DISTRIBUTED SOURCE CODING AUTHENTICATION OF IMAGES WITH AFFINE WARPING

Yao-Chung Lin, David Varodayan, and Bernd Girod

Information Systems Laboratory, Stanford University, Stanford, CA 94305
{yao-chung.lin, varodayan, bgirod}@stanford.edu

ABSTRACT

Media authentication is important in content delivery via untrusted intermediaries, such as peer-to-peer (P2P) file sharing. Many differently encoded versions of a media file might exist. Our previous work applied distributed source coding not only to distinguish the legitimate diversity of encoded images from tampering but also to localize tampered regions in an image already deemed to be inauthentic. In both cases, authentication requires a Slepian-Wolf encoded image projection that is supplied to the decoder.

We extend our scheme to authenticate images that have undergone affine warping. Our approach incorporates an Expectation Maximization algorithm into the Slepian-Wolf decoder. Experimental results demonstrate that the proposed algorithm can distinguish legitimate encodings of authentic images from illegitimately modified versions, despite an arbitrary affine warping, using authentication data of less than 250 bytes per image.

Index Terms— Image authentication, distributed source coding, Expectation Maximization

1. INTRODUCTION

Media authentication is important in content delivery via untrusted intermediaries, such as peer-to-peer (P2P) file sharing or P2P multicast streaming. In these applications, many differently encoded versions of the original file might exist. Moreover, transcoding and bitstream truncation at intermediate nodes might give rise to further diversity. Intermediaries might also tamper with the media for a variety of reasons, such as interfering with the distribution of a particular file, piggybacking unauthentic content, or generally discrediting a distribution system. In previous work, we applied distributed source coding (DSC) to image authentication to distinguish the diversity of legitimate encodings from malicious manipulation [1] and demonstrated that the same framework can localize tampering in images deemed to be inauthentic [2]. In this paper, we extend our image authentication scheme to be robust to affine warping. Our approach lets the authentication decoder learn affine warping parameters using an Expectation Maximization [3] (EM) algorithm.

Section 2 reviews our image authentication system using distributed source codes [1]. In Section 3, we formalize the authentication problem with affine warping and introduce our extension for image authentication with parameter learning. The EM algorithmic details are given in Section 4. Simulation results in Section 5 show that the proposed scheme can distinguish between authentic encodings of images with affine warping and illegitimately modified versions.

This work has been supported, in part, by a gift from NXP Semiconductors to the Stanford Center for Integrated Systems and, in part, by the Max Planck Center for Visual Computing and Communication.

2. REVIEW OF IMAGE AUTHENTICATION WITH DSC

Fig. 1 is the block diagram for our earlier image authentication scheme [1] as well as the current work. We denote the source image as x. We model the image-to-be-authenticated y by way of the space-varying two-state lossy channel in Fig. 2. The legitimate state of the channel performs lossy JPEG2000 or JPEG compression and reconstruction with peak signal-to-noise ratio (PSNR) of 30 dB or better. The illegitimate state additionally includes malicious tampering. The channel state variable S_i is defined per nonoverlapping 16x16 block of image y. If any pixel in block B_i has been tampered with, $S_i = 1$; otherwise, $S_i = 0$.

We now review the authentication system. The left-hand side of Fig. 1 shows that a pseudorandom projection (based on a randomly drawn seed K_s) is applied to the original image x to produce projection coefficients X, which are quantized to X_q. The authentication data comprise two parts, both derived from X_q. The Slepian-Wolf bitstream $S(X_q)$ is the output of a Slepian-Wolf encoder based on rate-adaptive low-density parity-check (LDPC) codes [4]. The much smaller digital signature $D(X_q, K_s)$ consists of the seed K_s and a cryptographic hash value of X_q signed with a private key. The authentication data are generated by a server upon request. Each response uses a different random seed K_s, which is provided to the decoder as part of the authentication data. This prevents an attack which simply confines the tampering to the nullspace of the projection. Based on the random seed, for each 16x16 nonoverlapping block B_i, we generate a 16x16 pseudorandom matrix P_i by drawing its elements independently from a Gaussian distribution $\mathcal{N}(1, \sigma^2)$ and normalizing so that $||P_i||_2 = 1$. We choose $\sigma = 0.2$ empirically. The inner product $\langle B_i, P_i \rangle$ is an element of X, quantized to an element of X_q.

