No name manuscript No.
(will be inserted by the editor)

Lightweight Human Skin Encryption for Public Safety in Real-time Surveillance Applications

Amna Shifa · Muhammad Babar Imtiaz · Mamoona Naveed Asghar · Martin Fleury

Received: date / Accepted: date

Abstract Individual's privacy protection is the concerning issue in surveillance videos. Existing research work for individual's identification on the bases of their skin detection is focused either on different human skin detection techniques, or on protection. This research paper considers both lines of research and proposes a hybrid scheme for human skin detection and protection by utilizing color information in dynamically varying illumination and environmental conditions. For the purpose, the dynamic and explicit skin detection approaches are implemented simultaneously considering the multi-color-space i.e. RGB, perceptual (HSV) and orthogonal (YCbCr) color-spaces and then the human skin is detected by the proposed Combined Threshold Rule (CTR) based segmentation by considering the advantages of three multi-color-spaces. The comparative qualitative and quantitative detection results with average 93.73% accuracy imply that the proposed scheme gain considerable accuracy without incurring the training cost. Secondly, once skin detection has been performed, the detected skin pixels (including false positives) are encrypted with state-of-the-art Advance Encryption Standard with Cipher Feedback Mode (AES-CFB) rather than applying selective encryption on complete video. The proposed scheme preserves the behavior of the subjects within the video, hence can be useful for further image processing and behavior analysis and if required can be decrypted by the authorized user. The experimental results show that average encryption time is 8.268 sec and Encryption Space Ratio (ESR) with an average 7.25 % for HD cricket video (1280 x 720 pixels/frame) strongly imply that to protect a person within a video, the method of encrypting skin detection is preferable. Thirdly, performance comparison between the proposed method in term of Correct Detection Rate (CDR) with an average 91.5%, RGB with 85.86%, HSV with 80.93% and YCbCr with an average 84.8% imply that proposed method has high potential to detect the skin accurately. Furthermore, the security analysis performed confirms that proposed scheme could be a suitable choice for real-time surveillance applications working on resource constrained devices.

Keywords: Color-spaces; Human skin detection; Parallel processing; Privacy protection; Segmentation; Selected pixel encryption; Selective encryption

Amna Shifa
The Islamia University of Bahawalpur, Punjab, Pakistan.
E-mail: amnashifa@yahoo.com
Muhammad Babar Imtiaz
Software Research Institute, Athlone Institute of Technology, Ireland.
E-mail: b.imtiaz@research.ait.ie
Mamoona Naveed Asghar
The Islamia University of Bahawalpur, Punjab, Pakistan.
Software Research Institute, Athlone Institute of Technology, Ireland.
E-mail: masghar@ait.ie
Martin Fleury
School of CSEE, University of Essex, Colchester, Essex, UK.
E-mail: fleury.martin55@gmail.com

1 Introduction

Owing to the high complexity and the non-uniform characteristics of skin such as varying skin tone colors within different regions of the body, outdoor and indoor environmental factors such as illumination contrast, human skin detection in videos is still a changing and developing task. The basic idea behind the research is to set apart pixels containing human skin tones from those that do not. Human skin detection plays an indispensable role in various important image-processing applications such as in: Content Based Image Retrieval (CBIR) systems; face detection; face tracking; and human-computer interaction. As a result, the ability to detect human skin is a significant requirement in different disciplines, for instance within medical systems; defense systems; and robotics. Along with the detection of prominent features through their shape, skin tone can be provided as an additional piece of information to establish accurate face detection [1-3]. Therefore, plethora of skin detection algorithms are proposed aiming to gain more accuracy [4-9]. Although, recent proposed machine learning and convolutional neural networks (CNNs) based approaches for skin detection in a narrow set of imaging conditions outperform the conventional approaches in accuracy, but such approaches demand more resources and greater execution time [10]. Thus, these techniques along with the security measures make them unsuitable choice for the real-time applications, particularly, in current resource constrained Internet of Things (IoT) environment, which requires more simple and efficient solutions. As a result, in this work color-based segmentation for the skin detection is considered as most appropriate method of detection along with the security measure for the real-time surveillance application.

It is well established fact that the color-based segmentation is a powerful approach for skin classification, but due to its simplistic implementation it is considered to lack the robustness compared with the other advance machine learning and Convolutional Neural Network (CNN) approaches. However, these simple approaches along with the security implementation such as encryption are suitable for real-time surveillance applications and particularly emerging Internet of Multimedia Things (IoMT). The skin-color information has proven to be an effective tool if the process of representation, modeling, and classification of skin-color pixels is carefully performed. Though the majority of the research in this domain focuses on visible spectrum imaging, it remains a challenging task to detect skin in the visible spectrum due to the numerous factors that arise because of camera traits (e.g. illumination level, spectral reflectance and the sensor's sensitivity), acquisition conditions (e.g. variance in illumination levels, background, motion), interpersonal feature variations (e.g. diverse ethnic group with various skin color tones, age and gender), intra-personal feature variations (e.g. makeup, hairstyle, pose, expressions and so on). Besides, these factors, in the real-time video surveillance, the shape, size and appearance of the subject and Field of view (FOV) (e.g. distance, direction) as illustrated in Fig.1 (shown with 3 cameras) can affects efficacy of the detection of even advanced CNN based skin detection algorithms and ultimately protection schemes. The task become more challenging when the diverse facial expressions change the appearance and shape of the subject of interest and features may not be identified or incorrectly identified by the feature detection algorithms and thus, privacy could be compromised [11]. These aforementioned factors that can degrade performance of detection and consequently privacy scheme are described below with the examples.

Use case 1: Suppose the eyes are considering for localization in the feature detection algorithm (e.g in the dynamic skin detection algorithms), and for example a person is laughing or screaming and in that case, laughing/screaming may cause his/her eyes to close or deform the shape of eyes. Similarly, in another example the person may wear the glasses and due to this eyes can be occluded by glasses, hair and shad. Hence in the both cases the eyes coordinates will not be captured or incorrectly captured thus the detection algorithm will fail to detect the person and subsequently skin pixels of that person cannot be detected.

Use case 2: In another scenario the person may be using the cellphone and his/her face could be downward hence the eyes coordinated or even face coordinated cannot be captured by the

detection algorithm. Similarly, the angle of subject such as tilted, side angle of faces may degrade the performance of algorithms.

Use case 3: In the third use case, 1 mega pixels camera are considered to monitor the surveillance site. The 1 mega pixels capture the video in low quality hence the feature could not be clearly detected by the detection algorithms. Moreover, ambient environment factors such zooming or blurring can significantly impact on the performance of feature detection algorithm and consequently the protection of individual's privacy. Thus in that case the explicit threshold detection will overcome the miss detection rate.

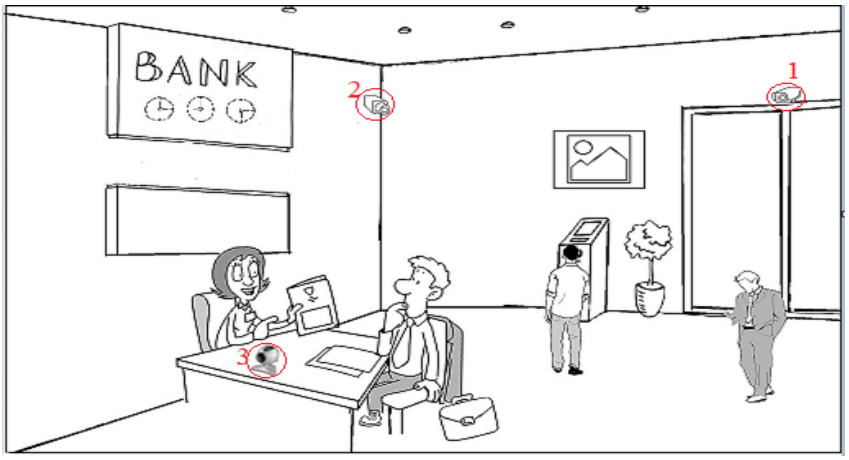


Fig. 1 Real-time scenario for surveillance videos

From the foregoing lines, one can conclude that it is generally very difficult for a single approach to achieve high performance in both robustness and accuracy. Thus, to avoid the problems of features localization and reduce the miss rate, in this paper adoptive color based skin detection scheme is employed by considering the dynamic and explicit thresholding skin detection methods to take the advantage of both and gain the maximum privacy protection in real-time surveillance application. Therefore, if feature cannot be detected the explicit threshold skin detection methods take into account to minimize the miss-rate and achieve robustness.

Security considerations are not new in the context of multimedia applications, however, the attributes of current pervasive and integrated environment presenting new security challenges. Over the past few years, many methods aimed to ensure data confidentiality of the visual data have been proposed by the researchers such as full scrambling (naive encryption) [12]. However, video surveillance systems augmented with the latest computer vision technologies equipped with Artificial Intelligence (AI) for individual's identification such as face recognition [13], activities tracking [14] such as kids monitoring (e.g. fall detection) [15] and abnormal behavior identification (e.g. robbery or fight between two people) [16] severely undermine the right of privacy of people [17]. Moreover, for large scale surveillance systems, it is unrealistic to fully encrypt and transmit the raw video data collected by surveillance nodes to a central server for processing due to huge amount of data and high communication bandwidth. This has led researchers to focus on the selective encryption (SE) to sufficiently secure the video content [18]. Nevertheless, privacy protection of those under surveillance has become an important safeguard [19-21]. Unfortunately, these approaches do not pay attention to the intelligibility, reversibility and hardware architectures with memory constrained and lower bandwidth altogether along with the security. The weakness of some methods other than encryption such as pixelation and blur are highlighted in [22]. Therefore, to alleviate the aforementioned limitation, in this work pre-processing is performed and only extracted features are transmitted to the central server where some data fusion and data association algorithm are running for further processing. The basic idea behind this research is to set apart pixels containing human skin tones from those that do not and encrypt these pixels with the standardized encryption algorithm AES to provide the sufficient security with low complexity with the aforementioned features for constrained visual devices such as Raspberry Pi cameras. Furthermore, the proposed methods will allow monitoring of activities and events within the video without revealing the identities of individuals. But, if required, the

protected region can also be decrypted with the secret key by the authorized user.

1.2. Contributions and Structure of Paper

In this paper, we propose detection, discrimination, and encryption of human skin within video when skin tone colors vary. This is done by transforming the frames within a video into RGB (red, green, blue), perceptual (Hue, Saturation, Value — HSV) and orthogonal (Luma, blue difference, red difference chroma components — YCbCr) color-spaces. The design goals of our proposed scheme are: (1) Efficient ROI (in our case skin) detection with proposed CTR segmentation by employing dynamic and explicit skin sampling; the explicit method is employed because the dynamic methods based on the feature localization require the space and geometry information of ROI along with its shape and size as discussed in section 1(see use case 1) suffers in term of efficiency of the detection thus explicit thresholding method resolve the problems as discussed in section 1 (use case,1,2 3), (2) Ensure the privacy of individuals within the surveillance video with behavior preservation so that further image processing algorithms could be performed on protected videos, (3) Achieve the higher efficiency in term of reduced encryption time and bitrate overhead, and (4) Reversibility so that if the authorized or law enforcement person want to view the encrypted region , they can extracted the protected region. The manifold contribution of the paper is elaborated as:

i. Privacy of individuals has been protected by encrypting the human skin pixels in such a way that the video remains watchable and intelligible for further processing but no one can breach the privacy of individuals.
ii. In order to protect the privacy, the human skin (including false positive) is encrypted with a state-of-the-art Advanced Encryption Standard with Cipher Feedback mode (AES-CFB) as a stream cipher.
iii. The protected video could be decoded with the decryption key by the authorized user to view the original content of the protected video for crime investigation.
iv. Our approach makes use of two skin detection schemes, dynamic and explicit for the skin pixels sampling, each of which suppresses the appropriate attributes while providing the ability to detect the maximum detection rate.
v. Combined Threshold Rules (CTR) based skin segmentation method has been proposed by utilizing the RGB, perceptual (HSV) and orthogonal (YCbCr) color-spaces to improve the human skin detection and hence the protection and robustness. The optimized threshold values for the explicit thresholding method are proposed for better skin pixel detection.
vi. Performance of three color-spaces and the proposed CTR based skin segmentation on different videos having camera traits, FOV, people of different skin color tone and illumination condition have been evaluated.
vii. The qualitative and quantitative comparison of proposed CTR with existing techniques is given in Table 1 and Table 3 (Fig. 10) respectively. The comparative analysis along with security analysis prove the significance of proposed skin encryption schemes.

The remainder of this paper is organized as follows. Section 2 provides related work and background overview in the area of skin detection and privacy protection schemes. Section 3 then details the research methodology and the proposed scheme. Section 4 provides empirical results and provides analysis of those results. Finally, Section 5 contains some concluding remarks and considerations for future research.

2 Context

This section provides an overview of this research area, particularly in respect to common color-spaces. In addition, in Section 2.3 provides a review of related research in privacy protection including research very recently published.

2.1. Skin Detection

2.1.1 Spectral range

It is a fact that assessing the human skin of peoples within the real-time surveillance

application remains a challenging task owing to acquisition conditions (e.g variant illumination and environmental conditions). The hitches and complications arising from visual spectrum imaging can be avoided by substituting the visual spectrum by imagery in the non-visual spectrum such as infrared [23] or through multi-spectral imaging [24]. Deploying in the non-visual spectrum reduces much of the overhead of detection. However, aside from this partial independence, deployment in the non-visual spectrum requires expensive sensors and demanding procedures, which has limited its scope to real-time applications. Therefore, in this paper, the focus will be on skin detection in the visible spectrum only.

2.1.2 Types of detections

The skin detection procedures can be further categorized into two major categories: region-based procedures and pixel-based procedures. Region-based procedures, also often known as Region of Interest (ROI) based methods, solely depend upon the spatial positioning of adjacent pixels, thus adding by their proximity to the skin or non-skin pixel classification. In ROI-based methods, some supplementary information such as texture is also required. ROI-based methods are actually primarily based on pixel-based procedures. In pixel-based approaches, each individual pixel is classified as a skin or non-skin pixel without taking neighboring pixels into account. As surveillance video is one application of the current paper's methods, ROI-based methods for skin detection are not an optimal choice. The reason for this is that multiple ROIs cannot normally be operated upon within a single video. Owing to this issue, pixel-based skin detection procedures were selected. The commonly used methods for pixel-based skin detection are parametric, non-parametric and explicit skin clustering. In parametric methods, Gaussian color distribution is used to distinguish between skin and non-skin color pixels. In non-parametric method, skin color distribution within multimedia data (image/video) is estimated by training dataset without using any explicit model for skin color such as Self Organizing Map (SOM) classifier, Bayes classifier, and histogram-based nonparametric skin model. The explicit skin clustering is used to explicitly establishing the boundaries between the skin pixels and non-skin pixels through multiple pre-defined threshold values in certain color-spaces [25]. The existing fixed human skin detection techniques have a greater probability of false positive skin region detection. Therefore, adaptive skin color models are constructed for robust skin pixel detection. The authors of [26] present a rule-based skin detection method in the YCbCr color-space. Correlation rules, under various illumination conditions, classify the skin pixels in the YCb and YCr sub-spaces into skin and non-skin pixels. The proposed method, which also considers the luminance component, depends on the size and shape of skin clusters computed dynamically by a statistical method. In [27], Dadgostar et al. achieve real-time skin detection by means of an adaptive hue threshold algorithm. They chose the hue channel for the skin segmentation because it is not easily affected by variation of environmental illumination intensities and skin tone. Naji et al. in [28] presents a sophisticated skin-pixel clustering method based on dynamic thresholding in HSV color-space. For efficient and accurate skin detection, skin segmentation is achieved under some constrained to reduce various illumination condition effects and image background complexity. In [29] authors suggested an automatic color-space switching system for better skin classification of skin color pixels in varying lighting and illumination conditions. They proposed three different algorithms based on Bayesian approaches to discriminate the skin and non-skin pixels in different color-spaces. However, the experimental results are achieved on color images only. The YCgCr color-space is a revised version of YCbCr utilized by the authors of [30]. The only variation it has with the previous version is that it replaces the blue difference (Cb) with the green difference (Cg), which has indeed improved the detection performance. However, in this paper we have considered the three RGB, HSV and YCbCr color-space because of their widespread adoption.

2.2. Choice of color-space

The 'choice of color-space' is the most immediate decision of a designer of a surveillance system involving skin detection, especially in situations where there are likely to be people of different ethnicity, such as at an international airport. In most cases, the default color-space is the well-known Red Green Blue (RGB), which can be converted to any other color-space via linear or non-linear transformations with the intention of diminishing the overlap between skin and non-skin pixels. For skin detection, it is common practice to increase the range of the luminance component because empirical observations imply that skin colors vary more in intensity rather than

in chrominance. Various color-spaces have been proposed and utilized in skin detection. The ones used most for skin detection are now reviewed.

2.2.1 RGB basic color-space

For the sole purpose of collecting and representing digital images, RGB is the most frequently utilized color-space because normally cameras export captured data as RGB. RGB is comprised of three primary colors, red, green and blue respectively. In another version of RGB, to diminish the effect of illumination, all the color components are normalized, causing the sum of all normalized components to be unity (r+g+b=1) using the following equations [31]:

$$r = Red/(Red + Green + Blue) \quad (1)$$
$$g = green/(Red + Green + Blue) \quad (2)$$
$$b = green/(Red + Green + Blue) \quad (3)$$

It is well-known that, under specific assumptions, variations in skin color pixels due to illumination levels and ethnicity are minimized when utilizing the normalized RGB color-space. Owing to those merits, RGB has been a well-recognized adoption for skin-detection, being utilized by [32] from many. On the other hand, as there lies a maximum association between the color components of RGB, according to the previous equations, due to intermixed chrominance (color information) and luminance (brightness measurement), the resulting data are the least recommended for color tone analysis and identification [29, 33].

