Secure Data-Hiding in Multimedia using NMF

Hafiz Malik[1], Farhan Baqai[2], Ashfaq Khokhar[1], and Rashid Ansari[1]

[1] Department of Electrical and Computer Engineering, University of Illinois at Chicago, MC 154, Chicago, IL, 60607
[2] Sony US Advanced Technologies Center, San Jose, CA 95134

Abstract. This paper presents a novel data-hiding scheme for multimedia data using non-negative matrix factorization (NMF). Nonnegative feature space (basis matrix) is estimated using the NMF-framework from the sample set of multimedia objects. Subsequently, using a secret key a subspace (basis vector) of the estimated basis matrix is used to decompose the host data for information embedding and detection. Binary dither modulation is used to embed/detect the information into the host signal coefficients. To ensure the fidelity of the embedded information for a given robustness, host media coefficients are selected for information embedding according to the estimated masking threshold. Masking threshold is estimated using the human visual/auditory system (HVS/HAS) and host media. Simulation results show that the proposed NMF-based scheme provides flexible control over robustness and capacity for imperceptible embedding.

1 INTRODUCTION

Digital watermarking refers to the process of imperceptible embedding of information (watermark) into a digital document (host data) to provide content protection and/or content authentication. Watermark embedding schemes can be classified into two major categories: (1) blind embedding, in which the watermark embedder does not exploit the host signal information during the embedding process, watermarking schemes based on spread-spectrum (SS) fall into this category, and (2) informed embedding, in which the watermark embedder exploits the properties of the host signal during the watermark embedding process. Watermarking schemes based on quantization index modulation belong to this category.

Existing watermark detectors may also be classified into two categories: (a) informed detectors, which assume that the host signal is available at the detector during the watermark detection process, and (b) blind detectors, which assume that the host signal is not available at the detector. Although the performance expected from a given watermarking system depends on the target application area [**?**] robust embedding schemes and efficient detection procedures are inherently desired.

This paper presents a secure data hiding scheme for multimedia data based on the non-negative matrix factorization (NMF) of the host signal. The NMF framework to estimate nonnegative feature matrix (or feature space) for the set of preselected multimedia documents. The host signal is projected into the selected nonnegative feature subspace for information embedding and detection. In order to improve the security of the proposed scheme, a secret key (or password), **K**, is used to select nonnegative feature basis vectors (or feature subspace) from the estimated feature matrix. The host signal is projected into the selected nonnegative feature subspace for information embedding and detection. For high capacity, the QIM-based framework is used for information embedding and detection. The proposed scheme exploits the human visual/auditory model to ensure the fidelity and robustness of the embedded information. The proposed NMF-based data hiding scheme is applicable to all media types, i.e. audio, video and images. However, in this paper report performance results using digital images as the host media for information embedding and detection.

2 NON-NEGATIVE MATRIX FACTORIZATION

The nonnegative matrix factorization or nonnegative matrix decomposition is an emerging method for dimensionality reduction, sparse nonnegative representation and coding, image coding, blind source separation (BSS), classification, clustering, data mining, etc. [2–4]. Paatero et al [2] introduced the NMF concept first time in 1994, however their proposed NMF scheme does not impose sparseness, smoothness or mutual independence (of the latent components) constrains on the observed data, the NMF framework was further investigated by many researches [3, 4]. Lee et al [3] introduced the NMF based on the notion of learning *parts-based linear representation* for nonnegative observed data. Nonnegativity is a natural constraint for many real-world applications, e.g., in the analysis of multimedia data i.e. images, video, audio, and text.

Existing dimensionality reduction schemes like PCA (principal component analysis), ICA (independent component analysis), and VQ (vector quantization) use additive and subtractive combinations of the basis vectors in order to reconstruct the original space, as there are negative entries in the basis vectors for PCA, ICA, and VQ used for original space reconstruction. The negative entries in the basis vectors for PCA, ICA, and VQ are not directly related to the original vector space in order to derive meaningful interpretation. Whereas, in case of NMF the basis vectors are nonnegative which allows only additive combinations of the basis vectors to reconstruct the original space. Lee at el [3] have shown that the NMF applied to face images yield features corresponding to institutive notion of face parts like lips, nose, eyes, etc. in contrast with the holistic representations learned by PCA and VQ [4].

