A NOVEL APPROACH TO ADAPTIVE IMAGE AUTHENTICATION

Paweł Korus and Andrzej Dziech

Department of Telecommunications, AGH University of Science and Technology
al. Mickiewicza 30, 30-059 Kraków, Poland

ABSTRACT

In this paper we address the issue of the trade-off between the tampering rate and the reconstruction quality of image authentication systems. We adopt the fountain coding paradigm and design an adaptive content reconstruction scheme. The scheme conforms the reconstruction quality of individual image fragments both to the local texture properties and to the specified requirements. Experimental evaluation confirms that a framework based on this approach is a valid and convenient model of the performance of the considered reconstruction problem.

Index Terms— image authentication, fragile watermarking, content reconstruction

1. INTRODUCTION

The capability of reconstructing the original content of tampered digital images is a compelling feature of image authentication systems. The general idea consist in generating a low quality reference image for reconstruction purposes and embedding that information directly into the protected image [1].

The quality of content reconstruction has been usually addressed as a secondary matter with the only requirement being that a human observer should be able to recognize the original content. Thus, the reference image is often ultimately compressed either by using JPEG-like schemes [1, 2] or by being reduced to a binary image [3]. Recently, it has been noticed that high reconstruction quality is in certain applications a strict requirement [4, 5].

An important performance criterion is the *tampering rate*, i.e. the area of the modified fragments for which the protected image looses the reconstruction capability. There is an obvious trade-off between the tampering rate and the reconstruction quality. This issue has not been addressed so far. In [4] the authors propose a scheme with lossless reconstruction ability and the tampering rate of 3.2%. An extensive restoration capability is presented in [2], where the maximal tampering rate of 59% is achieved at the cost of quality reduction to approximately 28 dB in terms of the Peak Signal to Noise Ratio (PSNR).

In this paper we address this issue. The main purpose is to design an adaptive image authentication scheme that allows to specify the desired reconstruction quality for each image fragment individually and, hence, control the quality-tampering rate trade-off. We propose to adopt the fountain coding paradigm [6] to encode the watermark payload and achieve the requested flexibility.

This paper is organized as follows. The proposed image authentication scheme is described in Section 2. Experimental evaluation results with respect to the tampering rate and the reconstruction quality are presented in Section 3. We compare the proposed scheme with existing ones and conclude in Section 4.

2. PROPOSED AUTHENTICATION SCHEME

The system consists of two modules: the encoder and the decoder (Fig. 1) responsible for image protection and tampering detection and reconstruction, respectively. Each of these modules is described in a dedicated subsection.

We use a system similar to the one from [2] as a basis and modify it to introduce additional functionalities. The system operates on 8x8 px image blocks and uses 3 least significant bit-planes to carry the watermark. Thus, the expected quality of a protected image is approx. 37 dB in terms of PSNR. We have redesigned the reference image generation algorithm in order to allow for adaptation and to conform the reconstruction quality to the distortion level introduced by the watermark.

We adopt the fountain coding paradigm which is based on an assumption that it should be possible to decode the message from arbitrarily selected fragments of a potentially limitless stream of data. It is only necessary to collect a certain portion of the encoded stream. Hence the analogy to filling a glass of water from a fountain.

One of the most important properties of fountain codes is that they are *rateless*, i.e. they can provide a limitless stream of symbols and as such they allow to control the desired redundancy. In this study, we have used the LT code which is a practical realization of the fountain coding paradigm [7].

The research leading to these results has received funding from the INDECT project funded by European Community's Seventh Framework Programme under grant agreement no. 218086 and from the European Regional Development Fund under INSIGMA project no. POIG.01.01.02-00-062/09. The latter has provided a fountain coding toolkit.

