Design of Enhanced Region Separation Algorithm for Effective Video Compression

Mahalakshmi Ramadoss , S.K Mahendran

Abstract: ***An important component of video communication system is video compression, which is the art of reducing the size of video file without the loss of visual quality. Apart from lossy and lossless compression, to increase the performance of compression algorithm, a ROI (Region of Interest) based compression algorithm is used. In ROI-based compression algorithms, a lossless algorithm is used to compress ROI region and lossy compression is used for non-ROI regions. In these types of algorithms, an important step is the accurate separation of ROI and non-ROI regions. In this research wok a region separation algorithm that extracts the ROI region effectively from its background is proposed. The ROI considered in this work is the human face. The proposed algorithm uses a dynamic hybrid color space algorithm along with Ant Colony Optimization and Kernel Principal Component Analysis algorithms to detect face skin regions, followed by the use of eye-mouth map features to detect face region in the video. A Kalman filter is used to track the detected ROI in the video. Experimental results shows that the proposed algorithm is efficient in identifying the face regions in terms of accuracy.***

Keywords: Face Detection, Face Skin Detection, Hybrid Color Space Model, Region Separation, ROI-Based Video Compression, Tracking.

I. INTRODUCTION

Communication is defined as the technology employed for transmitting messages from one person to another person using various mediums means like text, voice, video and image. In visual communication, usage of video has become more popular (Hussain and Ahmed, 2019).Video communication is the a field of research which studies the area concerned with the transmission of video information.One important part of this research is video compression, which is defined as an algorithm that is used to reduce the size of a video file, (i) without losing any important information and (ii) without degrading its visual quality. Compression of videos offers several advantages including efficient use of memory space and bandwidth availableand reducing the time required during transmission and downloading. Video compression consists of two tasks, namely, compression and decompression. Video compression algorithms can belong to two main categories. They are

Revised Manuscript Received on 30 May 2019.
***** Correspondence Author
Mahalakshmi Ramadoss, Research Scholar, Department of Computer Science, Bharathiar University.
S.K.Mahendran, Assistant Professor & HOD, Department of Computer Science, Govt Arts and Science College, Udhagamandalam, The Nilgiris.

1. Lossless compression where the reconstructed video is the exact replica of the original one. Examples include Antony and Ganapathy (2015), Wang *et al.* (2019).
2. Lossy compression where the reconstructed video has some loss that has minimum visibility the human eyes. Examples include Wei *et al.*, (2009), Pansare and Jadhav, (2018).

Many of the existing solutions apply the compression algorithm to the whole video, which often leads to huge number of artifacts, which also degrade the visual quality of the video(Pawłowski*et al.*, 2018). This issue is solved using a third of compression category called Region of Interest (ROI) compression (Ahmed *et al.*, 2018). These algorithms uses lossy compression on non-ROI regions and lossless compression on important ROI-regions, thus providing more quality to important regions (ROI) at the expense of reduced quality of other regions. The ROI-based video compression works on the principal that a viewer often gives more importance to certain regions over others and manipulation of this principal can produce better compression rate. It has several applications which includes real-time video transmission in surveillance, video conferencing / meeting, news telecasting application and medical videos. In several of these applications, human face plays an important role and therefore, is considered as the ROI in this work. Viewers are disturbed more by quality reductions in face rather than the background. Thus, the usage of a lossless video compression algorithm for face region and a lossy algorithm for other regions will be more beneficial while reducing the video file size. A ROI-based video compression algorithm consists of three main steps, namely, region separation algorithm to identify ROI and non-ROI regions, design of lossy compression algorithm for non-ROI regions and design of lossless compression algorithm for ROI-region (Xue*et al.*, 2016). The focal point of this paper is the design of a region separation algorithm. The rest of the paper is organized as follows. Section 2 describes the methodology used to design proposed region separation algorithm. Section 3 presents the results of the performance evaluation conducted to analyze the efficiency of the proposed algorithm, while Section 4 concludes the work with future research directions.

