Robust Image Hashing based Efficient Authentication for Smart Industrial Environment

[1]Muhammad Sajjad, [2]Ijaz Ul Haq, *Student Member, IEEE*, [3]Jaime Lloret, *Senior Member, IEEE*, [4]Weiping Ding, *Senior Member, IEEE*, [5]Khan Muhammad, *Member, IEEE*

***Abstract*—Due to large volume and high variability of editing tools, protecting multimedia contents and ensuring their privacy and authenticity has become an increasingly important issue in cyber-physical security of industrial environments, especially industrial surveillance. The approaches authenticating images using their principle content emerge as popular authentication techniques in industrial video surveillance applications. But maintaining a good trade-off between perceptual robustness and discriminations is the key research challenge in image hashing approaches. In this paper, a robust image hashing method is proposed for efficient authentication of keyframes extracted from surveillance video data. A novel feature extraction strategy is employed in the proposed image hashing approach for authentication by extracting two important features: the positions of rich and non-zero low edge blocks and the dominant DCT coefficients of the corresponding rich edge blocks, keeping the computational cost at minimum. Extensive experiments conducted from different perspectives suggest that the proposed approach provides a trustworthy and secure way of multimedia data transmission over surveillance networks. Further, the results vindicate the suitability of our proposal for real-time authentication and embedded security in smart industrial applications compared to state-of-the-art methods.**

***Index Terms*—Industrial surveillance, digital authentication, image hashing, embedded security**

I. INTRODUCTION

MULTIMEDIA content is considered to be one of the best sources for the delivery of information in many practical applications such as healthcare and industrial surveillance [1]. But the existence of editing tools for overwhelming diffusion of these multimedia contents, protecting their integrity and authenticity, and security of the systems from various undesired manipulations, has become an increasingly important issue. There are several elements for security and privacy consideration in an industrial setup. Taking industrial video surveillance application as an example, a mechanism is required for the authentication and protection of multimedia data from tampering and distortion during the transmission via different networks. The involvement of resource-constrained devices in surveillance setup for using low bandwidth brought some additional challenges in terms of computation and storage for authentication operations. In short, due to network-connected technologies [2], expertise is easily available for attackers, making surveillance networks vulnerable to different attacks and threats. Therefore, security has become an important issue in industrial surveillance, which may cause a huge economical damage.

The existing literature for the authentication of digital images can be grouped into three major classes: image forensic based [3, 4], watermark based [5-8], and image hashing based techniques [9-13]. Image hashing technology has a key role in ensuring the security and privacy of information during transmission, which contribute to the overall cyber-physical security system for smart industrial environments. Besides these authentication schemes, many encryption schemes are also used for industrial surveillance data [14].

Recently, the approaches authenticating images using their principle content, known as image hashing [15-24], emerge as the popular authentication techniques in video surveillance applications. Using image hashing approaches, the invariant features of the original image on the sender side are extracted and then represented as a numeric value called a hash. This hash value is then sent with the original image to the destination. In the authentication phase at the destination, the same hash generation algorithm is used to compute a hash value. Both hash values are compared to check the authenticity and integrity of the received image. Typically, a good image hash [9] should be reasonably short, robust to ordinary image manipulations, and sensitive to tampering. It should be also unique in the sense that the produced hash is fully dependent on the actual contents of the image and produces the same hash values for images that have a similar appearance to the human eye.

In industrial surveillance it is very challenging to meet all the requirements simultaneously, especially on balancing perceptual robustness and sensitivity to discrimination in an image hashing method. In other word, the generation of a hash that is independent of content-preserving manipulations like noise addition, JPEG compression and scaling, is a very difficult task. Also, an image hashing approach must be able to differentiate between content-preserving and content-degrading

Manuscript received March 19, 2019; Accepted: XXX, Published: XXXX. This work is partially supported by Six talent peaks project of Jiangsu Province, China (XYDXXJS-048), sponsored by Qing Lan Project of Jiangsu Province, China. (*Corresponding author: Khan Muhammad and Weiping Dong.*)

Muhammad Sajjad is with the Department of Computer Science Islamia College Peshawar,25000, Pakistan. (Email: muhammad.sajjad@icp.edu.pk).

Ijaz Ul Haq is with Intelligent Media Laboratory, Digital Contents Research Institute, Sejong University, Seoul 143-747, Republic of Korea. (Email: ijaz000007@gamil.com).

Jaime Lloret is with Instituto de Investigacion para la Gestion Integrada de Zonas Costeras, Universitat Politecnica de Valencia, Spain (Email: jlloret@dcom.upv.es).

Weiping Ding is with School of Information Science and Technology, Nantong University, Nantong 226019, China (Email: ding.wp@ntu.edu.cn).

Khan Muhammad is with the Department of Software, Sejong University, Seoul 143-747, South Korea. (Email: khan.muhammad@ieee.org).

manipulation. This leads to the fact that the features extracted for hash generation must be robust in terms of content-preserving manipulation. On the other hand, selection of a suitable threshold for balancing perceptual robustness with discrimination is another challenge in an image hashing method. Technically, the same images with different size and compression level or having noise could have different digital representation. So, maintaining a good trade-off between perceptual robustness and discriminations is the key research challenge.

In the present paper, we aim at proposing a new robust and secure image hashing based authentication method using keyframes extraction in industrial surveillance application. The objective is to provide a reasonably robust image hash with good perceptual robustness and sensitivity to discrimination. We use the keyframes extracted from industrial CCTV surveillance video data and convert it into standard size to ensure a fixed length for the hash. An edge detector is applied to the luminance component of the keyframe to obtain the real edges map. The keyframe is then divided into non-overlapping blocks of equal size, followed by their categorization into rich information and low information blocks, based on the edges map. The final hash is constructed from the dominant DCT coefficients of the rich information blocks and the difference in the positions of the corresponding rich and low information blocks in the sequence, which can be used for authentication. In addition to hash generation, our scheme automatically generates a key from the extracted features for encryption, considering its emphasis on the runtime environment for industrial video surveillance. Our key contributions are highlighted as follows:

- A novel three-fold robust and secure image hashing based authentication approach using salient structural feature for smart industrial surveillance applications is proposed. This authentication approach will guarantee the secure transmission of representative frames in industrial surveillance with interconnected vision sensors in networks.
- The edge detection mechanism in our proposed scheme generates a grayscale edge image from eight binary maps of different thresholds using Canny operator, in which edges are classified into strong and weak edges. By using this new strategy, content-preserving modified image by noise and compression will not have any effect on edges, and false edges can be neglected very easily, which increase the accuracy of detecting rich information blocks.
- The experimental results prove that the proposed solution can effectively improve the perceptual robustness with discrimination over existing state-of-the-art methods [15-18]. Our approach can play a vital role in cyber-physical security for image and video data authentication during transmission in industrial surveillance networks.