2.2.2 Perceptual Color-spaces (HSV)

The RGB color-space entirely lacks the ability to distinguish perceptual characteristics such as hue, saturation, and intensity. Hue (H) represents the color that dominates within a region or image. Saturation (S) defines the 'thickness' of the color, or one can say that it measures the brightness to colorfulness ratio. The intensity (V or I) is directly related to the amount of luminance. An RGB to HSV color-space transformation is achieved with various non-linear transformations such as in [34]. Such a transformation is:

$$Hue(H) = arccos \frac{\frac{1}{2}(2Red - Green - Blue)}{\sqrt{(Red - Green)^2 - (Red - Blue)(Green - Blue)}} \quad (4)$$

$$Saturation(S) = \frac{MAX(Red,Green,Blue) - MIN(Red,Green,Blue)}{MAX(Red,Green,Blue)} \quad (5)$$

$$Intensity\ Value(V) = MAX(Red, Green, Blue) \quad (6)$$

where Red, Green, and Blue are the original RGB values.

Raised intensities from ambient lighting and surface alignments in respect to light sources do not have an impact on RGB to HSV transformations. Hence, HSV is one of the most suitable color-spaces for skin detection. However, in practical terms, computation of this transformation can be time-consuming. Furthermore, when there are large swings in color information, high- and low-intensity level pixels will be ignored. To address this, issue orthogonal color-spaces are utilized, as is considered next. The HSV color-space has been employed by [35] from many.

2.2.3 Orthogonal Color-spaces (YCbCr)

These color-spaces extract independent components from RGB color channels. In particular, luminance and chrominance components are formed from RGB color channels in order to represent the visual information. The intensity of light is represented by luminance (Y) and chrominance is found by calculating the blue (Cb) and red (Cr) difference relative to luminance. The YCbCr color-space is recommended for skin detection due to the independence of its components. As YCbCr

space is one of the most accepted options for skin detention, it has been utilized by [28, 36] to cite a few from many. The values for the luminance (Y) and chrominance (Cb, Cr) components can be calculated as follows [35]:

$$Y = 0.299.Red + 0.587.Green + 0.114.Blue \quad (7)$$

$$\mathrm{Cb} = \mathrm{Blue} - \mathrm{Y} \quad (8)$$

$$\mathrm{Cr} = \mathrm{Red} - \mathrm{Y} \quad (9)$$

where Red, Green, and Blue are the original RGB values.

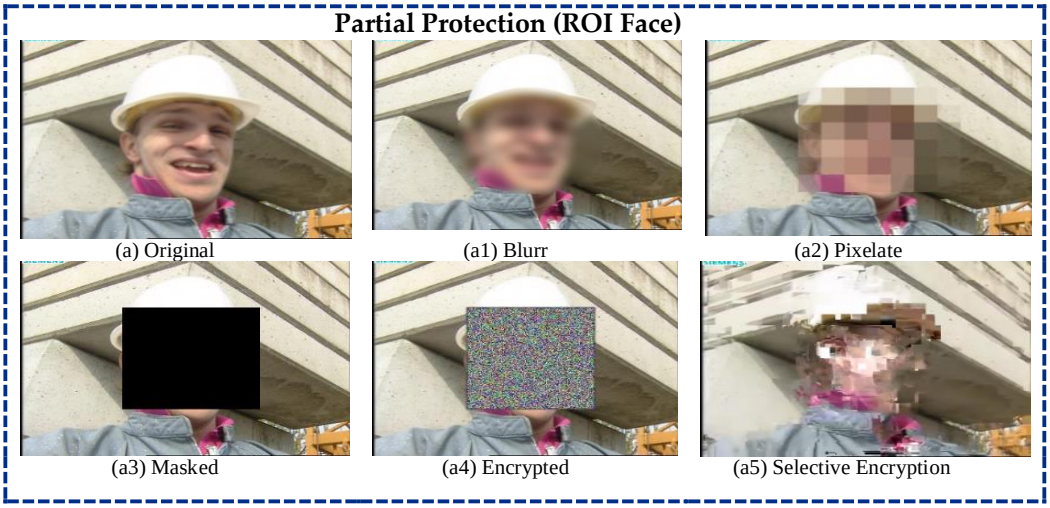
2.3. Privacy Protection

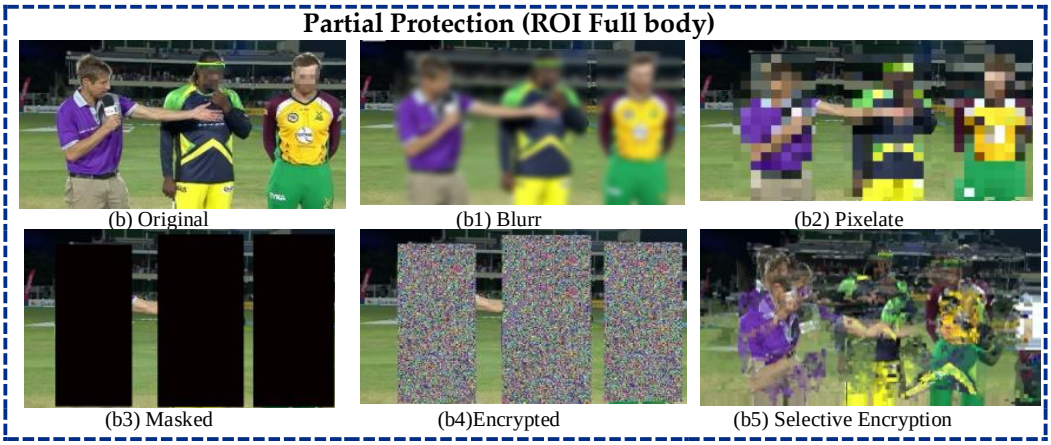
To mitigate the security and privacy concerns in video surveillance systems, most of the existing work targets ROI based partial encryption with the mild obfuscation techniques such as blurring or pixilation to ensure the privacy protection [21]. However, such approached cannot withstand privacy attacks [37], but still considered favorable for the real-time privacy protection in public places [38]. However, authors of [39] suggested the suitability of scrambling over afore mentioned simple techniques. Furthermore, based on existing literatures, the ROI consider for the privacy protection can be human face as illustrated in Fig.2(a), human body as illustrated in Fig.2(b) or whole frame as illustrated in Fig.2(c). In [40] the authors applied the complete encryption on the face region to destroy the ROI so that no one can reveal the identity of the person in surveillance. Though the proposed scheme does not scramble all the information however it complete destroy the ROI hence behavior of the person cannot remain perceptible. In contrast our scheme preserves the structure of protected sensitive region hence can be used for behavior analysis without revealing the identities of people. The authors of [41] considered the human (full body) as ROI for privacy protection. In the proposed scheme ROI protected it with the blurring to preserve the human behavior without identifying the person however, the proposed method is irreversible. Hence, if required the information protected cannot be recovered by the authorized users for future use. The authors of [42] achieved the privacy protection with the blurring on the whole frame. Recently, the authors of [43] presented a privacy filter framework in which human skin regions is detected by incorporating various state-of-the-art skin detection methods and detected skin region are removed from the video to achieve the privacy protection. In [44] the authors develop on-board digital signal processor PrivacyCam which encrypts the human faces for privacy preservation. In this system the human faces are detected by utilizing the background subtraction algorithms which are computationally complex. In contrast our proposed method uses the simple combined colors spaces threshold based skin detection methods, thus faster and cost effective and can be adopted for the constrained devices too. The authors of [45, 46] also employed the face obfuscated using state of the art face recognition algorithms for privacy protection.

Although most of the research has been focus on face detection and considered face de-identification sufficient to preserve privacy of individuals but skin detection in general may be an even stronger method. According to [47] face obscuring is not sufficient for the individuals' privacy protection, thus in this work encryption is employed over the human skin. A privacy mode for the camcorders in which skin color of individuals are manipulated and automatically obfuscating the defined region of interest (ROI) to avoid the ethnicity-based discrimination was patented by Sony [48]. Another work the authors of [49] proposed a privacy protection scheme in which ROI was removed by replacing it with the background pixels after that these pixels are embedded back with the reversible data hiding technique in the bitstream. However, this scheme incurs high bitrate overhead and degrades the quality of video hence couldn't be the efficient solution for the real-time surveillance system. In contrast, our approach has a negligibly small overhead and no quality degradation outside the ROI. Recently, the authors of [50] proposed reversible privacy protection for static images in which the original color information of entire frames is replaced with some other color palette information called false colors. It is a reversible technique and the original colors are reversed back to the original.

The comparative overview of the proposed method with the exiting skin detection and protection schemes is summarized in Table 1. Table 1 shows that most of these schemes either

propose skin detection or protection but do not combine together. The GPU-based, connected component labeling algorithm proposed in [51] provides adaptive skin detection for people with various skin colors but, alas from our point of view, did not make a performance comparison for different skin tones. The authors of [52] contributed a robust and computational cost efficient skin detection algorithm to detect different skin types under a variety of illumination conditions and backgrounds but privacy concerns were outside their remit. Our proposed scheme focuses on both approaches: Skin detection and privacy protection. This research also provides an empirical performance comparison of the algorithm developed in the RGB, YCbCr, HSV color-spaces with the proposed CTR scheme, as well as with selective encryption in general. An application for which these findings could be applied to is video surveillance at an international airport, when there is also an obligation to protect the privacy of those under surveillance. However, the research is widely applicable to video surveillance in general.





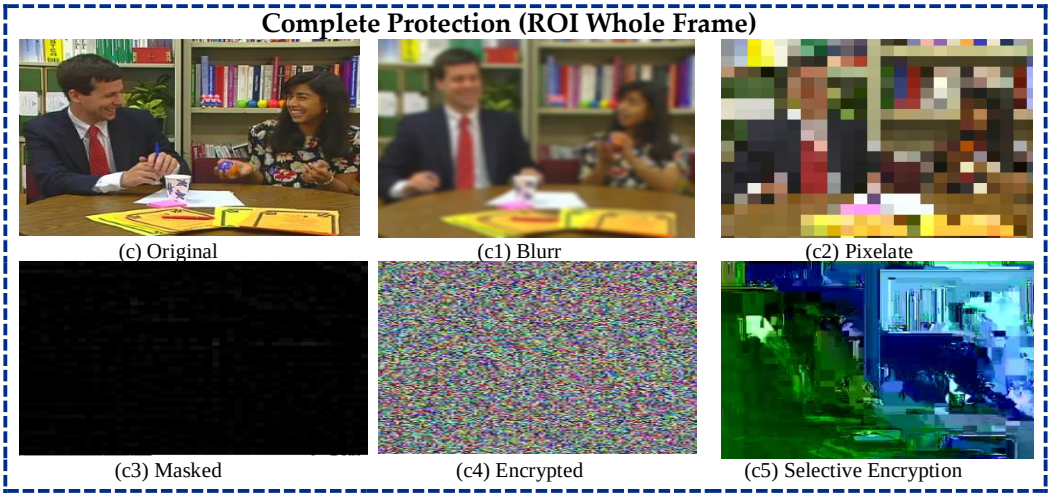


Fig 2. Sensitive ROI for privacy protection (a) Human Face as sensitive ROI (b) Human full body ROI (c) Full frame ROI

Table 1. Comparative overview of Proposed Scheme with existing schemes

Sr.No	Proposed Scheme	ROI	Skin Classifier	Selected Color-space			Skin Detection Methods	ROI Detection Approach	Skin Detection Model	Coding Standards/ Video Format	Protection Technique	Encryption Applied	Encryption Algorithm
				RGB	HSV	YCbCr							
1.	S. Bianco, F. Gasparini, R. Schettini [7]	Face+ Body	Pixel Based+ Face and Body based	No	Yes	No	Explicit Skin Cluster Definition Method	Adaptive Single Gaussian (ASG)	Multiple Model	Not Specified	No	No	No
2.	WL. Hoo, A. Miron , et al. [43]	Person+ +Motion +Skin	Pixel Based	Yes	No	No	Parametric	Fusion-based Algorithm + Random Forest (RF)	Multiple Model	HD	Yes (privacy Filters) Removing skin from the video	No	No
3.	W. Song, D. Wu ,et al. [51]	Face and Hand Gesture	Pixel Based + Motion Based	Yes	No	No	Non-parametric	Threshold-based Segmentation + GPU-based Connected Component Labeling Algorithm	Multiple Model	Not Specified	No	No	No
4.	S. Bilal, R. Akmeliawati, et al. [52]	Face or Hand	Face and Hand Based	Yes	Yes	Yes	Non-Parametric	Haar-like Features and Ada Boost algorithm	Appearance-based method	AVI	No	No	No
5.	J. Guo, J.Xu, J. Bao [53]	Human Face	Pixel Based	No	No	Yes	Parametric	Gaussian Distribution	Gaussian Model	H.264	Yes	Selective Video Encryption	Exclusive-or (XOR)
6.	**Proposed Scheme**	**Human Skin**	**Pixel Based**	**Yes**	**Yes**	**Yes**	**Dynamic and Explicit Skin Cluster Definition Method**	**Combined Threshold Rules Based Segmentation (CTR) (Proposed)**	**Adoptive (Dynamic and Explicit)**	**CIF,QCIF, HD**	**Yes**	**Specific Human skin pixel Encryption**	**AES-CFB**

3 Methodology

The proposed scheme consists of two modules 1) Extraction Module and 2) Encryption module. After extracting the skin, the detected skin pixels (including false positives) are encrypted with a state-of-the-art Cipher Advanced Encryption Standard with Cipher Feedback mode (AES-CFB) in order to protect the privacy. The proposed scheme comprised of following two modules is illustrated in Fig.3. Implementation of each module and their functionalities are described in detail below.

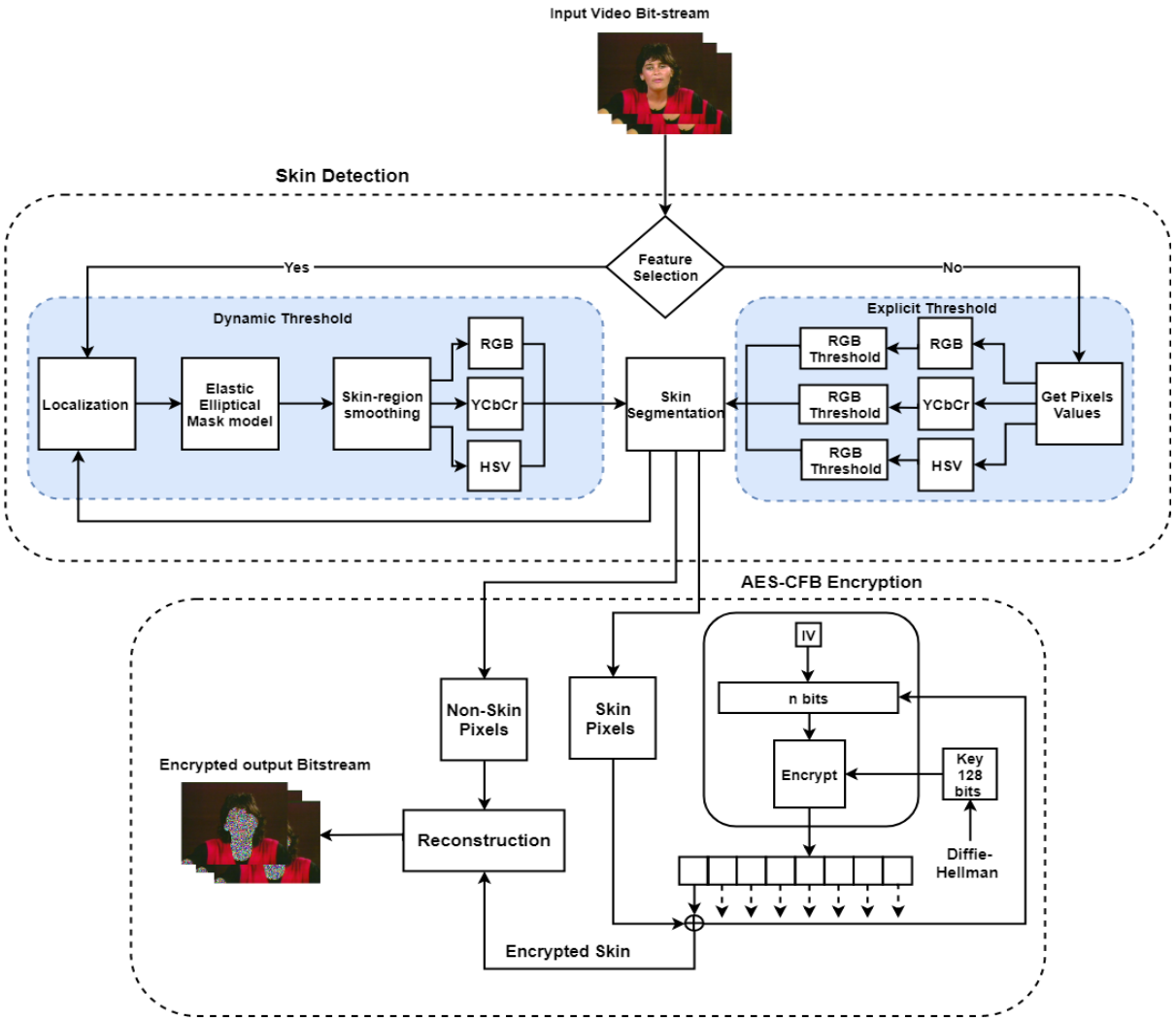


Fig. 3 Architecture of proposed privacy protection scheme for surveillance videos

3.1. Module 1: Skin detection

This module consists of three phases. (1) Preprocessing (2) Skin sampling (3) and (3) CTR based skin segmentation.

