Protecting Ownership Rights of Videos Against Digital Piracy: An Efficient Digital Watermarking Scheme

Himanshu Agarwal[1], Farooq Husain[2]

[1]Department of Computer Science and Engineering, Moradabad Institute of Technology, Moradabad (U.P.)-India
[2]Director, Shivadan Singh Institute of Technology and Management, Aligarh (U.P.)-India

***Abstract*:** Violation of one's intellectual ownership rights by the others is a common problem which entertainment industry frequently faces now-a-days. Sharing of information over social media platforms such as Instagram, WhatsApp and twitter without giving credit the owner causes huge financial losses to the owner and hence needs an immediate attention. Digital watermarking is a promising technique to protect owners' right against digital piracy. Most of the state-of-the-art techniques does not provides adequate level of resilience against majority of video specific attacks and other commonly applied attacks. Therefore, this paper proposes a highly transparent and robust video watermarking solution to protect the owners rights by first convert each video frame into YCbCr color components and then select twenty five strongest speeded-up robust features (SURF) points of the normalized luminance component as points for both watermark embedding and extraction. After applying variety of geometric, simple signal processing and video specific attacks on the watermarked video meticulous analysis is performed using popular metrics which reveals that the proposed scheme possesses high correlation value which makes it superior for practical applications against these attacks. The scheme also proposes a novel three-level impairment scale for subjective analysis which gives stable results to derive correct conclusions.

***Keywords*:** SURF, Feature Points, Frame Normalization, Three-Level Impairment Scale, Color Video Watermarking.

1. Introduction

Due to the massive use of internet and easy availability of sophisticated multimedia processing tools, anybody having basic knowledge of these tools can easily manipulate and distribute the digital contents without considerable loss of quality. Simple processing up to a level is acceptable if it is done in a positive manner i.e. making contents more attractive with due respect and permission of creator, but if these tools are used by nasty peoples with the intension of unauthorized manipulation, reproduction and distribution then it creates heavy financial losses to the producer. To provide a safeguard against these financial losses, digital watermarking proved itself to be an effective weapon since last two decades.

Many digital watermarking schemes [1,2,3] have been developed which mainly focuses on two domain of digital watermarking: spatial domain and transform domain. A comprehensive survey of digital watermarking techniques can be found in [4,5]. Spatial domain techniques are used to protect the unauthorized processing of digital data by directly altering the spatial pixels [6]. It is normally done by changing the least significant bit of the pixel value resulting insignificance change in the digital data. The main advantage of altering the pixels values directly is low processing complexity but these techniques can be easily broken down by basic signal processing attacks and sometimes fails to maintain the synchronization between vital information which makes the watermark unreadable.

Facing the above problems, the momentum was shifted towards transform/frequency domain watermarking scheme. Frequency domain techniques decompose the digital data in to various frequency components and use the suitable frequency components for embedding the watermark pattern [7]. Discrete Wavelet Transform (DWT) [8], Discrete Cosine Transform (DCT) [9] are popularly used frequency domain transformation techniques. Transformation of digital data in frequency components for watermark embedding purpose offers better result against the attacks that process the signal when compared to the techniques that locate the pixels in spatial domain. Abdallah and Alsodairi [10] developed a hybrid technique to get the advantages of combining the mathematical aspects of complex wavelet transform with SVD and attain some improvements over simple frequency domain watermarking techniques.

The motive of an adversary is to reduce the energy of watermark signal embedded in the host data by applying various signal processing techniques called attacks. In general, attacks on watermarked data can be classified into three major categories, first category belongs to the geometric distortion which includes rotation, cropping etc., simple signal processing attacks such as compression, histogram stretching and addition of various types of noise belong to the second category. Video specific attacks like format conversion, frame rate change, swapping, averaging and insertion of frames fits in the third category of attacks.

All the methods discussed above were normally provide protection against some attacks belongs to either geometric distortion or simple signal processing attacks but now a days, the attackers tries to disturb the watermarked information by reducing the synchronization between the vital information using image and video specific intentional attacks, which belongs to the third category of attacks. Due to the poor selection of watermarking location specialized intentional attacks will succeed to reduce the energy between the original and the extracted watermark.

The strength of the most modern computer science areas such as artificial intelligence, pattern recognition and computer vision can be exploited for the feature extraction. We can use these features for watermarking purpose either as reference orientation or by directly modify the features for embedding the watermark in the host data [11,12]. Therefore, we presents a video watermarking scheme in which embedded watermark not only survive under geometric distortion but at the same time also effectively works against simple signal processing and video specific attacks, based upon the use of semantically meaningful features of the data. Not all the data features are used for watermarking purpose. The semantically meaningful data features that are suitable for the purpose of watermark embedding must be invariant to noise, co-variant to geometric distortions and hold localization property i.e. alteration of data in one part does not affect in other parts of the data [13]. In images, semantically meaningful data features are the corners, edges and textured areas. Carefully selected semantically meaningful data features proved to be effective because they will maintain the energy between the original and the extracted watermarks.

In this paper, we proposed an efficient video watermarking scheme that first geometrically transforms an input video frame into its corresponding standard version by applying normalization process and then use the meaningful features of the standard version as the points to embed the watermark and to resolve the problems and challenges faced by spatial and frequency domain watermarking techniques.

Typically, watermarking is most suitable technique to prove creator's ownership to a court in case of unauthorized manipulation, reproduction and distribution of digital contents and hence act as a safeguard against financial losses to the creator. Now-a-days, entertainment industry is the key industry which normally faces the problem of unauthorized manipulation and distribution of one's intellectual property by the others. Movies, pictures and popular television shows are intentionally distributed over internet without taking the permission of owner. Pirating such contents causes huge financial losses to creator and hence an immediate solution of such problem needs to be addressed. Most of the solutions developed by researchers [9,10,11] to embed the watermark pattern in digital data are weakened by the adversary through reducing the synchronization between the vital information required to extract the embedded watermark by processing the digital contents.

Numerous proposals have been suggested by many researchers [14,15] that claimed a sufficient amount of robustness and transparency to the embedded watermark but they failed to provide a safeguard due to the poor selection of watermarking location against many signal processing attacks. Even though, the fundamental architecture of images is similar to video frames but videos are not identical to images and are prone to video specific attacks. Very few authors [16,17] have developed video watermarking schemes to make the watermark robust against video specific attacks. Recently, Agarwal et al. [18] used Harris corner detection method to embed the watermark information into the video. In particular, their main objective was to improve the robustness by considering corner points as suitable features but their schemes was not resistant to frame rate change and format conversion like advanced video processing attacks.

Presently, no successful watermarking solution exists in the literature that provides adequate level of resilience against most of the mainstream video specific attacks along with simple signal processing and geometric transformations. In this paper, a transparent and robust video watermarking scheme has been proposed to resolve above given challenges. The proposed video watermarking scheme successfully deals issues related to geometric distortions, simple signal processing and video specific attacks.

This paper proposes a normalization and SURF based video watermarking scheme which takes care of transparency as well as robustness of the watermark efficiently. Normalization of video frame makes it invariant to geometric distortions while careful selection of perceptually semantically meaningful features of each frame using SURF makes it transparent and rotation invariant. Considering not only simple signal processing but video specific attacks are also taken into account. The scheme proposed in this paper accurately extracts the watermark signal even after applying watermark attacks intentionally or unintentionally. In fact, this paper introduces the novel idea of subjective transparency assessment by the experts, which proves that scheme measures the high degree of visual transparency.

