Research article

A video watermark algorithm based on tensor decomposition

Shanqing Zhang\textsuperscript{1}, Xiaoyun Guo\textsuperscript{1}, Xianghua Xu\textsuperscript{1}, Li Li\textsuperscript{1,*}, Chin-Chen Chang\textsuperscript{2}

\textsuperscript{1} School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou 310018, China
\textsuperscript{2} Feng Chia University, Department of Information Engineering and Computer Science, Taiwan

*Correspondence: Email: lili2008@hdu.edu.cn; Tel: +86-136-6663-2945.

Abstract: Since most of the previous video watermark algorithms regard a video as a series of consecutive images, the embedding and extraction of watermark are performed on these images, and the correlation and redundancy among frames of a video are not considered. Such algorithms are weak in protecting against frame attacks. In order to improve the robustness, we take into consideration the correlation and redundancy among the frames of a video to propose a blind video watermark algorithm based on tensor decomposition. First, a grayscale video is represented as a 3-order tensor, and the core tensor is obtained by tensor decomposition. Second, the watermark embedding position is selected based on the stability of the maximum value in the core tensor because the core tensor represents the main energy of a video. Then, the watermark is embedded by quantifying the maximum value in the core tensor. Finally, the watermark is uniformly distributed across frames of a video by inverse tensor decomposition. The experiments show that our algorithm based on tensor decomposition has better imperceptibility and robustness against common video attacks.

Keywords: Video watermark, tensor, tensor decomposition, tucker decomposition, parity quantization.

1. Introduction

Copyright protection is more and more important as digital videos become popular. Video watermark technology is the digital watermark technology with video being the carrier. It embeds confidential information into carrier based on the video’s redundancy. Video watermark technology is invisible, able to resist malicious attacks, and has achieved video copyright protection or video content authentication [1].
Video watermark technology is classified into spatial domain watermark technology and transform domain watermark technology. Kalker [2] introduced spread spectrum into video watermark and proposed a classic video watermark algorithm for video broadcast monitoring. The algorithm regards a video as a sequence of images and embeds the watermark into the video frames in the spatial domain. The algorithm works well with broadcast transmission signal processing and produces a detection process with low complexity. But it is not robust against common attacks. The spatial video watermark algorithm proposed by Hartung [3] converts the original video image into a one-dimensional signal and modulates the watermark into a pseudo-random sequence. This watermark is embedded into the one-dimensional signal. This classical spatial domain algorithm has disadvantages in robustness against attacks such as video compression and filtering. Karybali [4] proposed an effective spatial watermark algorithm to improve robustness by perceptual masking and watermark blind extraction.

A spatial watermark algorithm directly modifies pixels of image in spatial domain. The advantages are good transparency and low complexity, whereas one main disadvantage is watermark loss after image compression or geometric attack [5–7]. A transform domain watermark algorithm transforms an image into a domain by Discrete Wavelet Transform (DWT), Discrete Fourier Transform (DFT), or Discrete Cosine Transform (DCT), and so forth. Watermark embedding is performed after the transform. The transform domain watermark technology is robust against attacks such as filtering and image compression. In 1995, Koch proposed a watermark algorithm in DCT domain, which is robust against compression and filtering attacks [8]. In 1997, Cox [9] summarized and analyzed the existing transform domain watermark algorithms and proposed to embed watermark into the low frequency coefficients of an image. The robustness of the watermark was effectively improved. In order to further improve anti-attack capability, Chandra [10] first used Singular Value Decomposition (SVD) for digital watermark in 2001, and embedded watermark image into singular value of the carrier image. The digital watermark algorithms with SVD can improve anti-attack capability, especially against geometric attacks.

Most video watermark algorithms so far regard a video as a series of consecutive images and embed watermark into the images. However, the correlation and redundancy among the frames of a video are not considered. They are not robust against frame attacks such as frame addition, frame deletion, or frame averaging.