The rest of this paper is organized as follows. Section II and III present the proposed methodology and experimental results, respectively. Conclusions and future research directions are given in Section IV.

II. The Proposed Frame work

The proposed scheme for image authentication in industrial surveillance consists of three steps: 1) pre-processing, 2) feature extraction, and 3) hash/key generation, and these are briefly discussed separately in the subsequent sections. Fig. 1 illustrates the proposed hashing scheme diagrammatically.

2.1 Pre-processing

To generate a uniform size for the hash value from images with different dimensions, the input image I_o is first resized to M×M pixels using bilinear interpolation, which is computationally efficient and has higher quality index score than other interpolation techniques [25], resulting in a standard-sized image I_o'. Next, a Gaussian low-pass filter is applied to the resized image I_o' as a convolution mask, to avoid the effect of noise on the final hash value using Eq. (1) as follows:

$$G(i,j)=\frac{1}{\sum_{i=1}^{m}\sum_{j=0}^{n}g(i,j)}\cdot e^{\frac{-(i^2+j^2)}{2\delta^2}} \tag{1}$$

where δ is the standard deviation of the distribution for coordinates (i, j). After this, the noise-free image is transformed from the RGB to the HSV color model. As the luminance information of HSV color space contains rich information about structural and geometric features, it is retrieved as a secondary image I

2.2 Feature Extraction

The feature extraction phase involves the conversion of the secondary image I into the salient edge structure and the retrieval of the positions of rich and low information blocks through edge detection and selective sampling. These steps are explained in subsections A and B.

A. Salient Edge Detection

In our proposed scheme, a double-threshold edge detector known as a Canny operator is used for edge detection. The Canny operator is applied to the secondary image I, through which a binary map B is obtained. The Canny operator uses the magnitude of the intensity gradient values handled by two threshold values, i.e., the low threshold value $\boldsymbol{T_L}$ and the high threshold value $\boldsymbol{T_H}$. The edge pixel gradient values are divided into three categories based on the aforementioned thresholds as follows: i) edge pixels with a gradient greater than $\boldsymbol{T_H}$ are marked as strong pixels, representing the low edge sensitivity; ii) edge pixels with a gradient less than $\boldsymbol{T_H}$ but greater than $\boldsymbol{T_L}$ are marked as weak edge pixels, which represent high edge sensitivity; and iii) edge pixels with a gradient smaller than $\boldsymbol{T_L}$ are neglected. A complete edge image is then obtained by connecting strong pixels with weak pixels to avoid false edges [26, 27]. The selection of suitable thresholds for edge detection is a complex task since low threshold values are sensitive to false edges detection and do not work for noisy images. On the other hand, the selection of high threshold values results in some true edges being neglected. The image hashing schemes

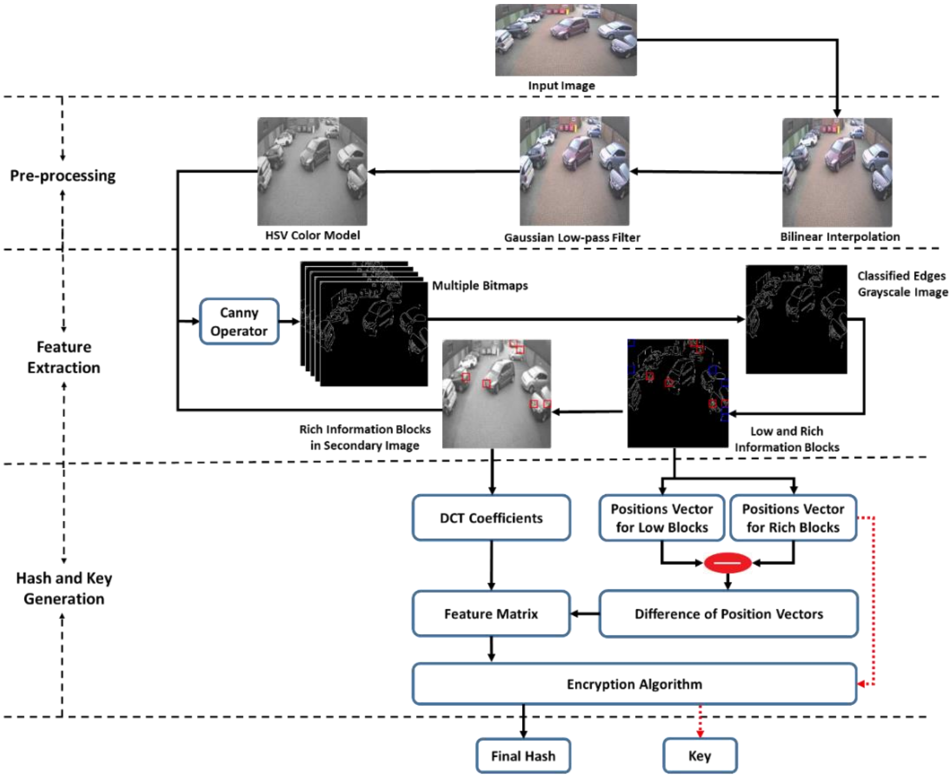

Fig. 1: Framework of our proposed system

given in [28, 29] also use the Canny operator for edge detection, but use a single pair of thresholds; due to this, the positions of rich information blocks are not the same for a given input image and its content-preserving modified image.

To avoid this problem, we introduce a new concept using eight threshold values for edge detection; from these, we obtain eight different binary maps, and we then construct a single edge grayscale image E_{img} from these eight binary maps, representing the classified edges of the secondary image I using Eq. 2.

$$E_{img}(i,j) = 32 \times \sum_{i=0}^{m}\sum_{j=0}^{m}\left[B_{T1}(i,j) + B_{T2}(i,j) + ... + B_{T8}(i,j)\right] \quad (2)$$

where B_{T1}, B_{T2}, …, B_{T8} are binary maps of the secondary image I constructed by the Canny operator using thresholds $T1$, $T2$.. …, $T8$. The mechanism for constructing the classified edges grayscale image E_{img} is described as follows. The presence of the edge pixel at the same position in all the eight binary edges maps is first counted; this total is then multiplied by 32 to obtain a single pixel value for the grayscale edge image using raster scanning. Hence, an edge pixel that is present in all eight binary maps results in a dark edge pixel in the grayscale edge image and is referred as a strong or true edge pixel. In this way, edges are classified into eight different bands according to their presence in the binary maps. Thus, in the grayscale edge image E_{img}, all non-zero pixels indicate classified edges of the contents of the secondary image I, while zero pixels indicate the background of the secondary image I. An example of the construction of the binary maps and grayscale edge image E_{img} from the secondary image I and the binary maps, respectively, is shown in Fig. 2. The overall mechanism is given in Algorithm 1.