Here we considered following NMF model in order to estimate nonnegative basis vectors, $\mathbf{b}_i : \; i = 1, \cdots, m$, from the data matrix $\mathbf{X} \in \mathcal{R}^{m \times N}, \; \mathbf{X} = \mathbf{BS}$ where $\mathbf{B} \in \mathcal{R}^{m \times n}$ is known as mixing matrix that contains *basis vectors* or feature space, $\mathbf{S} \in \mathcal{R}^{n \times N}$ is the coefficient matrix containing the underlying hidden components, $\mathbf{s}_i, i = 1, \cdots, n$ and $\mathbf{X}, \; \mathbf{B}, \;$ and $\mathbf{S}$ obey nonnegativity constraint [3].

The nonnegative matrix factorization with sparseness constraints can also be used to learn parts-based features of observed multimedia data. The sparseness constraints for the NMF helps to find an improved decomposition of the observed data, especially when Lee at al's proposed NMF scheme [3] fails to do so [4]. Hoyer [4] has been shown that sparseness constrained NMF can find qualitatively different parts-based representations that are more compatible with the sparseness assumptions instead of sparsifying the results of standard unconstrained NMF. In this paper, Hoyer's *non-negative sparse coding (NNSC)* [4] is used for learning basis vectors (or feature space) of the image data set (the observed data). Fig. 1 shows the basis vectors estimated from on the natural images using NNSC software package available at [5]. The basis vectors given in Fig. 1 are estimated using 40 natural images with following settings of NNSC software package, 1) total 15000 image segments was used, 2) each segment consists of 16×16 samples, 3) maximum number of iteration was set to 20000, 4) sparseness of the estimated coefficients was set to 0.85, 5) unconstrained sparseness for the basis vectors, and 6) number estimated sources was set to 72.

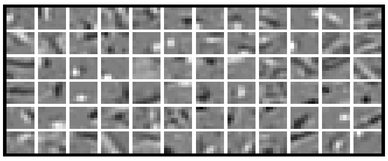

Fig. 1. Basis Vectors from Natural Images Estimated using NMF with Sparseness Constraints using NNSC software Package [5]

3 DATA EMBEDDING

The proposed data embedding process consists of two stages (a) the host image decomposition using selected basis vectors $\mathbf{B}_{sb}$ from the estimated based using the NMF with sparseness constraints, and (b) input message $\mathbf{M}$, embedding (encoding) by modifying (dithering) image coefficients in the selected features subspace using QIM. The nonnegative basis matrix $\mathbf{B}$ or nonnegative feature space is estimated based on the sparseness constrained NMF using a set of preselected images (see section 2). The host image is then projected to the feature subspace selected from the feature space $\mathbf{B}$ using a secret key $\mathbf{K}$ (or password). The input message $\mathbf{M} = \{m_1, \cdots, m_l\}$ is embedded into the host image by modifying image coefficients in the selected feature space using the binary dither modulation (a special case of QIM) [1]. In order to meet the fidelity requirement of the embedded information, estimated masking threshold or *just noticeable difference, (JND)* from the host image based on the human visual system (HVS) [7] is used. To this end, the estimated JND from the host image in the selected feature space is used to select image coefficients suitable for information embedding for a given quantization step size $\varDelta$ (used for information encoding/decoding). The quantization step size $\varDelta$ or embedding strength determines how much data can be embedded into a given host image. Therefore, stronger embedding can be achieved at the cost of lower embedding capacity and embedding distortion and vice versa (simulation results given in section 5 also highlight this fact).

The proposed scheme uses binary dither modulation (BDM) for information encoding/decoding. Low complexity is the main reason of using BDM for the simulation results presented in this paper. However higher dimensional QIM-based schemes with better capacity performance can also be used for information encoding/decoding. The binary dither modulation is quantization process based on two grids corresponding to the value of the message bit $m_i \in \{0, 1\}$. Fig. 2 illustrates the concept of binary dither modulation, in Fig. 2 the set Q_0 ('O') is defined by uniform quantizer with quantization step size $\varDelta$ which is used to map the host signal coefficient value s_i to a watermarked signal value $\hat{s}$ to encode $m_i = 0$. Similarly, the set Q_1 ('X') is another uniform quantizer with quantization step size $\varDelta$ and an offset of $\varDelta/2$ and used to encode $m_i = 1$. In Fig. 2 $*$ represents the selected image coefficient $s \in \mathcal{R}$ in the feature subspace for information encoding using quantizer Q_0 or Q_1 depending on the embedding message $m = 0$ or $m = 1$.