In particular, the major contributions of this paper are as follows:

- ✔ The paper analyzed the issues and challenges faced by the entertainment industry to develop safeguard against digital piracy in the era of social media.
- ✔ Transforming the input frame into its corresponding standard version using normalization technique which reduces the overall effect of geometric distortion without affecting accuracy of retrieval.
- ✔ Affine invariant technique known as Speeded-up robust features for semantically meaningful data features selection has been used.
- ✔ To be widely acceptable as realistic solution, the proposed scheme consider both standard definition (SD) and high definition (HD) videos as testing contents.
- ✔ Sufficient number of geometric, simple signal processing and video specific attacks was applied to test the transparency and robustness of the watermark.
- ✔ For ensuring the subjective quality of our scheme, the method proposes a novel three-level impairment scale oppose to five-grade impairment scale suggested by double-stimulus impairment scale (DSIS) which gives more stable results and drive more correct conclusions.

The rest of the paper is organized as follows: Overview of related work is presented in Section 2. To improve the understanding with the proposed scheme, in Section 3, we briefly introduce the key logical and mathematical concepts of frame normalization and SURF method. In Section 4, we explain the implementation scheme for embedding and extraction of watermark. Setup and results are presented in Section 5 which demonstrate the performance of our scheme. Finally, the conclusions of our study and suggestions for future work are discussed in Section 6.

2. Related Work

Digital content can be watermarked by embed special patterns in the host data without damaging its vital information. The embedded watermark can be extracted later to prove her/his ownership against unauthorized manipulation, reproduction and distribution of contents in the court. Due to rapid changes and improvement in digital technologies, now-a-days attacks against watermarking technologies become more sophisticated. Hence digital watermarking enters into a new era of technologies where we should consider advance aspects for development of watermarking techniques which proves them as an effective tool and provide a safeguard against technological improved attacks.

Many researchers [19,20] developed striking schemes to protect the unauthorized processing of digital data by directly altering the spatial pixels. Apart from spatial domain, considering frequency components of different regions in the image has also shown an important role in increasing the robustness of a digital watermarking based- ownership protection system in the literature. Abdallah and Alsodairi [10] have proposed a hybrid watermarking scheme based on CWT and singular value decomposition. Their main idea is to decompose the luminance part of the video frame into four frequency sub-bands of CWT and shaped the singular values of LL sub-band with the singular values of watermark image. The main drawback of their scheme was that this scheme does not incorporate newer type of attacks and limit itself to rotation, resizing and cropping. The scheme also gives poor results when exposed to noise.

Kutter et al. [13] have proposed that semantically features of the data have ability to maintain their energy even if watermarked data is attacked with the aim to destroy the embedded information. They embed the watermark using data features with scale interaction technique based on wavelet transform. The method proposed by them shows some degree of resilience against simple attacks like translation and cropping but fails when the attack changes the locations of some pixels. In response to the challenges raised using the scheme propose by [13], Tang and Hang [21], proposed a watermarking scheme that uses the combination of feature extraction and image normalization with the aim of improving the initial scheme. They use only 16-bit watermark sequence to embed in the original image. The scheme proposed by them survives against low-quality JPEG compression, filtering and linear geometric transformations but the scheme was not flexible for specialized signal processing attacks.

Zhao et. al. [22] presented a feature-based image fusion approach by decomposing the host and watermark images with GHM multiwavelet and fuse the watermark in the transform domain using phase congruency features. Simulation results of their schemes proved it resilience against compression but its performance is very poor for additive noise and completely ignores the rotation and translation attacks which are important aspect for developing efficient watermarking technique. Further, Lu and Chung [23] presented a scale interactive model-based filter image watermarking scheme. They embed the watermark in the local region of the related feature point. Their scheme is limited to noise addition, scaling and rotation attacks only.

Most of the methods available in the literature focuses either on reduces the energy of the embedded watermark by applying lossy compression, passing through filters, addition of noise or to devastate the synchronization information between the watermark bits using some basic type of signal processing attacks. Ayubi et al. [24] presented a watermarking scheme using 2D DWT and contourlet transforms utilizes the concept of singular value decomposition. Acceptable performance of their scheme was obtained after applying some geometric and non-geometric attacks but their scheme fails when applied to a range of video specific attacks. Kong and Zhou [25] proposed watermarking algorithm that employed a multi-level lifting wavelet transform (LWT) and embed the watermark into DC area of last level LWT based on amplitude modulation. Their scheme is effective against high compression ratio but fall short for most of the other attacks.

On account of designing an effective watermarking scheme, numerous schemes that combines transform domain and meaningful features of the data were suggested. In fact, these techniques slightly improve transparency and robustness of the embedded watermark. Umaa and Thanushkodi [26] proposed a digital watermarking technique that uses Harris Laplacian detector to detect the feature and use them to form a primary feature set where they embed the watermark. They employ six types of predefined attacks, to test the robustness of the extracted feature values. Resilience against the attacks is improved by extending the primary features set using some auxiliary features but their algorithm fails when the image was cropped and exposed to noise. The method proposed by Wang et al. [27] designed a robust image watermarking scheme that work against geometric distortions using exponent moments (EMs). They firstly applied non subsampled contourlet transform (NSCT) on host image and then compute the exponent moments of the low-pass NSCT sub-band. Finally, the method embeds the digital watermark by quantizing the modulus of the selected EMs. They achieve satisfactory level of robustness against geometric distortions, but their scheme fails for almost all other attacks.

The scheme based on candidates I-frames for copyright protection proposed by Ahuja and Bedi [28] for watermark a video utilize the MPEG-2 standard with DCT coefficients. They verified the results by simulating various intentional and non-intentional attacks. Rakesh and Sarabjeet [29] proposed a semi blind robust video watermarking scheme by inserting the watermark into some high frequency coefficients DCT blocks. The scheme is tested against various types of geometric and video specific attacks like cropping, rotation, frame inserting, deleting, swapping and averaging. But the scheme fails when exposed to format conversion and multiple cascading attacks.

Boris et al. [30] presents an image watermarking technique using a brightness model and the Hermite Transform (HT). They have taken advantage of the masking characteristics of Human Visual system (HVS) to generate a watermark that cannot be detected by a human observer. Proposed scheme is found robust against most of the common signal processing attacks and deals with the geometric distortions using image normalization. But the scheme fails to deal with high degree of transparency and robustness against specialized intentional attacks. Technique given by Agarwal et al. [18] facilitates the use of strongest corner points obtained using Harris method as a reference for embedding watermark information in the video. Their method shows some level of resilience against rotation and other affine transformations but due to the use of Harris corner detection their method fails to achieve scale invariance. The authors also do not consider frame rate change and format conversion like attacks which normally exist when videos are processed.

To design an effective video watermarking scheme which provides both transparency and robustness to the watermark against geometric distortions and signal processing transformations is still a challenging area, some newer type of attacks like format conversion, frame rate change, frame insertion, frame deletion, frame averaging are also need to be addressed immediately. Because most of the above-mentioned attacks are able to inadequately devastate the synchronization of vital information among the watermark bits, similar to the affine transformations. In this paper, we proposed an efficient video watermarking scheme which successfully protect the ownership rights against digital piracy by utilizing the features of the data as embedding points rather than fixed coordinates to embed the watermark and to resolve all the issues and challenges mentioned above.

3. Key Techniques

This section describes the logical and mathematical aspects of key techniques used to develop an imperceptible and robust watermarking scheme. Section 3.1 explains how we can transform an input frame into its corresponding standard version which becomes invariant under translation, scaling, skew, and rotation. Section 3.2 describes the method to calculate the Hessian matrix-based interest points, that provides better performance in terms of repeatability and robustness when compared to Harris-based counterparts.