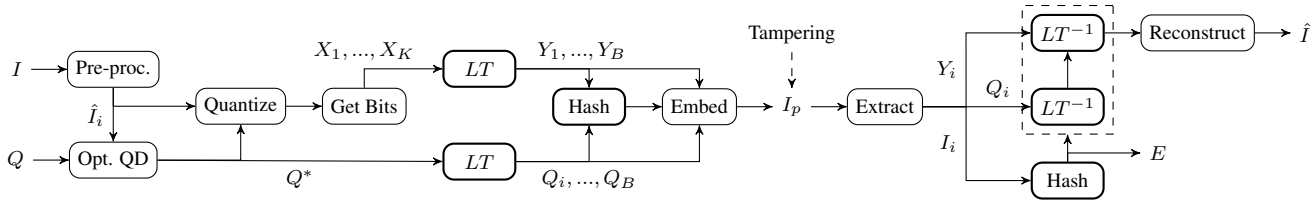


Fig. 1: Operation of the image authentication scheme; I - the cover image, I_p - the protected image, $\hat{I}$ - the reconstructed image, E - tampering map, Q - quality descriptor; Operations dependent on the secret key Z are marked by thick borders

2.1. The Encoder

The encoder accepts three input parameters: the input image I, a quality descriptor Q and a secret key Z. The quality descriptor defines the requested reconstruction quality for individual fragments of the image. Since the image is divided into blocks, the quality descriptor is a simple mapping $Q : \{1, \ldots, B\} \rightarrow \mathbf{Z}$ where B is the number of resulting blocks of the image.

We distinguish 8 quality levels. The maximal level corresponds to retaining 7 first frequency sub-bands according to the zig-zag order. Along with the drop of the quality level we eliminate successive sub-bands. The two lowest levels correspond to retaining only the DC coefficient and to completely removing it's reconstruction capability, respectively. Along with additional meta-data irrelevant for this study, we use 4 bits per block, i.e. in total $4B$ bits for the quality descriptor.

The quality descriptor takes into account the user's preferences towards the reconstruction quality. The encoder optimizes this descriptor to take into account the factual detail level of each individual block. The resulting *effective quality descriptor* is defined as $Q^*(i) = min(Q(i), L(i))$. $L(i)$ denotes the number of frequency sub-bands where the sum of coefficient magnitudes exceeds 1.

The protection process starts with basic image preprocessing. We discard 3 least significant bit-planes and denote the i^{th} block of the resulting low dynamic range image as I_i. Each block is then transformed into the DCT domain to produce $\hat{I}_i$.

The obtained spectrum is used to calculate the effective quality descriptor Q^*. Each block is then quantized according to the specified quality level. The coefficients resulting from block data quantization are converted into a bit-stream. Concatenation of the bit-streams from all blocks yields the reconstruction stream.

The reconstruction stream is divided into 144-bit symbols $X_1, \ldots, X_K$. We use the LT code to produce B output symbols $Y_1, \ldots, Y_B$ - one per block. The ratio $\lambda = \frac{K}{B}$ determines the robustness against tampering. The code is defined by a pseudo-random matrix, which is generated using the specified secret key Z.

The remaining 48 bits of available block capacity are used to store the quality descriptor and the data hash which allows for identification of tampered image fragments. The effective

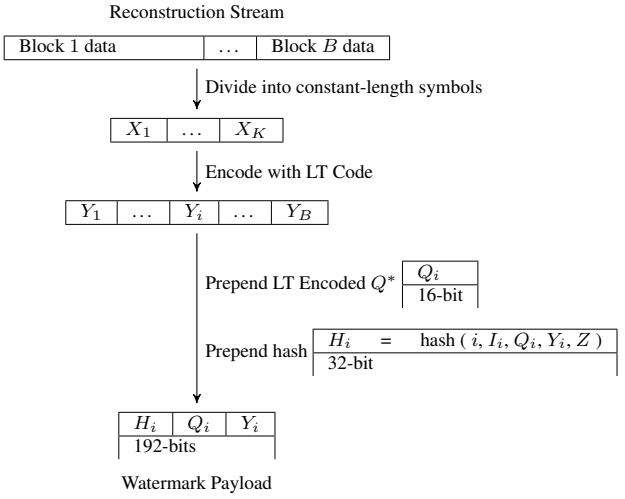


Fig. 2: Construction of the watermark payload. Each block contains a 32-bit data hash, a 16-bit symbol describing Q^* and a 144-bit symbol describing the reconstruction stream

quality descriptor Q^* is necessary for content reconstruction in the decoder. The $4B$ bits that define the descriptor are divided into 16-bit symbols and encoded using the LT code to produce B output symbols $Q_1, \ldots, Q_B$ for each block.