We introduce tensor into video watermark in order to solve the above problems. A tensor is a multi-dimensional array and has advantages in representing multi-dimensional data. Tensor computation has been successfully used in face recognition [11], visual tracking [12] and action classification [13]. Tensor decomposition [14–16] has become a landmark in video-related research. However, published studies on tensor-based video watermark are rare [17,18]. In [17], Abdallah proposed a tensor-based video watermarking algorithm, but it is non-blind. Recently, Xu [18] represented a color image as a third-order tensor, and proposed a robust watermark algorithm, but that is only suitable for color image. In this study, we represented a grayscale video as a 3-order tensor and proposed a blind video watermark algorithm based on tensor decomposition. The flow chart of our algorithm is shown in Figure 1. The core tensor is obtained by Tucker decomposition of the 3-order tensor. The watermark is embedded by parity quantization [19] of the maximum value of core tensor. The modified core tensor is uniformly distributed across frames of a video by inverse Tucker decomposition. The watermark extraction is simply the inverse process of embedding. Experiments show that our algorithm is robust against common video attacks and imperceptible.
The main contributions of this study are as follows:

1. We introduce tensor into video watermark. The correlation and redundancy among the frames of a video are considered to enhance the robustness against frame attacks.

2. The stability of the algorithm is guaranteed because Tucker decomposition is reversible, and the core tensor represents the main energy of the original video and is relatively invariant.

3. The modified core tensor is uniformly distributed among the frames of a video by inverse Tucker decomposition, so that the video quality and the imperceptibility of watermark are guaranteed.

This paper is organized as follows. The basics of tensor are introduced in Section 2, our watermark embedding and extraction algorithm is described in Section 3, and the experiment results are shown in Section 4.

![Diagram](image)

**Figure 1.** Video watermark framework based on tensor decomposition.
2. Tensor

The next notations are used.
Scalar: \((a, b, \text{ etc.})\); Vector: \((\mathbf{a}, \mathbf{b}, \text{ etc.})\); Matrix: \((A, B, \text{ etc.})\); High-order tensor: \((\mathcal{A}, B, \text{ etc.})\).

A tensor is a high-order matrix, an extended form of a matrix toward the higher dimension. Vector and matrix are first-order tensor and second-order tensor, respectively. A N-order tensor \(\mathcal{A}\) is defined as \(\mathcal{A} \in \mathbb{R}^{I_1 \times \cdots \times I_N}\). A video sequence can be regarded as a 3-order tensor, and its three dimensions are the width, height and length of the video.

2.1. Tensor unfolding

Tensor has many advantages in representing multi-dimension data. For example, if a video is considered as a tensor, the properties of the original video can be preserved to the maximum extent. However, high-order tensor results in higher-level computation. So, tensor is usually unfolded to matrices for easy computation [20].

\(\mathcal{A}\) is N-order tensor \(\mathcal{A} \in \mathbb{R}^{I_1 \times \cdots \times I_N}\), tensor \(\mathcal{A}\) is unfolded into matrices \(A_1, \cdots, A_N, A_k \in \mathbb{R}^{I_k \times (I_1 \times \cdots \times I_{k-1} \times I_{k+1} \times \cdots I_N)}\). The unfolding of a 3-order tensor is shown in Figure 2.

![Figure 2. The unfolding of a 3-order tensor.](image)

2.2. The mode-\(n\) product of tensor and matrix

The mode-\(n\) product of a N-order tensor \(\mathcal{A} \in \mathbb{R}^{I_1 \times \cdots \times I_N}\) and a matrix \(U \in \mathbb{R}^{J \times I_n}\) is noted as \(\mathcal{A} \times_n U\), where \(\times_n U \in \mathbb{R}^{I_1 \times \cdots \times I_{n-1} \times J \times I_{n+1} \times \cdots \times I_N}\), the entries are given by:
The mode-1 product of a 3-order tensor and a matrix is shown in Figure 3.

\[
[(\mathcal{A} \times_n U)_{i_1 \times \cdots \times i_{n-1} \times j \times i_{n+1} \times \cdots \times i_N} = \sum_{i_{N+1}=1}^{l_n} a_{i_1 \cdots i_N u_{j i_N}]
\]  

(1)

Figure 3. The mode-1 product of a 3-order tensor and a matrix.