3.1 Frame Normalization

Normalization of video frames prior to embedding the watermark is the first step of our scheme. Normalized frame serves as input to the extraction of features using SURF. In our scheme, we implement the normalization method described by Wang and Zhao [31] for image understanding and pattern matching. We can be effectively utilized this method in digital watermarking to overcome the effect of geometric distortions to watermarked images. The process of normalization is a well-designed pre-processing method that geometrically transform an input pattern (video frame) into its corresponding standard version in such a way that the standard version becomes invariant under translation, scaling, skew, and rotation. It is implemented through the use of parameters estimated from the geometric moments of the input pattern. Because moments are algorithmically simple, fast to compute and less vulnerable to manipulations, they provides attractive solutions to video watermarking problems if implemented efficiently.

The frame normalization method used by the proposed scheme consists of two major phases –implementation of compact algorithm and rotation of compact frame clockwise by an angle obtained by central moments. The mathematical calculations for creating the standard version of a frame are given as follows:

The covariance matrix of the video frame F(j, k) is calculated as given in Eq.(3.1),

$$\mathrm{Cov}\left[F(j,k)\right]=\begin{bmatrix} C_{20} & C_{11} \\ C_{11} & C_{02} \end{bmatrix} \qquad (3.1)$$

where, the elements C_{pq}; p, q ϵ N as order indices, of covariance matrix are the central moments of the frame and can be defined in Eq.(3.2),

$$C_{pq}=\frac{\sum_j\sum_k (j-m_x)^p (k-m_y)^q F(j,k)}{\sum_j\sum_k F(j,k)} \qquad (3.2)$$

where, m_x and m_y are the frame center and can be obtained as,

$$m_x=\frac{\sum_j\sum_k j\cdot F(j,k)}{\sum_j\sum_k F(j,k)} \text{ and } m_y=\frac{\sum_j\sum_k k\cdot F(j,k)}{\sum_j\sum_k F(j,k)}.$$

The eigenvalues λ_1 and λ_2 and associated eigenvectors $\begin{bmatrix} e_{ix} \\ e_{iy} \end{bmatrix}$ of Cov[F(j, k)] for i = 1,2 are calculated as,

$$\begin{bmatrix} e_{ix} \\ e_{iy} \end{bmatrix}=\begin{bmatrix} \dfrac{C_{11}}{\sqrt{(\lambda_i-C_{20})^2+C_{11}^2}} \\ \dfrac{\lambda_i-C_{20}}{\sqrt{(\lambda_i-C_{20})^2+C_{11}^2}} \end{bmatrix} \qquad (3.3)$$

where, eigenvalues λ_i for i = 1,2 is the solution of the matrix $\begin{vmatrix} C_{20}-\lambda & C_{11} \\ C_{11} & C_{02}-\lambda \end{vmatrix}=0$, thus

$$\lambda_1=\frac{C_{11}+C_{22}+\sqrt{(C_{11}-C_{22})^2-4C_{12}^2}}{2}$$ and

$$\lambda_2=\frac{C_{11}+C_{22}-\sqrt{(C_{11}-C_{22})^2-4C_{12}^2}}{2}.$$

Since, the matrix Cov[F(j, k)] is real and symmetric so, both eigenvectors are orthonormal to each other,

$$\left.\begin{aligned} e_{1x}e_{1y}+e_{2x}e_{2y}&=0 \\ e_{1x}^2+e_{1y}^2&=1 \\ e_{2x}^2+e_{2y}^2&=1 \end{aligned}\right\} \qquad (3.4)$$

By using the conditions given in Eq.(3.4), we can calculate e_{2x} = -e_{1y}, e_{2y} = e_{1x}, then rotational matrix 'R' can be described as,

$$R=\begin{bmatrix} e_{1x} & e_{1y} \\ -e_{1y} & e_{1x} \end{bmatrix} \qquad (3.5)$$

Furthermore, we can calculate the scaling matrix 'S_M' according to the eigenvalues of Cov[F(j, k)] as,

$$S_M=\begin{bmatrix} l/\sqrt{\lambda_1} & 0 \\ 0 & l/\sqrt{\lambda_2} \end{bmatrix} \qquad (3.6)$$

where, $l = (\lambda 1\lambda 2)^{1/4}$ [31].

The eigenvectors of Cov[F(j, k)] matrix corresponds to the major (longest eigenvector)and minor axes of the frame intensity values. So, the orientation can be thus extracted from the angle φ given by the Eq.(3.7),

$$\varphi=\frac{1}{2}\tan^{-1}\left(\frac{2C_{11}}{C_{20}-C_{02}}\right) \qquad (3.7)$$

To get the normalized frame we, transform the original coordinate system by first translating the origin to the image centre, multiply by matrix 'R', then scaling the matrix through multiply by 'S_M' we get the compact image. Further rotating the compact image clockwise by an angle φ we get normalized frame. Let $[j' \quad k']^T$ is new coordinate system of the normalized frame corresponding to the original frame coordinates (j, k) and can be obtained as,

$$\begin{bmatrix} j' \\ k' \end{bmatrix}=\begin{bmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{bmatrix}\begin{bmatrix} l/\sqrt{\lambda_1} & 0 \\ 0 & l/\sqrt{\lambda_2} \end{bmatrix} \qquad (3.8)$$
$$\begin{bmatrix} e_{1x} & e_{1y} \\ -e_{1y} & e_{1x} \end{bmatrix}\begin{bmatrix} j-m_x \\ k-m_y \end{bmatrix}$$

Equation (3.8), represents our new coordinate system which is invariant to translation, scaling, skew and rotation.

3.2 Feature detection using Speeded-Up Robust features (SURF)

Several affine-invariant features point detection [32,33] methods have been proposed in the state-of-the-art to cope with the changes in viewpoint. But most of them are suffered either with high computation time or with the robustness to commonly occurring deformations. In the proposed scheme, we used rotation invariant hessian based detector called SURF for detection of meaningful features. This method was originally used in Bay et al. [34] which performs better than their Harris-based counterparts with respect to repeatability and robustness.

The SURF method, as described [34], comprises of two major stages – interest point detection and interest point description. The scheme we developed for watermark embedding is majorly concerned with the interest point detection and use them as reference points to embed the watermark pattern. The computational steps for interest point detection, includes creation of integral video frame, scale space feature detection by a fast Hessian matrix and extracting the key interest points sub-stages and is given as:

Step 1. The first sub stage for the interest point detection is integration of the image/frame. The output is the integral frame/image, which is used as the basis of the subsequent scale-space analysis. Integral images allows fast computation of box type convolution filters.

The integral image $I_\Sigma(Z)$ at a location $Z = (x, y)$ can be defined in Eq.(3.9) as,

$$I_\Sigma(Z) = \sum_{j=0}^{x} \sum_{k=0}^{y} I(j,k) \qquad (3.9)$$

It is represented by the sum up of all the pixels in the input image I within a rectangular region formed by the origin and Z.

Step 2. Once the computation of integral image is completed, then the second sub stage is the calculation of Hessian matrix-based interest points.

(a) SURF feature points detection method uses the determinant of Hessian matrix as a discriminant to look for local maximum value. For the point $Z = (x, y)$ in an image I_m the Hessian matrix is defined in Eq.(3.10) as,

$$Hess(Z,\sigma) = \begin{bmatrix} L_{xx}(Z,\sigma) & L_{xy}(Z,\sigma) \\ L_{xy}(Z,\sigma) & L_{yy}(Z,\sigma) \end{bmatrix} \qquad (3.10)$$

Here, the responses of the input image I_m at point (x, y) under the convolution at scale σ in x direction is denoted as $L_{xx}(Z,\sigma) = I_m * \frac{\partial^2}{\partial x^2} g(\sigma)$ of the Gaussian second order partial derivative and similarly for xy and y direction as $L_{xy}(Z, \sigma)$ and as $L_{yy}(Z, \sigma)$ respectively.

