Robust and Discriminative Image Authentication Based on Sparse Coding

Luntian Mou[1,2], Tiejun Huang*[3], Yonghong Tian[3], Shiguo Lian[4], Xilin Chen[1]
[1]Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, CAS, Beijing, China
[2]Graduate University of Chinese Academy of Sciences, Beijing, China
[3]Institute of Digital Media, Peking University, Beijing, China
[4]France Telecom R&D Beijing, China
{ltmou, xlchen}@jdl.ac.cn, {tjhuang, yhtian}@pku.edu.cn, shiguo.lian@ieee.org

***Abstract*—Image authentication is usually approached by checking the preservation of some invariant features, which are expected to be both robust and discriminative so that content-preserving operations are accepted while content-altering manipulations are rejected. However, most of existing features have not obtained convincing performance due to insufficiency of experiments and over biasing of robustness. Motivated by the sparse coding strategy discovered in primary visual cortex, we explore the possibility of using sparse coding coefficients for image authentication. Through extensive experiments, we discover that the proposed feature bears great discrimination as well as robustness, which indicates the effectiveness of sparse coding as a new invariant feature for image authentication.**

Keywords-image authentication; sparse coding; robustness; discrimination; similarity

I. INTRODUCTION

With increasing accessibility of Internet and proliferation of powerful multimedia processing and manipulating tools, multimedia content protection is challenged by two facets of information security: confidentiality and authenticity [1]. Only when both facets are simultaneously resolved, can secure multimedia distribution be achieved [2][3]. While the first facet has been well tackled by encryption, the second remains a hot research topic. Image authentication is the technology employed in image communication to ensure the authenticity of an image, which is usually achieved by sender authentication and content integrity verification. For sender authentication, the actual identity of the sender and sometimes also its non-repudiability must be ensured. Unlike traditional message integrity verification, content integrity verification seeks to verify the preservation of the meaning of an image instead of its specific binary representation. Therefore, a general principle of image authentication is to accept content-preserving operations while rejecting content-altering manipulations.

Image authentication has been actively studied recently mainly in content-based approach [4]. Specifically, at sender side, a feature with certain kind of invariance is computed from an original image, and then embedded back into it as a watermark or appended to it as a signature; at receiver side, a new feature is calculated in the same way from the received image, which is then compared with the original feature based on certain distance metrics to decide whether the received image is authentic or tampered. In order to avoid impairing image quality by watermark embedding, we derive an authentication scheme (see Figure 1) from the traditional cryptographic digital signature by replacing the cryptographic hashing function involved in digital signature schemes with a content-based feature extraction algorithm. Naturally, sender authenticity and its non-repudiability can be ensured. For content integrity, the distance-based similarity between the new feature and the original feature conveyed in the signature is compared against a predefined threshold instead of verifying bit-by-bit equivalence required by cryptographic digital signature. Note that, a secret One-Time Session Key can be used to randomize the extracted feature, which would make signature forgery impossible. Obviously, it is the feature extraction that largely determines the performance of image authentication. Ideally, the feature should be robust to various content-preserving operations while strongly discriminative to content-altering manipulations.

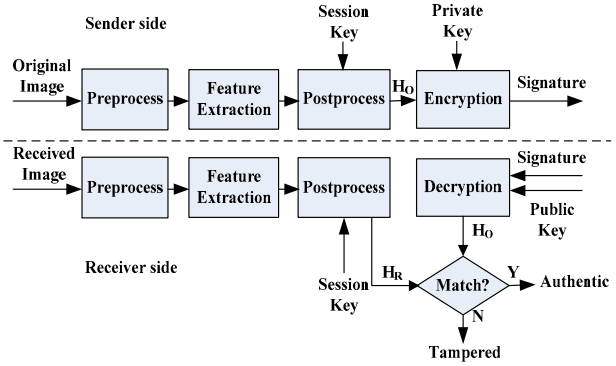


Figure 1 The derived content-based authentication scheme.