The salient steps of the proposed data embedding process are outlined as,

- Basis matrix $\mathbf{B}$ estimation from the set of preselected images using the NMF.
- Feature subspace, $\mathbf{B}_{sb}$, selection from the estimated feature space $\mathbf{B}$ using $\mathbf{K}$.
- The host image, $\mathbf{I}$ projection onto the selected feature subspace $\mathbf{B}_{sb}$.
- Host image coefficient selection based on the estimated *JND* i.e. $s(i, j) = f(\varDelta,\ JND(i, j))$ in order to achieve target robustness or vice versa.
- The channel encoded binary message, $M = \{m_1, \cdots, m_n\}$ embedding into the selected images coefficients ,$\mathbf{s}(i, j)$ using binary dither quantisers $Q_0\cdot$, or $Q_1\cdot$ corresponding to the embedding message m.
- Watermarked image I_{w} reconstruction by using the modified and unmodified coefficients.

The block diagram of the proposed NMF-based data embedding scheme is illustrated in Figure 3.

4 DATA DETECTION

The proposed NMF-based detector does not require the host image at the detector for information decoding/detection therefore falls into the category of blind detection. However, the encoder parameters i.e. codebook and quantization step size $\varDelta$, nonnegative basis matrix, and feature subspace selection key $\mathbf{K}$ are assumed to be available at the detector. Security of the proposed scheme depends on the following parameters 1) set of images used for nonnegative feature space estimation, 2) estimated feature space, and 3) feature space selection key. The

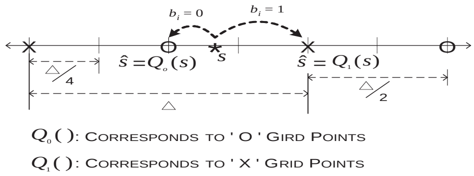

Fig. 2. Illustration of QIM using Binary Dither Modulation for Encoding Message bit $m = 0$ and $m = 1$ using Quatizers $Q_0(\cdot)$ and $Q_1(\cdot)$ respectively, into the Selected Coefficient s represented by $*$

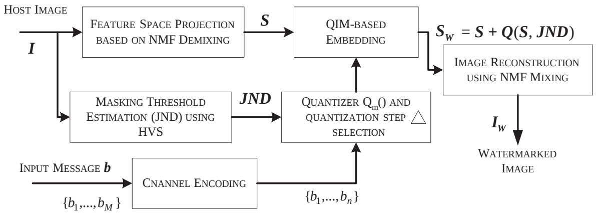
Host Image
I
Feature Space Projection based on NMF Demixing
S
QIM-based Embedding
$S_W = S + Q(S, JND)$
Image Reconstruction using NMF Mixing
Masking Threshold Estimation (JND) using HVS
JND
Quantizer $Q_m()$ and quantization step $\triangle$ selection
I_W
Watermarked Image
Input Message b
$\{b_1,\ldots,b_M\}$
Cnannel Encoding
$\{b_1,\ldots,b_n\}$

Fig. 3. Block Diagram of the Proposed NMF-based Secure Data Embedding

proposed scheme is reasonably secure as long as security of the estimated feature space ensured. In case security of estimated feature space is breached the security of the proposed scheme is determined from secret key or password. Lets assume that estimated feature space is known to attacker then active attacker can guess feature subspace for information decoding with probability $P_c = 1/r$ where $r = fracp!(p - h)$, here p is dimensionality of estimated feature space and h is dimensionality of subspace used for information encoding/decoing. Let us consider that 64–dimensional subspace of 72–dimensional estimated feature space is used for data encoding/decoding, in this case correct decoding probability $P_c =\approx 10^{-100}$. However questions such as how many feature vectors does an attacker require to achieve target decoding probability, needs further investigation.

The information detection process consists of decomposing the watermarked image subjected to attack-channel distortion, $\tilde{I}_{\mathrm{w}}$ using selected feature subspace. The JND estimated based on the watermarked image is used to select the watermarked image coefficients in the nonnegative feature subspace as the potential information carriers. For information decoding from the selected coefficients of the watermarked image subjected to attack-channel distortion, $\tilde{\mathbf{s}}_{\mathrm{w}}$, the *nearest neighborhood decoding* using predefined threshold and the *maximum a posterior (MAP)* based decoding can be used. The nearest neighborhood decoding is the simplest decoding for the QIM-based schemes. The nearest neighborhood decoding requires the knowledge of the codebook used for information encoding and its robustness depends on the quantization step size $\varDelta$ [1, 6].