A 32-bit data hash H_i is calculated for the i^{th} block using the block ID i, I_i, Q_i, Y_i and Z. Modification of each of these components, e.g. block reordering, image tampering, invalidates the hash in the decoder and classifies the transmitted symbol as erased. In this study, we have used the MD5 hashing algorithm, shortened to produce 32-bit output. The watermark structure is shown in Fig. 2.

The last step is to embed the resulting bit-stream into the image using bit substitution.

2.2. The Decoder

The operation of the decoder begins with the extraction of the embedded watermark. Then, the decoder recalculates the data hashes H_i^* and compares the results with the recovered H_i. The result of this step is an erasure or tampering map $\{0, 1\} \ni E(i) : i \in \{1, \ldots, B\} \wedge E(i) = 1 \Leftrightarrow H_i \neq H_i^*$.

This erasure map is relayed along with the extracted Q_i and Y_i to the appropriate LT decoders. Firstly, the decoder recovers the quality descriptor Q^*. Along with the erasure map, this information is used to determine, which symbols from $X_1, \ldots, X_K$ are in fact needed for successful content reconstruction. Having all of this information, the LT decoder recovers the necessary reconstruction stream symbols.

In the final step, the decoder renders the resulting bitstream and zeroes or reconstructs the DCT spectra of the tampered blocks. After an inverse DCT transform of the restored fragments, the resulting authenticated image is ready.

3. SYSTEM EVALUATION

In this study we address two aspects of the considered image authentication scheme. Firstly, we provide a theoretical boundary between successful and unsuccessful reconstructions. Secondly, we present exemplary reconstruction results and assess the achievable reconstruction quality.

3.1. Tampering Rate

Adoption of the LT code into the reconstruction process allows to directly assess the expected tampering rate γ_{max} as the necessary amount of correctly transmitted symbols (1) [7].

$$\gamma'_{max} = 1 - \frac{K}{B} - c\frac{\sqrt{K}}{B} ln^2(K/\delta) = 1 - \lambda - \theta \quad (1)$$

θ represents the overhead of the LT code over an ideal fountain code. c is a linear constant and δ is the probability of LT decoding failure. In this study, we use $c = 0.04$ and $\delta = 0.05$.

Since the reconstruction success is predicated on the knowledge of $Q*$, an additional bound needs to be taken into account (2). It stems from the ability to decode the quality descriptor.

$$\gamma''_{max} = 1 - \frac{1}{4} - \frac{c}{\sqrt{4B}} ln^2(B/4\delta) \approx 0.72 \quad (2)$$

Considering both factors, $\gamma_{max} = min(\gamma'_{max}, \gamma''_{max})$. The γ''_{max} bound can be easily improved as the quality descriptor can be efficiently compressed prior to encoding.

We have experimentally validated the applicability of this bound. From a test-set of 15 images, we randomly select an image and generate a random quality descriptor. The encoder produces a protected image which is tampered in randomly selected areas. This approach generates sufficiently diverse points on the $\lambda \times \gamma$ plane. The results of this experiment are shown in Fig. 3a. Reconstruction success cases are marked with diamonds and failures and marked with crosses. Since different images are expected to produce different reconstruction quality for the same reconstruction stream length, we show in Fig. 3b the achieved results for commonly known images.