In practical, Gaussians have to discretised and cropped which leads the loss of repeatability property, so we use approximation of Hessian matrix with box filters of size 9 x 9. The Hessian approximation matrix at point Z in x, y and xy direction is given by in Eq.(3.11) as,

$$Hess_{Approx}(Z,\sigma) = \begin{bmatrix} D_{xx}(Z,\sigma) & D_{xy}(Z,\sigma) \\ D_{xy}(Z,\sigma) & D_{yy}(Z,\sigma) \end{bmatrix} \qquad (3.11)$$

where, $D_{xx}(Z, \sigma)$, $D_{yy}(Z, \sigma)$ *and* $D_{xy}(Z, \sigma)$ denotes the convolution of the integral image with standard deviation σ =1.2 in x, y and xy direction[34].

(b) The determinant of $Hess_{Approx}$ is calculated as,

$$\left|Hess_{Approx}\right| = D_{xx}(Z, \sigma)\, D_{yy}(Z, \sigma) - \left(wD_{xy}(Z, \sigma)\right)^2$$

To balance the expression for Hessian determinant and to maintain the energy conservation between approximated Gaussian kernel and Gaussian kernel we use a relative weight factor ‘w’ of filter response whose value is approximated as 0.9 [35].

Since Hessian matrix represents local curvature so the $|Hess_{Approx}|$ represents the blob response at location Z in the image. The maxima of the $|Hess_{Approx}|$ as given in Eq.(3.12) is than interpolate by the scheme explained by Brown and Lowe[36], represents our interest point.

4. Proposed Implementation Scheme

In this section we explain the procedure for the selection of watermark embedding location and also the two watermarking stages i.e., watermark embedding and watermark extraction using proposed scheme. Unlike using binary image as a watermark in our scheme we use gray scale watermark ‘cameraman.tif’ because grayscale watermark is consider to be more significant in practical situations and can carry greater amount of watermarking contents [37]. Watermark extraction scheme identify the locations of each frame where embedding was done and subtracting their pixel value by the pixel value of obtained using the same procedure on original unwatermarked video. The block diagram of watermark embedding and extraction scheme is shown in Figure 4.1 and Figure 4.2.

4.1 Obtaining Watermark Embedding Locations

The video is first normalized and then the semantically meaningful features of the each video frame are detected by using SURF method. The twenty five strongest points are used reference points for watermark embedding and extraction purpose. The proposed scheme uses YCbCr color space instead of RGB and use its luminance (Y) part for embedding and extraction purpose which makes the scheme resilience against high lossy compression [38]. Because human visual system is highly sensitive to the modifications in the ‘Y’ component the embedding process perceptually shaped the watermark in all the frames considering only the twenty five strongest points of each frame of the video. Limiting the number of points to only twenty five will reduce the effect of the embedded watermark on luminance component and at the same time it improves the quality of the watermarked frame [38].

Algorithm 1, gives the description of obtaining watermark embedding locations based on the values of strongest interest points obtained using SURF as explained in Section 3.2.

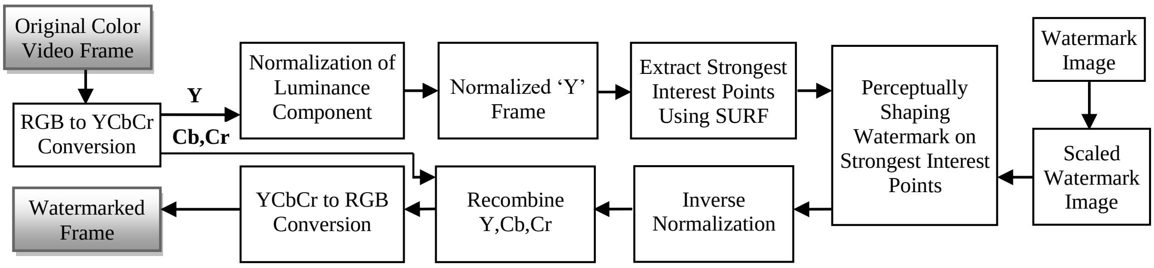

Figure 4.1. Frame normalization and SURF based color video watermark embedding scheme

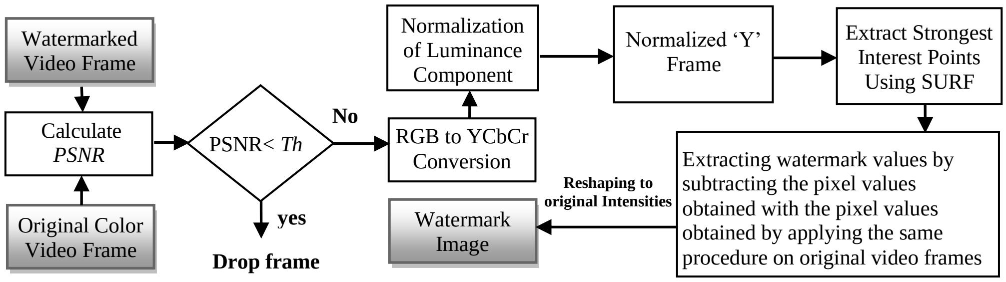

Figure 4.2. Frame normalization and SURF based color video watermark extraction scheme

Algorithm 1: Obtaining Watermark Embedding Locations

Input: Original Color Video (OCV)
Output: Watermark Embedding Locations (R_{EL})
Begin:

1. for i=1 to number of frames
   // Convert the frame F_i in YCbCr and decompose each component
   (F_{iY}, F_{iCb}, F_{iCr}) ← rgbtoycbcr(F_i)
2. // Normalize F_{iY} as per the method discussed in section 3.1
   $(F_{iY})_{norm}$ = normalized(F_{iY})
3. // Detecting interest points (IPs) using SURF
   IPs = SURF_IPs[$(F_{iY})_{norm}$]
4. // Calculation of twenty five strongest IPs
   SIPs = strongest (IPs, 25)
5. // Selecting embedding locations
   R_{EL} = (SIPs)
6. End of for loop

End

4.2 Watermark Embedding Scheme

Step 1. Read the watermark and scale it using a scaling factor say 'α' to maintain the balance between transparency and robustness is maintained by scaled the watermark. In our scheme we have taken the value of 'α' as 0.1.
Step 2. Video is watermarked by perceptually shape the scaled watermark on the embedding locations obtained from Algorithm 1.
Step 3. Inverse normalization is applied to each modified part of luminance component (Y') on watermarked frame.
Step 4. Reconstruct the watermarked frame from inverse normalized frame by convert the Y'CbCr component into RGB color components.

Algorithm 2: Watermark Embedding Scheme

Input: Watermark Embedding Locations (R_{EL}), Normalized frame $(F_{jY})_{norm}$, A gray scale watermark $(Wat)_{rxc}$ where r and c are the numbers of rows and columns respectively, Scaling factor (α)
Output: Watermarked Luminance Part of Frame (F_{jY})
Begin:

1. Read Watermark (Wat)
2. flag =1
3. // read the normalized luminance component of video frames $(F_{iY})_{norm}$ as discussed in Algorithm 1
   for j=1 to number of frames
4. for i =1 to 25
5. // Selecting the coordinate locations of watermark.
   x=floor (flag-1/c) +1
   y= abs (flag % c)
   if (y == 0) then y = c
6. // Scale the value of watermark by factor α
   value = Wat(x, y) x α
7. // Embedding watermark value in locations obtained through Algorithm 1
   $$\left[\left(F_{jY}\right)_{norm}\right]_{R_{EL}} = \left[\left(F_{jY}\right)_{norm}\right]_{R_{EL}} + value$$
8. flag ++
9. if (flag > (r x c))
   then flag = 1
10. End of for loop
11. //Inverse normalize the frame
12. F_{jY} = Inv_normalized $(F_{jY})_{norm}$
13. End of for loop

End

Algorithm 2, gives the description of embedding the watermark in the locations of normalized luminance component of video frame obtained through **Algorithm 1**. Pixel by pixel each coordinate locations of watermark is selected and then scaled by a factor 'α' before shaping it on the obtained locations. At the end each frame is inverse normalization which gives the watermarked frame. Algorithm uses a variable flag whose value is initialized as 1 because it is used to extract the location of the watermark. It should be important to note that in MATLAB software numbering is start from 1 instead of 0 (as in most of the other programming languages).