II. METHODOLOGY

The proposed region separation algorithm is the synergistic combination of four algorithms, namely, Frame Sequence Generator, Foreground/Background Region Separator, ROI Detector in Foreground Region and ROI Tracking Algorithm. The goal here is to design an algorithm that can detect background region and two foreground regions, (regions which do not have face object and regions which has the face object) along with a ROI-region tracking algorithm (Figure 1).

The working of each of these steps is described in this section.

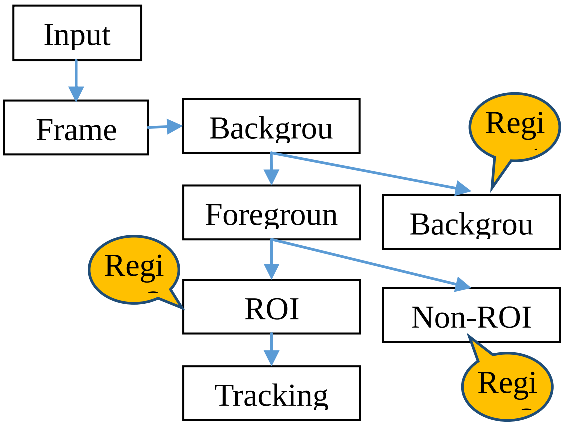

Figure 1 :Design of Region Separation Algorithm

2.1. Frame Sequence Generator

The aim of frame sequence generator is to separate the various frames (or images) in a video. This algorithm converts a video into a series of images with 640 x 480 resolution. The frame rate considered is 30 fps (frames per second).

2.2. Foreground/Background Region Separator

The region separator algorithm is used to divide a video frame into three regions, namely, background, foreground with ROI region and foreground with non-ROI regions. Initially, the frame differencing algorithm (https://en.wikipedia.org/wiki/Foreground_detection#Using _ frame_differencing) is used to detect the background and foreground regions. The next step is to detect the ROI or face in the foreground region. The automatic face detection algorithm is designed as a binary classification problem and has two major steps, namely, face detection and localization & extraction. Face detection is the process of analyzing a video frame to determine whether it consists of human faces or not. If the video frame has face, then the second step is used to locate and extract the face regions. The ROI-Region Detection algorithm performs face detection in three steps as listed below.

(i) Face localization – Using enhanced skin region detection algorithm
(ii) Feature Extraction – Location of eye(s) and mouth
(iii) Face Detection – Using boundaries enclosing eye and mouth detected

2.2.1. Face Localization

Face localization, in this work, is performed using Face Skin Region Detection (FSRD). FSRD is a challenging and important step during face detection, where several researches have been conducted (Jing *et al.*, 2019). One of the important step during FSRD is the use of an effective color space model. Despite the substantial research efforts in this area, choosing an optimal color space in terms of skin and face detection has remained an open issue. During the last decade, several conventional color spaces such as RGB, YCbCr, HSI, HSV, CIEXYZ, YUV, etc., have been used in many image and video analysis applications (Gowda and Yuan, 2019). Each of these color space models have its own advantages and drawbacks.In order to obtain a sophisticated color space model, hybrid models have been proposed as alternative to conventional color space models (Oghazet *al.*, 2015; 2017).These hybrid models combine color components from different models to obtain maximum advantage during skin and face detection.

In this work, a multiple color model is proposed which combines 18 color space models using an color component selection algorithm. This model is termed as Hybrid Color Model Created using Dynamic Color Component Selection Algorithm (HCD2CS).The basic idea of the proposed method is to combine several color components from different color spaces in order to

- To select optimal color components that can increase the discriminate effectively the skin and non-skin colors
- to reduce the average correlation rate between color components
- separate the intensity and chrominance components to increase the likeness
- unity of the skin tones of different human ethnic groups
- To improve the accuracy face identification

The 18 color systems used during the design of the proposed color space model belong to four major categories and are listed below.