B. Selective Sampling of Rich and Low Information Blocks

In our proposed scheme, we not only concentrate on the salient region but also protect non-salient regions. For this purpose, we divide the secondary image I and the grayscale edge image E_{img} into non-overlapping blocks of equal size.

Algorithm 1: Edge Grayscale Image Construction Algorithm

```
Inputs: Secondary image $I$, Threshold $T_n = [T_L, T_H]$ for Canny operator
1.  Set initial threshold $T_I$ as $T_L=0.1$ and $T_H=0.3$
2.  For i=1 to 8
        Apply Canny operator on $I$ with threshold $T_i$ to obtain binary map $B_i$
        Increment $T_L$ by 0.2
        Increment $T_H$ by 0.03
3.  Combine all 8 binary maps using equation (2) to get edge grayscale image $E_{img}$
Output: Edge grayscale image $E_{img}$
```

The total number of blocks for an image of size M×M pixels and block size k×k pixels can be calculated using $(M/k)^2$. The block representation is given in Eqs. (3) and (4).

$$I = \begin{bmatrix} IB_{1,1} & \Lambda & IB_{1,M/k} \\ \mathrm{M} & \mathrm{O} & \mathrm{M} \\ IB_{M/k,1} & \Lambda & IB_{M/k,M/k} \end{bmatrix} \quad (3)$$

$$E_{img} = \begin{bmatrix} EB_{1,1} & \Lambda & EB_{1,M/k} \\ \mathrm{M} & \mathrm{O} & \mathrm{M} \\ EB_{M/k,1} & \Lambda & EB_{M/k,M/k} \end{bmatrix} \quad (4)$$

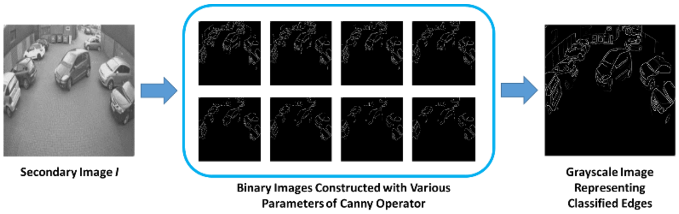
Secondary Image I
Binary Images Constructed with Various Parameters of Canny Operator
Grayscale Image Representing Classified Edges

Fig. 2: Illustration of salient edge detection using a Canny operator (with eight different thresholds) and construction of grayscale edge image.

To categorize the $EB_{i,j}$ blocks of E_{img} into salient and non-salient edge information blocks, we sum all the pixel values of each $EB_{i,j}$ block using the raster-scanning order of the blocks, to get the structure information value $S_{i,j}$ using Eq. (5).

$$S_{i,j} = \varphi\left(EB_{i,j}\right) = \sum_{i=1}^{k}\sum_{j=1}^{k} p_{i,j} \tag{5}$$

where φ (.) is a function which sums all the pixel values in the $EB_{i,j}$ block, and $\mathcal{P}_{i,j}$ refers to the pixel values of the $EB_{i,j}$ block. In this way, we get a salient edge structure information matrix S_{info} of size $M/k \times M/k$, as given in Eq. (6).

$$S_{\text{info}} = \begin{bmatrix} S_{1,1} & \Lambda & S_{1,M/k} \\ \text{M} & \text{O} & \text{M} \\ S_{M/k,1} & \Lambda & S_{M/k,M/k} \end{bmatrix} \tag{6}$$

A high value of $S_{i,j}$ represents rich information of the edges that exist in the corresponding $EB_{i,j}$ block, while a low value refers to low information; these blocks must be non-zero. Then, a number N of corresponding rich information blocks $IB_{i,j}$ are sampled in the secondary image I on the basis of the high values of $S_{i,j}$ in the edge structure information matrix S_{info}, which contain richer structural information. Fig. 3 shows selective sampling using the grayscale edge image E_{img} taken from CCTV surveillance. Fig. 3(a) is the classified edge image E_{img}, illustrating the salient edges. In Fig. 3(b), the red and blue squares indicate the rich and low edge information blocks $EB_{i,j}$, respectively, based on the six largest and six lowest non-zero values of $S_{i,j}$ in the S_{info} matrix (N = 6). In Fig. 3(c), six corresponding rich information blocks $IB_{i,j}$ are sampled in the secondary image I.

In order to get N rich and low information blocks, we re-arrange all the non-zero edge blocks $IB_{i,j}$ of the secondary image I in descending order on the basis of the edge information matrix S_{info}, as Di (i = 1, 2, …, v). Here, v is the total number of non-zero edge blocks (i.e., v = total blocks – no. of zero edge blocks) and a smaller value of i indicates a richer information block in Di. The first and last N blocks in D represent the richest and lowest edge information blocks, respectively. Next, the positions of the corresponding $\boldsymbol{N}$ sampled blocks in the secondary image I are retrieved. i.e., the abscissa i and the ordinate j of $IB_{i,j}$, symbolized by $p1$ and $p2$ respectively, to get the position matrices of the rich information blocks PR_{mat} and the low information blocks PL_{mat} using Eqs. (7) and (8).

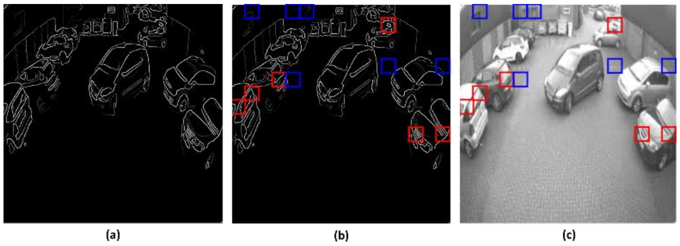

Fig. 3: Illustration of the concept of selective sampling using the grayscale edge image with N = 6: (a) grayscale edge image; (b) grayscale edge image with six rich and six low information blocks; (c) secondary image with corresponding sampled blocks. Red indicates rich information blocks, while blue color blocks contain low information

$$PR_{mat} = [pr_n]^t = [p1_n, p2_n] \quad \text{n = 1, 2, . . . } \boldsymbol{N} \tag{7}$$

$$PL_{mat} = [pl_n]^t = [p1_n, p2_n] \quad \text{n= v, v-1, . . . v-}\boldsymbol{N} \tag{8}$$