4.3 Watermark Extraction Scheme

Step 1. Extract the frame from the original and the watermarked video and calculate the value of PSNR. If the calculated PSNR value is less than a threshold say *'Th'* than drop the frame and repeat the procedure until we get a frame having threshold value greater than or equals to the value of *'Th'* (say 10 in our case).

Step 2. Convert the RGB video frame into YCbCr components by standard equation as discussed in watermark embedding process.

Step 3. Normalization is applied to the luminance component of each watermarked frame.

Step 4. Consider the normalized luminance component of each frame and calculate the key interest points using SURF.

Step 5. Select the twenty five strongest interest points from the set of key interest points. Watermark is extracted by subtracting their pixel value by the pixel value obtained using the same procedure on original unwatermarked video. Repeat the procedure for all other frames.

Step 6. Divide the extracted watermark values by α (scaling factor) to get the original intensity values of the watermark.

5. Setup and Results

All the experiments are simulated using MATLAB software by taking one standard definition (SD) color video file named "akiyo.avi" having 252 x 388 pixels and one high definition (HD) video file 'pedestrain.avi' 1090 x 1080 pixels each containing 295 frames with frame rate of 30 fps. Video is watermarked separately using grayscale image 'cameraman.tif' of size 85 x 85 pixels. The proposed consider only 25 watermark embedding locations per frame which means that 7375 watermark bits can be easily embedded in a video of 295 frames without affecting its imperceptibility. Increasing the video frames to some higher values further increases the size of the watermark. Standard definition video frames for the videos 'akiyo.avi' along with high definition video frame for the video 'pedestrian.avi' are shown in the **Fig. 5.1**.

(a)'akiyo.avi'

(b) 'pedestrian.avi'

Fig. 5.1 Standard Definition and High Definition Videos (a, b respectively)

Figure 5.2 shows the original, luminance component along with the normalized luminance frame with detected SURF interest points on SD video 'akiyo.avi'. Watermarked frame of the video and the corresponding extracted frame after applying video compression using motion JPEG 2000 with a compression ratio (CR) of 05 using the proposed scheme is shown in Figure 5.3. For testing the transparency and robustness of watermark three categories of attacks were applied to the watermarked video, i.e., geometric, common signal processing and video specific attacks. Geometric attacks include rotation and cropping while common signal processing consist of addition of salt & pepper noise and speckle noise. Insertion, swapping and averaging of frame including format conversion and change in frame rate are main testing agents under video specific attacks.

(a) Original Color Video Frame

(b) Luminance part of Video Frame

(c) Normalized Luminance component with detected Interest Points Using SURF

Figure 5.2 Original and Luminance component along with its Normalized video frame

(a) Watermarked Frame

(b) Attacked Frame (Format Conversion from .avi to .mj2 with CR=5)

Figure 5.3 Watermarked and Attacked Video frame under format conversion attack

In this paper, careful selection the distortion intensity is taken into account, so that the attacks actually meet the practical aspects of video processing applications. We use 0.5° and 1° angles for rotation, 5% and 10% uniform cropping of frame. So, that a synchronization error is generated between embedded and extracted values of watermark but the commercial value of the video is not affected by the attacks. The main testing contents for common signal processing attacks includes various types of noise such as salt and pepper noise with density of 0.001 and 0.0005 and speckle noise having a variance of 0.001 and 0.0005. For the attacks belonging to video specific category, we test the video against insertion of any number of frames, swapping and averaging of 5% to 10 % frames as well as changing the frame rate from 30 t0 45 & 30 to 15 and format conversion of video using motion JPEG 2000 compression i.e., conversion from .avi to .mj2 format. Changing the format to reduce the storage requirements are commonly applied unintentional attacks on videos, while averaging and deletion of frames belongs to the group of intentionally applied attacks. In the proposed scheme we use Motion JPEG 2000 compression with different compression ratio (CR = 5 and CR = 10) in which each frame is coded independently which makes it scalable and more resilient to propagation of errors over time. Considering a CR value of 05 means 80% space is saved to store the video.

Commercial advertisements are inserted unintentionally in the video but it disturbs the synchronization among the video frames and hence shows a major impact during the extraction of watermark. The proposed scheme deals with such problem in a very effective manner by simply drops all the frames having PSNR value less than a threshold say *'Th'* (say *'Th =10' in our case*). Value of calculated PSNR less than 10 specifies that the watermarked frame has no similarity with the frame under consideration and hence it should be rejected.

5.1 Testing for Transparency

Transparency assessment of watermarked video is an important performance parameter. There exists no single test procedure that can exactly quantify the quality of watermarked video. For better understanding the transparency of embedded watermark in the video both subjective and objective evaluation methods are used as test procedures in the proposed scheme.

5.1.1 Subjective Transparency Assessment using Three-Level Impairment Scale

Subjective evaluation judges the transparency of watermarked video based on Human Visual System (HVS). Performing subjective assessment test is the ultimate method to evaluate the quality of the video carrying a watermark, because this method can evaluate the watermarked video with high and intermediate imperceptible quality. In this paper we propose a three-level impairment scale oppose to five-grade impairment scale as suggested by double-stimulus impairment scale (DSIS) because it is usually found that the stability of the results is greater for small number of impairments levels than for large number of impairments levels or grades. This is because lowering the number of levels reduces the complexity and at the same time it avoids ambiguity in the meaning of the quality scale values (5 imperceptible, 4 perceptible, but not annoying, 3 slightly annoying, 2 annoying, 1 very annoying, as suggested by DSIS). In the proposed subjective evaluation procedure observers are expected to give evaluations on a scale of 10 with clear division between the level of impairments (10 Transparent, 5 Intermediate and 0 Visible) which also proves that at least the observers keenly scrutinize their task and they are not giving random votes. Furthermore, in case of DSIS a grand mean score (averaged overall judgments made in the experiment) close to 3 should be aimed on a scale of 5 to specify the perceptibility. It actually specifies "slightly annoying" corresponds to the five-grade impairment scale which could be hazardous, and even wrong to derive conclusions.

The proposed evaluation procedure is completed in the following two steps:

1. Video is labeled from transparent to visible by different observers and a score for each label is assigned [39, 40]. To score each level of impairment, **Table 5.1.1** is used as a scale of reference:

Table 5.1.1 Proposed Three-Level Impairment Scale Method

Level of Impairment	Score
Transparent	10
Intermediate	5
Visible	0

2. Based on the assigned scores a Percent Mean Opinion Score (MOS) will be calculated. The Percent MOS having a value greater than 80 is considered as transparent while 60 -79 is perceived as intermediate quality. The Percent MOS less than 60 is considered as visible and hence judges as a failure of watermarking scheme under transparency.