2.3. Tensor decomposition

Tensor decomposition is generalization of matrix singular value decomposition in high dimensions [21]. An image \( F \) with size \( I_1 \times I_2 \) can be decomposed by SVD as follows:

\[
[F_{I_1 \times I_2} = U_{I_1 \times I_1} \cdot S_{I_1 \times I_2} \cdot V_{I_2 \times I_2}^T]
\]  

(2)

where \( U \) and \( V \) are the left and right singular matrices of \( F \), respectively. \( S \) is the diagonal matrix composed of the singular values of \( F \).

A matrix is regarded as a 2-order tensor. According to the definition of mode product of tensor and matrix, the singular value decomposition SVD of a matrix can be represented as a 2-order tensor \( S \) mode product of a matrix \( U^{(1)} \) and a matrix \( U^{(2)} \) sequentially:

\[
[F = S \times_1 U^{(1)} \times_2 U^{(2)}]
\]  

(3)

with

\[
[F = U^{(1)} \cdot D_1 \cdot V_1^T]
\]  

(4)

\[
[F^T = U^{(2)} \cdot D_2 \cdot V_2^T]
\]  

(5)

For a tensor of high-order \((k>2)\), CP (CANDECOMP/PARAFAC) decomposition [22] and Tucker decomposition [23] are often used. CP decomposes a tensor into a finite sum of rank-1 tensors. CP guarantees the uniqueness of the decomposition result, but its rank solution is an NP problem. Tensor is decomposed into the mode product of the core tensor and factor matrices by Tucker decomposition. The core tensor contains the main information of the original tensor. Tucker decomposition is used in our study. High-Order Singular Value Decomposition (HOSVD) is the
classic algorithm for Tucker decomposition. Given a tensor $\mathcal{A}$ with size $M \times N \times K$, Tucker decomposition with HOSVD [24] is as follows:

$$[\mathcal{A} = S \times_1 U \times_2 V \times_3 W] \quad (6)$$

with

$$[A_1 = U \cdot D_1 \cdot V_1^T] \quad (7)$$

$$[A_2 = V \cdot D_2 \cdot V_2^T] \quad (8)$$

$$[A_3 = W \cdot D_3 \cdot V_3^T] \quad (9)$$

where $A_1, A_2, A_3$ are the unfolding matrices of $\mathcal{A}$ in three directions, respectively. And the core tensor $S$ writes:

$$[S = A \times_1 U^T \times_2 V^T \times_3 W^T] \quad (10)$$

Tucker decomposition of a 3-order tensor is shown in Figure 4, where $1 \leq R_1 \leq M$, $1 \leq R_2 \leq N$, $1 \leq R_3 \leq K$.

![Figure 4. Tucker decomposition of a 3-order tensor.](image_url)

3. **A watermark algorithm based on tensor decomposition**

In our watermark algorithm, a video is initially represented as a tensor, which contains the relevance and redundancy among the frames of a video. Then, a core tensor is obtained by Tucker decomposition. Finally, a watermark is embedded into the core tensor by parity quantization.

3.1. **The process of watermark embedding**

The resolution of a video $V$ is $M \times N$, and the size of a watermark $B$ is $m \times m$. To make full use of the relevance and redundancy among the frames of a video, $K$ frames of a grayscale video are grouped as a 3-order tensor. The size of tensor $A_i (1 \leq i \leq m^2)$ is $M \times N \times K$. The core tensor $S_i (1 \leq i \leq m^2)$ and 3 factor matrices $U_i, V_i, W_i$ are obtained through Tucker decomposition with HOSVD. The process of watermark embedding is as follows.

