JOURNAL OF INFORMATION PROCESSING SYSTEMS JIPS

Reversible Multipurpose Watermarking Algorithm Using ResNet and Perceptual Hashing

Mingfang Jiang[1] and Hengfu Yang[2,*]

Abstract

To effectively track the illegal use of digital images and maintain the security of digital image communication on the Internet, this paper proposes a reversible multipurpose image watermarking algorithm based on a deep residual network (ResNet) and perceptual hashing (also called MWR). The algorithm first combines perceptual image hashing to generate a digital fingerprint that depends on the user's identity information and image characteristics. Then it embeds the removable visible watermark and digital fingerprint in two different regions of the orthogonal separation of the image. The embedding strength of the digital fingerprint is computed using ResNet. Because of the embedding of the removable visible watermark, the conflict between the copyright notice and the user's browsing is balanced. Moreover, image authentication and traitor tracking are realized through digital fingerprint insertion. The experiments show that the scheme has good visual transparency and watermark visibility. The use of chaotic mapping in the visible watermark insertion process enhances the security of the multipurpose watermark scheme, and unauthorized users without correct keys cannot effectively remove the visible watermark.

Keywords

Deep Residual Network, Multipurpose Watermarking, Perceptual Hashing, Reversible Visible Watermarking

1. Introduction

With the rapid development of Internet communication technology and the wide application of digital multimedia, digital watermarking technology has become a hot issue in the field of digital copyright protection. Digital watermarking technology is a new information security method that is different from traditional encryption methods. It protects digital media by hiding secret information (watermarks) in digital media (image, voice, video, etc.). As an important method of copyright protection for digital media, digital watermarking technology has received extensive attention and development [1,2]. However, common digital watermarking schemes generally only have a single copyright protection or data authentication function, which cannot meet the strong demand for copyright protection, data authentication, violation tracking, and other aspects in practical communication applications [3]. Multipurpose watermarking refers to embedding multiple watermarks of different properties in the same digital image to achieve multiple objectives such as copyright protection, authentication, and user

tracking, which can meet the simultaneous requirements of multiple different objectives in the actual communication environment of digital media. A multipurpose watermark embedding strategy should consider the characteristics of each type of watermark and combine multiple watermarks with different properties. Sometimes, it is necessary to weaken the performance of a certain watermark to optimize the overall watermarking scheme. Therefore, multipurpose digital watermarking is not a simple addition of two or several watermarks with different properties. It is necessary to comprehensively balance the various requirements of the multipurpose watermarking system to achieve optimal overall performance. Lu et al. [4] put forward the concept of multipurpose watermarking earlier. Their scheme achieved audio protection and authentication by simultaneously inserting a robust watermark and a fragile watermark. In 2012, Lei and Soon [5] proposed a multipurpose audio watermarking scheme in which the watermark data are inserted in the low frequency largest significant discrete cosine transform (DCT) coefficients for robustness. Peng et al. [6] presented a multipurpose watermarking scheme for vector maps that embeds a robust watermark into the feature points of the objects and a fragile watermark into the nonfeature points. Sheidani and Eslami [7] proposed a blind multipurpose watermarking algorithm. They want to simultaneously achieve multiple security objectives by inserting a single watermark into the host image.

These existing multipurpose watermarking schemes are not sufficiently secure to prevent unauthorized users. To achieve a comprehensive balance of different watermark components in a multipurpose watermarking system, this paper proposes a reversible multipurpose image watermarking scheme based on Resnet and perceptual hashing. The main contributions are as follows:

- A reversible multipurpose watermarking algorithm is proposed by embedding a lossless visible watermark and a digital fingerprint in the two non-overlapping regions of the host image.
- It has good security because only authorized users with the correct key can remove the visible watermark.
- Experiments show that it has good visual transparency and watermark visibility.

2. Proposed Multipurpose Image Watermarking Methods

To achieve hierarchical browsing and violator tracking of visual media data, a multipurpose digital image watermarking algorithm based on ResNet is designed. In our multipurpose watermarking algorithm, a perceptual image hash sequence is first generated from the host image, and then the digital fingerprint is produced by combining the user identity with the perceptual image hash. Subsequently, the embedding strength of the digital fingerprint is computed using ResNet. Finally, both the visible watermark and digital fingerprint are embedded in different regions to achieve copyright notice and violation tracking simultaneously.