Table 5.1.2 Score of Subjective Evaluation by Individual Observer

Observer	Label	akiyo.avi		pedestrian.avi	
		Assigned Score	Maximum Score	Assigned Score	Maximum Score
1	Transparent	10	10	10	10
2	Transparent	10	10	10	10
3	Intermediate	5	10	10	10
4	Intermediate	5	10	10	10
5	Transparent	10	10	10	10
6	Transparent	10	10	10	10
7	Intermediate	5	10	10	10
8	Transparent	10	10	10	10
9	Transparent	10	10	10	10
10	Transparent	10	10	10	10
Sum of Assigned and Maximum Scores		**85**	**100**	**100**	**100**

In the proposed scheme we evaluate the subjective quality by enrolling a group of ten observers. Our observers are academic professors and industry experts of related field. They were shown an original video without any watermark first and then another video which is watermarked by our scheme. It was asked observers to judge the watermarked video in comparison to original un-watermarked video and ranked it as transparent, intermediate level and visible by assigning a score of 10, 05 and 00 respectively. The results of our evaluation scheme are given in the Table 5.1.2.

Percentage of Mean Opinion Score (MOS) from the assigned scores is calculated as,

Percent Mean Opinion Score

$$= \frac{\text{Sum of Assigned Scores}}{\text{Sum of Maximum Score}} \text{x}100 \qquad (5.1)$$

The MOS of the proposed scheme is calculated using Eq.(5.1) is 85 for SD video 'akiyo.avi' and 100 for HD video 'pedestrian.avi' i.e., labeled as transparent, which shows the high subjective quality of our scheme.

5.1.2 Objective Transparency Assessment

Because subjective assessment method is based on the availability of experts and it also varies from person to person, so we use the objective assessment method to judge the quality of watermarked method. Peak Signal to Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) is commonly employed objective assessment methods.

(a) Peak Signal to Noise Ratio (PSNR)

As we insert watermark into the frame the contents gets distorted. To ensure the quality of the watermarked frame higher value of PSNR is desirable which means the frame is not distorted significantly. The value of PSNR is measured using Eq.(5.2), whereas the Mean Square Error (MSE) between the watermarked frame W_f and the original frame O_f can be computed using Eq.(5.3).

$$\text{PSNR} = 10\log_{10}\frac{In^2}{\text{Mean Square Error}} \quad (5.2)$$

$$\text{Mean Square Error} = \frac{\sum_{j=1}^{M}\sum_{k=1}^{N}\left[O_f(j,k) - W_f(j,k)\right]^2}{MxN} \quad (5.3)$$

where *In* is the maximum intensity value of the frame under testing, in our case *In* has a value of 255. M and N are the high and the wide of the host frame respectively.

Table 5.1.3 Obtained PSNR and SSIM values under different types of attacks considering 'cameraman.tif' as watermark.

Type of Attack	akiyo.avi		pedestrian.avi	
	PSNR	**SSIM**	**PSNR**	**SSIM**
No Attack	50.52	0.9997	66.72	0.9999
Rotation (0.5°)	37.78	0.9593	39.26	0.9468
Rotation (1°)	32.47	0.9210	34.35	0.9217
Uniform Cropping (5%)	20.45	0.9113	19.73	0.9324
Uniform Cropping (10%)	17.78	0.8578	16.58	0.8891
Salt & Pepper Noise (With a Density of 0.001)	33.37	0.9575	33.98	0.9487
Salt & Pepper Noise (With a Density of 0.0005)	36.64	0.9738	36.88	0.9675
Speckle Noise (With a Variance of 0.001)	36.41	0.9654	38.67	0.9437
Speckle Noise (With a Variance of 0.0005)	39.55	0.9771	41.27	0.9697
Frame Insertion (Any number of frames)	50.52	0.9997	66.72	0.9999
5% Frame Swapping	46.82	0.9259	64.45	0.9899
10% Frame Swapping	45.27	0.8997	62.52	0.9713
5% Frame Averaging	47.19	0.9310	63.45	0.9887
5% Frame Averaging	45.07	0.8982	61.49	0.9675
Frame Rate Change (30 fps to 45 fps)	50.52	0.9997	66.72	0.9999
Frame Rate Change (30 fps to 15 fps)	50.52	0.9997	66.72	0.9999
Format Conversion (.avi to .mj2) with CR = 5	41.46	0.9710	41.56	0.9659
Format Conversion (.avi to .mj2) with CR = 10	39.78	0.9781	40.26	0.9598

(b) Structural Similarity Index Metrics (SSIM)

SSIM is used to measures the block-wise perceptual similarity between the frames of original and watermarked frames. Its value varies from -1 to +1, here +1 is the shows that the matched video frames are completely identical.

The SSIM between the original and watermarked frame can be obtained using Eqn.(5.4) as,

$$SSIM(O_f, W_f) = \left[\begin{array}{c} Lu_{com}(O_f, W_f)^{\delta}.Con_{com}(O_f, W_f)^{\eta} \\ .Str_{com}(O_f, W_f)^{\xi} \end{array}\right] \quad (5.4)$$

where Lu_{com}, Con_{com} and Str_{com} are the luminance comparison, contrast comparison and structure comparison function respectively. To adjust the relative importance of luminance, contrast or structure component, positive value δ, η and ξ parameters are always considered.

Wang et al. [41] defined the luminance comparison, contrast comparison and structure comparison given Eqn. (5.5) to Eqn. (5.7) as:

$$Lu_{com}(O_f, W_f) = \frac{2M_{O_f}M_{W_f} + K_1}{M_{O_f}^2 + M_{W_f}^2 + K_1} \quad (5.5)$$

where, $M_{O_f} = \frac{1}{N}\sum_{i=1}^{N} O_{f_i}$ and $M_{W_f} = \frac{1}{N}\sum_{i=1}^{N} W_{f_i}$ are the mean intensity of original and watermarked frame respectively. K_1 is small constant included to avoid instability when sum of $M_{O_f}^2$ and $M_{W_f}^2$ is very close to zero.

$$Con_{com}(O_f, W_f) = \frac{2\sigma_{O_f}\sigma_{W_f} + K_2}{\sigma_{O_f}^2 + \sigma_{W_f}^2 + K_2} \quad (5.6)$$

where, $\sigma_{O_f} = \left(\frac{1}{N-1}\sum_{i=1}^{N}\left(O_{f_i} - M_{O_{f_i}}\right)^2\right)^{1/2}$ and $\sigma_{W_f} = \left(\frac{1}{N-1}\sum_{i=1}^{N}\left(W_{f_i} - M_{W_f}\right)^2\right)^{1/2}$ are the estimated standard variations of original and watermarked frames, respectively. K_2 is small constant included to avoid instability when sum of $\sigma_{O_f}^2$ and $\sigma_{W_f}^2$ is very close to zero.

$$Str_{com}(O_f, W_f) = \frac{\sigma_{O_f W_f} + K_3}{\sigma_{O_f}\sigma_{W_f} + K_3} \quad (5.7)$$

where, $\sigma_{O_f W_f} = \frac{1}{N-1}\sum_{i=1}^{N}\left(O_{f_i} - M_{O_f}\right)\left(W_{f_i} - M_{W_f}\right)$ is the co-relation coefficient between original and watermarked frames. $K_3 = K_1/2$ is small constant with $K_1 << 1$ [41].

The summary of the PSNR value obtained by applying different types of attacks on the watermarked videos 'akiyo.avi' and 'pedestrian.avi' considering 'cameraman.tif' as watermark is shown in Table 5.1.3. Simulation results shown through PSNR and SSIM, explains that a good degree of transparency and similarity between the original and watermarked videos are obtained under all three categories of attacks which further verify the results obtained through subject assessment method. The embedded watermark signal remains invisible which proves that proposed method is useful for the applications where transparency of watermark is utmost requirement.