1. Arnold Scrambling. In order to eliminate spatial correlation among the binary watermark
pixels, B becomes $B'$ through Arnold transformation. It can be defined by the following equation:

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} 1 & a \\ b & ab + 1 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix} \mod (m) \tag{11}$$

where $(x, y)$ is the coordinate of the original watermark pixel, $(x', y')$ is the transformed coordinate of $(x, y)$ with Arnold, and $m$ is the width of the matrix. In the experiment, $a=1$ and $b=1$. The Arnold transformation is performed for $t$ times on the original watermark, and $t$ is saved as a key for watermark extraction.

(2) Tucker decomposition with HOSVD. Tucker decomposition is performed for each $\mathcal{A}_i(1 \leq i \leq m^2)$ to obtain the core tensor $S_i$.

$$S_i = \mathcal{A}_i \times_1 U_i^T \times_2 V_i^T \times_3 W_i^T \tag{12}$$

where $\mathcal{A}_i \in R^{M \times N \times K}$ is the original video tensor, $U_i \in R^{M \times m_1}$, $V_i \in R^{N \times m_2}$, $W_i \in R^{K \times m_3}$ are factor matrices.

(3) Quantification and modification of the core tensor. Parity quantization is used to embed watermark into the core tensor. For each tensor $\mathcal{A}_i$, $S_i(1,1,1)$ is the maximum value of the core tensor $S_i$, and is donated as $\sigma_i$.

a. Quantify the maximum value $\sigma_i$ of each core tensor, denoted as $\lambda_i = \text{round} \left( \frac{\sigma_i}{Q} \right)$, where $Q$ is the quantization intensity, and the value of $Q$ is discussed below.

b. The maximum value of each core tensor $S_i$ is modified to embed watermark.

$$[\sigma'_i] = \begin{cases} (\lambda_i + 0.5)Q & \text{if } (\lambda_i + B'_i) \text{ is odd} \\ (\lambda_i - 0.5)Q & \text{if } (\lambda_i + B'_i) \text{ is even} \end{cases} \tag{13}$$

(4) Reconstruction of the watermarked video. The watermarked video $\mathcal{A}'_i$ is reconstructed by inverse Tucker decomposition with the modified core tensor $S'_i$.

$$[\mathcal{A}'_i = S'_i \times_1 U_i \times_2 V_i \times_3 W_i] \tag{14}$$

3.2. The process of extracting watermark

Watermark extraction is the inverse process of watermark embedding. The specific steps of watermark extraction are as follows.

(1) Tucker decomposition is performed on each watermarked video tensor $\mathcal{A}'_i$ to obtain the core tensor $S'_i$.

$$[S'_i = \mathcal{A}'_i \times_1 U'_i \times_2 V'_i \times_3 W'_i] \tag{15}$$

(2) The extracted watermark is determined according to the maximum value of the core tensor $S'_i$.

a. Quantify the maximum value $\sigma'_i$ of each core tensor $S'_i$, denoted as $\lambda'_i = \text{floor} \left( \frac{\sigma'_i}{Q} \right)$.

b. Determine the extracted information according to the parity of $\lambda'_i$. $B'_i$ is 1 when $\lambda'_i$ is even; $B'_i$ is 0 when $\lambda'_i$ is odd.

(3) Perform inverse Arnold transformation on $B'$ to obtain the original watermark B.
4. Experiment results

4.1. Metrics

The follow metrics are used to measure the robustness and imperceptibility of our video watermark algorithm based on tensor decomposition. The imperceptibility of the watermark is evaluated with Peak Signal to Noise Ratio (PSNR) and Mean Square Error (MSE).

\[
\text{MSE} = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} [l(i,j) - l'(i,j)]^2
\]

where M and N are the height and width of a single-frame image, I and I’ are the original video frame and watermarked video frame. The smaller the MSE value, the smaller the difference between the single-frame watermark image and the original image is. PSNR is calculated by MSE as follow:

\[
\text{PSNR} = 10 \log_{10} \left( \frac{255^2}{\text{MSE}} \right)
\]