The authentication decoder, on the right-hand side of Fig. 1, seeks to authenticate the image y with authentication data $S(X_q)$ and $D(X_q, K_s)$. It first projects y to Y in the same way as during authentication data generation. A Slepian-Wolf decoder reconstructs ${X_q}'$ from the Slepian-Wolf bitstream $S(X_q)$ using Y as side information. Decoding is via joint bitplane LDPC belief propagation [5] initialized according to the known statistics of the legitimate channel state at the worst permissible quality for the given original image. Then the image digest of ${X_q}'$ is computed and compared to the image digest, decrypted from the digital signature $D(X_q, K_s)$ using a public key. If these two image digests are not identical, the receiver declares image y to be inauthentic. If they match, then X_q has been recovered. To confirm the authenticity of y, the receiver verifies that the empirical conditional entropy $H_{\text{emp}}(X_q|Y)$ (based on the legitimate channel model) is less than a certain threshold.

Since this second-pass comparison uses all available information, the threshold for $H_{\text{emp}}(X_q|Y)$ specifies how statistically similar the image-to-be-authenticated must be to the original to be de-

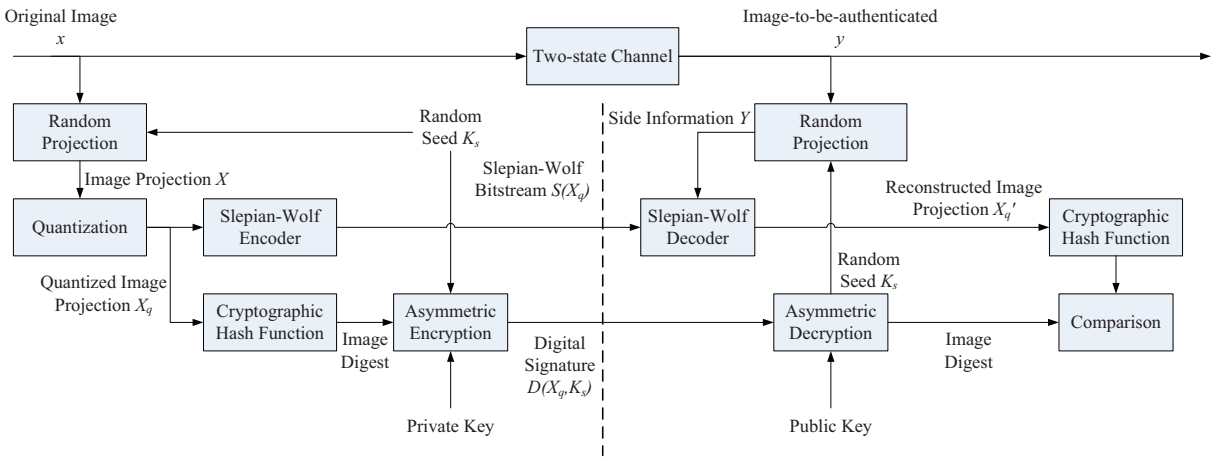


Fig. 1. Image authentication system based on distributed source coding.

clared authentic. But the rate of the Slepian-Wolf bitstream $S(X_q)$ determines whether the quantized image projection X_q is recovered at all [6]. Accordingly, at the encoder, we select a Slepian-Wolf bitrate just sufficient to successfully decode with both legitimate 30 dB JPEG2000 and JPEG reconstructed versions of x. At the decoder, we choose a threshold for $H_{\text{emp}}(X_q|Y)$ for the second-pass comparison to distinguish between the different joint statistics induced in the images by the legitimate and illegitimate channel states.

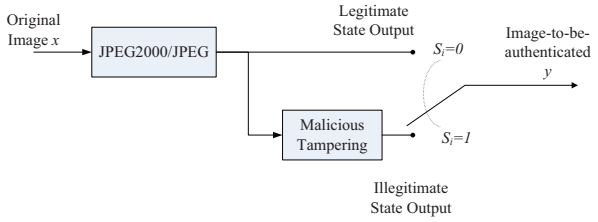


Fig. 2. Space-varying two-state lossy channel.