2.1 Deep Residual Network Model for the Quantization Step of Digital Fingerprints

In the existing digital fingerprint embedding methods, most of them use manual methods or HVS characteristics to determine the embedding strength, which makes it difficult to obtain the optimal embedding strength and cannot guarantee good transparency and robustness of the digital fingerprint [8]. To enhance robustness, this paper proposes the use of the depth residual network model to determine the embedding intensity (here is the quantization step). This scheme extracts complex features such as image

scale, brightness, texture, and so on using the deep network, and it adaptively obtains the quantization step of the digital fingerprint. The deep residual network takes the cover image as the input and the corresponding quantization step as the label to learn the relationship between the image and the quantization step. After the model is trained, the quantization step is predicted according to the cover image. To improve the fitting speed of the network and increase the number of samples, the cover image is divided into 32×32 as input.

2.1.1 Network structure

The network model is shown in Fig. 1. There are 19 layers, including the convolution layer, residual block, global average pooling layer, and full connection layer. The network input is host images with size 32×32, first pass through a convolution layer (convolution core size is 3×3, the quantity is 16); Then it passes through nine residual blocks (convolution kernel size is 3×3, the number is 16, 32 and 64, respectively). To extract more abundant feature information, the number of convolution cores increases with the deepening of network layers. The end connection of the deep residual network is the global average pooling layer and the full connection layer, and the final output size is 1×1 quantization step Δ. The residual network directly down-samples through the convolution layer with a step size of 2. To maintain the size of the feature map consistent with the input, both the step size and edge filling are set to 1.
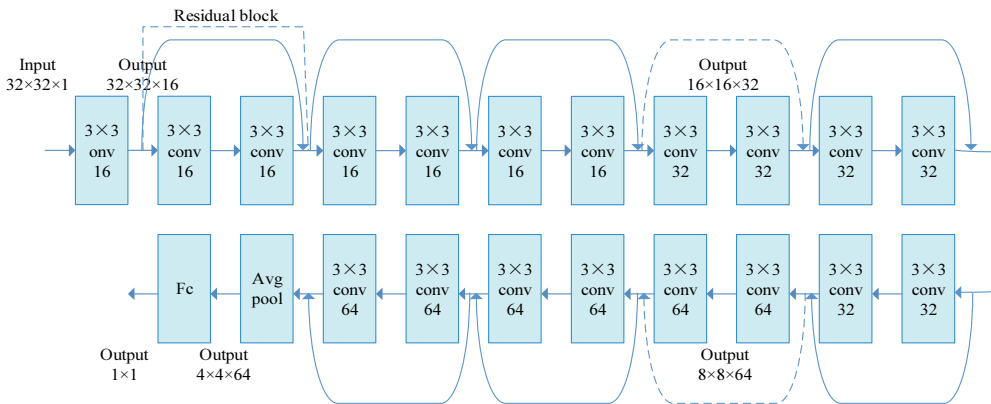

Fig. 1. Deep residual network model.

2.1.2 Network training

Model training is used to minimize the loss function. Therefore, the deep residual network takes the mean square error of the predicted quantization step and the tag quantization step as the loss function:

$$\ell(W) = \frac{1}{n} \sum_{i=1}^{n} \|f(W, x_i) - y_i\|^2, \tag{1}$$

where n is the number of training images, W is the weight parameter set, x_i is the ith image, y_i is the quantization step of the corresponding x_i. If $f(W, x_i)$ is the quantization step size of the predicted i^{th} image.

The experiment employs 200 grayscale images from the USC-SIPI image database [9], with an image size is 512×512. Split the test image into 32×32 image blocks which are used as the data set, 40,000

image blocks are selected as the training set, and 11,200 image blocks are selected as the test set. In training, the random gradient descent method is used to minimize the loss function. After each convolution and before the activation function, batch normalization is employed. The batch size is set to 64, i.e., 64 images, and corresponding label quantization steps are randomly selected from the sample image as a batch. The network has trained 200 epochs and the learning rate starts at 0.1. When the error stops, the learning rate is divided by 10.

2.2 Digital Fingerprint Generation using Perceptual Hashing

First, a digital fingerprint dependent on the user's identity is generated by employing perceptual image hashing [10]. The detailed process is described as follows:

Step 1: Check user identification via user login, and only authorized users can produce the digital fingerprint sequence.