To extract further structure features from the secondary image I, we apply DCT to each of the N sampled rich information blocks $IB_{i,j}$ of size $k \times k$. As a result, we get two dominant DCT coefficients, as given in Eqs. (9) and (10).

$$Cv(1,2) = \frac{\sqrt{2}}{k} . \sum_{m=1}^{k}\sum_{n=1}^{k} Di(m,n)\cos\left[\frac{(2n+1)\pi}{2k}\right] \tag{9}$$

$$Cv(2,1) = \frac{\sqrt{2}}{k} . \sum_{m=1}^{k}\sum_{n=1}^{k} Di(m,n)\cos\left[\frac{(2m+1)\pi}{2k}\right] \tag{10}$$

Here Cv denotes the DCT coefficient matrix of size k×k for the block Di. Since the elements in the top left-hand corner of the DCT matrix represent high information for the whole matrix, we therefore retrieve only the second coefficient in the first row Cv(1, 2) and the second coefficient of the first column Cv(2, 1) as $qv1$ and $qv2$, respectively, to get a DCT coefficient matrix Q of size 2×N, as given in Eq. (11).

$$Q = Q_v = [q1_v, q2_v] = [C_v(1,2), C_v(2,1)] \quad \text{v= 1, 2, . ., N} \tag{11}$$

2.3 Hash and Key Generation

As a result of the previous steps, we get three matrices, i.e., PL_{mat}, PR_{mat} and Q, which are used in the generation of a key and a hash value. In our proposed scheme, the key is generated automatically, since our algorithm is designed for real-time industrial surveillance systems with minimum human intervention. We use a novel approach for key generation using the position information of the rich edge blocks PR_{mat}, which is used to encrypt the final hash as well as provide feature information in the authentication phase. The flowchart for key and hash generation from these matrices is illustrated in Fig. 4. Firstly, the position matrix of the rich edge information PR_{mat} is subtracted from the position matrix of low edge information matrix PL_{mat}. The resultant position difference matrix PD_{mat} is then concatenated with the DCT coefficient matrix Q to get a feature matrix F of size $N{\times}4$ using Eq. (12). Next, the feature matrix F is converted into a one-dimensional array. In Eq. (13) h_k denotes the k-th element of feature matrix F, which is traversed in raster scanning order (k=1,2, . . ,4N) to get an intermediate hash H' of length 4N, as given in Eq. (13).

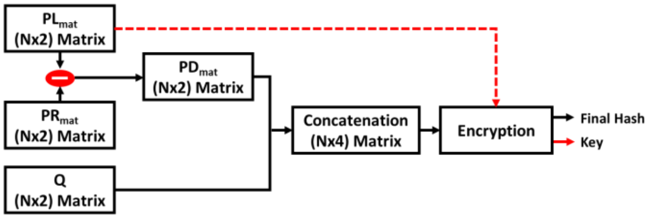

Fig. 4: Flowchart for key and hash generation

$$F=\left[PD_{mat},Q\right] \tag{12}$$

$$H'=\left[h_1,h_2,....h_{4N}\right] \tag{13}$$

To generate the key, the position information matrix of the rich information blocks PR_{mat} is converted into a one- dimensional array and is then encrypted to increase the security of the overall framework. Next, we encrypt the intermediate hash H' using the generated key to get the final encrypted hash H, using Eq. (14).

$$H=\psi\left(H',K\right) \tag{14}$$

Here, Ψ is a function that encrypts the intermediate hash H' using key K. The encryption algorithm encrypts the position information matrix of the rich information blocks PR_{mat} to obtain a key, and then encrypts the intermediate hash H' to increase the security and robustness of our method. Finally, we get a key K and encrypted hash H. The generated key is sufficiently sensitive that any minor alteration in the key leads to completely different hash; this is an important distinguishing characteristic of our algorithm. The encryption method is given in Algorithm 2.

In the authentication phase, the key is used to encrypt the hash generated from the image under consideration, for comparison with the received hash. The key is first decrypted using the decryption algorithm, and details of the rich information blocks are extracted for hash generation. This reduces the computational process of sampling the rich information blocks since the positions have already been extracted from the key. This novel mechanism increases the suitability of the proposed framework for deployment in real CCTV surveillance systems for authentication.

III. Experimental Results

We conducted various experiments to evaluate the performance of our method and achieve the required trade-off between perceptual robustness and discrimination. In our experiments, bilinear interpolation was used to resize the input image to *512×512* pixels, i.e., the parameters are set to *M=512* and the block size of $EB_{i,j}$ and $IB_{i,j}$ is set to *32×32* pixels with *k=32*. The number of selective blocks $IB_{i,j}$ of the secondary image I is set to *N=25*. Hence, the final hash of an input image consists of *4×25= 100* decimal numbers, i.e., two position vector differences and two dominant DCT components. The thresholds for the Canny operator are set to 0.1and 0.3 for low threshold and high threshold to construct the first binary image. The interval of 0.02 for low threshold and 0.03 for high threshold is set for the next seven binary images. Further details of the conducted experiments are given in the subsequent sections.

Algorithm 2: Encryption Algorithm

```
Inputs: Position information matrix of rich information blocks $PR_{mat}$, intermediate hash $H'$
1.  Convert $PR_{mat}$ into one-dimensional array.
2.  Convert $PR_{mat}$ and $H'$ into binary to get keybits $K_{bits}$ and hashbits $H_{bits}$, respectively.
3.  Perform bitxor operation of $K_{bits}$ and logical 1.
4.  Shuffle $K_{bits}$ by replacing first four bits with last four bits in every eight-bit combination.
5.  Perform bitxor operation of $H_{bits}$ and logical 1 using $K_{bits}$
        If $K_{bits}$ (i) equals 1,
            Perform bitxor($H_{bits}$(i), logical 1)
        End
6.  Convert $K_{bits}$ and $H_{bits}$ into decimal to get key $K$ and final hash $H$.
Output: key $K$ and final hash $H$
```

3.1 Hash Similarity Measurements

In the literature, many typical measurements have been reported for comparing the similarity of hash values, such as Euclidean distance, L2-norm, Hamming distance, and cosine similarity. In our experiments, we used the correlation coefficient (CC) to measure the similarity of the hashes of two images. The motivational reason for this is its widespread usage in measuring the linear correlation of two variables. Furthermore, it can identify a statistical relationship between two random variables or observed datasets [23,24]. We assume that $H^{(1)}$ = [h1$^{(1)}$, h2$^{(1)}$, … , h$q^{(1)}$] and $H^{(2)}$= [h1$^{(2)}$, h2$^{(2)}$, … , h$q^{(2)}$] are two hashes of length q for two input images. Then, the CC between $H^{(1)}$ and $H^{(2)}$ can be calculated using Eqs. (15) and (16) as follows.