3. AFFINE WARPING MODEL

In this paper, we replace the two-state lossy channel in Fig. 2 with the one in Fig. 3. Now both the legitimate and illegitimate states of the channel are affected by affine warping. In the legitimate state, we model the channel as

$$y(\mathbf{m}) = x(\mathbf{n}) + z(\mathbf{m}), \text{ where } \mathbf{m}, \mathbf{n} \in R^2$$
$$\text{with } \mathbf{n} = A\mathbf{m} + \mathbf{b}, \text{ where } A \in R^{2\times 2} \text{ and } \mathbf{b} \in R^2.$$

A and $\mathbf{b}$ are transformation and translation parameters, respectively, and z is noise introduced by compression and reconstruction. To keep y the same size as x, it is padded with black pixels (arbitrarily) and cropped. Fig. 4(a) shows a source image "Lena" at 512x512 original resolution. The legitimate y in Fig. 4(b) is first rotated by 5 degrees around the image center, then cropped to 512x512, and finally JPEG2000 compressed and reconstructed at 30 dB PSNR. Here, $A = \begin{bmatrix} 0.996 & -0.087 \\ 0.087 & 0.996 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 23 \\ -21 \end{bmatrix}$. The illegitimate y in Fig. 4(c) additionally includes malicious tampering.

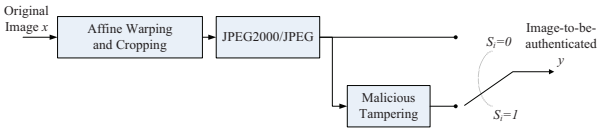


Fig. 3. Space-varying two-state lossy channel with affine geometric transform.

Fig. 4(d) shows the illegitimate y realigned to the original, with channel states S_i labeled red if illegitimate and blue if cropped out in y. The remainder are legitimate cropped-in states.

The image authentication system described in Section 2 cannot authenticate legitimate images that have undergone affine warping, because the side information is not aligned with the corresponding authentication data. Approaches suggested in the prior art to overcome this problem involve generating affine-invariant features that serve as authentication data [7, 8]. These features are usually derived from large portions of the image or even the whole image. Therefore, the authentication is less sensitive; i.e. a small amount of tampering cannot be detected. We instead propose that the authentication decoder estimate the affine warping parameters directly from the Slepian-Wolf bitstream $S(X_q)$ and the image-to-be-authenticated y using an EM algorithm. This combination of unsupervised learning with distributed source decoding is closely related to the learning of motion vectors in distributed video coding [9].

4. EXPECTATION MAXIMIZATION

The introduction of learning to the system in Fig. 1 requires a modification of the Slepian-Wolf decoder block from a joint-bitplane LDPC decoder [5] to the affine-learning Slepian-Wolf decoder shown in Fig. 5. As before, it takes the Slepian-Wolf bitstream $S(X_q)$ and the image-to-be-authenticated y and yields the reconstructed image projection $X_q{}'$. But it now does this via an EM algorithm. The E-step updates the *a posteriori* probability mass functions (pmf) $P_{\text{app}}(X_q)$ using the joint bitplane decoder and also estimates displacement vectors for a subset of reliably-decoded projection pixels. The M-step updates the affine warping parameters based on the displacement vector distributions, denoted $P_{\text{app}}(\mathbf{v})$ in

(a) (b) (c) (d)

Fig. 4. Test image "Lena" (a) x original, (b) y in legitimate state, (c) y in illegitimate state, (d) channel states S_i (red: illegitimate, blue: cropped-out) associated with the 16x16 blocks of realigned output in (c).

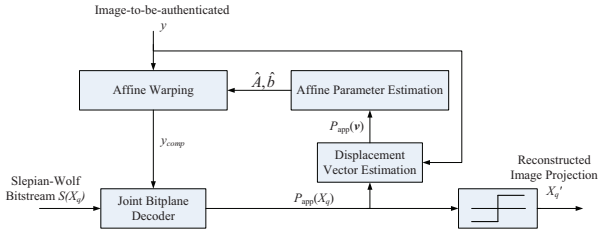


Fig. 5. Slepian-Wolf decoder with affine warping parameter learning.