Step 2: After user authentication, user identification data D is obtained.

$$D = \{d(i)|d(i) = 0,1\}, \ i = 1, \cdots, l_1 \tag{2}$$

Step 3: Read the host image I of size $m_1 \times n_1$ and the binary visible watermark image W of size $m_2 \times n_2$.

Step 4: Divide the host image I into sub-blocks of size $s \times s$, and calculate the average value of each sub-block. Let q_{1i} be the number of pixels larger than the average value in the ith sub-block and q_{2i} be the number of pixels smaller than the average value, then the identification bit of the ith sub-block can be obtained by

$$t_i = \begin{cases} 1 & if \ q_{1i} \geq q_{2i} \\ 0 & else \end{cases} \quad i = 1,2, \cdots, l_2 \tag{3}$$

where $l_2 = \frac{m_1 \times n_1}{s \times s}$.

Step 5: The identification bits of all sub-blocks form the digital image hash $G = \{g(i), i = 1, \cdots, l_2\}$. A binary sequence with the length $l = l_1 + l_2$ is generated by concatenating the image hash and the user identification data. The binary sequence is used as the digital fingerprint H and is registered in the user fingerprint database. It can be written as follows,

$$H = G||D = \{h(i), i = 1, \cdots, l\}, \tag{4}$$

where $||$ denotes the concatenating operator.

2.3 Visible Watermark Embedding based on a Chaotic Map

To enhance the security of visible watermark embedding, a chaotic map is used to modulate the visible watermark embedding strength. The specific steps are as follows.

Step 1: The region of interest [11] of the host image is extracted as the visible watermark embedding region.

Step 2: Resize the visible watermark V to the size of the specified embedded region to obtain the watermark signal V'. If the pixel value "1" in the binary watermark image represents the

background and "0" represents the foreground, the watermark embedding can only adjust the gray level of the pixel corresponding to the foreground of the watermark image. The watermark embedding strategy can be described as follows:

$$I_1(x,y) = \begin{cases} f(I(x,y)) & if\ v'(x,y) = 0 \\ I(x,y) & if\ v'(x,y) = 1 \end{cases},$$ (5)

where I_1 refers to the hidden image after embedding the visible watermark, and the embedding function f is a bijective grayscale mapping function. To ensure better watermark visibility, different embedding strategies are used in different brightness areas of the image. For any gray level t, function f is defined as follows:

$$f(t) = \begin{cases} t + 128 & if\ t < 128 \\ 255 - t & if\ t \ge 128 \end{cases}.$$ (6)

Step 3: The logistic map is a classical one-dimensional chaotic map [12], which is defined as follows.

$$x_{n+1} = \mu x_n(1 - x_n),$$ (7)

where μ is a control parameter, $\mu \in (0,4]$, $x_n \in (0, 1)$. When parameter $\mu > 3.57$, the system will be in a chaotic state.

where, the user *key* is used as the initial value, and the chaotic map is used to generate the pseudo-random sequence $R = \left\{r(x,y) \middle| r(x,y) \in \left[0.85\ ,\ 0.98\right]\right\}$, $x = 1, \cdots, m_3, y = 1, \cdots, n_3$.

Step 4: Modulate the image pixels according to the pseudo-random sequence R. Then the coordinates of two vertices on the main diagonal of the designated visible watermark embedding region are losslessly hidden in the host image using the difference expansion method [11]. Therefore, the sego-image I_2 is produced.

$$I_2(x,y) = \begin{cases} I_1(x,y) \times r(x,y) & if\ v'(x,y) = 0 \\ I_1(x,y) & if\ v'(x,y) = 1 \end{cases},$$ (8)

2.4 Digital Fingerprint Insertion using ResNet

To track the violation of authorized users, a digital fingerprint is embedded in the non-watermark region of the host image after embedding the visible watermark. The detailed procedure for digital fingerprint insertion is as follows.

Step 1: The non-watermark region of the stego image is mapped into a one-dimensional sequence in row-scanning order, and l pixel values are selected pseudo-randomly to form a one-dimensional vector $P = \{p(i), i = 1, \cdots, l\}$.