The invariant features proposed in the literature can be roughly classified into three categories: statistical features, transform domain features and low-level visual features. Statistical features include intensity histogram [5] and moments [6]. There is a main drawback for this category: it is possible to modify images without changing their statistical features. Transform domain features are extracted by exploiting the invariance in transform domains, which include relative magnitude relationship between two transform coefficients [7] and dominant components which preserve coarse image content [8]. While robustness to certain content-preserving operations is highlighted, discrimination ability of such transform domain features is not clear. The third category of low-level visual features includes edges [9] and feature points

[10]. Their robustness is not good due to their intrinsic sensitivity even to some content-preserving operations such as scaling, blurring and high quantization. In fact, most of proposed features have not convincing performance since usually only one or two operations are targeted and only several images are involved in the experiments. Furthermore, discrimination is often not sufficiently addressed due to over biasing of robustness.

Motivated by the sparse coding strategy discovered in primary visual cortex [11], we explore the implication of the sparseness in this paper. Since sparse coding coefficients can be viewed as the responses of corresponding neurons to the stimuli of image patches, it seems reasonable to expect some robustness and discrimination from these coefficients. This speculation can be partially supported by a previous work which distinguishes imitation drawings from authentic artworks by applying sparse coding to the quantification of artistic style [12]. Through extensive experiments, we further confirm that sparse coding coefficients possess excellent properties of robustness and discrimination. Therefore, we propose sparse coding as a new invariant feature for image authentication.

The rest of the paper is organized as follows. Section II presents the sparse coding based image authentication scheme by first reviewing the sparse coding theory, then bringing forward the feature extraction algorithm, and finally establishing a corresponding similarity measurement. The scheme's performance is evaluated in Section III. In Section IV, conclusions are drawn and the future work is given.

II. Image Authentication Based On Sparse Coding

As a neural coding strategy, sparse coding draws great attention for its potential application in high-efficiency visual information representation and encoding [13]. Here, we explore the possibility of applying it to image authentication. As the flow of image authentication has already been illustrated and described in Section I, this section will be dedicated to sparse coding theory, feature extraction of sparse coding and the corresponding image similarity measurement.

A. Sparse Coding

When presented to a scene, only a small number of early visual neurons out of a large set will be activated [14]. To simulate this property of simple cells in the primary visual cortex, the sparse coding theory is proposed to extract the intrinsic structure of natural images [11]. The theory assumes that an image patch is a linear superposition of a set of basis functions:

$$I_n(x,y) = \sum_i a_i \phi_i(x,y), \tag{1}$$

where a_i is the response from the ith neuron, and can also be viewed as the contribution of the basis function $\phi_i(x,y)$ to the image patch. This coefficient can be computed by its corresponding filter function:

$$a_i = \sum_{x,y} G_i(x,y) I_n(x,y), \tag{2}$$

where $G_i(x,y)$ is the pseudo inverse of $\phi_i(x,y)$.

Adopting Independent Component Analysis (ICA) [15], we learn a set of basis functions that yields a sparse representation of natural images. Since color reduction is usually accepted as a content-preserving operation, 64 basis functions are learned from 50,000 8x8 gray image patches randomly extracted from natural images (see Figure 2). Thereafter, 64 filter functions can be obtained as each being the pseudo inverse of corresponding basis function.

Figure 2 64 basis functions for gray level images.

B. Feature Extraction

For the same set of basis functions, similar sets of sparse coding coefficients will be acquired among perceptually similar patches, while randomness is to appear at coefficients for perceptually different images. Thus, the set of coefficients are assumed to possess properties of robustness and discrimination to some extent, which implies the possibility of being used as a new feature in image authentication.

Feature extraction of sparse coding is performed both at sender side and receiver side (see Figure 1). It actually consists of three stages, namely, preprocessing, feature extraction and postprocessing. The specific steps involved at each stage are briefly described as follows.