Fig. 5. This loop of EM iterations terminates when hard decisions on $P_{\text{app}}(X_q)$ satisfy the constraints imposed by $S(X_q)$.

In the E-step, we fix the parameters A and $\mathbf{b}$ at their current hard estimates. The inverse transform is applied to the image y to obtain a compensated image y_{comp}. If the affine warping parameters are accurate, y_{comp} would be closely aligned to the original image x in the cropped-in region. We derive intrinsic pmfs for the image projection pixels X_q as follows. In the cropped-in region, we use Gaussian distributions centered at the random projection values of y_{comp}, and in the cropped-out region, we use uniform distributions. Then, we run three iterations of joint bitplane LDPC decoding on the intrinsic pmfs with the Slepian-Wolf bitstream $S(X_q)$ to produce extrinsic pmfs $P_{\text{app}}([X_q]_i = x_q)$.

We estimate displacement vectors for those projection pixels for which $\max_{x_q} P_{\text{app}}([X_q]_i = x_q) > T = 0.995$, denoting this set of reliably-decoded projection indices as $\mathcal{C}$. We also denote the maximizing reconstruction value x_q to be $[x_q^{\max}]_i$. (To guarantee that $\mathcal{C}$ is nonempty, we make sure to encode a small portion of the quantized image projection X_q with degree-1 syndrome bits. The decoder knows those values with probability 1 and includes their indices in $\mathcal{C}$.) We obtain displacement vector pmfs $P_{\text{app}}(\mathbf{v}^{(i)})$ for these pixels by maximizing the following log-likelihood function:

$$L(A,\mathbf{b}) \equiv \sum_{i\in\mathcal{C}} \log P([x_q^{max}]_i, \mathbf{n}^{(i)}, y; A, \mathbf{b})$$
$$= \sum_{i\in\mathcal{C}} \log \left(\sum_{\mathbf{m}^{(i)}} P([x_q^{max}]_i, \mathbf{n}^{(i)}, y|\mathbf{m}^{(i)}; A, \mathbf{b}) P(\mathbf{m}^{(i)}) \right),$$

where $\mathbf{n}^{(i)}$ is the set of top-left co-ordinates of the 16x16 projection blocks B_i in the original image x, and the latent variable $\mathbf{m}^{(i)}$ represents the corresponding set of co-ordinates in y. The latent variable update can be written as

$$Q_i(\mathbf{m}) := P(\mathbf{m}^{(i)} = \mathbf{m}|[x_q^{max}]_i, y, \mathbf{n}^{(i)}; A, \mathbf{b})$$
$$= P(\mathbf{v}^{(i)} = \mathbf{m} - \mathbf{n}^{(i)}|[x_q^{max}]_i, y, \mathbf{n}^{(i)}; A, \mathbf{b})$$
$$\equiv P_{\text{app}}(\mathbf{v}^{(i)} = \mathbf{m} - \mathbf{n}^{(i)}).$$

In this way, we associate a displacement vector pmf $P_{\text{app}}(\mathbf{v}^{(i)})$ with each projection pixel $[X_q]_i$ in $\mathcal{C}$, in a process similar to learning motion vectors in distributed video coding [9]. For the projection pixel $[X_q]_i$, we produce the pmf $P_{\text{app}}(\mathbf{v}^{(i)} = \mathbf{v})$ by matching the pixel to projections obtained from y through vectors $\mathbf{v}$ over a small search window. Specifically, $P_{\text{app}}(\mathbf{v}^{(i)} = \mathbf{v})$ is proportional to the integral over the quantization interval of $[x_q^{\max}]_i$ of a Gaussian centered at the projection of a block displaced by vector $\mathbf{v}$ in the image y.