3.1.1 Preprocessing

In this phase, at the first, the input video sequences were transformed into the RGB, HSV and YCbCr color-space. However, from the point-of-view of usage of image-processing software libraries, transformation to one of those color-spaces cannot be a direct operation. Firstly, the original video sequence is transformed to YUV, which itself is a color representation that takes account of the human response to color. Furthermore, in this phase human faces are detected with

the Viola–Jones (V–J) face detector [54]. The V–J cascades a sequence of classifiers based on a set of Haar-like features. If the faces are detected within the frame, then the dynamic thresholding method will be considered for the skin detection otherwise explicit thresholding will work.

3.1.2 Skin Sampling

In the proposed methods for skin sampling two schemes are employed. The dynamic scheme is deployed to achieve higher detection accuracy due to their feature of dynamically adaptability to the changes of skin tones because of the environmental conditions. The explicit method is employed because the dynamic methods require the space and geometry information for accurate detection along with its shape and size as discussed in section 1, hence suffers in term of efficiency of the detection (falsely detection or miss detection). However, the problems identified in the section 1 (Use case 1-3), are solved by employing the explicit detection method for skin.

- Dynamic Sampling

Once the face within the frame is found, firstly, to get the color information of the detected face region, eye coordinated are calculates using the Machine Perception Toolbox as described in [54]. However, detected face region contains pixels that are not belong to skin colour such as pixels of eyebrows, lips and mouth area, which can lead to generate the false skin color threshold range in the respective color-space. Thus, to obtain the dynamically threshold ranges (max-min) that fall within the detected face region for each color-space (RGB, *HSV* and *YCbCr*), online dynamic approach given in [55] has been performed on each frame separately. Following the online sampling approach [55], after localization, elliptical face boundary is generated using the elliptical mask model which is followed by skin-region smoothing. The skin-region smoothing has been performed using Sobel detector to further filter out non-skin pixels. The elliptical mask model and sobel filters are considered due of their fast execution and simplicity. Further the detected edge pixels are to expand using dilation operation. Afterwards, the resultants obtained region is converted into *RGB, YCbCr* and *HSV* colors space as discussed in section 2.2. However, in this work trivariate histogram is considered with the smoothing densities. Let, *D(m,n)* is the color vector of the *(m,n)th* pixel of the detected face region *D*. The corresponding three variables for each color-space are *R, G* and *B, H, S* and *V* and *Y, Cb* and *Cr* respectively. Then, *R(i,j), G(i, j)* and *B(i,j) , H(i,j), S(i, j)* and *V(i,j)* and *Y(i,j), Cb(i, j)* and *Cr(i,j)* are the color intensities values of (m,n)th pixel for each color component of respective color-space. Hence,

$$D(m,n) = R(i,j), G(i,j), B(i,j) \tag{10}$$
$$D(m,n) = H(i,j), S(i,j), V(i,j) \tag{11}$$
$$D(m,n) = Y(i,j), Cb(i,j), Cr(i,j) \tag{12}$$

The corresponding histogram of D be denoted by *DHist and h1(a1, b1*, c1) indicate the number of pixels having the R value as a1, G value as *b1* and B value as c1, *h2(a2, b2*, c2) indicate the number of pixels having the H value as a2, S value as *b2* and V value as c2 and *h3(a3, b3*, c3) indicate the number of pixels having the Y value as a3, Cb value as *b3* and Cr value as c3. After that the smoothing operation is performed for every (*i,j,k*) of *DHist*, to normalized the histogram as:

$$\big(h1(a1, b1, \mathrm{c1})\big)` = \frac{1}{27} \sum_{i=a1-1}^{a1+1} \sum_{i=b1-1}^{b1+1} \sum_{i=c1-1}^{c1+1} \big(\mathrm{DHist}(i,j,k)\big) \tag{13}$$

$$\big(h1(a2, b2, \mathrm{c2})\big)` = \frac{1}{27} \sum_{i=a2-1}^{a2+1} \sum_{i=b2-1}^{b2+1} \sum_{i=c2-1}^{c2+1} \big(\mathrm{DHist}(i,j,k)\big) \tag{14}$$

$$\big(h1(a2, b2, \mathrm{c2})\big)` = \frac{1}{27} \sum_{i=a3-1}^{a3+1} \sum_{i=b3-1}^{b2+1} \sum_{i=c3-1}^{c3+1} \big(\mathrm{DHist}(i,j,k)\big) \tag{15}$$

The new normalized histogram is represented as *(DHist)`* and (*h1(a1, b1,* c1))`, (*h2(a2, b2,* c2))` and (*h3(a3, b3,* c3))` for the respectively for each color. The minimum (min) and maximum (max) values each component of RGB, HSV and YCbCr are represented as *l1, l2, l3*. Where *l1= R* or *G* or *B*, *l2= H* or *S* or *V* and *l3 = Y* or *Cb* or *Cr*. Note that min*(R) < a1 < max(R), min(G) < b1 < max(G),* and *min(B)<c1< max(B).*Similarly, *min(H) < a2 < max(H), min(S) < b2 < max(S),* and *min(V)<c2< max(V) and min(H) < a3 < max(H), min(S) < b3 < max(S),* and *min(V)<c3< max(V)* respectively. After that two sided 95% confidence interval normal distribution *N(*μ*,* σ*),* is carried out for each color component of normalize histogram to generate the dynamic threshold ranges (min and max boundaries) where μ is the mean deviation and σ is standard deviation and calculated as :

$$T1_{(RGB)} = \pm 2\,\sigma \quad (16)$$

$$T1_{(HSV)} = \pm 2\,\sigma \quad (17)$$

$$T1_{(YCbCr)} = \pm 2\,\sigma \quad (18)$$

Where, *T1* is the dynamic threshold value range each color component of the respective color-space are considered to segment the skin pixels and non-skin pixels within the frame for each detected face. The lowest value is considered as lower bound while the highest value is considered as upper bound value. The process is repeated for every face region detected in a frame to construct a dynamic threshold and finally, all the detected face regions are merged to generate the final skin region threshold values.

- Explicit Threshold

After inputting video, the color-space of the video was converted to the desired format i.e. *RGB, HSV* or *YCbCr* as defined in section 2.3. This color-space conversion was achieved employing OpenCV's function *cvtColor (`input, flag`)*, where `flag` establishes which type of conversion is to be done, as there are otherwise more than 150 color-space conversions. After successful color-space conversion, an appropriate threshold was applied in order to filter pixels on the basis of the threshold range of pixel values.

Furthermore, in order to achieve the maximum detection with minimum complexity explicit thresholding scheme is employed for the skin pixels detection. For this purposes, the OpenCV image-processing library from Intel was imported into the source code. Thresholding was performed by utilizing OpenCV's basic thresholding function *InRange(`input, lower bound, upper bound, output`)*, where the bounds determine the threshold values. Thus, for the chosen *RGB, HSV* or *YCbCr* color-spaces, three threshold values were needed for each. For *YCbCr*, to start with maximum and minimum luminance and chrominance values were applied, namely *Y(0,255), Cb (80, 120), Cr (133,173).* Later on, the *YCbCr* performance was optimized by replacing those threshold values by *Y (80, 255), Cb (85, 135), Cr (135, 180)* to get the best results. It is important to notice that the proposed optimized thresholds are obtained by means of the chrominance *Cb* and *Cr* histograms analysis suggested by [56]. The readers are referred to [56] for details of the approach to parameter selection. Similarly, for HSV three threshold values were selected by using the histogram method [56], again maximum and minimum values, namely *H (20, 170), S (40, 210) and V (20, 170).* The appropriate threshold ranges (upper bound and lower bound) of *RGB, HSV*, and *YCbCr* are given below, where the bounds determine the minimum and maximum threshold values. After determining the optimized threshold for each color-space, the threshold rules are defined for combined thresholding rules based skin segmentation. The explicitly defined the threshold bound rules T for each color-space are given as:

For RGB

$$T\_min_{(RGB)} = (R > 95)\ \&\&\ (G > 40)\&\&\ (B > 20)\ \&\&(|R - G|) > 15)\&\&(R > G)\&\&(R > B) \quad (19)$$

$$T\_max_{(RGB)} = (R > 220)\&\&(G > 210)\&\&(B > 170)\&\&(|R - G|) \leq 15)\&\&(R > B)\&\&\ (G > B) \quad (20)$$

$$T_{(RGB)} = T\_min_{(RGB)} \cup T\_max_{(RGB)} \quad (21)$$

For HSV

$$T\_min_{(HSV)} = (H < 20)\&\&(S < 40\ )\&\&(V < 20) \quad (22)$$
$$T\_max_{(HSV)} = (H > 170)\&\&(S > 210)\&\&(V > 170) \quad (23)$$
$$T_{(HSV)} = T\_min_{(HSV)} \cup T\_max_{(HSV)} \quad (24)$$

For YCbCr

$$T\_min_{(YCbCr)} = (Y < 80)\&\&(Cb < 85\ )\&\&(Cr < 135) \quad (25)$$
$$T\_max_{(YCbCr)} = (Y > 255)\&\&(Cb > 135)\&\&(Cr > 180) \quad (26)$$
$$T_{(YCbCr)} = T\_min_{(YCbCr)} \cup T\_max_{(YCbCr)} \quad (27)$$

Combined Threshold Rules based Skin Segmentation

After obtaining the threshold bound values that combined threshold rule (CTR) based segmentation is implemented by considering the advantages of broadly adopted *RGB, HSV* and *YCbCr* color-spaces (refer to section 3). The purpose of the CTR segmentation is to improve the probability of skin pixel detection and reduce the false positives. In CTR based segmentation rather than combining skin distribution of all the color-spaces as in [55] different combination of *RGB, HSV* and *YCbCr* color-spaces have been explored to improve the accuracy of skin pixels and non-skin pixels' classification and avoid data over-fitting. By taking all thresholds bound rules the current pixels value of the input video frame is classify as skin pixel if and only if when two or more than two threshold rules vote for it, otherwise it will be classified as non-skin pixels. The threshold rules for the decision of skin segmentation are given below:

$$R\_Vote_{(min)} = Vote\_2min(T_{RGB}, T_{YCbCr}, T_{HSV}) \quad (28)$$

And

$$R\_Vote_{(min)} = Vote\_2min(T1_{RGB}, T1_{YCbCr}, T1_{HSV}) \quad (29)$$
$$R\_Vote_{(min)} = True \rightarrow Skin\_Pixels \quad (30)$$
$$R\_Vote_{(min)} = False \rightarrow Non\_Skin\_Pixels \quad (31)$$

The segmentation performance is measured in terms of the correct detection rate (CDR), the false detection rate (FDR), and the overall classification rate (CR). The CDR is the percentage of skin pixels correctly classified; the FDR is the percentage of non-skin pixels incorrectly classified; the CR is the percentage of pixels classified as skin pixels. Each output frame generated by a CTR was compared pixel wise with the corresponding skin segmented ground-truth. The monochrome images in Fig. 4 and Fig.5 show the visual analysis of the proposed CTR scheme for skin segmentation at the optimized threshold bounds for the Miss-America, Foreman, Paris, Cricket and MOT17-11 video frames with people of different skin tones and camera directions. Additionally, the comparative analysis with the other proposed schemes demonstrated in Table 2 validates that proposed CTR outperform the other colors space in CTR on all test videos thus a favorable choice for real-time multimedia applications.

3.2. **Module 2: Encryption**

In this work the two types of encryption schemes are implemented and compared. In the first type, ROI based lightweight encryption scheme has been proposed. In the proposed method after skin detection, the pixel values at the locations indicated by a bitmap were encrypted with the Advance Encryption Standard in Cipher Feedback Mode (AES-CFB) [57]. Among the AES encryption modes (CBC, ECB, CTR, OCB, and CFB) [58], the CFB mode is used because it employs a block cipher, namely AES, as a stream cipher which is feasible for encryption of real-time streaming surveillance data. Furthermore, CFB mode holds the property of self-synchronization and chaining dependency. Consequently, in this mode any change in the plain bitstream or the initialization vector *(iv)* reflected in the preceding encrypted output bitstream. Thus there is no need to keep *iv* secret. Furthermore, in the proposed scheme the skin pixels are

encrypted independently, hence do not generate significant encryption bitrate overhead. The encryption process is performed as:

$$Ki = Ke(Ei - 1) \quad (32)$$

$$Ei = Si \oplus \text{Ki} \quad (33)$$

where *Ki* are the generated keys stream, and initially, CFB uses *iv* and afterword, the previously encrypted bitstream is XORed with current plain bitstream to generate the current encrypted bitstream as output. The symbol $\oplus$ represent the XOR operator, *Si* is the input bitstream with the skin detection, and *Ei* is the skin encrypted output bitstream (output of the encryption algorithm). The 128-bits encryption key is generated and distributed through the through Diffie-Hellman key exchange method [59]. Afterwards, the skin encrypted output bitstream and non-skin pixels are reconstructed to generate privacy protected output bitstream. The proposed lightweight encryption encrypts the skin pixels independently without infringementing the structure of video frame and make the privacy sensitive area (in our case skin) unrecognizable. Furthermore, as the proposed methods the do not destroy structure of protected sensitive region thus behavior of the person can be observed and without revealing the identities of person. Therefore, proposed scheme can be suitable for the real-time application where the further processing such as video and behavior analysis are required to perform without breaching the personal privacy of individuals.

In the second type, acting as a default, light-weight method of encryption for surveillance video, each video frame without skin detection was selectively encrypted using the authors' own scheme [60]. Selective encryption takes place during the compression process, which was with the Scalable Video Coding (SVC) extension of the H.264/AVC codec acting in single-layer mode. That is to say encryption took place as part of the compression process using H.264/SVC in a mode that corresponds to H.264/AVC. While in the first form of encryption, detected skin pixels were encrypted, with the locations within each video frame indicated by a bitmap. In this case, encryption was with AES-CFB, which acted as the stream cipher to encrypt only the skin color pixels. Notice that AES-CFB is also the method of encrypting selected parameters in the authors' selective encryption scheme [60]. In the selective encryption scheme of [60] the parameters that are selected are those that impose no overhead in a statistical sense and additionally maintain decoder format compatibility i.e. do not breach the H.264/AVC standard. In summary, acting as a default method of protecting surveillance video, the selectively-encrypted bit-streams do not select for skin pixels but for coding parameters, whereas in proposed method the pixels that have been detected as skin are encrypted, after which the same compression as for selective encryption is applied. Thus method provides the sufficient privacy protection by encrypting the skin only and keep the rest of pixels unencrypted hence the protected surveillance videos preserves the sufficient information and can be utilized for further real-time processing without breaching the individuals. Furthermore, when required the authorized users can retrieve the encrypted videos with the same 128-bit secret key i.e is used to generate the key-stream *Ki.*