- Normalize an image. Decode a compressed image into a YUV image, extract the one-channel Y image, and resize it to 64x64. This preprocessing step targets for improving the robustness to color reduction as well as scaling and resizing.
- Extract sparse coding coefficients. Divide the resized image into 64 8x8 blocks (patches), index them in the order of "top to bottom, left to right", and filter each block by Eq. (2) to get 64 sparse coding coefficients. The result is a feature of 4096 (64x64) sparse coding coefficients for the image:

$$\{a_i\} = \left\{ \sum_{x,y} G_i(x,y) I_n(x,y) \right\} (0 \le n, i \le 63). \tag{3}$$

- Quantize the feature. For each block, a string of 64 bits is achieved in the following way:

$$h_n(i) = \begin{cases} 1 & |a_i| \ge |a_{i+1}| \\ 0 & |a_i| < |a_{i+1}| \end{cases} (0 \le i \le 62) \tag{4}$$

$$h_n(i) = \begin{cases} 1 & |a_i| \ge |a_0| \\ 0 & |a_i| < |a_0| \end{cases} (i = 63).$$

Thus, a feature of 4096 bits is obtained for an image by concatenating all feature bits block by block:

$$H = h_0(0) \,\|\, h_0(1) \,\|\, \ldots \,\|\, h_0(63) \,\|\, h_1(0) \,\|\, \ldots \,\|\, h_{63}(63), \tag{5}$$

which results in a feature space of the size 2^{4096}.

C. Similarity Measurement

Accordingly, the Hamming distance between quantized features extracted from two blocks at the same location of two images can be computed as:

$$Dis\,(I_n(x,y), I'_n(x,y)) = \sum_{i=0}^{63} XOR\,(h_n(i), h'_n(i)). \quad (6)$$

Therefore, the similarity between two images is calculated as follows:

$$Sim\,(I(x,y), I'(x,y)) = 1 - \frac{\sum_{n=0}^{63}\sum_{i=0}^{63} XOR\,(h_n(i), h'_n(i))}{64 * 64}. \quad (7)$$

Generally, for two random strings of 4096 bits, the expected number of different bits is 2048, which implies an expected similarity value of 0.5. Thus, ideally, the similarity value assessed by the proposed similarity measurement should be about 0.5 between two different images, while is close to 1 between an original image and its slightly modified version.

III. Performance Evaluation

Experiments are carried out mainly on two datasets. The original images for the first dataset is 1000 high definition JPG photographs selected from NOVA "Art Explosion 800000" [16], which covers ten categories from art, landmark, flowers, sports, and humans to wildlife, plus 10 super high resolution raw images [17]. The other set of original images includes 3600 images extracted from evaluation datasets of TRECVID [18]. By performing 7 content-preserving operations and 8 complicated manipulations respectively to the two sets of original images, we obtain one dataset of 7020 images and the other of the size of 32400.

Due to the fact that the trade-off between robustness and discrimination is purely determined by practical application requirements, the performance of the proposed feature is evaluated separately in terms of robustness and discrimination. For comparison of the features themselves, a DCT based similarity measurement is performed by only substituting 1 DC and 3 ACs in zig-zag scanning order for the 64 sparse coding coefficients.

A. Robustness Evaluation

Seven content-preserving operations are carried out respectively on 1000 JPG photographs from NOVA [16] to produce modified versions, with the exception of JPEG compression which is performed on 10 super high resolution raw images [17]. SPC (short for sparse coding) and DCT based similarity measurements are respectively applied to assess the similarity between each original image and its modified versions. Great robustness can be observed for both SPC and DCT from Table I.

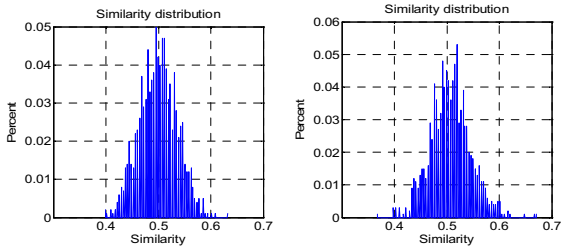
B. Discrimination Evaluation

By arranging the same 1000 original images into a ring and measuring each pair of two neighboring images, SPC achieves an average similarity value of 0.5001, while DCT achieves 0.5096. Distribution of similarity is shown in Figure 3, with 4 pairs of similar but different images given for intuitive illustration (see Figure 4). It is observed that SPC bears excellent discrimination. In particular, the resulted average similarity value between different images matches very well with the expected similarity value of 0.5.