Step 2: Calculate the adaptive quantization step of the digital fingerprint using the deep residual network. The deep residual network is used to extract the characteristics of image brightness, edge, and texture sensitivity, and to train the relationship model between the image and digital fingerprint quantization step. The quantization step Δ is obtained using the trained deep residual network to predict the image.

Step 3: Insert the digital fingerprint $h(i)$ by quantizing pixel values $p(i)$ and obtain the final stego image I_3:

$$p'(i) = \begin{cases} \left[\frac{p(i)}{\Delta(i)}\right] \times \Delta(i) + \frac{\Delta(i)}{4} & if\ h(i) = 0 \\ \left[\frac{p(i)}{\Delta(i)}\right] \times \Delta(i) + \frac{3}{4}\Delta(i) & if\ h(i) = 1 \end{cases}, \tag{9}$$

where $[\bullet]$ denotes the rounding operator.

2.5 Visible Watermark Removal, Image Authentication, and Traitor Tracking

2.5.1 Visible watermark removal

Watermark removal is the reverse process of watermark embedding. The detailed process is as follows.

Step 1: Read the stego image I_3 and the binary visible watermark image V.

Step 2: The vertex coordinates of the main diagonal of the watermarking region are restored losslessly.

Step 3: Resize the visible watermark V to the size of the watermarking region and obtain the watermark signal V'. A pseudo-random sequence R is generated by the user key, and then the pixel values of the watermarking region are recovered according to the watermark signal V'.

$$I_1(x, y) = \begin{cases} \frac{I_2(x,y)}{r(x,y)} & if\ v'(x, y) = 0 \\ I_2(x, y) & if\ v'(x, y) = 1 \end{cases}. \tag{10}$$

We know that the stego image is the image containing the digital fingerprint. Because the digital fingerprint and the visible watermark are embedded in different regions, respectively, the pixel values are consistent for the stego image I_3 and I_2 in the watermark regions. At the same time, users without the correct keys will not be able to effectively remove the visible watermark.

Step 4: Because function f is a bijective function, the original pixel value in the watermark region can be recovered by

$$I(x, y) = \begin{cases} f^{-1}(I_1(x, y)) & if\ v'(x, y) = 0 \\ I_1(x, y) & if\ v'(x, y) = 1 \end{cases}, \tag{11}$$

where

$$f^{-1}(t) = \begin{cases} 255 - t & if\ t < 128 \\ t - 128 & if\ t \geq 128 \end{cases}. \tag{12}$$

2.5.2 Image authentication and traitor tracking

During the image authentication and violation tracking process, we first produce the perceptual hash G and extract the perceptual image hash G' from the stego image. The similarity discrimination of two hash codes is used for content authentication and traitor tracking.

Step 1: According to the same method for the digital fingerprint embedding process, construct image hash G according to the digital image with the visible watermark removed.

Step 2: Pseudo-randomly select l pixel values in the fingerprint embedding region of the stego image to generate a one-dimensional vector P', and calculate the adaptive quantization step $\Delta(i)$ using the trained deep residual network.

Step 3: Extract the digital fingerprint H' from the pixel sequence P'.

Step 4: Separate the user information D' and the perceptual image hash G' from the digital fingerprint.

Step 5: The user's illegal behavior will be tracked through a consistent comparison between the user information D' and the user information D registered in the user's digital fingerprint database.

Step 6: Complete image content authentication by comparing the similarity between the produced image hash G and the extracted image hash G'.
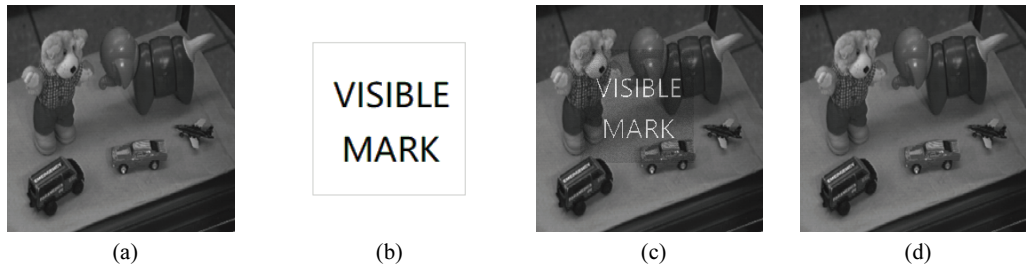

Fig. 2. An example of multipurpose watermarking: (a) test image 'Toys,' (b) binary watermark, (c) stego-image, and (d) recovered image.