Pseudo Code for Skin Detection and Encryption

```
Input: Video bitstream
Output: Skin encrypted bitstream
Void main ()
Face_detect(); // for face detection
Skin_detect(); // Skin detection and segmentation
Skin_Encryption (); // for skin pixels encryption

// For face detection
Face_detect()
{
var cascade = new Accord.Vision.Detection.Cascades.FaceHaarCascade();
var detector = newObjectDetector(cascade, minSize: 50, searchMode:  ObjectDetectorSearchMode.NoOverlap);
Bitmap bmp = Accord.Imaging.Image.Clone(frame);
Rectangle [] rectangles = detector.ProcessFrame(bmp);
// detect all the face within the frame
Pen pen = new Pen(Color.White);
Graphics g = Graphics.FromImage(bmp);
for (int i = 0; i < rectangles.Length;i++ )
   g.DrawRectangle(pen, rectangles[i])
   if (rectangles.Length <= -1)
      MessageBox.Show("No Human face in this image");
      Return (Face==0);
   End if
   Else { MessageBox.Show("There are "+rectangles.Length +" Human face in this image frame");
   Return(Face = = 1);
   End Else
  End for
pictureBox.Image = bmp;
bmp.Save("C:*.jpg", ImageFormat.Jpeg);
Bitmap frame = (Bitmap)pictureBox2.Image;
Bitmap BMP = new Bitmap(img.Width, img.Height);
}

//For Skin detection
Skin_Detect();
{
  If (Face==true)
      Eye_Localization(); // detect eye landmarks using Machine Perception Toolbox
      Frame_current= LoadFrame("*.jpg"); // load input frame with detected face
      float * landmarks = landmark_detect(frame_current, bbox, model,landmarks)
      Return(x, y coordinates of detected landmarks) ;
      Elliptical face boundary ();
          Elliptical mask model();
      Return(Face_region_detected)
      Sobel_detector ();
      Getpixl(use, Face_region_detected).Save("C:*.Face_region.jpg", ImageFormat.Jpeg);
      Face_region.jpg=D(i,j)
      Return (D(i,j));  ) // the frame with the face region that contain skin pixels
      Get_color_values (); //Get the color vectors value (max,min) from the smooth Face region
      Get_detected_face_region=D(i,j)
          RGB_Histogram();
          Color GetRgb();
       for (int i = 0; i < D.Width; i++)
         for (int j = 0; j < D.Height; j++)
             DRGB(i,j,)=R(i,j),G(i,j),B(i,j)
         End for
      End for
      Return (RGB((byte)(R), (byte)(G), (byte)(B));
      CovertColor(HSV);
      HSV_Hotogram();
       Color GetHSV(int H, int S, int V)
       for (int i = 0; i < D.Width; i++)
          for (int j = 0; j < D.Height; j++)
            DHSV(i,j)=H(i,j),S(i,j),V(i,j)
           End for
        End for
      Return (HSV ((byte)(H), (byte)(S), (byte)(V));
      CovertColor(YCbCr); //Color-space conversion
      YCbCr_Hotogram()
        Color GetYCbCr(int Y, int Cb, int Cr)
          for (int i = 0; i < D.Width; i++)
            for (int j = 0; j < D.Height; j++)
DYCbCr(i,j)=Y(i,j),Cb(i,j),Cr(i,j)
          End for
        End for
      Return (YCbCr((byte)(Y), (byte)(Cb), (byte)(Cr));
  Histogram_Normalization();
  Get_dynamic_threshold_range()  // Get Threshold range dynamically
  if((T1<=R<=T2 && T3<=G<=T4 && T5<=Br<=T6)
     Return (T1(R,G,B) );
  End if
  if ((T1<=H<=T2 && T3<=S<=T4 && T5<=V<=T6)
    Return (T1(H,S,V) )
  End if
    if (T1<=Y<=T2 &&T3<=Cb<=T4 && T5<=Cr<=T6)
     Return (T1(Y.Cb,Cr) );
    End if
 Else if (Face== 0) //get the skin pixel using explicit threshold
Video Frame, Explicit Threshold values = T1,T2,T3,T4,T5, T6
    for (int i = 0; i < frame_curent.Width; i++)
       for (int j = 0; j < frame_current.Height; j++)
          Get_Histogram_values_RGB();
          Return(R,G,B)
          Get_Histogram_values_HSV();
          Return(H,S,V)
          Get_Histogram_values_YCbCr();
          Return(Y,Cb,Cr)
       End for
    End for
    if (T_min(RGB)  =(R>T1)  && (G>T2)&& (B>T3)  &&(|R-G|)>15)&&(R>G)&&(R>B))
        Getpixl (T_min(RGB));
    End if
  if (T_max(RGB)=(R>T4)&&(G>T5)&&(B>T6)&&(|R-G|)≤ 15)&&(R>B)&& (G>B) )
      Getpixl (T_min(RGB));
  End if
  T(RGB) = T_min(RGB)  ∪T_max(RGB)
  Return(Getpixl(T(RGB));
    if (T_min(HSV)  = (H< T1)&&(S< T2 )&&(V< T3)
        Getpixl(T_min(HSV) );
    End if
    if (T_max(HSV)= (H > 170)&&(S > 210)&&(V > 170)
        Getpixl(T_max(HSV));
    End if
  T(HSV) = T_min(HSV)  ∪T_max(HSV)
 Return(Getpixl(THSV);
    if (T_min(YCbCr)  = (Y< 80)&&(Cb< 85 )&&(Cr< 135)
       Getpixl(T_min(YCbCr));
    End if
    if (T_max(YCbCr)= (Y > 255)&&(Cb > 135)&&(Cr>180)
        Getpixl(T_max(YCbCr));
    End if
   T(YCbCr) = T_min(YCbCr)  ∪T_max(YCbCr)
   Return(Getpixl(T(YCbcr);
// for skin pixels and non_skin pixels segmentation
CTR_Segementation()
    R_Vote(min)  = (Vote_2min(T(RGB), T(HSV), T(YCbCr) )||( Vote_2min(T1(RGB), T1(HSV), T1(YCbCr) ))
   if (R_Vote(min)  ==1)
   Bitmap newBitmap = new Bitmap(source.Width, source.Height); //make an empty bitmap the same size as scrBitmap to store the skin pixels values detected
 int index = 0;
        for (int i = 0; i < scrBitmap.Width; i++)
          for (int j = 0; j < scrBitmap.Height; j++)
                store_point.Add(index);
                int x, y;
                for (int i = 0; i < store_point.Count; i++)
```

```
134.                    x = (store_point[i] % newBitmap.Height);
135.                    y = (store_point[i] / newBitmap.Height);
136.                    }
137.                    End for
138.                End for
139.               End for
140.      Return(Skin_pixels_bitmap)
141.      Else (R_Vote(min)  == 0)
142.           Return (Non-Skin_Pixels_bitmap)
143.     End Else
144.     End if
145.     index++Output: skin pixels Detected

146.   //for skin pixels' encryption
147.   Skin_Encryption ()
148.   {
149.    int DH-Key ()  //Diffie-Hellman secret  Key Exchange
       Method
150.   Input: Public_key1, Publick_key2, Private_Key1,
       Private_Key2
151.       long int X=PK_1;
152.       long int Y=PK_2;
153.       long int s=Private_Key1;
154.       long Int t= Private_Key2;
155.       A=Y^s modX;
156.       B=Y^ tmodX;
157.       temp = A;
158.       A= B;
159.       B=temp;
160.       Ks=Y^s modP;
161.       Kt=X^t modP;
162.       Z=ks=kt;
163.   Return (Z);
164.   Output: Secret-key (Z)
165.   }

166.  int Enc_AES_CFB( )   //AES-CFB
      Skin_Encryption/Decryption(mbedTLSLibrary)
167.    Input: Detected_skin_bitstream, Initialization Vector (iv),
      Secret-key, key-size =128;
168.    #if defined(CIPHER_MODE_CFB)
```

```
169.     int aes_SK(aes_Process *prc, const unsigned char* Secret-
       key, unsigned int key-size)
170.    if (key-size =128)
171.    prc->rounds  = 10;
172.    aes_encrypt_cfb();
173.    break;
174.     int aes_encrypt_cfb(aes_Process *prc, int mode-type, size_t
        l-size, size_t *iv_off, unsigned char iv, const unsigned char
        *skindetectedvideo, unsigned char *outvid )
175.      int i;
176.      size_t s = *iv_off;
177.      if(mode-type == ENCRYPT_AES_CFB)
178.          while (l-size -- )
179.                if(s == 0 )
180.     aes_encrypt_cfb (prc, ENCRYPT_AES, iv, iv );
181.     iv[l] = *output++ = (unsigned char)( iv[s] ^
        *skindetectedvideo++ );
182.                 s = ( s + 1 ) & 0x0F;
183.               End if
184.             End while
185.          End Else
186.       *iv_off = s;
187.      End if
188.     Else
189.     while(l-size -- )
190.      if( s == 0 )
191.     aes_encrypt_cfb (prc, ENCRYPT_AES, iv, iv);
192.          c = *input++;
193.          *outvid++ = (unsigned char) ( i ^ iv[s] );
194.           iv[s] = (unsigned char) i;
195.          n = ( s + 1 ) & 0x0F;
196.        End if
197.       End while
198.      End Else
199.      Return(0);
200.     Output: Skin EncryptedVideo bit-stream
201.     }
202.     Reconstruct( ); // for reconstruction the skin encrypted
        bitstream and non-skin bitstream
203.     Output: Output bitstream with skin encrypted
```

4 Evaluation

In order to evaluate the performance of the proposed scheme, the experiments were actually performed with Quarter CIF (QCIF) (176 × 144 pixels/frame), Common Intermediate Format (CIF) standard (352 × 288 pixels/frame) as well as High-definition (HD) (1280 x 720 pixels/frame) video bit-streams. However, in this paper CIF resolution results are reported for reasons of space and the ease of making repeated experiments. Sample experimental results for skin detection on selected video frames are shown in above in Figs. 4 and Fig. 5. The MOT17-11 was taken from URL https://motchallenge.net/data/MOT17/ while the other videos (Miss-America, Foreman, and Paris) were available at the URL https://media.xiph.org/video/derf/. JSVM 9.19 reference encoder (with single layer) was used to encode test video bit-streams for the experimental results of default SE (See Fig. 8 & 9). The input video bit-streams were at a frame rate of 30 fps with Group-of-Pictures (GOP) size 16 in an IBBP… frame structure. The proposed SE for skin encryption over videos (Fig. 8 & 9) is implemented in C++ language with our own developed software.

4.1 Experimental Results on Videos for Skin Detection

The detection results for each test video dataset are illustrated in Figs.4 and Fig.5 respectively. The performance of each color-space and the proposed scheme has been evaluated through the Ground Truth (GT) of the selected test video. However, due to unavailability of the GT on the selected test videos, all GTs are in-house generated for tested videos through RoboRealm v2.87.25 (http://www.roborealm.com/) and Adobe Photoshop CS6 for testing purpose. The monochrome images in Figs. 4 and Fig.5 show the visual results of skin detection through HSV (Fig. 4-5(a2, b2, c2, d2, e2)), YCbCr (Fig. 4-5(a3, b3, c3, d3, e3)), RGB (Fig. 4-5. (a4, b4, c4, d4, e4)) and proposed CTR (Fig. 4-5. (a5, b5, c5, d5, e5)) color-spaces, with the skin as white areas while the non-skin areas are shown as black. In addition, some pixels/areas have

been falsely detected as skin areas, as the reader should notice to gain a subjective impression of the impact of the different color-spaces for the same video frames, especially across the differing skin tones. In the visual results the reduction in false positives in the proposed CTR is apparent. Thus, the proposed scheme achieved the better performance for skin segmentation without incurring the higher computation cost required for other statistical and advanced machine learning algorithms.

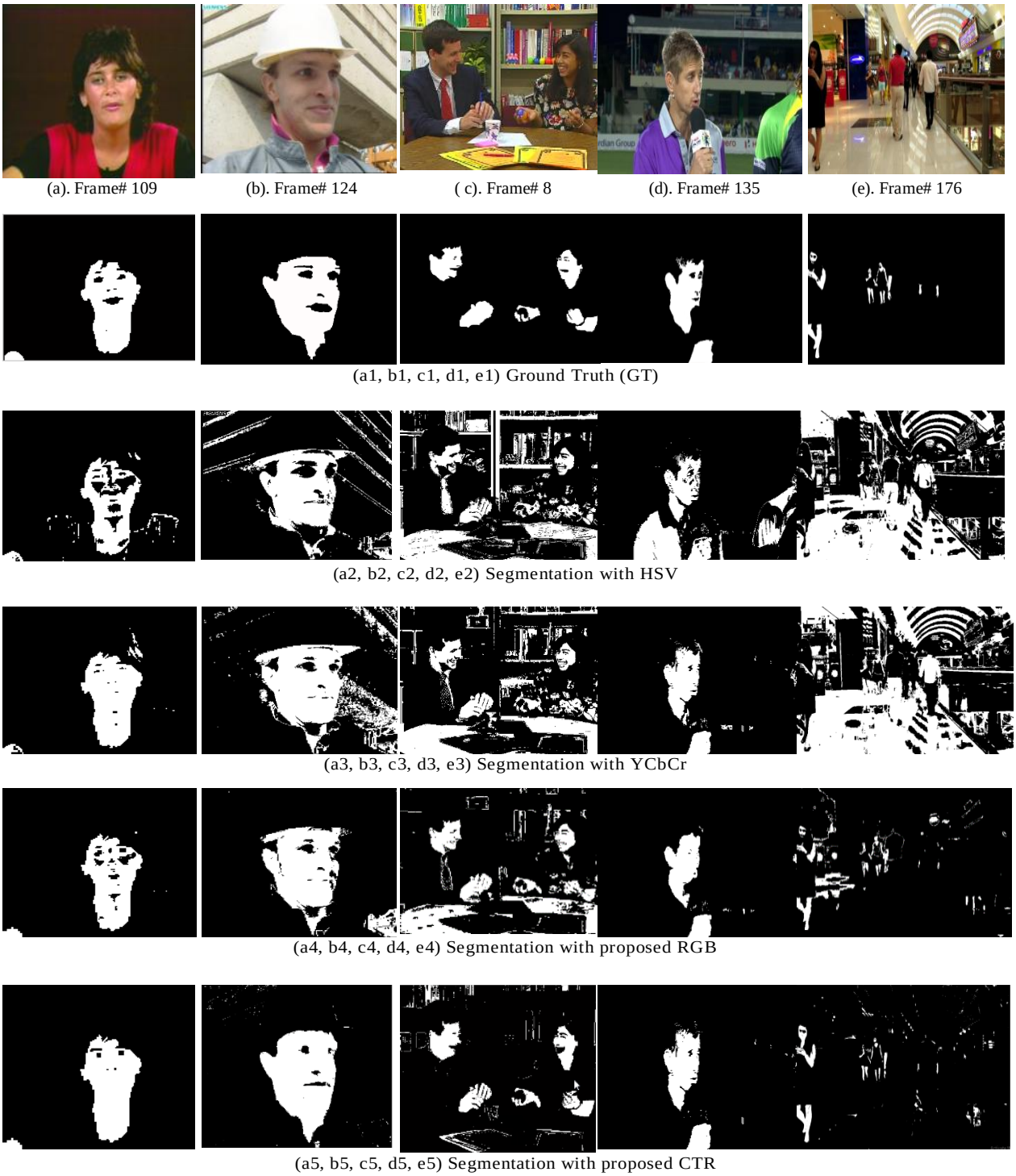
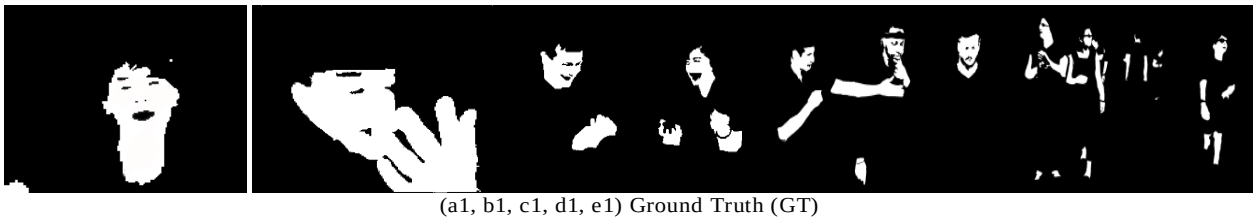


Fig. 4 Visual comparative result of skin detection on (a) Miss-America (b) Foreman (c) Paris (d) Cricket and (e) MOT17-11 videos when segmented with HSV, YCbCr, RGB and proposed CTR color-spaces.

(a). Frame# 120 (b). Frame# 154 (c). Frame# 43 (d). Frame# 202 (e). Frame# 851



(a1, b1, c1, d1, e1) Ground Truth (GT)

(a2, b2, c2, d2, e2) Segmentation with HSV

(a3, b3, c3, d3, e3) Segmentation with YCbCr

(a4, b4, c4, d4, e4) Segmentation with proposed RGB

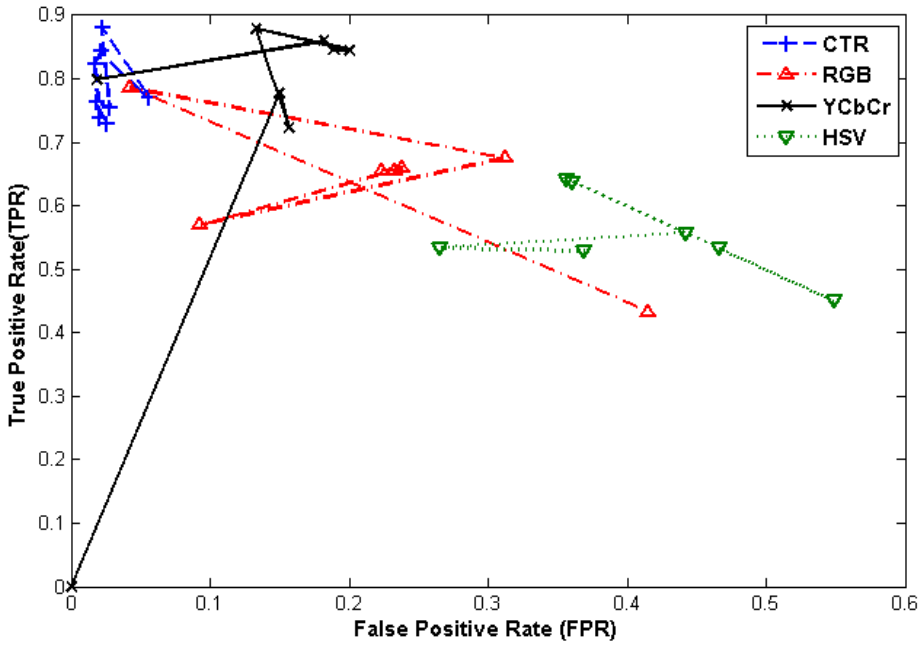
(a5, b5, c5, d5, e5) Segmentation with proposed CTR

Fig. 5 Visual comparative result of skin detection on (a) Miss-America (b) Foreman (c) Paris (d) Cricket and (e) MOT17-11 video when segmented with HSV, YCbCr, RGB and proposed CTR color-spaces.

