mathbf{42.95}$, $\mathbf{38.84}$ and $\mathbf{36.31}$ dB, respectively.

after applying the proposed anti-forensic algorithm, the SR drops to 0, demonstrating the effectiveness of our proposed method. In this table, we also present the PSNR results of the anti-forensically modified images, for both our method and the random dither addition approach. It can be observed that our proposed method outperforms the random addition approach by a big margin.

TABLE I
Performance of the forensic detector [3]. The results for random addition are given inside parentheses.

τ	SR Before	SR After	Min dB	Max dB	Averaged dB
1	100%	0%	49.91 (4.24)	52.09 (28.86)	50.80 (10.60)
2	100%	0%	45.15 (4.07)	47.60 (25.70)	45.97 (10.48)
3	100%	0%	42.41 (3.43)	44.61 (23.08)	43.01 (10.25)
4	100%	0%	40.05 (2.71)	42.47 (23.15)	40.71 (10.13)
5	100%	0%	38.22 (3.78)	40.93 (20.61)	38.90 (10.03)

V. Conclusions

We have proposed a novel anti-forensic framework to erase the compression evidence of lossy predictive image coding. We have shown that a prediction-direction preserving strategy is very effective in suppressing the error propagation effect, during the process of adding dither noise in the PE domain. Experimental results have been provided to demonstrate that high-quality, anti-forensically modified image can be obtained.

References

[1] H. Farid, "A survey of image forgery detection," *IEEE Signal Process. Mag.*, vol. 26, no. 2, pp. 16–25, 2009.

[2] Z. Fan and R. L. de Queiroz, "Identification of bitmap compression history: JPEG detection and quantizer estimation," *IEEE Tran. Image Process.*, vol. 12, no. 2, pp. 230–235, 2003.

[3] W. S. Lin, S. K. Tjoa, H. V. Zhao, and K. J. R. Liu, "Digital image source coder forensics via intrinsic fingerprints," *IEEE Trans. Inf. Forensics Security*, vol. 4, no. 3, pp. 460–475, 2009.

[4] M. Stamm and K. J. R. Liu, "Anti-forensics of digital image compression," *IEEE Trans. Inf. Forensics Security*, vol. 6, no. 3, pp. 1050–1065, Sept 2011.

[5] G. Valenzise, M. Tagliasacchi, and S. Tubaro, "The cost of JPEG compression anti-forensics," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, May 2011, pp. 1884–1887.

[6] W. Fan, K. Wang, F. Cayre, and Z. Xiong, "JPEG anti-forensics with improved trade-off between forensic undetectability and image quality," *IEEE Trans. Inf. Forensics Security*, vol. 9, no. 8, pp. 1211–1226, 2014.

[7] G. Valenzise, M. Tagliasacchi, and S. Tubaro, "Detectability-quality trade-off in jpeg counter-forensics," in *Proc. IEEE Int. conf. Image process.*, Oct 2014, pp. 5337–5341.

[8] M. Stamm, W. Lin, and K. Liu, "Temporal forensics and anti-forensics for motion compensated video," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 4, pp. 1315–1329, Aug 2012.

[9] J. Zhou, X. Wu, and L. Zhang, "ℓ_2 restoration of ℓ_∞ decoded images via soft-decision estimation," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4797–4807, Dec. 2012.

[10] X. Wu and P. Bao, "ℓ_∞-constrained high-fidelity image compression via adaptive context modeling," *IEEE Trans. Image Process.*, vol. 9, no. 4, pp. 536–542, Apr. 2000.

[11] X. Wu and N. Memon, "Context-based, adaptive, lossless image coding," *IEEE Trans. Commun.*, vol. 45, no. 4, pp. 437–444, Apr. 1997.

[12] M. Weinberger, G. Seroussi, and G. Sapiro, "The loco-i lossless image compression algorithm: principles and standardization into JPEG-LS," *IEEE Trans. Image Process.*, vol. 9, no. 8, pp. 1309–1324, Aug 2000.