In the M-step, we re-estimate the parameters A and $\mathbf{b}$ by holding the displacement vector pmfs $P_{\text{app}}(\mathbf{v}^{(i)})$ fixed and maximizing a lower bound of the log-likelihood function $L(A,\mathbf{b})$:

$$(A,\mathbf{b})$$
$$:= \arg\max_{A,\mathbf{b}} \sum_{i\in\mathcal{C}} \sum_{\mathbf{m}^{(i)}} Q_i(\mathbf{m}^{(i)}) \log P([x_q^{max}]_i, \mathbf{n}^{(i)}, y|\mathbf{m}^{(i)}; A, \mathbf{b})$$
$$= \arg\max_{A,\mathbf{b}} \sum_{i\in\mathcal{C}} \sum_{\mathbf{m}^{(i)}} Q_i(\mathbf{m}^{(i)})$$
$$\left(\log P(\mathbf{n}^{(i)}|\mathbf{m}^{(i)}, [x_q^{max}]_i, y; A, \mathbf{b}) + \log P([x_q^{max}]_i, y|\mathbf{m}_k^{(i)}) \right).$$

The lower bound is due to Jensen's inequality and concavity of $\log(.)$. Note also that $P([x_q^{max}]_i, y|\mathbf{m}^{(i)})$ does not depend on the parameters A and $\mathbf{b}$ and can be thus ignored in the maximization. We model $P(\mathbf{n}^{(i)}|\mathbf{m}^{(i)}, [x_q^{max}]_i, y; A, \mathbf{b})$ as a Gaussian distribution, i.e., $(\mathbf{n}^{(i)} - A\mathbf{m}^{(i)} - \mathbf{b}) \sim \mathcal{N}(0, \sigma^2\mathbf{I})$. Taking partial derivatives with respect to A and b, and setting to zero, we obtain the optimal updates:

$$\begin{bmatrix} A_{11} & A_{21} \\ A_{12} & A_{22} \\ b_1 & b_2 \end{bmatrix} := E(G^TG)^{-1}E[G^T] \begin{bmatrix} \vdots & \vdots \\ n_1^{(i)} & n_2^{(i)} \\ \vdots & \vdots \end{bmatrix},$$

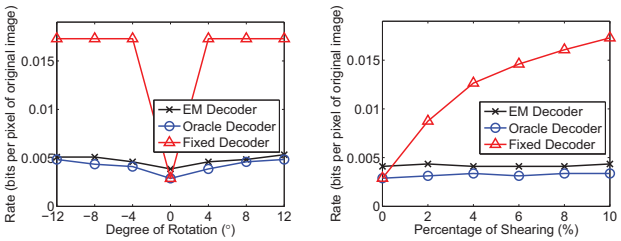
where

$$G = \begin{bmatrix} \cdots & m_1^{(i)} & \cdots \\ \cdots & m_2^{(i)} & \cdots \\ \cdots & 1 & \cdots \end{bmatrix}^T.$$

5. SIMULATION RESULTS

Our experiments use "Barbara","Lena","Mandrill," and "Peppers" of size 512x512 at 8-bit gray resolution. The two-state channel in Fig. 3 applies a geometric transformation to the images and crops them to 512x512. Then JPEG2000 or JPEG compression and reconstruction is applied at 30 dB reconstruction PSNR. In the illegitimate state, the malicious attack overlays a 20x122 pixel text banner randomly on the image. The text color is white or black, whichever is more visible, to avoid generating trivial attacks, such as white text on a white area. The image projection X is quantized to 4 bits, and the Slepian-Wolf encoder uses a 4096 bit LDPC code with 400 degree-1 syndrome nodes.

Fig. 6 compares the minimum rates for decoding $S(X_q)$ with legitimate test images using three different decoding schemes: the proposed EM decoder that learns the affine parameters, an oracle decoder that knows the parameters, and a fixed decoder that always assumes no geometric transformation. Fig. 6 (a) and (b) show the results when the geometric transformations are rotation around the image center and horizontal shearing, respectively. The EM decoder requires minimum rates only slightly higher than the oracle decoder, while the fixed decoder requires higher and higher rate as the geometric distortion increases.



(a) Rotation (b) Horizontal shearing

Fig. 6. Minimum rate for decoding the legitimate test image, "Barbara," using different decoders.