Whereas, the MAP-based decoding relies on the probabilistic framework for information decoding. For example, the MAP based decoding maximize the posterior probability ,$p(m_i|\tilde{s}_w)$, in order to estimate the embedded message $\hat{m}_i$ i.e. $\hat{m}_i = \max_{m_i}\{p(m_i|\tilde{s}_w)\}$. Bounkong et al [6] have shown that the decoding performance of the MAP-based decoding for QIM-based embedding directly depends on the probabilistic models for both the host signal in the selected nonnegative feature subspace and the attack-channel noise. Simulation results presented in this paper are based on the nearest neighborhood-based decoding. The nearest neighborhood decoder is used for information decoding from the watermarked image due to its simplicity over the MAP-based decoder. The block diagram of the proposed NMF-based detection scheme is given in Fig. 4.

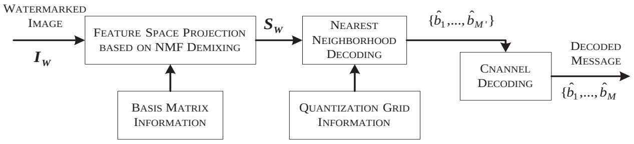
WATERMARKED IMAGE
$\boldsymbol{I}_w$
FEATURE SPACE PROJECTION BASED ON NMF DEMIXING
$\boldsymbol{S}_w$
NEAREST NEIGHBORHOOD DECODING
$\{\hat{b}_1,\ldots,\hat{b}_{M'}\}$
CNANNEL DECODING
DECODED MESSAGE
$\{\hat{b}_1,\ldots,\hat{b}_M$
BASIS MATRIX INFORMATION
QUANTIZATION GRID INFORMATION

Fig. 4. Block Diagram of the Proposed NMF-based Information Detection

5 SIMULATION RESULTS

In order to test performance of the proposed data hiding scheme in terms of fidelity, capacity, and robustness. Five 256 × 256 gray scale images are used. The 72–dimensional nonnegative feature space is estimated using 40 natural gray scale images. Hoyer's NNSC software package [5] is used for feature space estimation. The 64–dimensional feature subspace consisting is selected using secret key **K**. The secret key **K** consists of 20 alphanumeric characters is used as a seed to the pseudo-random number generator which iteratively generates random number between 1 and 72-(*iteration number*), which is used to select a feature from $72 - (\textit{iteration number})$ remaining vectors. Simulation results presented in this section are based on following setting: 1) quantization step size $\Delta \in \{1.0, 2.0, 3.0\}$, 2) no channel coding is used, and 3) embedding distortion is measured in terms of *peak signal to noise ratio (PSNR)* which is calculated as,$PSNR = 10\log_{10}(\frac{1}{m\times\ n}\sum_{i,j}^{m,n}(d^2(i,j)))$ where $\mathbf{d} = \mathbf{I}_w - \mathbf{I}$.

Fig. 5 shows the fidelity performance of the proposed NMF-based scheme at the quantization step size $\Delta = 2$ (only two images of different textures are presented here due to space limitations). Fig. 5 shows that ***Baboon*** image has rich texture compared to ***Bird*** image hence higher capacity for a given embedding strength and likewise higher distortion and results given in Table 5 aslo agree with this fact.

Experimental results presented in the Table 5 show that capacity of the proposed scheme depends on the embedding strength and the host image character-

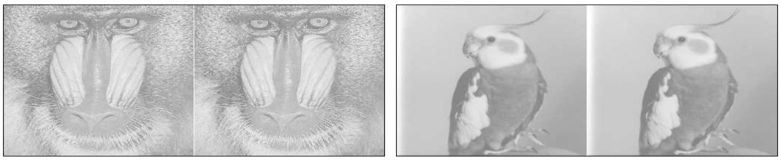

Fig. 5. Fidelity Performance of the Proposed Scheme: (from left to right) Original ***Baboon*** Image Data Embedded ***Baboon*** with PSNR = 34.8704, Original ***Bird*** Data Embedded ***Bird*** with PSNR = 53.76

istics. For example, for a given image, in order to achieve strong embedding we have to compromise capacity and vice versa. Similarly, the images with stronger texture have higher capacity over the low texture images for a given embedding strength, e.g. 3960 bits can be embedded in *Baboon* whereas only 55 bits of data in *Bird* for quantization step size $\Delta = 2$.