TABLE I. Average Similarity Values On Nova

Content-preserving	Parameters	SPC Sim.	DCT Sim.
Aspect ratio	4:3 → 16:9	0.9582	0.9930
Auto-level		0.9831	0.9970
Blur	3x3	0.9496	0.9922
Brightness change	+10%	0.9354	0.9759
Gaussian noise	4.0	0.9369	0.9866
JPEG compression	QF:80	0.9133	0.9664
Scaling	90%	0.9515	0.9917

By comparison, SPC shows better discrimination while DCT is more robust.



(a) (b)

Figure 3 Distribution of similarity values. (a) SPC; (b) DCT

SPC:	0.460938	0.515625	0.476563	0.570313
DCT:	0.492188	0.558594	0.578125	0.585938

Figure 4 Similarity values between similar but different images.

The performance of the proposed feature is further evaluated on a dataset derived from TRECVID [18], which intends to simulate the complicated real world manipulations. According to the modifications contained therein, the eight tasks of copy detection are roughly classified into two categories: content-preserving and content-altering. The category of content-preserving includes insertion of pattern (T3), strong re-encoding (T4), change of gamma (T5) and decrease in quality (T6). Whilst, the category of content-altering consists of simulated camcording (T1), picture in picture (T2), post production (T8) and combination of 3 randomly chosen transformations (T10). The tasks of T7 and T9, which are similar to T6 and T8 respectively but are more complicated, have been dropped by TRECVID. Examples of

these two categories of tasks are shown in Figure 5. It should be noted that the objective of image authentication and that of copy detection are different: the former is to accept moderately modified versions of original images, while the latter attempts to identify even severely distorted copies. Thus, for image authentication, it would make sense to reject T1 as a kind of counterfeit attack, while accepting T3 if the degree of modification is not severe with respect to content integrity. Experimental results in Table II show that DCT achieves better robustness as indicated by the larger similarity values by DCT at content-preserving tasks, while SPC bears better discrimination, which can be seen from the smaller similarity values by SPC at content-altering tasks. Generally speaking, SPC will be preferred at image authentication because better discrimination implies better sensitivity at tamper detection, which is especially highlighted by systems of which security is the first priority.

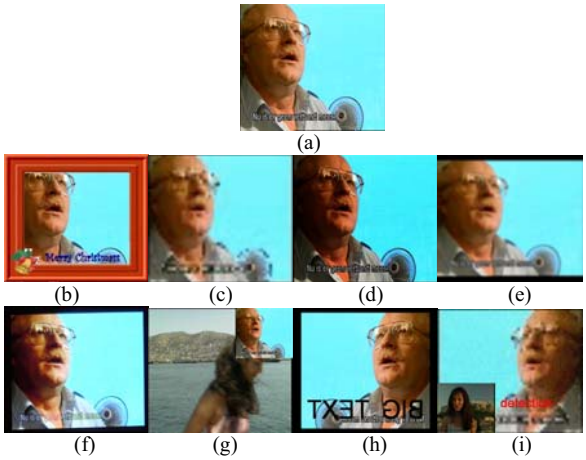


Figure 5 Examples of copy detection tasks, with the tasks on the second row viewed as content-preserving, and the third row as content-altering.

. (a) Source image; (b) T3; (c) T4; (d) T5; (e) T6; (f) T1; (g) T2; (h) T8; (i) T10.

TABLE II. AVERAGE SIMILARITY VALUES ON TRECVID

Content-preserving	SPC Sim.	DCT Sim.	Content-altering	SPC Sim.	DCT Sim.
T3	0.7892	0.8312	T1	0.5769	0.6586
T4	0.6870	0.8567	T2	0.5066	0.5077
T5	0.8587	0.9262	T8	0.5530	0.5647
T6	0.6743	0.8001	T10	0.5235	0.5436

IV. CONCLUSIONS

We propose a new feature of sparse coding for robust and discriminating image authentication. By exploiting the characteristics of sparse coding that it extracts intrinsic structure of natural images, we suggest representing an image by a set of sparse coding coefficients and measuring image similarity based on it. Experimental results demonstrate its excellent performance of robustness and discrimination, thus indicate the effectiveness of sparse coding as a potentially ideal feature for image authentication.