$$\phi\left(H^{(1)},H^{(2)}\right)=\frac{\sum_{i=1}^{l}\left(h_i^{(1)}-\mu_1\right)\left(h_i^{(2)}-\mu_2\right)}{\sqrt{\sum_{i=1}^{l}\left(h_i^{(1)}-\mu_1\right)\sum_{i=1}^{l}\left(h_i^{(2)}-\mu_2\right)+\xi}}\ ,\in\ [-1,1] \tag{15}$$

$$\mu_{i,j}=\frac{\sum_{i=1}^{q}h_i^{(j)}}{q},\quad j=1,2 \tag{16}$$

where $\boldsymbol{\phi}$ is a function providing the CC between $H^{(1)}$ and $H^{(2)}$ and $\boldsymbol{\mu}1$ and $\boldsymbol{\mu}2$ are their mean values individually. ξ is a constant nearly equal to zero, maintaining the denominator in Eq. (15) as not equal to zero. The range of $\boldsymbol{\phi}$ lies between −1 and 1. For the interpretation of images, the CC score is treated as follows: a CC score greater than a pre-determined threshold T indicates that the two images are visually identical, while if the score is less than the threshold T, the images under consideration are treated as different or modified.

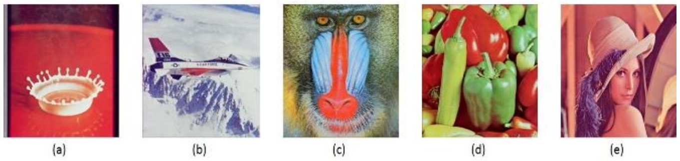

Fig. 5: Five test images from the USC-SIPI data set: (a) Splash (b) Airplane (c) Baboon (d) Peppers (e) Lena

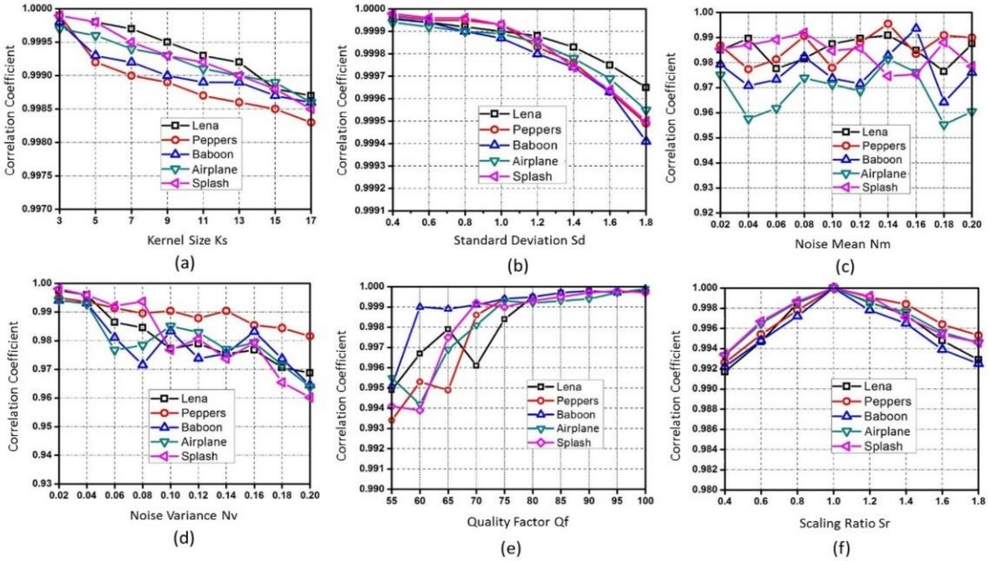

Fig. 6: Perceptual robustness test against different image manipulations: (a) average filtering; (b) Gaussian filtering; (c) Gaussian noise; (d) speckle noise; (e) JPEG compression; (f) scaling

3.2 Performance Evaluation

A. Perceptual Robustness

The robustness of an image hashing method is measured by its ability to withstand different attacks, such as noise, compression and blurring. A perceptual robustness test is therefore conducted to evaluate the perceptual robustness of our scheme in this section. The experiments are conducted using a standard dataset (USC-SIPI Image Database Volume 3: Miscellaneous) of standard color images. Overall, 20 images were selected for testing, of which five are given in Fig. 5. Six different image manipulation techniques were applied to these 20 images to obtain 1120 modified images for testing. The details of these six manipulation techniques with their parameters are given in Table II.

To collect the results, we generate hash values for all 20 test images and their manipulated versions obtained using the six manipulation techniques listed in Table I. Then, the similarity between the original image hash and its corresponding processed image hash is judged using the CC based on Eq. (15). For ease of understanding, we present the results of this perceptual robustness test in two ways: i) the results of five standard test images of Fig. 5 are shown graphically in Fig. 6; and ii) the overall results for the 20 test images are given in Table II. In each sub-graph of Fig. 6, each single point represents the CC of the original image hash and the corresponding processed version by the parameter given on the abscissa. The ordinate in the graphs shows the range of the CC. In Table II, the results are displayed as the maximum, minimum and mean CCs between the original image hash and the processed version image hash for all 20 test images. It is notable that the maximum value of all the CCs for any image manipulation is closely equal to 1.00, the minimum mean value of any image manipulation is 0.95, and all CCs are greater than 0.85 for any image manipulation. This proves the robustness of our method against various image processing operations such as average and Gaussian filtering, Gaussian and speckle noise, JPEG compression and scaling. Our results also suggest that a suitable threshold *T* can be chosen as 0.85 for a CC, based on which we can classify visually similar and visually different images as authentic and non-authentic, respectively, in surveillance applications. Fig. 7 demonstrates video frames obtained from industrial surveillance and their manipulated versions with detected rich information blocks.