– Primaries color systems (4) - RGB, nRGB, XYZ. LSLM
– Luminance-chrominance color systems (8) - YUV, YIQ, YCbCr, YPbPr, YCgCr, YQCr, HSV, HSI
– Perceptual color systems (2) - L*a*b, Luv
– Non-correlated color systems (2) - $I_1I_2I_3$, TSL
– Existing Hybrid Models (2) -YES, RIQ

The 18 selected models together has 56 color components, which have several repeating components. In order to eliminate redundancy, only non-repeating components are selected.This yields 36 unique color components, which are used to form HCD2CS. Figure 2 shows the 56 color components and thereduced 18 unique color components (indicated by #).

1	$R_{RGB}^{\#}$	19	$Y_{YCBCR}^{\#}$	37	$L_{LAB}^{\#}$
2	$G_{RGB}^{\#}$	20	$CB_{YCBCR}^{\#}$	38	$A_{LAB}^{\#}$
3	$B_{RGB}^{\#}$	21	$CR_{YCBCR}^{\#}$	39	$B_{LAB}^{\#}$
4	$R_{nRGB}^{\#}$	22	Y_{YPBPR}	40	L_{LUV}
5	$G_{nRGB}^{\#}$	23	$PB_{YPBPR}^{\#}$	41	$U_{LUV}^{\#}$
6	$B_{nRGB}^{\#}$	24	$PR_{YPBPR}^{\#}$	42	$V_{LUV}^{\#}$

7	X_{XYZ}#	25	Y_{YCGCR}	43	$i1_{i1i2i3}$#
8	Y_{XYZ}#	26	CG_{YCGCR}#	44	$i2_{i1i2i3}$#
9	Z_{XYZ}#	27	CR_{YCGCR}#	45	$i3_{i1i2i3}$#
10	L_{LSLM}#	28	Y_{YQCR}	46	T_{TSL}
11	S_{LSLM}#	29	Q_{YQCR}	47	S_{TSL}#
12	LM_{LSLM}#	30	CR_{YQCR}	48	L_{TSL}#
13	Y_{YUV}	31	H_{HSV}	49	Y_{YES}
14	U_{YUV}	32	S_{HSV}	50	E_{YES}#
15	V_{YUV}	33	V_{HSV}	51	S_{YES}#
16	Y_{YIQ}	34	H_{HSI}#	52	R_{RIQ}
17	I_{YIQ}#	35	S_{HSI}#	53	I_{RIQ}
18	Q_{YIQ}#	36	I_{HSI}#	54	Q_{RIQ}

Figure 2 : Original and Unique Components(#) of 18 Selected Color Space Models

A color space model with 36 components will raise the issue of curse of dimensionality. Thus, it is imperative to find a smaller combination setup of color components that can improve the identification of human skin. For this, a dynamic color component selection and dimensionality reduction algorithm is proposed (Figure 3).

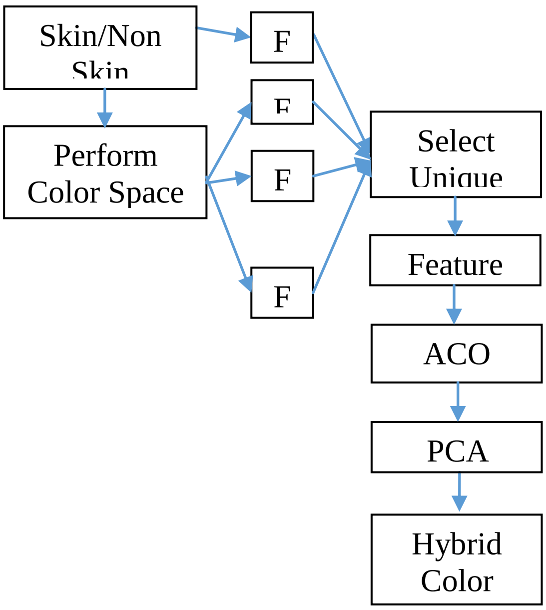

Figure 3 : Proposed HCD2CSColor Space Model

The dynamic selection is performed using Ant Colony Optimization (ACO) algorithm (Manosij*et al.*, 2019) while the dimensionality reduction is performed using enhanced Kernel Principal Component Analysis (KPCA) (Kung, 2014) algorithm. The output of ACO-based selection algorithm is a nine color component set {V_{HSV}, G_{nRGB}, B_{RGB}, Y_{YUV}, Q_{YIQ}, CB_{YCBCR}, CR_{YCBCR}, E_{YES}, $I3_{i1i2i3}$}. This color space model having nine color components also suffers from high dimensionality. This problem is solved through the use of KPCA algorithm. The first three significant principal components are used to form the new 3-component color model. The proposed 3-component color model has the components, namely, Y_{YUV}, CB_{YCBCR}, CR_{YCBCR}.