A smaller PSNR means that the distortion of the watermarked frame is more serious. In addition, the bit error rate (BER) and normalized correlation coefficient (NC) are used to evaluate the robustness of the watermark. The equations are as follows:

\[
\text{BER} = \frac{\sum_{i=1}^{m} \sum_{j=1}^{m} |B'(i,j) - B(i,j)|}{m \times m}
\]

\[
\text{NC} = \frac{\sum_{i=1}^{m} \sum_{j=1}^{m} (B'(i,j) \cdot B(i,j))}{\sqrt{\sum_{i=1}^{m} \sum_{j=1}^{m} (B'(i,j))^2 \cdot \sum_{i=1}^{m} \sum_{j=1}^{m} (B(i,j))^2}}
\]

where \(m\) is the size of the watermark, B and B’ are the original watermark and extracted watermark, respectively. The robustness of watermark increases as NC increases.

4.2. Experiment parameters

The size of the test video is 352 × 640 and there are 2268 frames in total. The size of the watermark is 18 × 18. A bit of watermark is embedded into a group of \(K\) frames. The size of the tensor is 352 × 640 × \(K\), \(K\) being 7 in our experiment. The number of scrambling \(t\) is 15. The relationship between quantization strength \(Q\) and watermark BER is shown in Figure 5. BER decreases as \(Q\) increases. The watermark is correctly extracted when \(Q \geq 1000\). \(Q\) is set to 2000 in order to ensure the robustness of the algorithm and video quality. The PSNR of the first 100 frames of a video when \(Q = 2000\) is shown in Figure 6. The PSNR of the watermarked video is over 40dB. The examples of our algorithm are shown in Figure 7.
Figure 5. The relationship between quantization strength $Q$ and BER.

Figure 6. The PSNR of the first 100 frames of the video.

Figure 7. (a) original video, (b) watermarked video, (c) original watermark, (d) extracted watermark.
4.3. Robustness test

Different attacks are used to testify the robustness of our watermark algorithm, including frame swapping, zooming, cropping, filtering, noising, and black-border filling. Our tensor-based watermark algorithm is robust against frame attacks, zooming, rotation, cropping in the experiments. The NC of the extracted watermark through the frames swapping in a group is shown in Table 1. The NC of the extracted watermark remains high even if about 50% frames in a group are swapped. The modified maximum value of the core tensor has uniform effect on each frame, because every element in the core tensor takes part in the mode product of the factor matrices by inverse Tucker decomposition.

<table>
<thead>
<tr>
<th>Number of frames replaced</th>
<th>Extracted watermark</th>
<th>NC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><img src="image" alt="Watermark" /></td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td><img src="image" alt="Watermark" /></td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td><img src="image" alt="Watermark" /></td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td><img src="image" alt="Watermark" /></td>
<td>0.8055</td>
</tr>
</tbody>
</table>

The relation between the video’s zooming and NC of the extracted watermark is shown in Figure 8. Watermark is extracted irrespective of a video being zoomed in or out as long as the video is restored at the same resolution. The NC of the extracted watermark is as high as 0.8098 even when the video is zoom out to 0.1 of its original size.

![Figure 8](image)

Figure 8. The relationship between zooming and NC.
Watermark is extracted correctly when black-border is filled around a borderless video because the core tensor represents the main energy of a video and the energy from these zero-valued pixels in a video is tiny. The result is shown in Figure 9.

Figure 9. (a) The watermarked video filled with black-border (b) Extracted watermark.

The watermark is also extracted correctly as long as the direction of the video remains unchanged. The relationship between the rotation angle and the maximum value of the core tensor is shown in Figure 10. The maximum value of the core tensor changes periodically as the rotation angle changes. The watermark is extracted correctly from the watermarked video after rotation correction [25,26]. Some experiment results for rotation attack are shown in Table 2.

Figure 10. The relationship between the rotation angle and the maximum value of the core tensor.
Table 2. Results for rotation attack.