TABLE I
TYPES AND PARAMETERS OF IMAGE MANIPULATIONS

Manipulation	Description	Parameters
Average filter	Kernel size K_S	3, 5, ...,17
Gaussian filter	Standard deviation S_d	0.4, 0.6, ...,1.8
Gaussian noise	Noise mean N_m	0.02, 0.04 . . ., 0.2
Speckle noise	Noise variance Nv	0.02, 0.04 . . ., 0.2
JPEG compression	Quality factor Q_f	55, 60 . . . . ., 100
Scaling	Scaling ratio S_r	0.2, 0.4 . . . . ., 2.0

B. Hash Discrimination Test

To prove that our proposed scheme produces significantly different image hashes for two visually distinct images, a test is conducted called a hash discrimination test. For this test, a dataset of 200 different images was used, which included 30 images collected by the authors using a digital camera, 70 from the UCID dataset, and 100 from the Video Surveillance Online Repository (VISOR) [30]. The size of images ranged from 256×256 to 1024×1024 pixels. Firstly, the hash value was calculated for all 200 images using the proposed scheme, and then the hash similarity was calculated between each hash pair, giving 19,900 results for correlation coefficients, as shown in Fig. 8. In Fig. 8, the abscissa presents the image pairs and the ordinate shows the values of the corresponding correlation coefficients. In the results, only three cases of correlation coefficients (i.e., 0.8023, 0.8191, and 0.8261) were observed

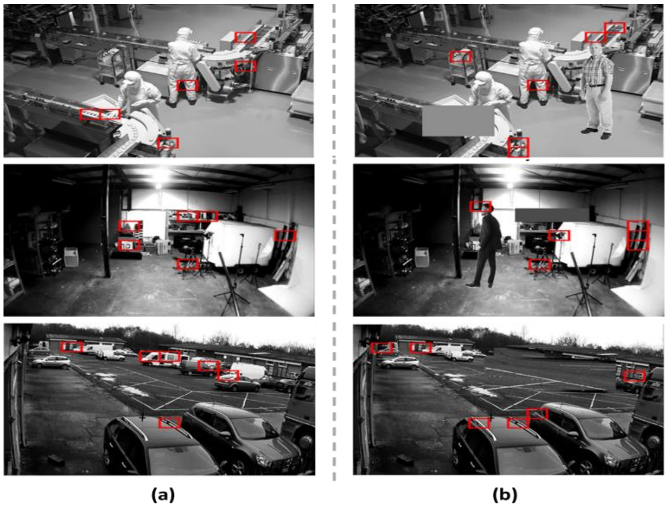

Fig. 7: Sample frames from industrial surveillance with six rich information blocks. (a) original frames, (b) manually manipulated frames. The difference can be noticed based on changed positions of rich information blocks.

that exceeded the limit of 0.8; however, these scores are still less than the predetermined threshold T=0.85. Hence, it is proved that our proposed scheme balances perceptual robustness and discrimination capability, since all CC values are less than the specified threshold T=0.85.

TABLE II
MEAN, MAX AND MIN CORRELATION COEFFICIENTS OF TWENTY IMAGES FOR DIFFERENT IMAGE PROCESSING OPERATIONS

Manipulation	Parameters	Correlation Coefficient		
		Mean	Max.	Min.
Average filter	Ks=3	0.9994	1.000	0.9991
Average filter	Ks=17	0.9947	1.000	0.9439
Gaussian filter	Sd = 0.4	1.0000	1.000	1.0000
Gaussian filter	Sd = 1.8	0.9997	1.000	0.9983
Gaussian noise	Nm = 0.04	0.9948	0.999	0.9521
Gaussian noise	Nm = 0.20	0.9878	0.999	0.8564
Speckle noise	Nv = 0.04	0.9968	0.999	0.9893
Speckle noise	Nv = 0.20	0.9521	0.998	0.8798
JPEG	Q_f=55	0.9999	1.000	0.9992
JPEG	Q_f=100	1.0000	1.000	1.0000
Scaling	Sr=0.4	1.0000	1.000	1.0000
Scaling	Sr =1.0	1.0000	1.000	1.0000

3.3 Performance Comparison with State-of-the-Art Methods

We compared our proposed method with four state-of-the-art techniques using the standard dataset images used in Section 3.2 (A). In Fig. 10, the six subfigures correspond to six different kinds of image manipulations, in which the abscissa of each subfigure presents the parameter value of applied manipulation and the ordinate shows the average score for 20 CCs between the original image and its processed version. It is clearly visible from Fig. 10 that the CC scores for our method are greater than those of the four competing schemes [15-18]. This dominance is due to the unique characteristics of our approach in terms of preprocessing, features extraction and hash generation. In addition to the above experiments, the computational complexity and hash length of our algorithm were compared with the schemes in [15-18]. The average time consumed by each hashing scheme was calculated using a computer with the following specifications: 2.40 GHz Intel i3 processor, 4.00 GB memory, and Windows 8.1 (64bit) operating system.

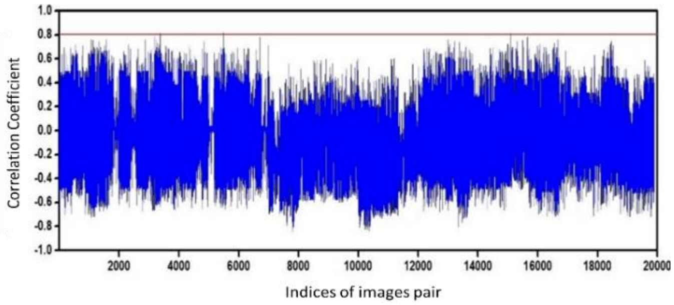

Fig. 8: Hash discrimination test based on 200 images

The implementation tool used was MATLAB 2016a. Time complexity and hash length comparisons are given in Table III, demonstrating the strength of our method in balancing perceptual robustness and discrimination with a reasonable hash length and fast execution speed. This makes the proposed system suitable for embedded devices, particularly, for the authentication of keyframes in industrial surveillance systems.

TABLE III
COMPARISON BASED ON HASH LENGTH AND AVERAGE RUNNING TIME

Schemes	Hash Length	Average Time (seconds)
Scheme in [15]	64	0.728
Scheme in [16]	180	0.749
Scheme in [17]	64	1.125
Scheme in [18]	84	1.4
Proposed scheme	100	0.616

3.4 Key-Dependent Security

In this section we verify the sensitivity of the key used in the proposed method. For this analysis we generated 1000 random wrong keys, given as the abscissa in Fig. 9, and average values of twenty CCs between the hash pairs of images of Section 3.2 (A), represented by the ordinate of Fig. 9; the latter were collected using correct and wrong secret keys. It is evident from Fig. 9 that most of the CCs are in the interval 0.4, −0.4. This indicates that it is difficult for an adversary to guess or identify the correct key. Hence, this analysis verifies the heavy