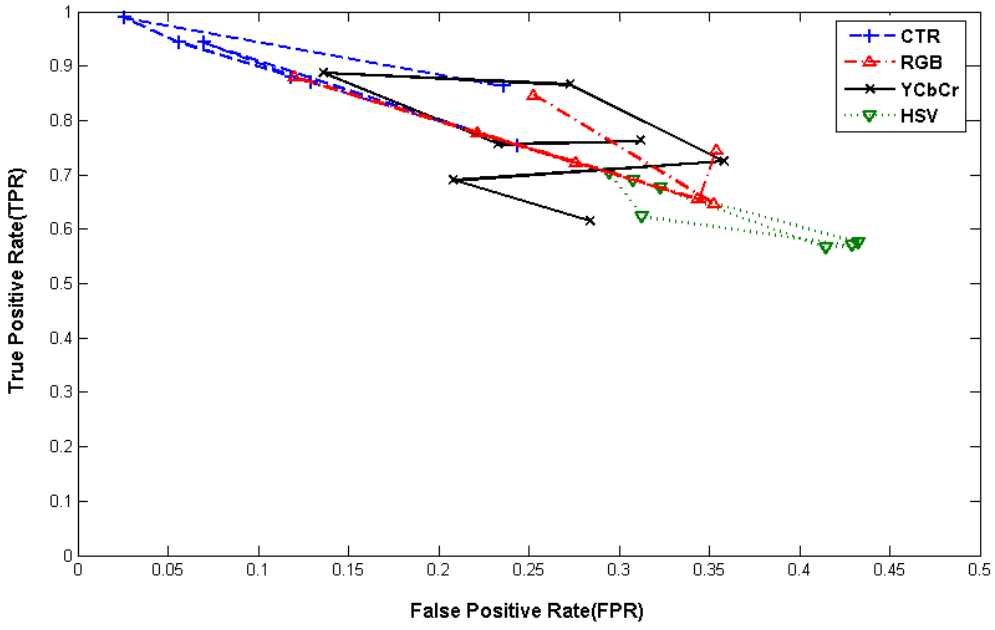
Table 2 Comparison between three color-spaces (HSV, YCbCr, RGB) and proposed method (CTR)

Sr. No.	Video	Average CR (%)				Average CDR (%)				Average FDR (%)			
		HSV	YCbCr	RGB	CTR (Proposed)	HSV	YCbCr	RGB	CTR (Proposed)	HSV	YCbCr	RGB	CTR (Proposed)
1.	Miss America	88.2	96.9	93.6	91.2	84.6	83.4	85.0	89.1	3.5	13.5	8.6	2.3
2.	Foreman	102.9	118.8	101.9	98.4	89.1	84.9	85.9	89.9	13.9	33.8	16.45	8.6
3.	Paris	329.9	281.9	227.5	142.5	76.1	85.0	85.2	87.16	254.3	195.5	142.4	55.79
4.	Cricket	129.5	109.9	99.2	93.0	69.99	85.5	86.9	87.99	44.6	24.8	14.2	11.42
5.	MOT17-11	223.2	124.7	107.6	103.5	84.9	85.2	86.3	91.15	138.9	38.7	22.8	13.81

Additionally, for the performance evaluation of the CTR and against each color-space, Receiver Operating Characteristic (ROC) curves for the test videos are shown in Fig.6. ROC curves determine the association among true positives and false positives. The value of true positive near to 1 and value of false positive near to 0 represents that approach gain better results. From the visual results of Fig. 6 it is noticed that ROC curve of CTR methods is closer the y-axis which indicates that proposed scheme obtained high true positive rate (i.e. less true negative rate) comparative to other color-spaces, thus performance of proposes scheme is better for all the test video dataset.



(a).



(b).

Fig. 6 Comparative ROC curves analysis for False-positive rate versus true-positive rate for (a) Foreman video bitstream (b). Cricket video bitstream

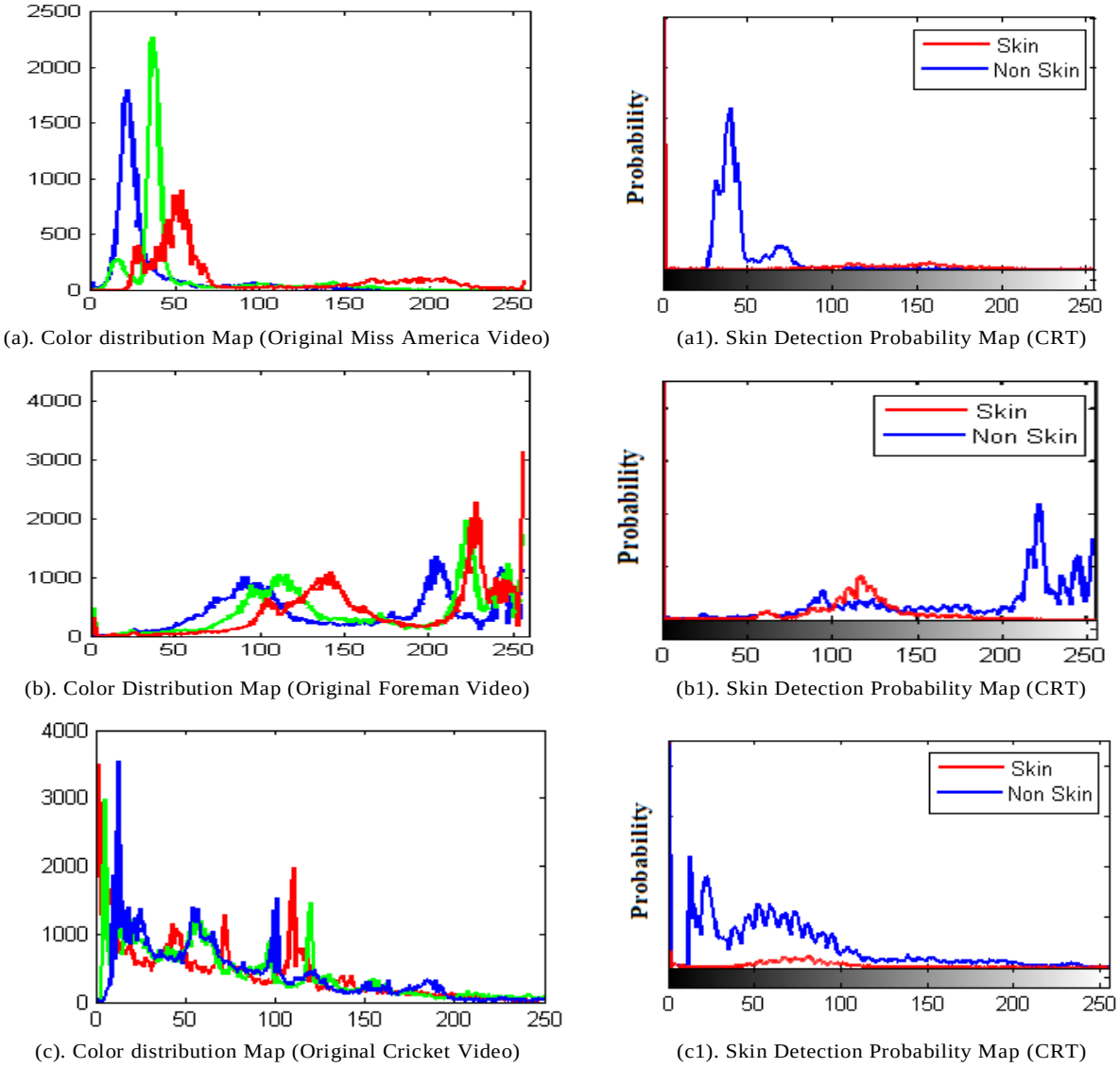
Fig.7 illustrates the color distribution map of input videos bitstream. Fig 7(a, b, c) and the estimated skin probability distribution map in the video with CTR (Fig 7(a1,b1,c1) to show the performance of the proposed scheme. Suppose, *ctr(x,y)* is the color components of the combined colors

spaces then skin probability distribution map can calculated as:

$$ps(i,j) = pds(i)|ctr(i,j) = n \quad (34)$$

Where *ps* is the skin probability distribution, *n* is one vector of *ctr* and *(i,j)* is the coordinates position in *ctr*. The *pds* is the dynamic skin probability model which is calculated as:

$$pds = exp\left[-\frac{1}{2}(n-\mu)^t \ \Sigma^{-1}(n-\mu)\right] \quad (35)$$



(a). Color distribution Map (Original Miss America Video)

(a1). Skin Detection Probability Map (CRT)

(b). Color Distribution Map (Original Foreman Video)

(b1). Skin Detection Probability Map (CRT)

(c). Color distribution Map (Original Cricket Video)

(c1). Skin Detection Probability Map (CRT)

Fig. 7 Skin and non-skin Probability distribution Map.

Furthermore, the experimental results illustrate that YCbCr is less sensitive to illumination conditions, thus perform better as compared to HSV and RGB for real-time surveillance systems. Moreover, due to perceptual non-uniformity and high correlation between the channels (R, G, B) within the video data the YCbCr is a favorable choice for real-time multimedia data analysis as compared to RGB. On further investigation, interestingly, the results imply that under same illumination conditions both YCbCr and HSV produces different results on different skin tones. Fig.8 that YCbCr perform better on dark tones (encircled with the red) as compared to HSV under the same illumination conditions (compare Fig.8(a1, b1, c1 vs. a2, b2, c2)) while Fig. 9 shows that the perform HSV produce better results on white skin tones (compare Fig.9 (a1,b1,c1 vs. a2, b2, c2)) while the performance of YCbCr is poor. This indicates that choice of color-space may be important when dealing with videos showing people of differing ethnicity. However, our proposed CTR method has advantage by considering the all color-spaces and which produces better detection accuracy will be candidate for the skin segmentation.

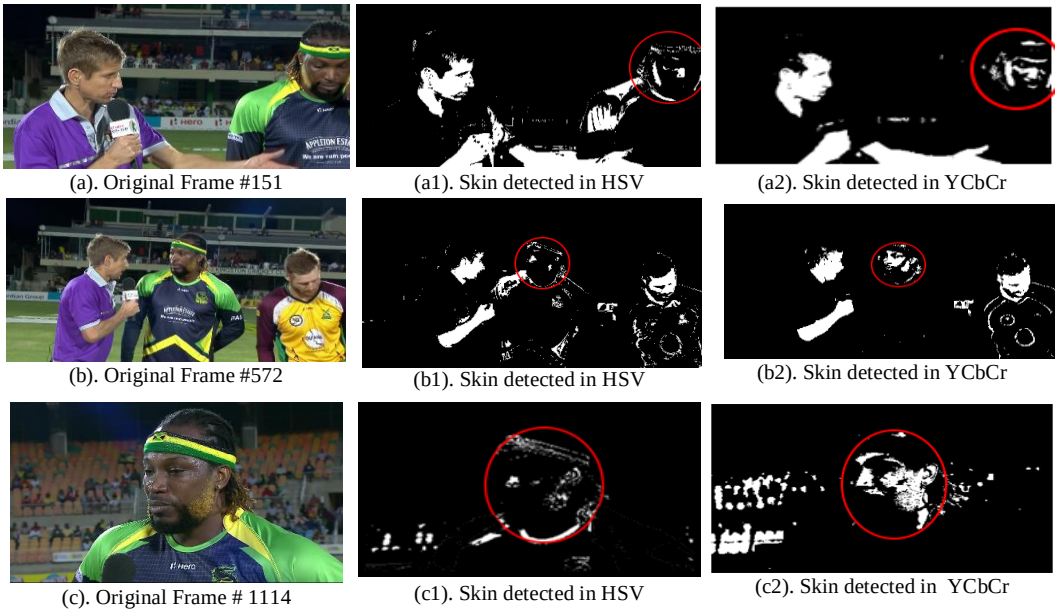


Fig.8. Impact of different color-space on different skin tones. (a,b,c) Original input video sequence. (a,b1,c1) Skin detected in HSV color-space. (a2,b2,c2) Skin detected in YCbCr color-space

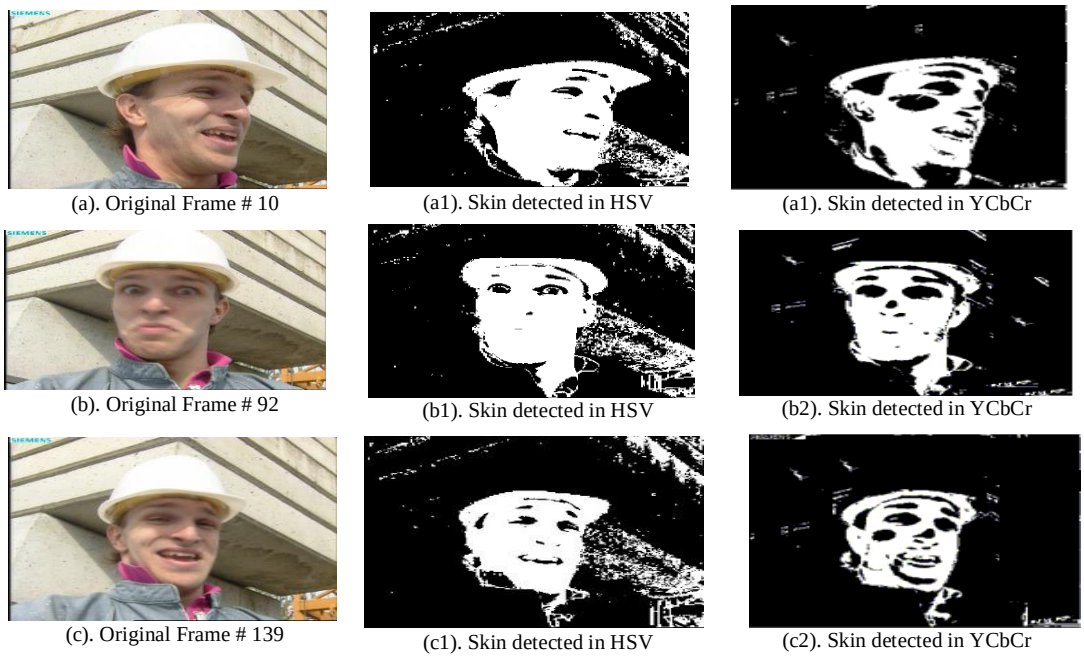


Fig.9. Impact of different color-space on different skin tones. (a,b,c) Original input video sequence. (a,b1,c1) Skin detected in HSV color-space. (a2,b2,c2) Skin detected in YCbCr color-space

Our findings imply that choice of color-space may be important when dealing with videos showing people of differing ethnicity. Clearly, people in a video may well have different skin complexions and, therefore, improving the detection response for face detection is important, especially during video surveillance.

Further to evaluate the performance of the proposed detection algorithm has been evaluated by comparing it with other proposed algorithms considering TDSD dataset [61] and the selected test dataset. The TDSD data set has been selected because it contains multiple ethnicity groups, face poses and body orientation, and the dataset is acquired under various illumination conditions. Tables 3 shows the quantitative comparisons of the proposed method with quantitative evaluation indexes including Precision, Recall accuracy and F-measure. The higher value of precision and recall indicated the algorithm returns more pertinent results as compared to irrelevant results. However, precision measures exactness of the results whereas recall measures completeness. The F-measure indicates that algorithms attain the

higher recall without scarifying the precision. Let TP represents true positive, TN represents true negative, FP represents false positive and FN is the false negative then the quantitative evaluation indexes are computed as:

$$Precision = \frac{TP}{TP+FP} \tag{36}$$

$$Recall = \frac{TP}{TP+FN} \tag{37}$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{38}$$

$$F-measure = \frac{2Recall.Precision}{Recall+Precision} \tag{39}$$

Table 3 Quantitative comparison of proposed method with other existing methods for skin detection in TDSD database.

Sr. No.	Method	Precision	Recall	Accuracy	F-measure
1.	A. Cheddad et.al [5]	0.4133	0.5570	0.8334	0.4745
2.	W.R. Tan,et al. [6]	0.5857	0.8592	0.8303	0.6966
3.	S. Bianco et.al [7]	0.7470	0.7504	0.9055	0.7487
4.	Y. Luo, Y.-P. Guan [9]	0.8323	0.8742	0.9374	0.8528
5.	P. Yogarajah et.al [55]	0.5133	0.5725	0.8535	0.5922
6.	**Proposed**	**0.7269**	**0.8330**	**0.9178**	**0.7763**

The comparative results demonstrate the performance of the proposed scheme has been much improved as compared to the static scheme [5], fusion scheme [6] and dynamic scheme [55] and. The fusion scheme produces unsatisfactory results due to pose variation as it based on the landmarks detection, however, our purposed scheme utilized the advantages of both dynamics and static approach hence gain better classification. One can notice that scheme proposed in [9] gain more accuracy as compared to our proposed scheme however [9] is computational complex as Gaussian model is adopted and required large training dataset. As there is tradeoff between the accuracy and computational overhead hence, the proposed methods can be a suitable choice for real-time surveillance applications. Fig. 10 shows the comparison of proposed scheme with various schemes by calculating the sensitivity for the benchmark and test dataset. The results illustrate that the performance of the CTR is better to the existing schemes.

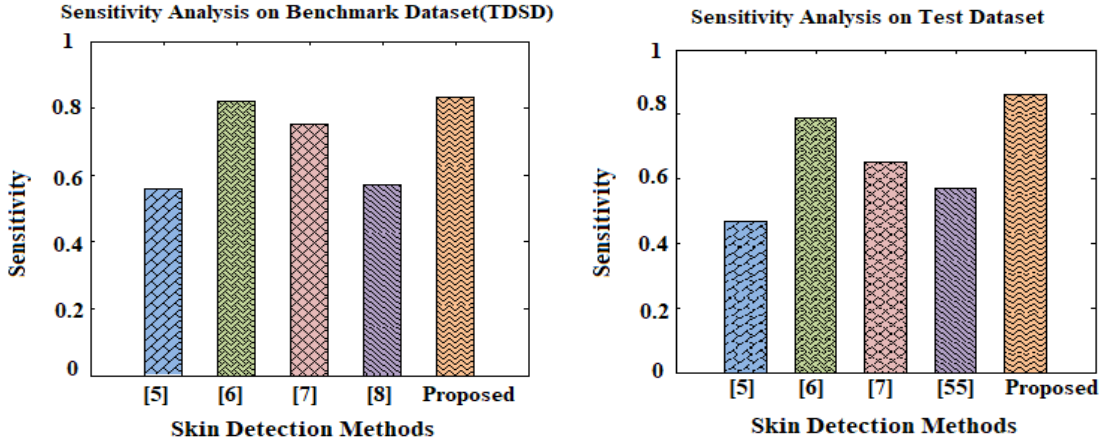


Fig.10. Comparative sensitivity based performance of proposed scheme with the existed

4.2 Experimental Results with Encryption on detected Skin for Privacy Protection

4.2.1 Perceptual assessment of encrypted bit-stream

For the performance assessment of the proposed scheme the encryption is applied with the default SE and proposed ROI based selective encryption on detected skin with CTR version of the test video, so

that privacy could be protected. Comparison of visual results of original video bit-stream with encrypted videos in CTR and the original with selectively encrypted video are shown in Figures 8 and 9 respectively for the test video streams. In the case of the selectively encrypted video bit-streams, the authors' own scheme [60] was employed in single-layer mode for the H.264/AVC codec. The encryption is based on encryption of selected parameters in the entropy coder's output, namely those parameters that impose no statistical overhead and maintain decoder format compatibility. The selectively-encrypted bit-streams in [60] do not select for skin pixels but for coding parameters, whereas the encrypted bit-streams in the aforementioned color-spaces and proposed in this work, encrypt pixels that have been detected as skin. Whatever data are encrypted with AES in CFB mode. Fig.11-12(a) shows the original video bit-stream and Fig.11-12 (b) shows the original with the selective encrypted video. Fig.11-12 (c), show the encrypted bit-stream of in proposed CRT scheme. Although selective encryption is an attractive method when seeking to reduce encryption time while streaming video, when video surveillance is undertaken. The skin pixels detection and their encryption are more attractive than by conventional means. This is because protecting the identity of characters, who may be actors in fictional video streams, is relatively less important than when protecting the privacy of real people [19] in surveillance video.