3. Experiments and Analysis

In the experiment, some gray-level images of size 256×256 are used as the host images, which can be obtained from the USC-SIPI image database [9], and binary images of the same size are used as the visible watermarks. Fig. 2 shows an example of visible watermark embedding. The adaptive presentation of visible watermarks does not significantly change the visual quality of the host images. At the same time, the visible watermark in the stego images has a high definition, which can effectively announce the copyright of the host image.

3.1 Transparency and Visibility

In addition to subjective evaluation of the visual effect of stego images (shown in Fig. 2), we also tested the transparency of different types of images under different sizes of visible watermark embedding. Fig. 3 lists the peak signal-to-noise ratio (PSNR) values of different types of stego images. It can be seen that the visual quality of stego images decreases with the visible watermark size. For different types of test images, the average PSNR (computed using Eq. 13) value is more than 24.30 dB when the watermark size is greater than 128×128. The test results show that our multipurpose watermarking scheme (MWR) has a good visual effect on different types of test images, and the translucent presentation of the visible watermark does not affect the user's visual browsing of images.

$$PSNR = 10 log_{10}\left(\frac{255^2 \times m_1 \times n_1}{\sum_1^{m_1}\sum_1^{n_1}[I(x,y)-I'(x,y)]^2}\right). \tag{13}$$

To further test the visibility of the watermark embedding, the difference images between the stego image and the original image are shown in Fig. 4. From Fig. 4, the visible watermarks can be identified. This indicates that the inserted visible watermarks have high visibility and can serve as a copyright notice.
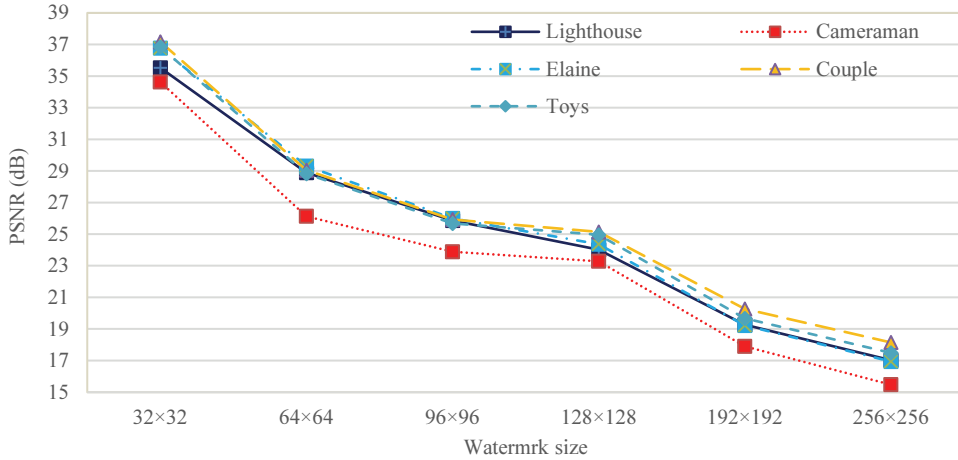
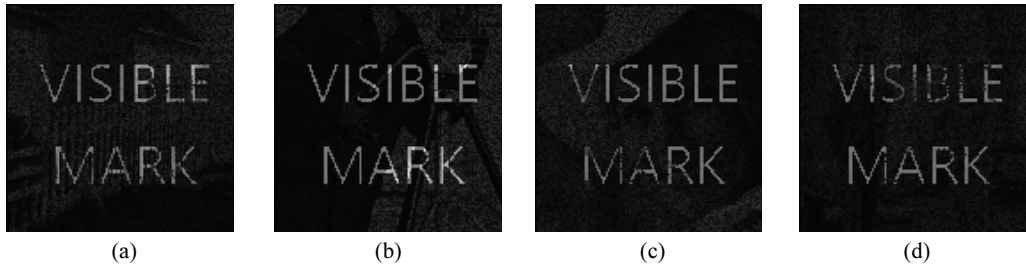**Fig. 3.** PSNR values for different types of stego images.

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

Fig. 4. Difference images before and after watermark embedding: (a) Lighthouse, (b) Cameraman, (c) Elaine, and (d) Couple.