The robustness performance of the proposed scheme is also tested against additive white Gaussian noise (AWGN) attack. To simulate this attack, white Gaussian noise added to the watermarked image. The resulting image is then applied to the proposed detector for information decoding. Robustness performance of the proposed scheme in terms of P_e for various SNR (dB) values is plotted in Fig. 6. The P_e plot given in Fig. 6 is obtained by averaging over 1000 independent simulations for each image listed in Table 5. Fig. 6 shows that for a given image, robustness performance of the proposed scheme improves stronger embedding but at the cost of embedding capacity and vice versa.

IMAGE	Δ	**Capacity**	*PSNR* (dB)	**IMAGE**	Δ	**Capacity**	*PSNR* (dB)
Baboon	1.0	15257	35.22	Bridge	1.0	8250	38.26
	2.0	3960	34.87		2.0	979	41.25
	3.0	1232	36.46		3.0	88	48.00
Lenna	1.0	3509	41.84	Bird	1.0	1133	46.70
	2.0	341	45.30		2.0	55	53.76
	3.0	55	50.23				
Hat	1.0	2805	42.81	Hat	2.0	132	49.00

Table 1. Performance of the proposed scheme in terms of Embedding Capacity (in bits) and Fidelity for a given Embedding Strength

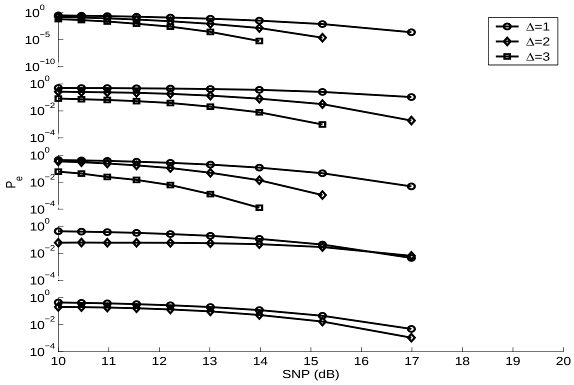

Fig. 6. Robust Performance of the Proposed Scheme against White Gaussian Noise Attack for Test Images ***Baboon***, ***Lenna***, ***Bridge***, ***Bird***, and ***Hat*** (top to bottom)

6 CONCLUSION

This paper presents a novel secure data hiding scheme for multimedia data based on the NMF. The nonnegative feature space is estimated using the sparseness constrained NMF-framework for a preselected set of natural images. The subspace is selected from the estimated feature space using secret key which is used to decompose the host image for information embedding and detection. Simulation results show that performance of the proposed scheme in terms of fidelity and capacity directly depend on Δ used for information embedding and detection.

References

1. B. Chen, and G. W. Wornell, Quantization Index Modulation: A Class of Provably Good Methods for Digital Watermarking and Information Embedding, *IEEE Trans. Information Theory,* vol. 47(4), pp. 1423–1443, May, 2001.
2. P. Paatero, and U. Tapper, Positive Matrix Factorization: A Non-Negative Factor Model with Optimal Utilization of Error Estimates of Data Values, *Environmetrics,* vol. 381, pp. 607–609, 1994.
3. D. D. Lee, and H. S. Seung, Learning the Parts of Objects by Non-Negative Matrix Factorization, *Nature,* vol. 401(6755), pp. 788–1718, 1999.
4. P. O. Hoyer, Non-Negative Sparse Coding, *J. Machine Learning Research,* vol. 5 pp. 1457–1469, 2004.
5. *http://www.cis.hut.fi/phoyer/code/*
6. S. Bounkong, B. Toch, D. Saad, and D. Lowe, ICA for Watermarking Digital Images, *J. Machine Learning Research,* vol. 1 pp. 1–25, 2002.
7. A. B. Watson, Visual Optimization of DCT Quantization Matrices for Individual Images, *Proc. AIAA Computing in Aerospace 9,* pp. 286–291, 1993.