The visual results of Fig. 11 and 12 clearly show that the people in the selectively-encrypted original video bit streams are more recognizable/decipherable as compared to the skin-pixel encrypted video bit-stream. Hence, encryption with proposed SE provides more privacy protection and data confidentiality as compared to the default selectively-encrypted original video and save the computational complexity as well.

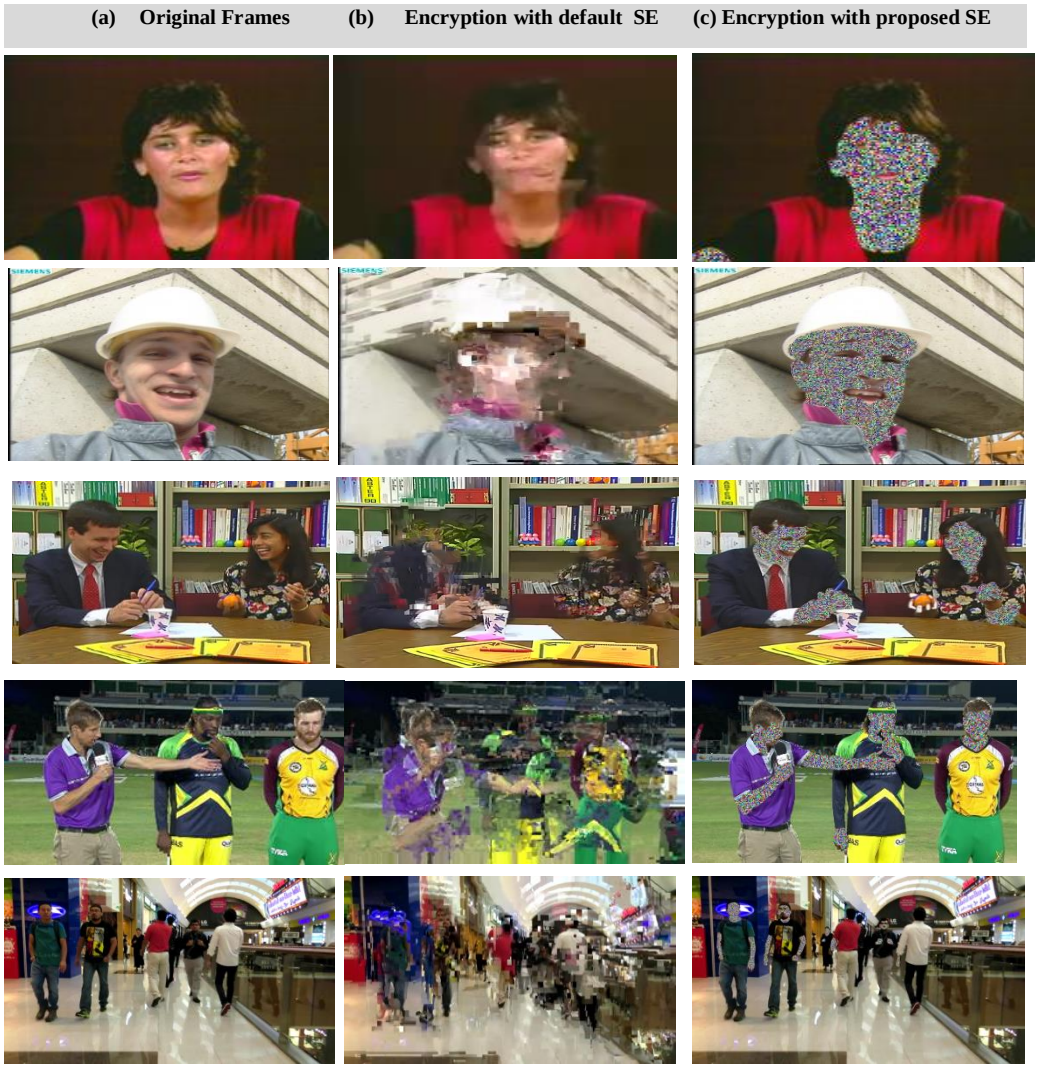


Fig. 11. Comparative visual results on tested videos with default and proposed SE.

Fig. 12. Comparative visual results on tested videos with default and proposed SE.

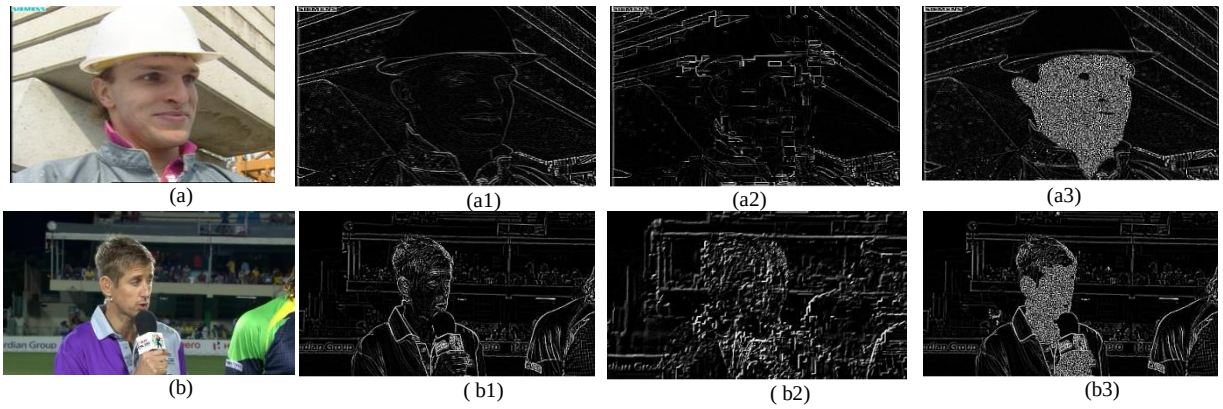


Fig. 13. Comparative Analysis of Foreman and Cricket video bit-stream with Laplacian edge detector. (a,b)Original video frame. (a1.b1)) Detected edges of Original video frame.(a2,b2) Default selectively-encrypted video bit-stream. (a3,b3) Skin encrypted bitstream with proposed SE

To further analyze the performance of selective encryption of both schemes, Laplacian edge detection [62] has been performed on the encrypted bit-streams. The detected edges of original video bit-stream are shown in Fig.13 (a1,b1) by using 3x3 Laplacian edge detector whereas detected edges of selectively-encrypted video bit-stream are illustrated in Fig. 13(a2, b2). Fig.13 (a3, b3) show the detected edges of skin encrypted bit-stream. The ratio for detection of edges are detected through the following

equation:

$$R = \frac{\sum_{x,y=1}^{n} |O(x,y) - O`(x,y)|}{\sum_{x,y=1}^{n} |O(x,y) + O`(x,y)|} \quad (40)$$

The comparative results illustrated in Fig. 13 (compare (a2,b2) with (a3,b3) show that default SE destroy the structure of the video frames hence behavior/activity of individuals cannot be preserved thus became useless for further any kind information processing while the proposed methods successfully encrypt the skin pixels without disturbing the overall structure of the video frame, subsequently can be utilized for further information processing and video analysis without breaching the privacy of individuals. Hence, the proposed methods can be adopted for the real-time surveillance application with the confidence that one's privacy will be protected.

4.2.2 Quality Assessment of skin-encrypted bit-stream

To further evaluate the security performance of proposed methods and to compare the results on default SE, Structural Similarity Index (SSIM) [63] video quality assessment metric was used at QP = 34. SSIM [63] is a video-quality metric that gauges the structural similarity between two video frames in a way that is more sensitive to the human visual system than PSNR. Thus, the Video Quality Experts Group (VQEG) Full- Reference Television (FR-TV) Phase II tests resulted in Pearson linear correlation coefficients (PCCs) with Difference Mean Opinion Score (MOS) from subjective tests of above 0.9. SSIM also has less computational complexity compared to the Video Quality Metric (VQM). The SSIM range is 0 to 1, values nearer to 1 mean more structural similarity between the original and encrypted video bit-stream, which means less protection is achieved. Therefore, the SSIM of the test video bit-streams was calculated to evaluate the results of the scheme. An average SSIM value of skin pixel encryption and default SE for the Foreman test video is reported in Fig.14 As it turns out, the latter method, in terms of privacy protection, falls considerably short of directly encrypting detected skin pixels, as the SSIM plots make clear.

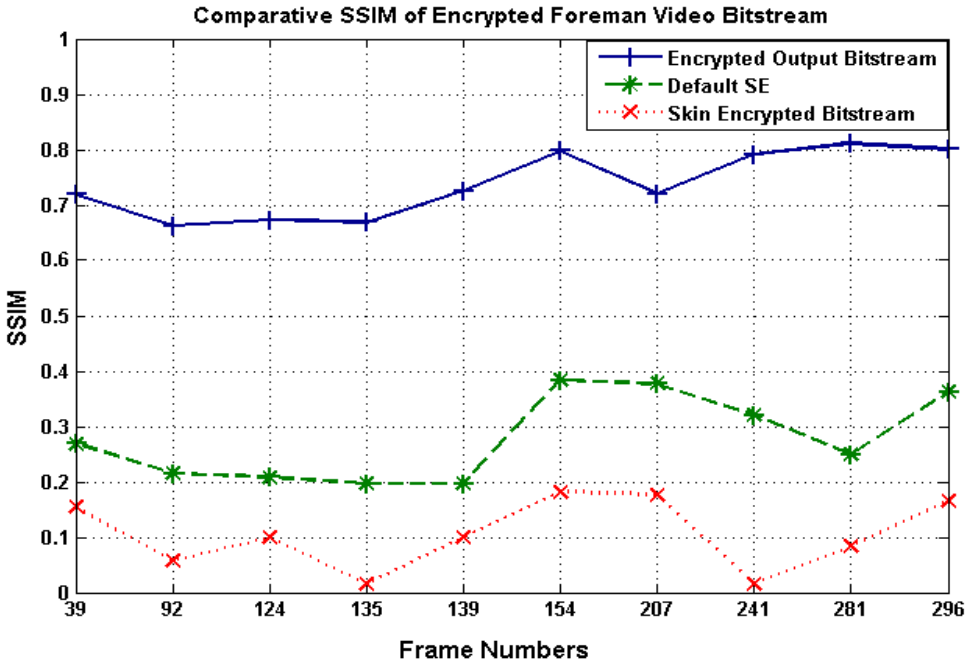


Fig. 14. Comparative SSIM of Skin encrypted bitstream and Default selectively-encrypted video bitstream

4.2.3 Performance Analysis

Computational Complexity

In order to evaluate the performance in term of computational complexity of each phase; skin detection and encryption for privacy protection, experiments were performed on a 64-bit operating system with 2.60 GHz Intel Core(TM) i7-3720 QM processor and 8 GB RAM. The total encoding time (T)

of the proposed scheme consists of: i) skin detection t_d and encryption time t_e. The skin detection time include the conversion time into color-spaces and segmentation time. The equation presents the total encoding time with the proposed method.

$$T = t_d + t_e \tag{41}$$

and

$$t_d = t_c + t_s \tag{42}$$

where t_c is the conversion time and t_s is the segmentation time.

Fig.15 shows the encoding time of the skin detection and encryption time per frame by the proposed methods in seconds. In fact, timings were calculated by finding the elapsed clock ticks in the Open CV software. Later 'count' method of 'duration' class is called to get elapsed time interval in seconds and converting clock ticks to the nearest second based on the known machine clock frequency. However, the number of seconds per video frame is hardware dependent. The average total execution time of the proposed method and default SE for the Foreman, Paris and Cricket videos is presented in Table 4. The results shows total sequential execution time decreases with the average of 90.58% with proposed scheme as compared to average total execution time of the default SE Miss America video. Similarly there is average of 92.31% and 91.78% decrease in total sequential execution time for Paris and Cricket videos respectively which verifies the effectiveness of proposed solution for the surveillance systems.

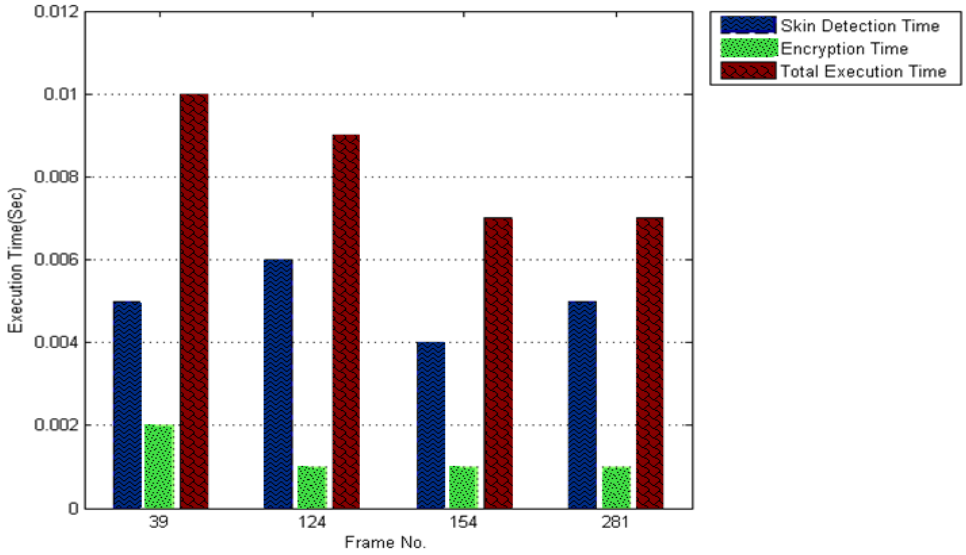


Fig.15. Encoding time for skin detection time, encryption time and total execution time for the Foreman sequence.

Table 4. Comparative Total Average sequential and average parallel execution time (sec) with proposed scheme and default SE

Sr.No	Video	Proposed SE			Default SE		
		Total Detection Time (Sec)	Total Encryption Time (Sec)	Total Execution Time (Sec)	Encoding Time (Without Encryption (Sec)	Encryption Time (Sec)	Total Execution Time (Sec)
1.	Foreman	5.525	2.634	9.159	86.471	13.863	100.335
2.	Paris	6.617	6.015	13.632	95.366	70.455	165.821
3.	Cricket	58.305	8.268	68.573	1227.068	234.996	1462.064

Encryption Space Ratio(ESR)

The performance of the proposed encryption is evaluated by calculating the encryption space ratio which indicates the amount of data encrypted in terms of percentages. ESR is directly proportional to the computational cost encryption. Thus the smaller ESR indicate the lower encryption cost and higher the efficiency of the encryption scheme over the streaming data. The ESR for test videos is shown in Table 5. The ESR is calculated on the bases of detected skin pixels. The results of table show that the average ESR for each video is very low, hence can meet the requirement of real-time surveillance application.

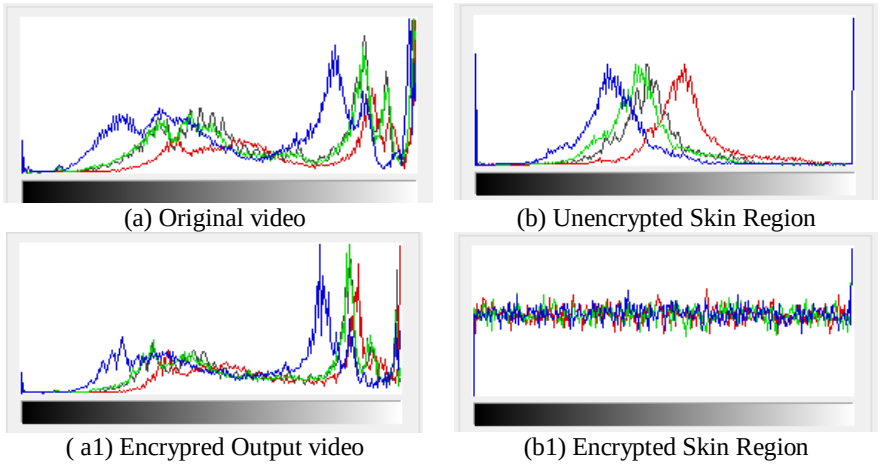
Table 5. Average ESR (%) of test videos with proposed scheme

Sr..No	Video	Total Size (MB) (YUV)		Average ESR per video (%)
		Before Encryption	After Encryption	
1.	Miss America	5.43	5.43	12.56
2.	Foreman	43.506	43.506	11.97
3.	Paris	154.446	154.446	8.87
4.	Cricket	1924.805	1924.805	7.25
5.	MOT17-11	667	667	3.83

4.2.4 Security Analysis

Histogram Analysis

Histogram shows the distribution of pixels within the frame. The even histogram illustrate there is greater uniform distribution among the pixels of encrypted frame and hence the encrypted scheme achieved higher security. The Fig 16 shows that histogram of skin encrypted and encrypted output is significantly different unencrypted skin and original frame (Compare Fig.16 (a) vs. Fig.16 (a1) and Fig.16 (b) vs. Fig.16 (b1)). The results imply that with there is adequate uniform distribution among the encrypted pixels hence the proposed method is secure enough to resist against the statistical attacks on the privacy protect region.