Using the above color space, a random forest classifier (https://en.wikipedia.org/wiki/Random_forest) is trained to detect skin and non-skin regions. The classifier is initially trained with a set of training images having samples from both skin and non-skin regions. A template database having 300 skin color patches was created using randomly selected 30 face images from the databases used during experimentation (Section 3). The skin color patches were constructed manually using Photoshop and some examples are shown in Figure 4.

Figure 4 : Example Skin Patches

The result of the random forest classifier still can be improved if two main issues can be solved. The first issue is that the skin regions detected may sometime contain disjoints or fragments with gaps and holes. The second issue is that the classifier sometimes wrongly label regions, that is a non-face skin may be identified as face-skin and non-skin region may be identified as skin-region and vice versa. In this work, both the issues are solved using morphological operators. A simple morphological dilation operation is applied on large regions to solve the first issue. The second issue is solved by applying the morphological opening operation, to open-up or separate connected skin regions. The open operator is derived from morphological erosion operation and is used to refine region boundaries and thus separate wrongly grouped regions. Apart from the above two solutions, to guarantee accuracy of the detected skin region to belong to human face, two more additional properties, namely, box ratio (width to height ratio of the region bounding box) and eccentricity (ratio of minor axis to major axis of a bounding ellipse), are used. The eccentricity ratio is more sensitive to the region shape and is able to consider various face rotations and poses. A detected region is considered to be a face region if the following two conditions are satisfied.

Condition 1 : $0.4 \leq$ Box Ratio ≥ 1.0
Condition 2 : $0.3 \leq$ Eccentricity ≥ 0.9

All the regions that fail with the post processing procedure are eliminated and the rest of the regions are considered as best possible face candidate regions.

2.2.2. Feature Extraction

The second step of face detection is feature extraction.Among the various facial features, eyes and mouth are the most suitable features for recognition. Eye regions are located by using binary template matching (Campadelli and Lanzarotti, 2002) and random forest classifier.

The mouth regions are located using the method proposed by Hsu et al. (2002). After extracting the eyes and mouth a triangle is drawn with the two eyes and a mouth as the three points in case of a frontal face. This results in an isosceles triangle (i j k) in which the Euclidean distance between two eyes is about 90-110% of the Euclidean distance between the centre of the right/left eye and the mouth. After getting the triangle, the coordinates of the four corner points that form the potential facial region is obtained, using which the face is segmented. The above mentioned method can detect face with a frontal view. When the image has one eye and one mouth, then a face from side view is obtained. This view has a right-angled triangle (i, j, k) and is identified if the Euclidean distance of (i k) is equal to twice that of the Euclidean distance of line (j k) and if the Euclidean distance of line (i j) is equal to 1.732 times that of the Euclidean distance of line (j k).A 25% deviation allowance is used as tolerance against imperfect binarization. Sometimes, there might be more than one set of eye-mouth features, in which case the number of triangles formed is more than one. In these situations, the correct feature triangle has to be obtained. This is done by analyzing the luma variance and average gradient orientation of eye-mouth along with the analysis of the geometry and orientation constraints of the triangle. A weighted score is estimated for each triangle and the triangle whose score is greater than a threshold is taken as the correct set. The weighted score is obtained as the average eye-mouth-triangle score weight. The eye pair score weight is the normalized average of the eye map value of the ith candidate around the two eyes by the eye map value of the most significant eye candidate. The mouth score weight is the normalized mouth map value of the kth candidate by the mouth map value of the most significant mouth candidate.

2.2.3. Face Detection

After identifying the precise triangle of the face, the face region is obtained using the coordinates of the four corner points. The procedure used to calculate the coordinates of frontal view face is as follows.