Due to the fact that the acceptable content-preserving operations vary greatly from one application to another, there lacks a benchmark dataset for image authentication. Our future work could be a contribution to this benchmark construction by defining different levels of acceptable modifications to cover the whole range of authentication requirements.

ACKNOWLEDGMENT

The work is supported in part by a grant from the Chinese National Natural Science Foundation under contract No. 90820003, in part by the CADAL project.

REFERENCES

[1] S. Lian, D. Kanellopoulos, and G. Ruffo. Recent Advances in Multimedia Information System Security. Informatica, Vol. 33, No. 1, 2009, pp. 3-24.

[2] S. Lian and Y. Zhang. Handbook of research on secure multimedia distribution. IGI Global (formerly Idea Group, Inc), March 2009.

[3] L. Mou, T. Huang, L. Huo, W. Li, W. Gao, X. Chen. A secure media streaming mechanism combining encryption, authentication, and transcoding. Signal Processing: Image Communication, Vol. 24, No. 10, Nov. 2009, pp. 825-833.

[4] A. Haouzia and R. Noumeir. Methods for image authentication: a survey, Multimedia Tools and Applications, Vol. 39, No. 1, pp. 1-46, Aug. 2008.

[5] M. Schneider and S. Chang. A Robust Content Based Digital Signature for Image Authentication, Proc. IEEE Int'l Conf. on Image Proc.(ICIP), Vol. 3, pp. 227-230, Sept. 1996.

[6] M. Alghoniemy, A. Tewfik. Geometric invariance in image watermarking. IEEE Trans. Image Process 13(2), 145–153 (2004).

[7] Q. Sun, S. Chang, A Robust and Secure Media Signature Scheme for JPEG Images. Journal of VLSI Signal Processing 41 (2005) 305-317.

[8] A. Swaminathan, Y. Mao and M.Wu. Robust and Secure Image hashing, IEEE Trans. on Information Forensics and Security, vol. 1, no. 2, pp 215 – 230, June 2006.

[9] J. Dittmann, A. Steinmetz, R. Steinmetz. Content-based digital signature for motion pictures authentication and content-fragile watermarking. Proceedings IEEE International Conference on Multimedia Computing and System, vol. 2, pp. 209–213 (1999).

[10] V. Monga and B.L. Evans. Perceptual image hashing via feature points: performance evaluation and tradeoffs, IEEE Transactions on Image Processing 15 (11) (2006), pp. 3453–3466.

[11] B. Olshausen, D. Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature, 1996, 381: 607~609.

[12] J. Hughes, D. Graham and D. Rockmore. Quantification of artistic style through sparse coding analysis in the drawings of Pieter Bruegel the Elder. Proceedings of the National Academy of Sciences USA, 107, 1279-1283, 2010.

[13] D. Graham and D. Field. Efficient coding of natural images. New Encyclopedia of Neuroscience, 2007.

[14] H. Barlow. Unsupervised learning. Neural Computation, 1989.

[15] J. Hateren and A. Schaaf. Independent component filters of natural images compared with simple cells in primary viusal cortex. Proc. R. Soc. Lond. B, 1998.

[16] NOVA. http://www.amazon.co.uk/Nova-ARW-Art-Explosion-800000/dp/B0001XWNSS/ref=pd_bbs_sr_1/203-3503298-4948756?ie=UTF8&s=software&qid=1183552443&sr=8-1, accessed by August 18, 2010.

[17] Image Compression. http://www.imagecompression.info/test_images/, accessed by August 18, 2010.

[18] TRECVID. http://www-nlpir.nist.gov/projects/trecvid, accessed by August 18, 2010.