<table>
<thead>
<tr>
<th>Rotation angle</th>
<th>Rotated video</th>
<th>Corrected video</th>
<th>Extracted watermark</th>
<th>NC</th>
</tr>
</thead>
<tbody>
<tr>
<td>10°</td>
<td><img src="image1" alt="Image" /></td>
<td><img src="image2" alt="Image" /></td>
<td><img src="image3" alt="Image" /></td>
<td>1</td>
</tr>
<tr>
<td>45°</td>
<td><img src="image4" alt="Image" /></td>
<td><img src="image5" alt="Image" /></td>
<td><img src="image6" alt="Image" /></td>
<td>1</td>
</tr>
<tr>
<td>60°</td>
<td><img src="image7" alt="Image" /></td>
<td><img src="image8" alt="Image" /></td>
<td><img src="image9" alt="Image" /></td>
<td>1</td>
</tr>
</tbody>
</table>

The experiment results of black-border video by our algorithm are shown in Table 3. Watermark is extracted correctly irrespective of the cropping being in the up or down, left or right side when the cropped part is included in a black-border because the contribution on the maximum value of the core tensor from black-border (0-value pixels) is very small. The watermark extraction of filtering and noise attacks is shown in Table 4.

5. Conclusions

A grayscale video is represented as a 3-order tensor in order to make full use of the relevance and redundancy among the frames of a video. The core tensor is obtained by Tucker decomposition, and the embedding and extraction of the video watermark is achieved using parity quantization of the maximum value of the core tensor. Watermark information is uniformly distributed across the frames of a video because of the reversibility and stability of Tucker decomposition, so that the video quality and the imperceptibility of watermark are guaranteed. It is robust against various video attacks, especially frame attacks. In our algorithm, only grayscale videos are used, color video watermarking method based on the tensor domain will be studied in the future.
### Table 3. Results for cropping attack on black-border video.

<table>
<thead>
<tr>
<th>Cropping attack</th>
<th>Attacked video</th>
<th>Extracted watermark</th>
<th>NC</th>
</tr>
</thead>
<tbody>
<tr>
<td>No attack</td>
<td><img src="image1" alt="Attacked video" /></td>
<td><img src="image2" alt="Extracted watermark" /></td>
<td>1</td>
</tr>
<tr>
<td>Cut 45 rows at the top</td>
<td><img src="image3" alt="Attacked video" /></td>
<td><img src="image4" alt="Extracted watermark" /></td>
<td>1</td>
</tr>
<tr>
<td>Cut 45 rows on the left</td>
<td><img src="image5" alt="Attacked video" /></td>
<td><img src="image6" alt="Extracted watermark" /></td>
<td>1</td>
</tr>
</tbody>
</table>

### Table 4. Results for other attacks.

<table>
<thead>
<tr>
<th>Attack type</th>
<th>Extracted watermark</th>
<th>NC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median filtering (3 × 3)</td>
<td><img src="image7" alt="Extracted watermark" /></td>
<td>0.9587</td>
</tr>
<tr>
<td>Mean filtering (3 × 3)</td>
<td><img src="image8" alt="Extracted watermark" /></td>
<td>0.9973</td>
</tr>
<tr>
<td>Gaussian noise (0,0.01)</td>
<td><img src="image9" alt="Extracted watermark" /></td>
<td>0.8675</td>
</tr>
<tr>
<td>Salt and pepper noise (0.01)</td>
<td><img src="image10" alt="Extracted watermark" /></td>
<td>0.9973</td>
</tr>
<tr>
<td>Poisson noise</td>
<td><img src="image11" alt="Extracted watermark" /></td>
<td>0.9783</td>
</tr>
</tbody>
</table>
Acknowledgments

This research was funded by the Public Welfare Technology and Industry Project of Zhejiang Provincial Science Technology Department (Grant No. LGG19F020016 and No. LGG18F020013) and the Key Research and Development Project of Zhejiang Province (Grant No. 2017C01022).

Conflict of interest

The authors declare no conflicts of interest.

References