3.2 Reversibility of Visible Watermarking

The example in Fig. 2 shows the watermark reversibility. The experimental principle of our scheme theoretically guarantees the reversibility of visible watermarking.

Proof. Note that the visible watermark and the digital fingerprint are embedded in different regions of the host image I, respectively, we can suppose that U is the visible watermarking region, whereas $\bar{U}$ is the digital fingerprint region, and u is any pixel in U, that is to say $I = U + \bar{U}$. To verify the reversibility of visible watermarks. We only consider workspace U.

(1) Watermark embedding process. The watermarking region U_1 containing the visible watermark can be obtained after the insertion of the visible watermark using Eq. (5). For any pixel $u_1 \in U_1$, we have

$$u_1 = \begin{cases} u + 128 & u < 128 \\ 255 - u & else \end{cases}. \tag{14}$$

To enhance security, the pseudo-random sequence R generated by the key is used to modulate pixels in the region U_1 as shown in Eq. (10), subsequently, the watermarked region U_2 is obtained.

(2) Watermark removal process. Given the watermarked region U_2, the authorized user can also generate the same pseudo-random sequence R using the user secret key. Because Eq. (8) is reversible, U_1 can be further recovered from the watermarked region U_2 according to Eq. (10). The embedding strategy shown in Eq. (6) makes the pixels smaller than 128 become pixels larger

than 128, and the pixels larger than 128 are mapped to pixels smaller than 128, so the original image watermarking region U can be recovered from U_1 according to Eq. (15), i.e.,

$$u = \begin{cases} u_1 - 128 & u \geq 128 \\ u_1 + 128 & else \end{cases}. \tag{15}$$

Therefore, the proposed MWR scheme is reversible.

3.3 Security

In the process of visible watermarking, the scheme uses the key to chaotically modulate the pixel in the watermarking region of the host image. Users without the correct key cannot effectively remove the visible watermark from the stego image. Therefore, the visible watermark can be used to declare copyright. The user key is a 15-bit precision floating-point number, so the user key space is 10^{15}.

Yang et al. [13] proposed a removable visible watermarking algorithm in the DCT domain based on a chaotic map. To test the superiority of the proposed MWR algorithm, we compared the watermark removal performance of the MWR algorithm with that of Yang et al.'s scheme. In the comparison test, we set the size of the visible watermark as 64×64. More test results of watermark removal for different images are listed in Table 1. The average PSNR value of recovered images by our MWR is as high as 69.09 dB, which is much higher than the 54.22 dB in [13] during the legal watermark removal process. When the watermark is removed illegally, the PSNR value of the recovered image by the MWR algorithm is only 24.71 dB on average. At this time, the average PSNR value of the recovered image by illegal removal using Yang et al.'s method is 28.25 dB. the average PSNR value of the recovered image by illegal removal using Yang et al.' This indicates that the proposed MWR algorithm has better security than Yang et al.'s algorithm. Unauthorized users with the wrong key cannot remove the visible watermark in the stego-image.

Table 1. Performance comparison of watermark removal

Images	Legal removal (dB)		Illegal removal (dB)	
	MWR	Yang et al. [13]	MWR	Yang et al. [13]
Lighthouse	69.46	55.06	24.16	26.77
Cameraman	69.18	53.71	23.04	26.23
Elaine	69.25	54.82	25.78	29.38
Couple	69.27	54.90	26.01	29.08
Lena	68.81	53.67	23.36	29.42
Boat	68.89	54.24	25.00	28.12
Goldhill	68.78	53.13	25.59	28.78
Average	69.09	54.22	24.71	28.25

4. Conclusion

In this paper, a new multipurpose image watermarking scheme is proposed using digital fingerprints and removable visible watermarking techniques. The generated stego image has a good visual effect, and the visible watermark presented in the stego image has a high definition. The embedding of the removable visible watermark controlled by a user key makes it impossible for unauthorized users to remove visible

watermarks, thus providing an effective explicit copyright notice for digital images. Authorized users can effectively remove visible watermarks to achieve high-definition browsing and the use of digital images. The quantization step of the digital fingerprint is computed using ResNet, which makes the embedded digital fingerprint highly robust and helps achieve image authentication and track violation behaviors of authorized users. Experimental results show that our scheme is effective and achieves a good balance between robustness, transparency, and security.