Let (X_i, Y_i), (X_j, Y_j) and (X_k, Y_k) be the three center points of i, j, and k that forms the isosceles triangle. Let (X_1, Y_1), (X_2, Y_2), (X_3, Y_3) and (X_4, Y_4) be the four corner points of the face region. Let D(i, k) be the Euclidean distance between the centers of blocks i and k. Here, the coordinates X_1 and X_4, X_2 and X_3, Y_1 and Y_2, Y3 and Y4 represents the same coordinates. The coordinates are calculated using Equation (5.34) – (5.37). After finding the four coordinates of the four corner points, a bounding box is drawn to indicate the detected face.

$$X_1 = X_4 = X_i - 1/3\ D(i, k) \quad (5.34)$$

$$X_2 = X_3 = X_k + 1/3\ D(i, k) \quad (5.35)$$

$$Y_1 = Y_2 = Y_i + 1/3\ D(i, k) \quad (5.36)$$

$$Y_3 = Y_4 = Y_j - 1/3\ D(i, k) \quad (5.37)$$

The 4 rules to get the face boundary for right side view are as follows (Equation 5.38 –5.41). After finding the four coordinates of the four corner points, a bounding box is drawn to indicate the detected face.

$$X1 = X4 = Xj - 1/6D(j, k) \quad (5.38)$$

$$X2 = X3 = Xj + 1.2D(j, k) \quad (5.39)$$

$$Y1 = Y2 = Yj + 1/4D(j, k) \quad (5.40)$$

$$Y3 = Y4 = Yj - 1.0D(j, k) \quad (5.41)$$

2.3. Tracking

Tracking of the face detected is defined as localization and association of features across series of frames.The main goal here is to establish a correspondence between face or face parts in consecutive frames and to extract temporal information about face such as path, posture, speed and direction. In this work, the detected face is treated as a whole object and tracking is performed frame by frame using Kalman Filter (Abdelali*et al.*, 2016).

2.4. EXPERIMENTAL RESULTS

Performance evaluation of the proposed algorithm was done using three datasets. The first dataset used is the color FERET dataset (https://www.nist.gov/itl/iad/image-group/color-feret-database), which has 14,126 face images collected from over 800 individuals. The second dataset was constructed by the authors using face images obtained from World Wide Web and has 20,000 images. This dataset is termed as Web dataset in this paper. The third dataset is the video dataset,having four videos (https://media.xiph.org/video/derf). Some sample frames from these videos are shown in Figure 5.

Foreman	Claire	News	Bowing

Figure 5 : Sample Frames from Video Dataset

Two stages of experiments were used during performance evaluation. The first stage focused on the proposed face detection algorithm. The second stage of experiments studied the effect of using the ROI separation algorithm during compression of videos.

Three performance metrics, namely, False Detection Rate (FDeR) (%), False Dismissal Rate (FDiR)(%) andAccuracy (%)were used to evaluate the algorithm in Stage 1 experiments. Compression ratio (%), Peak Signal to Noise Ratio (PSNR) (dB) and speed of compression were used as performance metrics by Stage 2 experiments. Four videos were used from Figures 6 to 8 shows the results of Stage 1 experiments for the selected three performance metrics.

The face detection algorithm that used the proposed HDC2CS model was compared with three other frequently used color space models, namely, RGB, HSV and YCbCr and two other hybrid models, namely, YES and RIQ. The classifier used, as mentioned earlier, is random forest classifier.

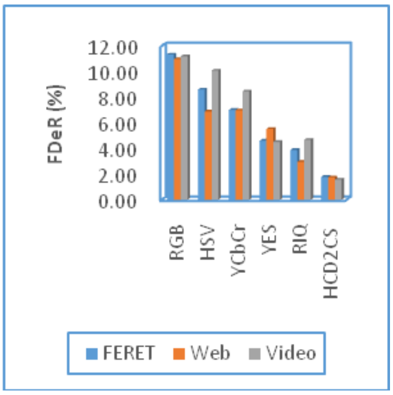

Figure 6 : FDeR (%)