(a) Original video (b) Unencrypted Skin Region

(a1) Encrypred Output video (b1) Encrypted Skin Region

Fig 16. Comparative histogram of the original and encrypted bitstream. (a) Histogram of original bitstream (a1) Histogram of encrypted output bitstream (b) Histogram of only unencrypted skin region (b1) Histogram of only encrypted skin .

Correlation Analysis

In order to demonstrate the strength of the encryption correlation coefficient analysis is performed in horizontal, vertical, and diagonal directions by randomly selecting 5000 pairs of adjacent pixels of original bitstream and the encrypted bitstream. The coefficient correlation score can range between [-1, 1]. The negative score reveals that the relationship between the pixels of original frames and encrypted frames is negatively correlated, whereas the positive score reveals relationship between the pixels of original and encrypted frames is positively correlated. The higher value indicates there high correlation among the pixels. Correlation coefficient can be mathematically calculated as:

$$r = \frac{\frac{1}{n}\sum_{i=0}^{n}(x_i - O(i))(y_i - O(i))}{\sqrt{\frac{i}{n}\sum_{i=0}^{n}(x_i - O(i))^2}\cdot\sqrt{\frac{i}{n}\sum_{i=0}^{n}(y_i - O(i))^2}}$$

(43)

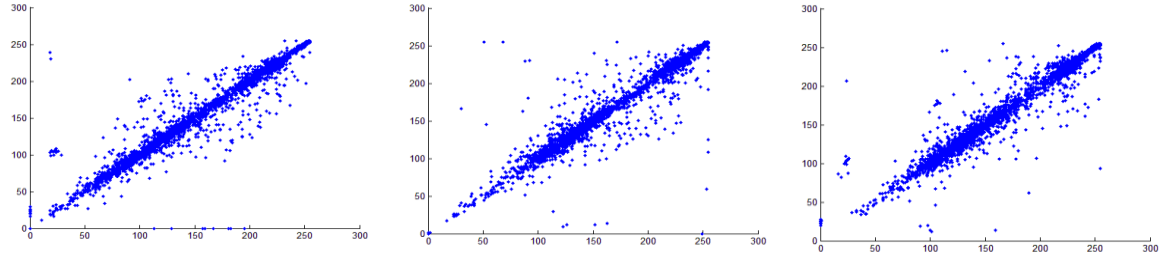
Figure.17 shows comparative correlation coefficients between the adjacent pixels of original frame and encrypted frame for Foreman video. The plot of Fig.17(a1-a3 and b1-b3) show the correlation curves of the adjacent pixels in horizontal, vertical diagonal direction for the unencrypted frame and the only skin detected frame are congregated strongly. This indicated that correlation between adjacent pixels in

the original video frame and only skin detected frames is high and its corresponding correlation coefficients are close to one. Whereas, Fig.17(c1-c3 and d1-d3) show that pixels in skin encrypted frames and the encrypted output frame are scattered over the entire plane thus imply that correlation among them is greatly reduced. This indicated that greater randomness has achieved on the encrypted skin pixels thus has greater potential to resist against the statistical attacks.

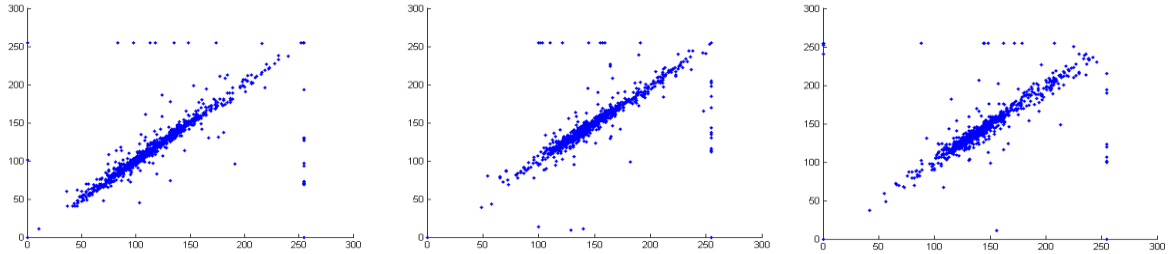
Average correlation coefficients values for the horizontal, vertical, and diagonal directions for test videos are shown in Table 6. The correlation among the pixels on 100 frames was calculated before averaging across all frames of the tested video bitstream.

Table 6. Average Correlation Coefficient of original vs. output skin encrypted videos (100 frames)

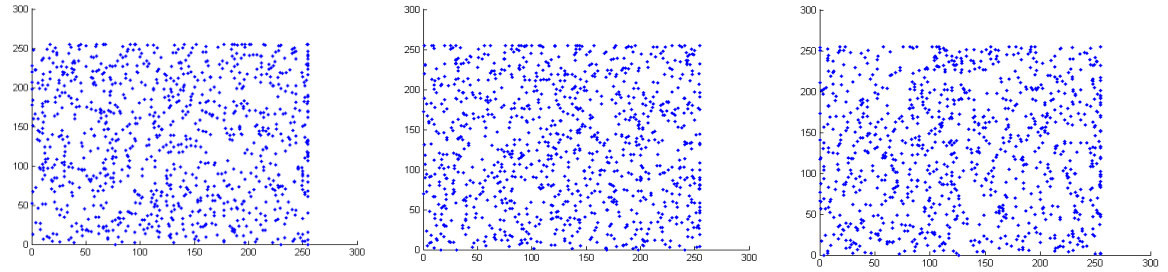
Sr.No	Video	Vertical		Horizontal		Diagonal	
		Original	Skin Encrypted	Original	Skin Encrypted	Original	Skin Encrypted
1.	Miss America	0.9624	0.6760	0.9520	0.6734	0.9101	0.6769
2.	Foreman	0.9647	0.6669	0.9679	0.6570	0.9717	0.6640
3.	Paris	0.9777	0.7783	0.9773	0.7864	0.9473	0.7367
4.	Cricket	0.9797	0.7537	0.9038	0.7480	0.9711	0.7526



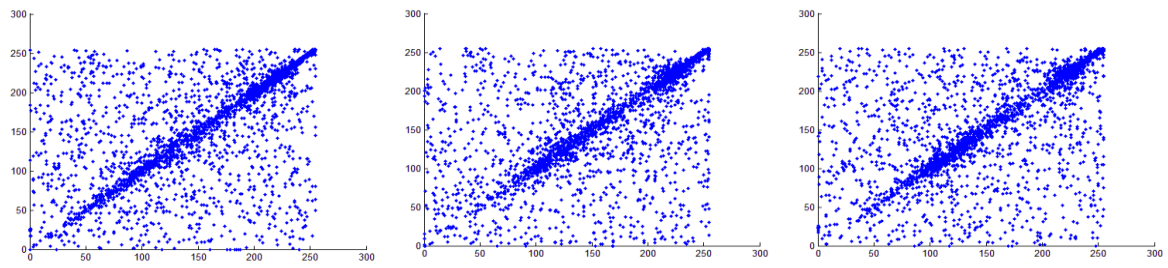
(a1,a2,a3).Original frame



(b1,b2,b3). Skin Detected



(c1,c2,c3).Only Skin encrypted



(d1,d2,d3).Encrypted output frame

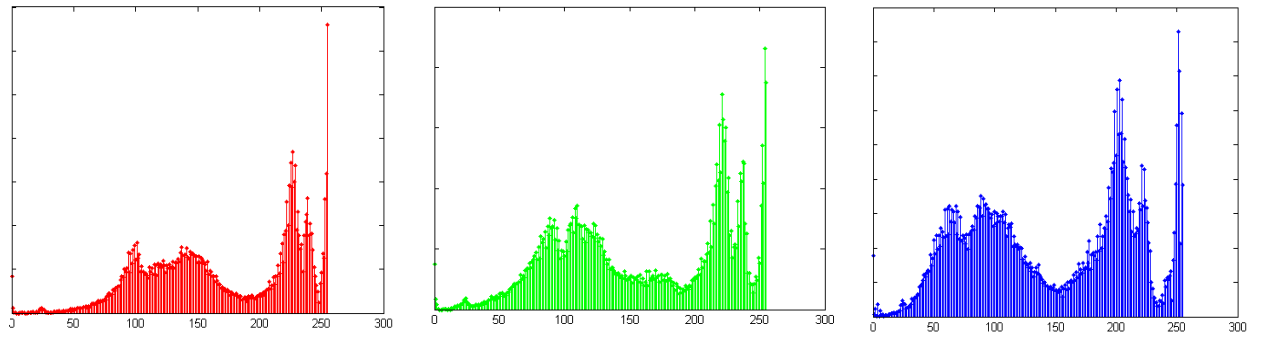
Fig.17. Correlation distributing of original and encrypted bitstream. (a1,a2,a3) correlation of horizontally, vertically and diagonally adjacent pixels of original frame. (b1,b2,b3) correlation of horizontally, vertically and diagonally adjacent pixels of only skin detected. (c1,c2,c3) correlation of horizontally, vertically and diagonally adjacent pixels of only skin encrypted. (d1,d2,d3) correlation of horizontally, vertically and diagonally adjacent pixels of encrypted output frame.

Contrast Analysis

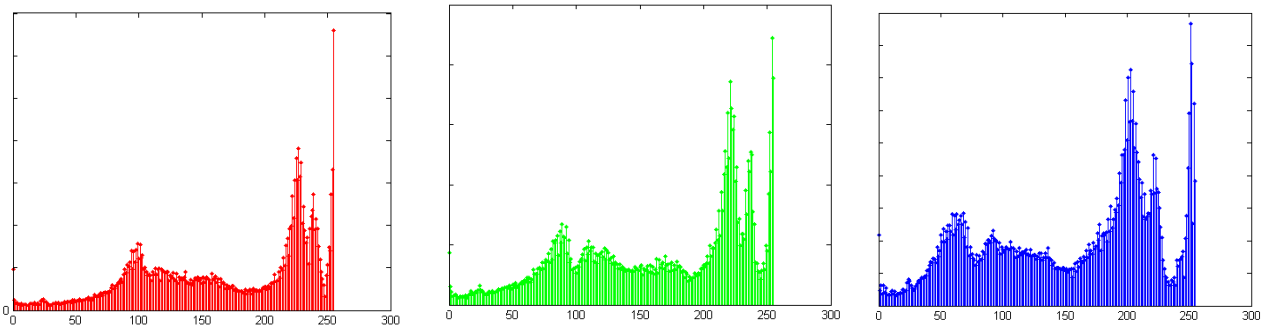
To further validate the strength of the proposed scheme contrast analysis has been performed. In contrast analysis the difference of intensities of pixels and its neighbor pixels is calculated as:

$$S = \sum(|m - n|^2) \times K(m, n) \tag{44}$$

The higher contrast value, the better will be the encryption scheme. Fig.18 shows that the proposed scheme has attain higher value of contrast between the original and encrypted frames, thus the proposed scheme can be considered robust.



(a,a1,a2) Original frame # 124



(b,b1,b2) Encrypted frame # 124

Fig 18. Contrast analysis of foreman video over red, green and blue channels

Differential attacks

Furthermore, in the CFB mode the chaining dependency exits and thus sensitive towards any change in the initialization vector (IV) generate different output encrypted bitstream for the same original bitstream.

Key strength analysis

Brute-force attack or sometime known as exhaustive attack are the attacks in which every possible key combination is tried to identify the correct key. The key strength against the brute force attack is mostly measured by key space. The key space of 2^{100} is considered resilient against brute force attacks. [64]. Moreover, 128-bit keys are considered unbreakable till 2020. Thus in the proposed scheme 128-bit key are utilized in the AES-CFB encryption.

Man-in middle attack

In this work Diffie-Hellman key exchange method [59] is employed to generate and distribute the keys so that the secret key can never be extracted as plaintext but only in encrypted form and hence

secure against man-in middle attack.

4.3 Parallelism

To fully exploit the smart surveillance system, parallelism with multiple computing nodes has been implemented. For the parallel computing standardized Message Passing Interface(MPI) has been utilized. The parallel time consists of processing time (t_{Pro}) and communication time/data distribution time (t_{comm}) among different nodes of the surveillance system.

$$T_p = t_{(Pro)} + t_{(Comm)} \quad (45)$$

The comparison of encoding time for sequential and parallel (with three processors) with proposed solution for Forman and Cricket video is provided in Table 7 and Table 8 respectively. Although the distributed computing increases the performance efficiency but there is a stark trade-off between number of processor and speedup for resource constrained smart systems. The speedup performance for skin detection, encryption and Total execution with the proposed scheme for cricket video is illustrated in Fig. 19. The results show that overall performance efficiency achieved is encouraging with the parallel computation; however the parallel performance decreases gradually as the number of processing nodes increases. This performance variation is due to communication overhead which dominates the processing time t_{Pro} of distributed devices. Moreover, the performance of parallelism is also architecture dependent, therefore, it may varies on different architectures.

Table 7. Sequential and parallel encoding time (sec) with proposed scheme for Foreman video

Frame No	Sequential Time (Sec)				Parallel Time (Sec)				Efficiency (%)
	Skin Detection Time	Encryption Time	Reconstr. Time	Total Sequential Time (T_s)	Skin Detection Time $t_{(Pro)}+t_{(Comm)}$	Encryption Time $t_{(Pro)}+t_{(Comm)}$	Reconstr. Time $t_{(Pro)}+t_{(Comm)}$	Total Parallel Time (T_p)	
39	0.005	0.002	0.003	0.010	0.00379	0.00148	0.00213	0.00741	45.016
92	0.005	0.002	0.001	0.008	0.0039	0.00128	0.00053	0.00571	46.718
124	0.006	0.001	0.002	0.009	0.00415	0.00055	0.00124	0.00594	50.527
154	0.006	0.002	0.002	0.010	0.00463	0.00127	0.00121	0.00711	46.901
207	0.004	0.001	0.002	0.007	0.00345	0.00057	0.00078	0.00479	48.63
281	0.005	0.001	0.001	0.007	0.00378	0.00055	0.00063	0.00497	46.951
Average	**0.0049**	**0.0016**	**0.0015**	**0.0084**	**0.00382**	**0.00084**	**0.001**	**0.00596**	**46.959**

Table 8. Sequential and parallel encoding time (sec) with proposed scheme for Cricket sequence

Frame No	Sequential Time (Sec)				Parallel Time (Sec)				Efficiency (%)
	Skin Detection Time	Encryption Time	Reconstr. Time	Total Sequential Time (T_s)	Skin Detection Time $t_{(Pro)}+t_{(Comm)}$	Encryption Time $t_{(Pro)}+t_{(Comm)}$	Reconstr. Time $t_{(Pro)}+t_{(Comm)}$	Total Parallel Time (T_p)	
135	0.005	0.002	0.001	0.008	0.00482	0.00122	0.00063	0.00568	46.977
202	0.006	0.001	0.002	0.009	0.00493	0.00068	0.00123	0.00684	43.860
265	0.005	0.001	0.003	0.009	0.00400	0.00055	0.00140	0.00595	50.443
555	0.005	0.002	0.001	0.008	0.00367	0.00107	0.00057	0.00532	50.164
769	0.006	0.002	0.002	0.01	0.00499	0.00107	0.00107	0.00713	46.779
1114	0.004	0.002	0.001	0.007	0.00307	0.00127	0.00063	0.00497	46.975
Average	**0.006**	**0.002**	**0.002**	**0.009**	**0.005**	**0.00097**	**0.00092**	**0.00598**	**47.533**

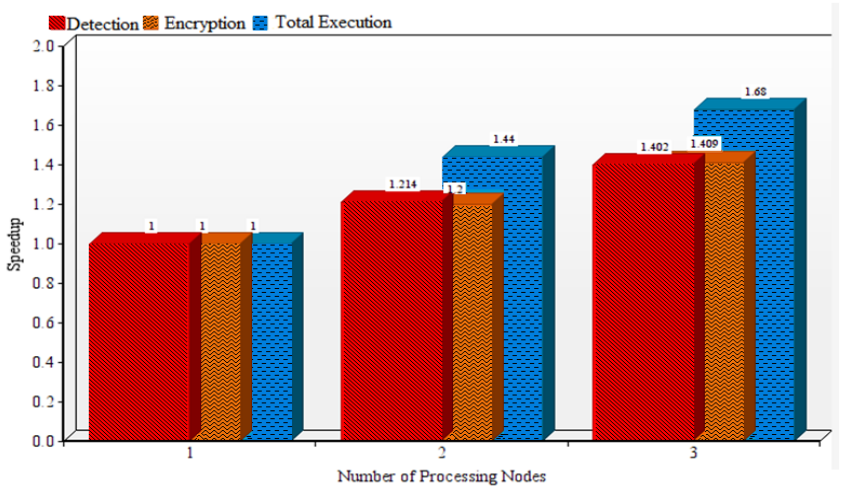


Fig. 19. Speedup with the parallel processing with proposed method for Cricket video

Furthermore, the good execution rate of our proposed scheme for the privacy protection make it good choice for the smart surveillance system. Table 9 presents execution rate (MB/sec) for test video datasets. The results shows that with the proposed scheme maximum 30.06MB/sec has been achieved for High-definition (HD) (1280 x 720 pixels/frame) video bit-streams which depict the suitability of the proposed method for surveillance systems.