Acknowledgement

References

[1] P. W. Wong and N. Memon, "Secret and public key image watermarking schemes for image authentication and ownership verification," *IEEE transactions on image processing*, vol. 10, no. 10, pp. 1593-1601, 2001. https://doi.org/10.1109/83.951543

[2] F. Y. Shih and S. Y. Wu, "Combinational image watermarking in the spatial and frequency domains," *Pattern Recognition*, vol. 36, no. 4, pp. 969-975, 2003. https://doi.org/10.1016/S0031-3203(02)00122-X

[3] S. Voloshynovskiy, S. Pereira, V. Iquise, and T. Pun, "Attack modelling: towards a second generation watermarking benchmark," *Signal Processing*, vol. 81, no. 6, pp. 1177-1214, 2001. https://doi.org/10.1016/S0165-1684(01)00039-1

[4] C. S. Lu, H. Y. M. Liao, and L. H. Chen, "Multipurpose audio watermarking," in *Proceedings 15th International Conference on Pattern Recognition (ICPR)*, Barcelona, Spain, 2000, pp. 282-285. https://doi.org/10.1109/ICPR.2000.903540

[5] B. Lei and I. Y. Soon, "A multipurpose audio watermarking algorithm with synchronization and encryption," *Journal of Zhejiang University (SCIENCE C)*, vol. 13, pp. 11-19, 2012. https://doi.org/10.1631/jzus.C1100085

[6] Y. Peng, H. Lan, M. Yue, and Y. Xue, "Multipurpose watermarking for vector map protection and authentication," *Multimedia Tools and Applications*, vol. 77, pp. 7239-7259, 2018. https://doi.org/10.1007/s11042-017-4631-z

[7] S. Sheidani and Z. Eslami, "Blind multipurpose watermarking with insertion of a single watermark: a generic construction based on verifiable threshold secret sharing," *IET Image Processing*, vol. 14, no. 17, pp. 4766-4773, 2020. https://doi.org/10.1049/iet-ipr.2019.1576

[8] F. Chaabane, M. Charfeddine, W. Puech, and C. B. Amar, "A two-stage traitor tracing scheme for hierarchical fingerprints," *Multimedia Tools and Applications*, vol. 76, pp. 14405-14435, 2017. https://doi.org/10.1007/s11042-016-3749-8

[9] University of Southern California, "The USC-SIPI image database," c2023 [Online]. Available: https://sipi.usc.edu/database/.

[10] H. Yang, J. Yin, and Y. Yang, "Robust image hashing scheme based on low-rank decomposition and path integral LBP," *IEEE Access*, vol. 7, pp. 51656-51664, 2019. https://doi.org/10.1109/ACCESS.2019.2911207

[11] L. Zhang and H. Wu, "Cosaliency detection and region-of-interest extraction via manifold ranking and MRF in remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1-17, 2022. https://doi.org/10.1109/TGRS.2021.3079441

[12] Z. Rahman, X. Yi, M. Billah, M. Sumi, and A. Anwar, "Enhancing AES using chaos and logistic map-based key generation technique for securing IoT-based smart home," *Electronics*, vol. 11, no. 7, article no. 1083, 2022. https://doi.org/10.3390/electronics11071083

[13] Y. Yang, X. Sun, H. Yang, and C. T. Li, "Removable visible image watermarking algorithm in the discrete cosine transform domain," *Journal of Electronic Imaging*, vol. 17, no. 3, article no. 033008, 2008. https://doi.org/10.1117/1.2952843

Mingfang Jiang https://orcid.org/0000-0003-3527-2113

She received an M.S. degree in Computer Application from Hunan University, China, in 2011. She is currently an associate professor at Hunan First Normal University, P.R. China. Her research interests include information security, information management, and multimedia signal processing.

Hengfu Yang https://orcid.org/0000-0002-8006-9715
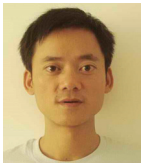
He received an M.S. degree in Computer Application from Guizhou University, China, in 2003, and a Ph.D. degree in Computer application from Hunan University, China, in 2009. He is currently a professor at the School of Computer Science, Hunan First Normal University, China. His research interests include information security, image processing, digital watermarking, and deep learning.