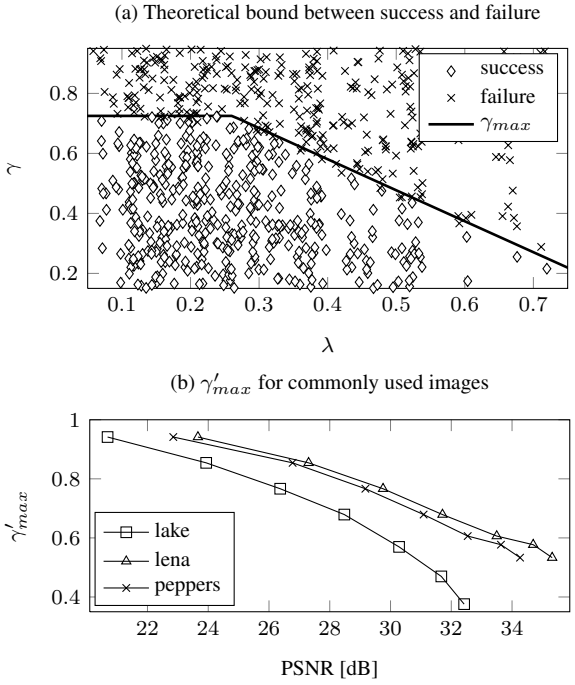


Fig. 3: Evaluation of the tampering rate

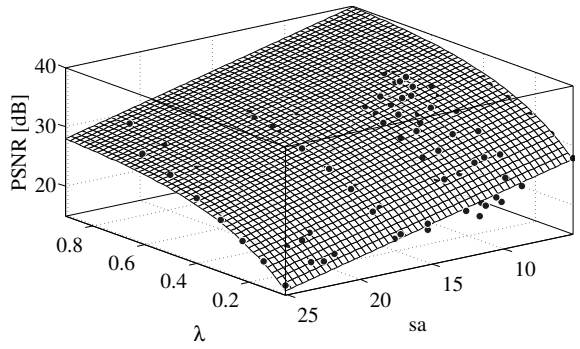


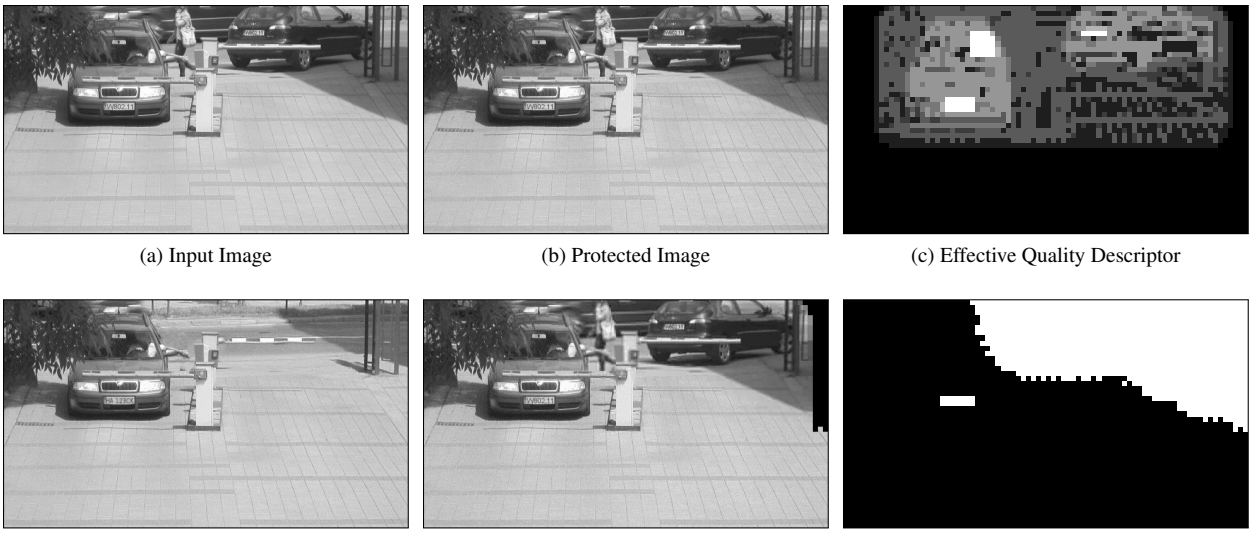
Fig. 4: Reconstruction quality versus λ and sa

3.2. Reconstruction Quality

One of the main advantages of the proposed scheme is the ability to adapt the reconstruction quality both to the needs of the user and to the content itself. The achievable quality depends heavily on the amount of details in the original image (Fig. 4). We measure the details using spatial activity sa, i.e. average standard deviation of $I_i : i = 1, \ldots, B$.