For the next experiment, we set the authentication data size to 220 bytes and measure false acceptance and rejection rates. The acceptance decision is made based on the empirical conditional entropy of X_q of the estimated cropped-in blocks. The channel settings remain the same except that transform parameters A_{11} and A_{22} are randomly drawn from [0.95, 1.05], A_{21} and A_{12} from $[-0.1, 0.1]$, and b_1 and b_2 from $[-10, 10]$. The JPEG2000/JPEG reconstruction PSNR is selected from 30 to 42 dB. With 4000 trials each on "Barbara", "Lena", "Mandrill", and "Peppers," Fig.7 shows the receiver operating characteristic curves created by sweeping the decision threshold of the empirical conditional entropy. The EM decoder performs very closely to the oracle decoder, while the fixed decoder rejects authentic test images with high probability. In the legitimate case, the EM decoder estimates the transform parameters A_{11}, A_{21}, A_{12}, A_{22}, b_1, and b_2, with mean squared error 6.0×10^{-7}, 4.1×10^{-6}, 4.2×10^{-7}, 1.6×10^{-6}, 0.06, and 0.69, respectively.

6. CONCLUSIONS

We have extended our image authentication system to handle images that have undergone affine warping. Our authentication decoder learns the affine warping parameters via an unsupervised EM algorithm. We demonstrate that an authentication Slepian-Wolf bitstream of 220 bytes is sufficient to distinguish between legitimate encodings of warped images and illegitimately modified versions. The work can be extended to other geometric transformations using an appropriate M-Step.

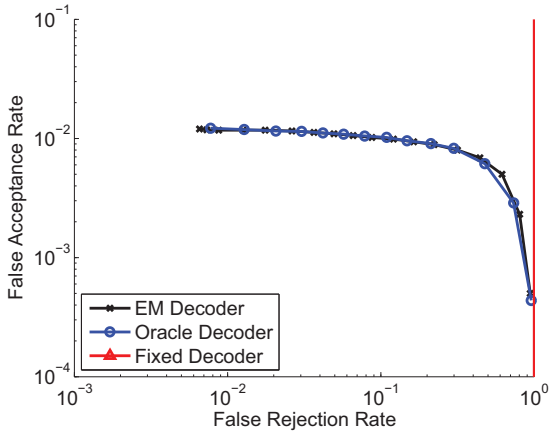


Fig. 7. Receiver operating characteristic curves for different decoders.

7. REFERENCES

[1] Y.-C. Lin, D. Varodayan, and B. Girod, "Image authentication based on distributed source coding," in *IEEE International Conference on Image Processing*, San Antonio, TX, Sep. 2007.

[2] Y.-C. Lin, D. Varodayan, and B. Girod, "Image authentication and tampering localization using distributed source coding," in *IEEE Multimedia Signal Processing Workshop*, Crete, Greece, Oct. 2007.

[3] L. E. Baum, T. Petrie, G. Soules, and N. Weiss, "A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains," *Annals of Mathematical Statistics*, vol. 41, no. 1, pp. 164–171, Oct. 1970.

[4] D. Varodayan, A. Aaron, and B. Girod, "Rate-adaptive codes for distributed source coding," *EURASIP Signal Processing Journal, Special Section on Distributed Source Coding*, vol. 86, no. 11, pp. 3123–3130, Nov. 2006.

[5] D. Varodayan, A. Mavlankar, M. Flierl, and B. Girod, "Distributed grayscale stereo image coding with unsupervised learning of disparity," in *IEEE Data Compression Conference*, Snowbird, UT, 2007.

[6] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Transactions on Information Theory*, vol. IT-19, no. 4, pp. 471–480, July 1973.

[7] F. Lefebvre, J. Czyz, and B. Macq, "A robust soft hash algorithm for digital image signature," in *International Conference on Multimedia and Expo*, Baltimore, Maryland, 2003.

[8] C. D. Roover, C. D. Vleeschouwer, F. Lefebvre, and B. Macq, "Robust image hashing based on radial variance of pixels," in *IEEE International Conference on Image Processing*, Genova, Italy, 2005.

[9] D. Varodayan, D. Chen, M. Flierl, and B. Girod, "Wyner-Ziv coding of video with unsupervised motion vector learning," *EURASIP Signal Processing: Image Communication Journal*, vol. 23, no. 5, pp. 369–378, June 2008.