Table 9. Comparative Average execution rate (MB/sec) with proposed scheme and default SE of Foreman, Paris and Cricket bit-streams

Video	Size of Video (MB) (YUV)	Total Number of Frames within a Video	Proposed SE	Default SE
			Execution Rate (MB/Sec)	Execution Rate (MB/Sec)
Foreman	43.506	300	4.75	0.511
Paris	154.446	1065	11.33	0.9314
Cricket	1924.805	1459	30.069	11.367

5 Conclusion

Two approaches for Skin detection and privacy protection are presented in this paper. The contribution of the research is to detect human skin within video bit-streams in the presence of different skin colors/complexions by transforming them into RGB, perceptual (HSV) and orthogonal (YCbCr) color-space. Moreover, in this work the dynamic and explicit threshold schemes are adopted. The idea of the proposed scheme is based on the assumption that under the same illumination conditions within the video frames, the dynamic thresholding schemes suffers due to localization, intra-personal features (e.g. physical appearance, makeup and hair style) and ambient environmental conditions (as discussed in use case 1, 2 and 3). Thus, in these cases the explicit threshold will work to provide the sufficient privacy protection. Furthermore, in this paper clustering method with CTR values for combined RGB, HSV and YCbCr color-spaces is proposed to improve detection in term of greater accuracy and precision. The experimental results affirm that proposed CTR based segmentation attain high detection rate without adding any computational complexity and memory requirement. This confirms that the performance of pixel-based skin classifiers improves significantly. Additionally, in order to protect the privacy of the subjects, once skin detection has been performed, the detected skin pixels (including false positives) are encrypted with a standard secure encryption algorithm (AES-CFB) in a stream cipher mode. Further, this method was compared to the default selective encryption with selected entropy coder parameters of the whole frame. It was found that the latter method, whatever its practical value for real-time video streaming, did not protect the privacy of the individual as much as first applying skin detection and then partially encrypting those detected pixels. The alternative, in terms of reducing encryption time by reducing the number of pixels or parameters encrypted, is to selectively encrypt a video bit-streams' frames. However, our results strongly imply that to protect a person within a video's privacy the method of encrypting skin detection is preferable. Apart from protection of the privacy the context of the surveillance required to be processed for the behavior analysis, thus the proposed scheme preserves the behavior so that can be used for behavior analysis without breaching the individual's privacy. However, identity protection of people is a first step as part of such a program. Similarly, privacy protection was the

first objective of this research and choice of color-space has an interesting impact, not so much for automatically determining the ethnicity of those under surveillance but much more so for providing the most effective form of privacy protection. There are many interesting lines of research stemming from this study. Key management governing access to encrypted regions is important. It is also possible that different regions, for example detected skin pixels and other areas could be differentially protected by encryption. Thus one could combine selective encryption of non-detected-skin pixels with the fully encrypted detected skin pixels, as in this scheme. Furthermore, the ESR and security analysis show that the selective encryption on the specific skin pixels provides good privacy protection without incurring considerable encryption overhead. Thus due to simplicity and efficiency the proposed scheme is good choice for the resources constrained surveillance devices.

Acknowledgment: This research paper is produced as part of a government-funded project (National Research Program for Universities (NRPU-2016)) with no: 6282/Punjab/NRPU/R&D/HEC/2016. We appreciate the support of the Higher Education Commission (HEC) of Pakistan for this project.

References

1. Pujol, F., Pujol, M., Jimeno-Morenilla, A., Pujol, M.: Face Detection Based on Skin Color Segmentation Using Fuzzy Entropy. Entropy. 19, 26 (2017). doi:10.3390/e19010026
2. Ban, Y., Kim, S.-K., Kim, S., Toh, K.-A., Lee, S.: Face detection based on skin color likelihood. Pattern Recognit. 47, 1573–1585 (2014). doi:10.1016/J.PATCOG.2013.11.005
3. Zafeiriou, S., Zhang, C., Zhang, Z.: A survey on face detection in the wild: Past, present and future. Comput. Vis. Image Underst. 138, 1–24 (2015). doi:10.1016/J.CVIU.2015.03.015
4. Chai, D., Ngan, K.N.: Face segmentation using skin-color map in videophone applications. IEEE Trans. Circuits Syst. Video Technol. 9, 551–564 (1999). doi:10.1109/76.767122
5. Cheddad, A., Condell, J., Curran, K., Mc Kevitt, P.: A skin tone detection algorithm for an adaptive approach to steganography. Signal Processing. 89, 2465–2478 (2009). doi:10.1016/J.SIGPRO.2009.04.022
6. Tan, W.R., Chan, C.S., Yogarajah, P., Condell, J.: A Fusion Approach for Efficient Human Skin Detection. IEEE Trans. Ind. Informatics. 8, 138–147 (2012). doi:10.1109/TII.2011.2172451
7. Bianco, S., Gasparini, F., Schettini, R.: Adaptive Skin Classification Using Face and Body Detection. IEEE Trans. Image Process. 24, 4756–4765 (2015). doi:10.1109/TIP.2015.2467209
8. Jairath, S., Bharadwaj, S., Vatsa, M., Singh, R.: Adaptive Skin Color Model to Improve Video Face Detection. Presented at the (2016)
9. Luo, Y., Guan, Y.-P.: Adaptive skin detection using face location and facial structure estimation. IET Comput. Vis. 11, 550–559 (2017). doi:10.1049/iet-cvi.2016.0295
10. Zuo, H., Fan, H., Blasch, E., Ling, H.: Combining Convolutional and Recurrent Neural Networks for Human Skin Detection. IEEE Signal Process. Lett. 24, 289–293 (2017). doi:10.1109/LSP.2017.2654803
11. Saini, M., Atrey, P.K., Mehrotra, S., Kankanhalli, M.: W3-privacy: understanding what, when, and where inference channels in multi-camera surveillance video. Multimed. Tools Appl. 68, 135–158 (2014). doi:10.1007/s11042-012-1207-9
12. Auer, S., Bliem, A., Engel, D., Uhl, A., Unterweger, A.: Bitstream-Based JPEG Encryption in Real-time. Int. J. Digit. Crime Forensics. 5, 1–14 (2013). doi:10.4018/jdcf.2013070101
13. Zhang, D., An, P., Zhang, H.: Application of robust face recognition in video surveillance systems. Optoelectron. Lett. 14, 152–155 (2018). doi:10.1007/s11801-018-7199-6
14. Cosar, S., Donatiello, G., Bogorny, V., Garate, C., Alvares, L.O., Bremond, F.: Toward Abnormal Trajectory and Event Detection in Video Surveillance. IEEE Trans. Circuits Syst. Video Technol. 27, 683–695 (2017). doi:10.1109/TCSVT.2016.2589859
15. Abdelhedi, S., Wali, A., Alimi, A.M.: Fuzzy Logic Based Human Activity Recognition in Video Surveillance Applications. Presented at the (2016)
16. Tripathi, V., Mittal, A., Gangodkar, D., Kanth, V.: Real-time security framework for detecting abnormal events at ATM installations. J. Real-Time Image Process. 1–11 (2016). doi:10.1007/s11554-016-0573-3
17. Qureshi, F.Z.: Object Video Streams: A Framework for Preserving Privacy in Video Surveillance. In: Intelligent Multimedia Surveillance. pp. 67–82. Springer Berlin Heidelberg, Berlin, Heidelberg (2013)
18. Asghar, M.N., Ghanbari, M., Fleury, M., Reed, M.J.: Confidentiality of a selectively encrypted H.264 coded

video bit-stream. J. Vis. Commun. Image Represent. 25, 487–498 (2014). doi:10.1016/J.JVCIR.2013.12.015
19. Rashwan, H.A., Solanas, A., Puig, D., Martínez-Ballesté, A.: Understanding trust in privacy-aware video surveillance systems. Int. J. Inf. Secur. 15, 225–234 (2016). doi:10.1007/s10207-015-0286-9
20. Rahman, S.M.M., Hossain, M.A., Hassan, M.M., Alamri, A., Alghamdi, A., Pathan, M.: Secure privacy vault design for distributed multimedia surveillance system. Futur. Gener. Comput. Syst. 55, 344–352 (2016). doi:10.1016/j.future.2014.10.019
21. Padilla-López, J.R., Chaaraoui, A.A., Flórez-Revuelta, F.: Visual privacy protection methods: A survey. Expert Syst. Appl. 42, 4177–4195 (2015). doi:10.1016/J.ESWA.2015.01.041
22. Erdélyi, Á., Winkler, T., Rinner, B.: Privacy protection vs. utility in visual data. Multimed. Tools Appl. 77, 2285–2312 (2018). doi:10.1007/s11042-016-4337-7
23. Kong, S.G., Heo, J., Abidi, B.R., Paik, J., Abidi, M.A.: Recent advances in visual and infrared face recognition—a review. Comput. Vis. Image Underst. 97, 103–135 (2005). doi:10.1016/J.CVIU.2004.04.001
24. Zhihong Pan, Healey, G., Prasad, M., Tromberg, B.: Face recognition in hyperspectral images. IEEE Trans. Pattern Anal. Mach. Intell. 25, 1552–1560 (2003). doi:10.1109/TPAMI.2003.1251148
25. Kakumanu, P., Makrogiannis, S., Bourbakis, N.: A survey of skin-color modeling and detection methods. Pattern Recognit. 40, 1106–1122 (2007). doi:10.1016/J.PATCOG.2006.06.010
26. Brancati, N., De Pietro, G., Frucci, M., Gallo, L.: Human skin detection through correlation rules between the YCb and YCr subspaces based on dynamic color clustering. Comput. Vis. Image Underst. 155, 33–42 (2017). doi:10.1016/J.CVIU.2016.12.001
27. Dadgostar, F., Sarrafzadeh, A.: An adaptive real-time skin detector based on Hue thresholding: A comparison on two motion tracking methods. Pattern Recognit. Lett. 27, 1342–1352 (2006). doi:10.1016/J.PATREC.2006.01.007
28. Naji, S.A., Zainuddin, R., Jalab, H.A.: Skin segmentation based on multi pixel color clustering models. Digit. Signal Process. 22, 933–940 (2012). doi:10.1016/J.DSP.2012.05.004
29. Gupta, A., Chaudhary, A.: Robust skin segmentation using color-space switching. Pattern Recognit. Image Anal. 26, 61–68 (2016). doi:10.1134/S1054661815040033
30. De Dios, J.J., Garcia, N.: Face detection based on a new color-space YCgCr. In: Proceedings 2003 International Conference on Image Processing (Cat. No.03CH37429). p. III-909-12. IEEE
31. Gonzalez, R., Woods, R.: Digital image processing. (2002)
32. Mahmoodi, M.R., Sayedi, S.M., Karimi, F.: Color-based skin segmentation in videos using a multi-step spatial method. Multimed. Tools Appl. 76, 9785–9801 (2017). doi:10.1007/s11042-016-3579-8
33. Chawla, D., Trivedi, M.C.: Difference in Lights and Color Background Differentiates the Color Skin Model in Face Detection for Security Surveillance. In: Networking Communication and Data Knowledge Engineering. pp. 127–135. Springer, Singapore (2018)
34. Shaik, K.B., Ganesan, P., Kalist, V., Sathish, B.S., Jenitha, J.M.M.: Comparative Study of Skin Color Detection and Segmentation in HSV and YCbCr Color-space. Procedia Comput. Sci. 57, 41–48 (2015). doi:10.1016/J.PROCS.2015.07.362
35. Prema, C., Manimegalai, D.: Survey on skin tone detection using color-spaces. Int. J. Appl. Inf. Syst. 2, 18–26 (2012)
36. Yan, C., Wang, Z., Zhou, X.: Design of a detection system of faces intercepted by video based on the skin color model. In: 2017 4th International Conference on Systems and Informatics (ICSAI). pp. 165–169. IEEE (2017)
37. Saini, M., Atrey, P.K., Mehrotra, S., Kankanhalli, M.: Adaptive Transformation for Robust Privacy Protection in Video Surveillance. Adv. Multimed. 2012, 1–14 (2012). doi:10.1155/2012/639649
38. Luo, Y., Cheung, S.S., Lazzeretti, R., Pignata, T., Barni, M.: Anonymous subject identification and privacy information management in video surveillance. Int. J. Inf. Secur. 17, 261–278 (2018). doi:10.1007/s10207-017-0380-2
39. Korshunov, P., Ebrahimi, T.: Towards optimal distortion-based visual privacy filters. In: 2014 IEEE International Conference on Image Processing (ICIP). pp. 6051–6055. IEEE (2014)
40. Ma, X., Yang, L.T., Xiang, Y., Zeng, W.K., Zou, D., Jin, H.: Fully Reversible Privacy Region Protection for Cloud Video Surveillance. IEEE Trans. Cloud Comput. 5, 510–522 (2017). doi:10.1109/TCC.2015.2469651
41. Agrawal, P., Narayanan, P.J.: Person De-Identification in Videos. IEEE Trans. Circuits Syst. Video Technol. 21, 299–310 (2011). doi:10.1109/TCSVT.2011.2105551
42. Thorpe, C., Li, F., Li, Z., Yu, Z., Saunders, D., Yu, J.: A Coprime Blur Scheme for Data Security in Video Surveillance. IEEE Trans. Pattern Anal. Mach. Intell. 35, 3066–3072 (2013). doi:10.1109/TPAMI.2013.161

43. Hoo, W.L., Miron, A., Badii, A., Chan, C.S.: Skin-based privacy filter for surveillance systems. In: 2015 International Conference on Systems, Signals and Image Processing (IWSSIP). pp. 269–272. IEEE (2015)
44. Chattopadhyay, A., Boult, T.E.: PrivacyCam: a Privacy Preserving Camera Using uCLinux on the Blackfin DSP. In: 2007 IEEE Conference on Computer Vision and Pattern Recognition. pp. 1–8. IEEE (2007)
45. Dufaux, F., Ebrahimi, T.: A framework for the validation of privacy protection solutions in video surveillance. In: 2010 IEEE International Conference on Multimedia and Expo. pp. 66–71. IEEE (2010)
46. Melle, A., Dugelay, J.-L.: Scrambling faces for privacy protection using background self-similarities. In: 2014 IEEE International Conference on Image Processing (ICIP). pp. 6046–6050. IEEE (2014)
47. Saini, M., Atrey, P.K., Mehrotra, S., Emmanuel, S., Kankan, M.: Privacy modeling for video data publication. In: 2010 IEEE International Conference on Multimedia and Expo. pp. 60–65. IEEE (2010)
48. Privacy mode for acquisition cameras and camcorders. (1998)
49. Cheung, S.S., Paruchuri, J.K., Nguyen, T.P.: Managing privacy data in pervasive camera networks. In: 2008 15th IEEE International Conference on Image Processing. pp. 1676–1679. IEEE (2008)
50. Ciftci, S., Akyuz, A.O., Ebrahimi, T.: A Reliable and Reversible Image Privacy Protection Based on False Colors. IEEE Trans. Multimed. 20, 68–81 (2018). doi:10.1109/TMM.2017.2728479
51. Song, W., Wu, D., Xi, Y., Park, Y.W., Cho, K.: Motion-based skin region of interest detection with a real-time connected component labeling algorithm. Multimed. Tools Appl. 76, 11199–11214 (2017). doi:10.1007/s11042-015-3201-5
52. Bilal, S., Akmeliawati, R., Salami, M.J.E., Shafie, A.A.: Dynamic approach for real-time skin detection. J. Real-Time Image Process. 10, 371–385 (2015). doi:10.1007/s11554-012-0305-2
53. Guo, J., Xu, J., Bao, J.: Region of interest based selective encryption scheme for privacy protection in H.264 video. J. Shanghai Jiaotong Univ. 19, 385–391 (2014). doi:10.1007/s12204-014-1513-7
54. Viola, P., Jones, M.J.: Robust Real-Time Face Detection. Int. J. Comput. Vis. 57, 137–154 (2004). doi:10.1023/B:VISI.0000013087.49260.fb
55. Yogarajah, P., Condell, J., Curran, K., Cheddad, A., McKevitt, P.: A dynamic threshold approach for skin segmentation in color images. In: 2010 IEEE International Conference on Image Processing. pp. 2225–2228. IEEE (2010)
56. Basilio, J.A.M., Torres, G.A., Pérez, G.S., Medina, L.K.T., Meana, H.M.P.: Explicit Image Detection using YCbCr Space Color Model as Skin Detection. Appl. Math. Comput. Eng. 123–128 (2011
57. Technology, N.I. of S. and: FIPS-197: Advanced Encryption Standard (AES), (2001)
58. Jayasinghe, D., Ragel, R., Ambrose, J.A., Ignjatovic, A., Parameswaran, S.: Advanced modes in AES: Are they safe from power analysis based side channel attacks? 2014 32nd IEEE Int. Conf. Comput. Des. ICCD 2014. 173–180 (2014). doi:10.1109/ICCD.2014.6974678
59. E. Rescorla: Diffie-Hellman Key Agreement Method, https://www.ietf.org/rfc/rfc2631.txt
60. Asghar, M.N., Ghanbari, M., Fleury, M., Reed, M.J.: Sufficient encryption based on entropy coding syntax elements of H.264/SVC. Multimed. Tools Appl. 74, 10215–10241 (2015). doi:10.1007/s11042-014-2160-6
61. Qiang Zhu, Kvvang Ting Cheng, Ching-Tung Wu: A unified adaptive approach to accurate skin detection. In: 2004 International Conference on Image Processing, 2004. ICIP '04. pp. 1189–1192. IEEE
62. Bhardwaj, S., Mittal, A.: A Survey on Various Edge Detector Techniques. Procedia Technol. 4, 220–226 (2012). doi:10.1016/j.protcy.2012.05.033
63. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image Quality Assessment: From Error Visibility to Structural Similarity. IEEE Trans. Image Process. 13, 600–612 (2004). doi:10.1109/TIP.2003.819861
64. Khan, J.S., Ahmad, J.: Chaos based efficient selective image encryption. Multidimens. Syst. Signal Process. 1–19 (2018). doi:10.1007/s11045-018-0589-x