Assuming a uniform quality descriptor is used, the average reconstruction quality for images with medium spatial activity, i.e. $sa < 15$, is 35 dB. For highly textured images, it reaches 30 dB.

In Fig. 5 we show a protection example. (a) and (b) show the original and the protected images, respectively. (c) is the effective quality descriptor which combines both user preferences and the local image texture. (d) shows a tampered



(a) Input Image (b) Protected Image (c) Effective Quality Descriptor

(d) Tampered Image (e) Reconstructed Image (f) Detected Tampering Map

Fig. 5: Images from an exemplary protection-tampering-reconstruction process

Table 1: Comparison of the performance of different fragile watermarking schemes.

Algorithm	**Typical Distortion (PSNR)**		**Tampering Rate**	**Payload Encoding**	**Embedding**
	Embedding	**Reconstruction**	γ_{max}		
Fridrich [1] - Method I	44 dB	21 dB	N/A	none	bit substitution
Fridrich [1] - Method II	33 dB	29 dB	N/A	none	bit substitution
Zhang [4]	29 dB	∞ dB	3.2%	N/A	difference expansion
Zhang [2]	37 dB	28 dB	59%	random linear code	bit substitution
Zhu [8]	35 dB	21 dB	N/A	irregular sampling	custom, sine transform
Cheddad [3]	42 dB	28 dB	N/A	not needed	custom, DWT domain
Proposed scheme	37 dB	flexible, up to 37 dB	flexible, up to 72%	LT Code	bit substitution

image with one of the cars removed from the scene and with the second's license plate number modified. (e) shows the result of content reconstruction. Notice the reduced reconstruction quality outside the restored car and missing reconstruction where requested (top right). (f) is the detected tampering map.

4. CONCLUSIONS

We have presented a new approach for designing and evaluating adaptive image authentication schemes. Our system is able to trade-off the reconstruction quality with the tampering rate by adapting the reconstruction both to user preferences and to the local texture of the protected image content.

Adoption of the fountain coding paradigm allows for straightforward design of flexible content reconstruction systems and delivers the tools for their analysis.

The performance of the proposed scheme, both in terms of the reconstruction quality and the tampering rate, is among the best compared to existing image authentication systems. Table 1 shows typical performance of existing schemes along with the incorporated coding and watermarking techniques.

5. REFERENCES

[1] J. Fridrich and M. Goljan, "Images with Self-correcting Capabilities," in *Proc. IEEE Int. Conference on Image Processing*, 1999, vol. 3.

[2] X. Zhang, S. Wang, and G. Feng, "Fragile Watermarking Scheme with Extensive Content Restoration Capability," in *Proc. of International Workshop on Digital Watermarking*, 2009.

[3] A. Cheddad, J. Condell, Kevin Curran, and P. Mc Kevitt, "A Secure and Improved Self-Embedding Algorithm to Combat Digital Document Forgery," *Signal Processing*, vol. 89, pp. 2324–2332, 2009.

[4] X. Zhang and S. Wang, "Fragile Watermarking with Error Free Restoration Capability," *IEEE Transactions on Multimedia*, vol. 10, no. 8, 2008.

[5] P. Korus, W. Szmuc, and A. Dziech, "A Scheme for Censorship of Sensitive Image Content with High-Quality Reconstruction Ability," in *Proc. 11-th IEEE International Conference on Multimedia and Expo*, 2010.

[6] D.J.C. MacKay, "Fountain codes," *IEE Proceedings Communication*, vol. 152, no. 6, 2005.

[7] M. Luby, "Lt codes," *Proc. 43rd Ann. IEEE Symp. on Foundations of Computer Science*, , no. 16-19, 2002.

[8] X. Zhu, A. Ho, and P. Marziliano, "A New Semi Fragile Image Watermarking with Robust Tampering Restoration Using Irregular Sampling," *Signal Processing : Image Communication*, vol. 22, no. 5, 2007.