5.2 Testing for Robustness

Robustness is the ability of watermark signal to withstand against modification of signal due to simple and specialized watermark removal attacks. For testing the robustness property offered by the of the proposed scheme the watermarked videos are attacked through many attacks (as mentioned in Table 5.2) and then watermark extraction procedure is used for watermark extraction. Extracted watermark is then compared with the original watermark to calculate the loss of information. If the loss of information is less, then watermark is supposed to be resilient against the attacks. Correlation coefficient (*CC*) between the original and the extracted watermark is the popularly used metric. If the correlation between original watermark (*O*) and extracted watermark (*E*) is close to one than extracted watermark is identical to original and hence scheme is resilient against the attacks. Dropping of correlation value close to zero reflects that than extracted watermark is dissimilar to original watermark. Correlation coefficient is computed using equation (5.8).

$$CC = \frac{\sum_m \sum_n \left(O_{mn} - \bar{O}\right)\left(E_{mn} - \bar{E}\right)}{\sqrt{\left(\sum_m \sum_n \left(O_{mn} - \bar{O}\right)^2\right)\left(\sum_m \sum_n \left(E_{mn} - \bar{E}\right)^2\right)}} \qquad (5.8)$$

where, O and E are the image matrices of the same size, $\bar{O}$ and $\bar{E}$ are the mean value of the original and the extracted watermark signals respectively.

	
(a) Original Watermark Cameraman.tif	(b) Recovered Watermark after Frame Insertion Attack

Figure 5.2. Original and Extracted Watermarks under Frame Insertion Attack (Any Number of Frames).

Table 5.2 Correlation value between the original and extracted watermark under different types of attacks considering 'cameraman.tif' as watermark.

Type of Attack	***CC* Value**	***CC* Value**
	akiyo.avi	**pedestrian.avi**
No Attack	0.9751	0.9572
Rotation (0.5°)	0.9218	0.9261
Rotation (1°)	0.9189	0.9221
Uniform Cropping (5%)	0.9751	0.9489
Uniform Cropping (10%)	0.9751	0.9175
Salt & Pepper Noise (With a Density of 0.001)	0.8132	0.8246
Salt & Pepper Noise (With a Density of 0.0005)	0.8538	0.8863
Speckle Noise (With a Variance of 0.001)	0.8992	0.9036
Speckle Noise (With a Variance of 0.0005)	0.9203	0.9256
Frame Insertion (Any number of frames)	0.9751	0.9572
5% Frame Swapping	0.8167	0.8864
10% Frame Swapping	0.7984	0.8014
5% Frame Averaging	0.8248	0.8953
10% Frame Averaging	0.7932	0.8216
Frame Rate Change (30 fps to 45 fps)	0.9751	0.9572
Frame Rate Change (30 fps to 15 fps)	0.9751	0.9572
Format Conversion (.avi to.mj2) with CR = 5	0.9558	0.9536
Format Conversion (.avi to.mj2) with CR = 10	0.9216	0.9321

Table 5.3 Comparison of Robustness between Proposed Scheme and State-of-the-Art Schemes

S. No.	Type of Attack	Agarwal et al. [18]	Rakesh and Sarabjeet [29]	Proposed
1	Rotation 5°	X	0.8541	0.8987
2	Speckle Noise (Variance = 0.001)	0.8989	0.9404 (Var. = 0.003)	0.8992
3	20% Uniform Cropping	0.9617 (40 columns from either side)	0.8656	0.9751
4	Frame Insertion (Any number of frames)	0.9193 (Insertion of Ten frames)	0.9418 (Insertion of Thirteen frames)	0.9751

The summary of the correlation value obtained between the original and extracted watermark by applying different types of attacks on the two different resolution watermarked videos considering cameraman.tif as watermark is shown in **Table 5.2**. Results obtained through the simulation demonstrate that the proposed scheme is effective under all three types of attacks i.e., geometric, common signal processing and video specific attacks. Due to the use of normalization the embedded watermark survives against geometric and signal processing changes. For efficiently extract the watermark, precisely locating the interest features points where information was embedded earlier is necessary. Adoption of SURF method to locate the exact feature points, avoids the influence of various intentional or unintentional attacks. High value of correlation coefficient between the embedded and extracted watermark further verify the worth of the proposed scheme.

5.3 Performance Comparison of Proposed Scheme with State-of-the-Art Schemes

To further verify the performance, the proposed scheme is compared with the already available video watermarking schemes [18, 29] published in recent years. From Table 5.3, it has been evident clearly that proposed scheme offers superior results when compared to the state-of-the-art video watermarking schemes which are based on Harris corner detector and based on intra-coding process in MPEG-2 style. Some of the entries in the **Table 5.3** are marked as ‘x’ intentionally, as the authors of such schemes do not test their schemes on the mentioned attacks.

6. Conclusions and Future Work

In this paper we proposed a highly transparent and robust color video watermarking scheme based on normalization and selected feature points of speeded-up robust features. Scheme is tested by embedded a gray scale image as watermark. Problem of geometric synchronization is solved by choosing twenty five strongest interest points per frame that serves as points used to embed and extract the watermark. Use of less number of interest points reduces the effect of the embedded watermark on luminance component and makes the scheme transparent. The repeatability property of SURF in comparison to all available state-of-the-art schemes makes the proposed scheme relevant in video watermarking applications. The proposed scheme is tested against various geometric, simple signal processing and video specific attacks. Subjective and objective evaluation is performed to evaluate the transparency of the watermarked video. Mostly employed correlation coefficient metric is used to judge the robustness of the watermark under proposed scheme. Experimental and simulation results show that proposed scheme is resilience against most of the mentioned attacks.

The performance of the proposed scheme can be enhanced further if we use schemes that provides more robust feature points. As video provides high watermark payload capacity, so designing schemes that can embed the color pattern as watermark in the video is also a future research area. Multiple cascading attacks and collusion attacks may be some good candidates that can be used to test the robustness of the video watermarking schemes. Thus, in future we can develop better extraction scheme under various video specific intentional and unintentional distortions.

Acknowledgements

This work is supported by Moradabad Institute of Technology, Moradabad under its faculty research promotion scheme. The authors want to thanks to anonymous professors and industry experts for their contribution in subjective assessment of the scheme.

References

[1] N. Nikolaidis, I. Pitas, “Robust image watermarking in the spatial domain,” Signal Processing, Vol. 66, No.3, pp. 385-40, 1998.

[2] J. Abraham and Varghese Paul, “An imperceptible spatial domain color image watermarking scheme,” Journal of King Saud University-Computer and Information Sciences , Vol. 31, No. 1, pp. 125-133, 2019.

[3] A. K. Singh, N. Sharma, M. Dave and A. Mohan, “A novel technique for digital image watermarking in spatial domain,” *In*: Proceedings of 2nd IEEE International Conference on Parallel Distributed and Grid Computing (PDGC), pp. 424-429, 2012.

[4] Yu, Xiaoyan, Chengyou Wang, and Xiao Zhou, “A survey on robust video watermarking algorithms for copyright protection. Applied Sciences,” Vol. 8, No.10, pp. 1891, 2018.

[5] K. Swaraja, , G. Karuna, Padmavathi Kora, and K. Meenakshi, “Video Watermarking Fundamentals and Overview,” *In:* Proceedings of International Conference on Intelligent Computing and Communication Technologies Springer, Singapore, pp. 379-385,2019.

[6] Nin, Jordi, and Sergio Ricciardi, “Digital watermarking techniques and security issues in the information and communication society,” *In:* Proceedings of 27th International Conference on Advanced Information Networking and Applications Workshops, pp. 1553-1558, 2013

[7] Berghel, Hal, and Lawrence O' Gorman, “Protecting ownership rights through digital watermarking,” Computer, Vol. 29, No.7, pp. 101-103,1996

[8] S. S. Bedi, Rakesh Ahuja, and Himanshu Agarwal, “Copyright Protection using Video Watermarking based on Wavelet Transformation in Multiband,” International Journal of Computer Applications, Vol. 66 No. 8, 2013.