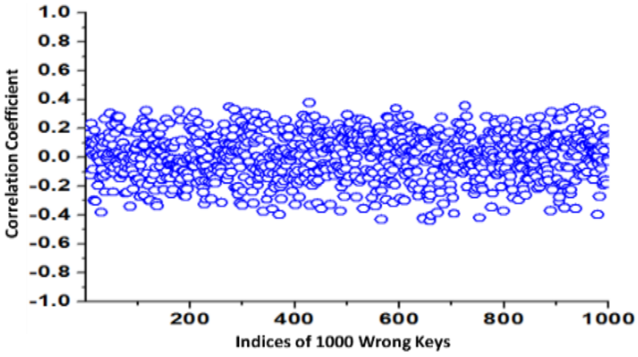

Fig. 9: Key sensitivity test based on 1000 wrong keys

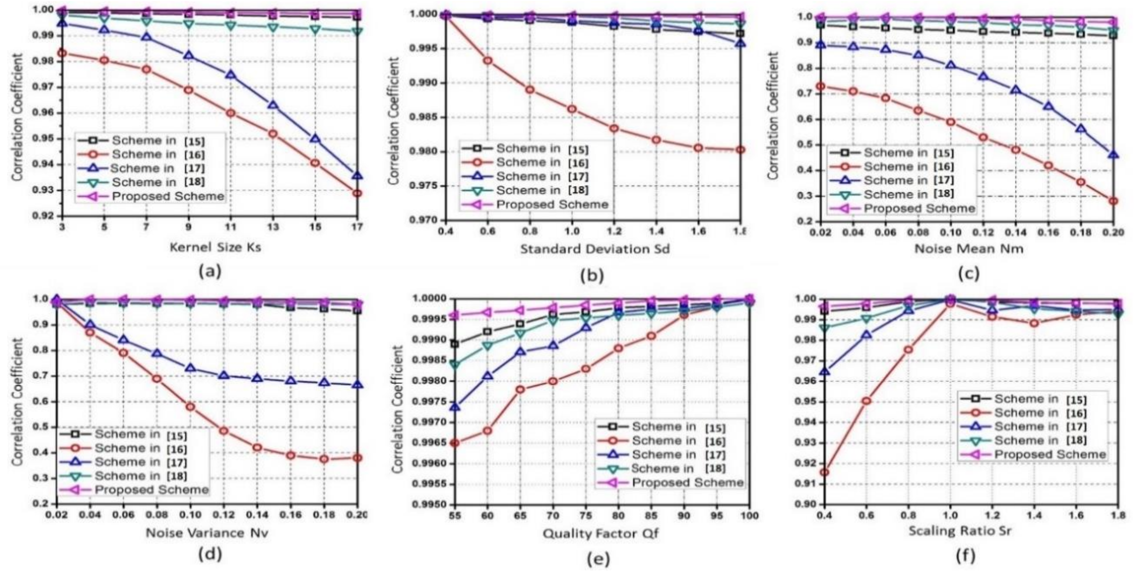

Fig. 10: Performance comparison of our method with state-of-the-art schemes for perceptual robustness: (a) average filtering; (b) Gaussian filtering; (c) Gaussian noise; (d) speckle noise; (e) JPEG compression; (f) scaling

dependence of our approach on the secret key, thus satisfying the security requirements dictated by Kirchhoff's principle [31, 32].

IV. Conclusions

This article proposes a secure efficient image hashing method for secure data dissemination for smart industrial surveillance networks, with a focus on balancing robustness and discrimination. Our method uses perceptual structure features extracted from the input image after pre-processing for image regularization. A single grayscale edge image, representing the classified edges is generated by combining eight binary edges images using Canny operator. Rich and non-zero low edge blocks are then sampled in the secondary image using the grayscale edge image. The final hash is then generated by combining the dominant DCT coefficients of the sampled rich edge image blocks and the difference of the rich and low edge block positions. To increase the security of industrial surveillance network, the final hash is encrypted using an automatically generated key from the rich edge block position vector. Through extensive experiments, it was concluded that our method can achieve better performance in terms of perceptual robustness and discrimination with a reasonable length of hash and running time, as compared with state-of-the-art image hashing methods. In future, we aim to investigate deep neural networks for detecting salient regions and generating hash codes from them with focus on efficiency for resource constrained devices, which can be easily adjusted in surveillance networks.

V. References

[1] P. Basanta-Val, "An efficient industrial big-data engine," *IEEE Transactions on Industrial Informatics,* vol. 14, pp. 1361-1369, 2018.

[2] M. Garcia, A. Canovas, M. Edo, and J. Lloret, "A QoE management system for ubiquitous IPTV devices," in *2009 Third International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies*, 2009, pp. 147-152.

[3] H. Farid, "Image forgery detection," *IEEE Signal processing magazine,* vol. 26, pp. 16-25, 2009.

[4] J. Fridrich, "Digital image forensics," *IEEE Signal Processing Magazine,* vol. 26, 2009.

[5] X.-L. Liu, C.-C. Lin, and S.-M. Yuan, "Blind dual watermarking for color images' authentication and copyright protection," *IEEE Transactions on Circuits and Systems for Video Technology,* 2016.

[6] X. Li, X. Sun, and Q. Liu, "Image integrity authentication scheme based on fixed point theory," *IEEE Transactions on image Processing,* vol. 24, pp. 632-645, 2015.

[7] Z. Wei, Y. Wu, R. H. Deng, and X. Ding, "A hybrid scheme for authenticating scalable video codestreams," *IEEE transactions on information forensics and security,* vol. 9, pp. 543-553, 2014.

[8] S. D. Roy, X. Li, Y. Shoshan, A. Fish, and O. Yadid-Pecht, "Hardware implementation of a digital watermarking system for video authentication," *IEEE transactions on circuits and systems for video technology,* vol. 23, pp. 289-301, 2013.

[9] Y. Zhao, S. Wang, X. Zhang, and H. Yao, "Robust hashing for image authentication using Zernike moments and local features," *IEEE transactions on information forensics and security,* vol. 8, pp. 55-63, 2013.

[10] B. Alomair and R. Poovendran, "Efficient authentication for mobile and pervasive computing," *IEEE Transactions on Mobile Computing,* vol. 13, pp. 469-481, 2014.

[11] S. A. H. Tabatabaei, O. Ur-Rehman, N. Zivic, and C. Ruland, "Secure and robust two-phase image authentication," *IEEE Transactions on Multimedia,* vol. 17, pp. 945-956, 2015.

[12] X. Wang, K. Pang, X. Zhou, Y. Zhou, L. Li, and J. Xue, "A visual model-based perceptual image hash for content authentication," *IEEE Transactions on Information Forensics and Security,* vol. 10, pp. 1336-1349, 2015.

[13] M. Li and V. Monga, "Twofold video hashing with automatic synchronization," *IEEE Transactions on Information Forensics and Security,* vol. 10, pp. 1727-1738, 2015.

[14] K. Muhammad, R. Hamza, J. Ahmad, J. Lloret, H. H. G. Wang, and S. W. Baik, "Secure surveillance framework for IoT systems using probabilistic image encryption," *IEEE Transactions on Industrial Informatics*, 2018.

[15] R. Davarzani, S. Mozaffari, and K. Yaghmaie, "Perceptual image hashing using center-symmetric local binary patterns," *Multimedia Tools and Applications,* vol. 75, pp. 4639-4667, 2016.

[16] C. De Roover, C. De Vleeschouwer, F. Lefebvre, and B. Macq, "Robust video hashing based on radial projections of key frames," *IEEE Transactions on Signal processing,* vol. 53, pp. 4020-4037, 2005.

[17] Z. Tang, X. Zhang, Y. Dai, and W. Lan, "Perceptual image hashing using local entropies and DWT," *The Imaging Science Journal,* vol. 61, pp. 241-251, 2013.

[18] C.-P. Yan, C.-M. Pun, and X.-C. Yuan, "Quaternion-based image hashing for adaptive tampering localization," *IEEE Transactions on Information Forensics and Security,* vol. 11, pp. 2664-2677, 2016.

[19] R. K. Karsh, A. Saikia, and R. H. Laskar, "Image authentication based on robust image hashing with geometric correction," *Multimedia Tools and Applications,* vol. 77, pp. 25409-25429, 2018.

[20] H. Yang, J. Yin, and M. Jiang, "Perceptual Image Hashing Using Latent Low-Rank Representation and Uniform LBP," *Applied Sciences,* vol. 8, p. 317, 2018.

[21] N. D. Gharde, D. M. Thounaojam, B. Soni, and S. K. Biswas, "Robust perceptual image hashing using fuzzy color histogram," *Multimedia Tools and Applications,* vol. 77, pp. 30815-30840, 2018.

[22] L. Ghouti, "Robust perceptual color image hashing using randomized hypercomplex matrix factorizations," *Multimedia Tools and Applications,* vol. 77, pp. 19895-19929, 2018.

[23] Z. Tang, Z. Huang, H. Yao, X. Zhang, L. Chen, and C. Yu, "Perceptual image hashing with weighted dwt features for reduced-reference image quality assessment," *The Computer Journal,* vol. 61, pp. 1695-1709, 2018.

[24] L. Du, Z. Chen, and Y. Ke, "Image Hashing for Tamper Detection with Multiview Embedding and Perceptual Saliency," *Advances in Multimedia,* vol. 2018, 2018.

[25] D. Han, "Comparison of commonly used image interpolation methods," *ICCSEE, Hangzhou, China,* pp. 1556-1559, 2013.

[26] J. Canny, "A computational approach to edge detection," *IEEE Transactions on pattern analysis and machine intelligence,* pp. 679-698, 1986.

[27] F. Yan, X. Shao, G. Li, Z. Sun, and Z. Yang, "Edge detection of tank level IR imaging based on the auto-adaptive double-threshold Canny operator," in *Intelligent Information Technology Application, 2008. IITA'08. Second International Symposium on,* 2008, pp. 366-370.

[28] C. Qin, X. Chen, J. Dong, and X. Zhang, "Perceptual image hashing with selective sampling for salient structure features," *Displays,* vol. 45, pp. 26-37, 2016.

[29] C. Qin, M. Sun, and C.-C. Chang, "Perceptual Hashing for Color Images Based on Hybrid Extraction of Structural Features," *Signal Processing,* 2017.

[30] R. Vezzani and R. Cucchiara, "Video surveillance online repository (visor): an integrated framework," *Multimedia Tools and Applications,* vol. 50, pp. 359-380, 2010.

[31] K. Muhammad, M. Sajjad, I. Mehmood, S. Rho, and S. W. Baik, "A novel magic LSB substitution method (M-LSB-SM) using multi-level encryption and achromatic component of an image," *Multimedia Tools and Applications,* vol. 75, pp. 14867-14893, November 01 2016.

[32] R. Hamza, "A novel pseudo random sequence generator for image-cryptographic applications," *Journal of Information Security and Applications,* vol. 35, pp. 119-127, 2017.

Muhammad Sajjad received his Ph.D. degree in Digital Contents from Sejong University, Seoul, Republic of Korea. He is now working as an associate professor at Department of Computer Science, Islamia College Peshawar, Pakistan. His primary research interests include computer vision, image understanding, pattern recognition, and robot vision and multimedia applications, with current emphasis on raspberry-pi and deep learning-based bioinformatics, video scene understanding, activity analysis, Fog computing, Internet of Things, and real-time tracking.

Ijaz Ul Haq (S'19) received the B.S degree in computer science from the Islamia College Peshawar, Peshawar, Pakistan. He is currently pursuing the M.S. degree with the Intelligent Media Laboratory, Sejong University, South Korea. His research interests include video summarization, image and video analysis, image hashing, steganography, and deep learning for multimedia understanding.

Jaime Lloret (M'07–SM'10) is an associate professor at Politechnic University of Valencia, Spain. He was Internet Technical Committee Chair during 2014–2015 and is the current Chair of IEEE 1907.1. He is the director of the Research Institute IGIC and head of the Innovation Group EITACURTE. He is co-Editor-in-Chief of Ad Hoc and Sensor Wireless Networks and Editor-in-Chief of Network Protocols and Algorithms. He has been General Chair of 36 international workshops and conferences.

Weiping Ding (M'16–SM'19) is Deputy Dean of the School of Information Science and Technology, Nantong University, China. He has published over 50 papers in flagship journals and conference proceedings as the first author, he has held ten approved invention patents in total over 18 issued patents. His current research interests include data mining, machine learning, and granular computing. He served/serves as an Associate Editor of the IEEE TRANSACTION ON FUZZY SYSTEMS, Information Sciences, and Swarm and Evolutionary Computation.

Khan Muhammad (S'16–M'18) is an assistant professor in the Department of Software, Sejong University, South Korea. His research interests include information security, video summarization, computer vision, and video surveillance. He has authored over 40 papers in peer-reviewed international journals such as IEEE TII, TIE, IoTJ, and TSMC-Systems, and is a reviewer of over 30 SCI/SCIE journals including IEEE Communications Magazine, IEEE Network, IEEE Internet of Things Journal, TIP, TII, TCYB, and IEEE Access. He is a member of the ACM.