[9] Musrrat Ali, Chang Wook Ahn and Millie Pant, “A robust image watermarking technique using SVD and differential evolution in DCT domain,” Optik, Vol. 125 No. 1, pp. 428-434, 2014.

[10] HA Abdallah and S. A. Alsodairi, “A novel watermarking technique based on hybrid transforms,” International journal of

computer network and information security, Vol. 12, No. 3, pp. 316-325, 2020.
[11] Jin S. Seo and Chang D. Yoo, "Localized image watermarking based on feature points of scale-space representation," Pattern Recognition, Vol. 37 , No. 7, pp. 1365-1375, 2004.
[12] Qi. Xiaojun and Ji, Qi, "A robust content-based digital image watermarking scheme," Signal Processing, Vol. 87 No. 6, pp. 1264-1280, 2007.
[13] Kutter, Martin, Sushil K. Bhattacharjee and Touradj Ebrahimi, "Towards second generation watermarking schemes," *In*: Proceedings of International Conference on Image Processing (Cat. 99CH36348), pp.320-323, 1999.
[14] Yavuz, Erkan, and Ziya Telatar, "Improved SVD-DWT based digital image watermarking against watermark ambiguity," *In:* Proceedings of ACM symposium on Applied computing, pp. 1051-1055, 2007.
[15] Abraham, Jobin, and Varghese Paul, " An imperceptible spatial domain color image watermarking scheme," Journal of King Saud University-Computer and Information Sciences, Vol. 31, No. 1, pp. 125-133, 2019.
[16] Himanshu Agarwal, Manoj Kumar Singh, Rakesh Ahuja and Anurag Panday, "DCT-SVD Based Hybrid Digital Video Watermarking Scheme for Ownership Protection,"MITIJCSIT,Vol. 6, No. 1, pp. 37-41, 2016.
[17] Kapil Juneja and Shaloo Bansal, "Frame selective and dynamic pattern based model for effective and secure video watermarking," International Journal of Computing, Vol. 18, No. 2, pp. 207-219, 2019.
[18] Agarwal, Himanshu, Farooq Husain, and Praveen Saini. Next Generation Noise and Affine Invariant Video Watermarking Scheme Using Harris Feature Extraction. *In:* Proceedings of International Conference on Advances in Computing and Data Sciences, 2019. p. 655-665.
[19] Qingtang Su and Beijing Chen, "Robust color image watermarking technique in the spatial domain," Soft Computing, Vol. 22, No. 1, pp. 91-106, 2018.
[20] Naseem Asad and Ismail Shayeb, "A Modification of Least Significant Digit (LSD) Digital Watermark Technique," International Journal of Computer Applications, Vol. 179, No. 32, pp.19, 2018.
[21] Chih-Wei Tang and Hsueh-Ming Hang, "A feature-based robust digital image watermarking scheme," IEEE transactions on signal processing, Vol. 51, No. 4, pp. 950-959, 2003.
[22] Jiying Zhao, Zheng Liu and Robert Laganiere, "Digital watermarking by using a feature-based multiwavelet fusion approach," *In:* Proceedings of Canadian Conference on Electrical and Computer Engineering, pp. 563-566, 2004.
[23] Wei Lu, Hongtao Lu and Fu-Lai Chung, "Feature based watermarking using watermark template match," Applied Mathematics and Computation, Vol. 177, No. 1, pp. 377-386, 2006.
[24] Ayubi, Peyman, Milad Jafari Barani, Milad Yousefi Valandar, Behzad Yosefnezhad Irani, and Reza Sedagheh Maskan Sadigh, "A new chaotic complex map for robust video watermarking," Artificial Intelligence Review, pp. 1-44, 2020
[25] Jian Quan Kong and De Quan Zhou, "Digital Image Watermarking Based on Second Generation Wavelet Transform," In Applied Mechanics and Materials, Vol. 241, pp. 3026-3029, 2013.
[26] A. Umaamaheshvari and K. Thanushkodi, "A robust digital watermarking technique based on feature and transform method," Scientific research and essays. Vol. 8, No. 32, pp. 1584-1593, 2013.
[27] Xiang-Yang Wang, Ai-Long Wang, Hong-Ying Yang, Yan Zhang and Chun-Peng Wang, "A new robust digital watermarking based on exponent moments invariants in nonsubsampled contourlet transform domain," Computers & Electrical Engineering, Vol. 40, No. 3 , pp. 942-955, 2014.
[28] Ahuja Rakesh and S. S. Bedi, "Robust Video Watermarking Scheme Based on Intra-Coding Process in MPEG-2 Style," International Journal of Electrical & Computer Engineering, Vol. 6, pp. 2088-8708, 2017.
[29] Rakesh Ahuja and Sarabjeet. S. Bedi, "Video watermarking scheme based on candidates I-frames for copyright protection," Indonesian Journal of Electrical Engineering and Computer Science, Vol. 5, No. 2, pp. 391-400, 2017.
[30] Boris Escalante-Ramírez and S. L. Gomez-Coronel, "A perceptive approach to digital image watermarking using a brightness model and the hermite transform," Mathematical Problems in Engineering, pp. 1-19. 2018.
[31] Wang Xiaohong and Zhao Rongchun, "A new method for image normalization," *In:* Proceedings of International Symposium on Intelligent Multimedia Video and Speech Processing, (IEEE Cat. No. 01EX489), pp. 356-359, 2001.
[32] T. Lindeberg, "Feature detection with automatic scale selection," International journal of computer vision, Vol. 30, No. 2, pp. 79-116, 1998.
[33] David G. Lowe, "Object recognition from local scale-invariant features," *In:* Proceedings of the seventh IEEE international conference on computer vision, pp. 1150-1157, 1999.
[34] Herbert Bay, Andreas Ess, Tinne Tuytelaars and Luc Van Gool, "Speeded-up robust features (SURF)," Computer vision and image understanding, Vol. 110, No. 3, pp. 346-359, 2008.
[35] P. Burt, E. Adelson, "The Laplacian pyramid as a compact image code," IEEE Trans. on Communications, Vol. 31, No.4, pp. 532–540, 1983.
[36] Matthew Brown and David G. Lowe, "Invariant features from interest point groups," BMVC, Vol. 4, No. 253-262, 2002.
[37] Gaurav Bhatnagar, Q. M. Jonathan Wu, Balasubramanian Raman, "Robust gray-scale logo watermarking in wavelet domain," Computers and Electrical Engineering, Vol. 38, No. 5, pp. 1164-1176, 2012.
[38] Narong Mettripun, Thumrongrat Amornraksa and Edward J. Delp, "Robust image watermarking based on luminance modification," Journal of Electronic Imaging, Vol. 22, No.3, pp. 33-39, 2013.
[39] Marini, Enrico, Florent Autrusseau, Patrick Le Callet, and Patrizio Campisi, "Evaluation of standard watermarking techniques," In Security, Steganography, and Watermarking of Multimedia Contents, Vol. 9,pp. 6505-24, 2007.
[40] Xinhong Zhang, Fan Zhang and Yuli Xu, "Quality evaluation of digital image watermarking," *In:* Proceedings of International Symposium on Neural Networks, Springer, Berlin, Heidelberg, pp. 241-250, 2011.
[39] Z.Wang, A. C. Bovik, , H. R.Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," IEEE Transactions on Image Processing, Vol. 13, No. 4, pp. 600-612, 2004.

Note:
The concept and mathematical foundation of Frame Normalization and SURF techniques have connections with many other developments. We have only mentioned the details related to the development of our method. All developments are pointed out in the original papers which leads to a bibliography of large set of items.