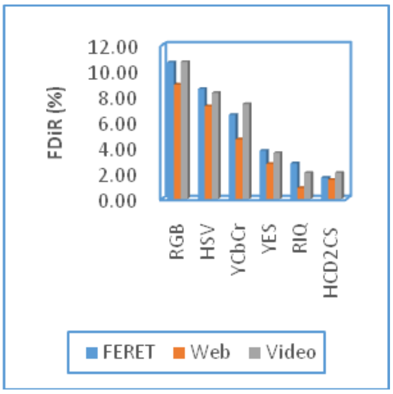

Figure 7 : FDiR (%)

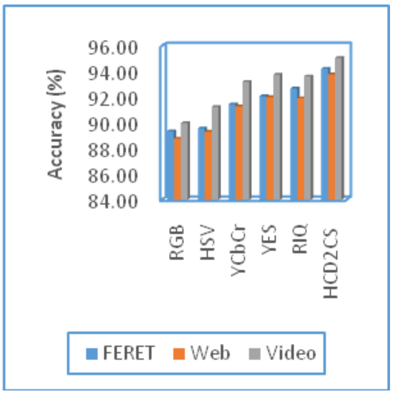

Figure 8 : Accuracy(%)

From the above results, it is clear that the usage of the dynamic hybrid model, HCD2CS, has improved the performance of ROI region detection with respect to all the selected performance metrics. This trend envisaged is the same with all the three datasets. An average accuracy of 94.35% was obtained by HCD2CS model during face detection, which has improved the face detection process by 5.29%, 4.57%, 2.52%, 1.84% and 1.71% when compared with RGB, HSV, YCbCr, YES and RIQ models respectively. This shows that the proposed face detection algorithm is efficient than the existing algorithms.

Figures 9 and 10 show the compression rate (%) and PSNR (dB) of the algorithms while using JPEG 2000 lossy algorithm (JL) for compressing non-ROI regions and JPEG LS (Lossless) algorithm (JLL) for compressing ROI region respectively. The ROI-based compression algorithm is termed as ROI_JL_JLL in this paper. In order to evaluate the effect of ROI separation algorithm during compression, the compression algorithm are applied to the whole of the video and compared with compression after using the proposed algorithm.

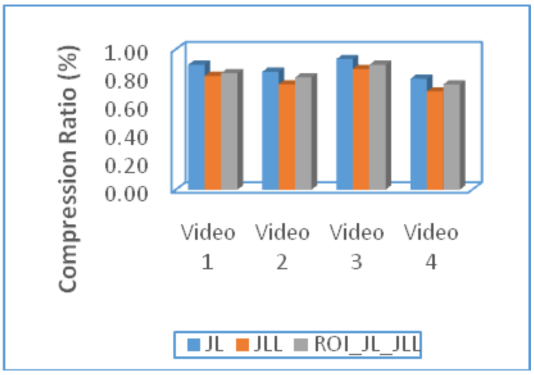

Figure 9 : Compression Ratio

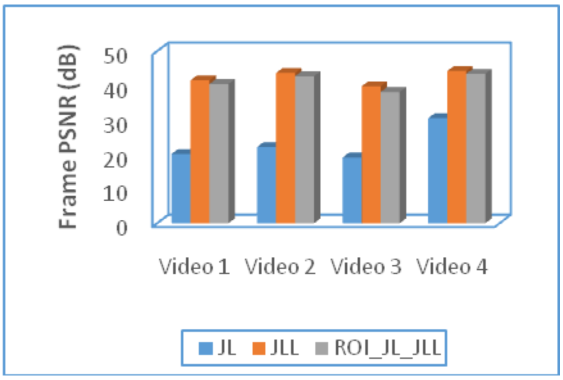

Figure 10 : PSNR

From the Figures, it is clear that the proposed ROI-JL-JLL algorithm can maintain an optimal balance between the preservation of required visual quality while achieving high compression ratio. This is evident from the PSNR value which is close to the JL algorithm, while the compression ratio is near to the JLL algorithm.

III. CONCLUSION

In this paper, a ROI-separation algorithm that can be used to ROI-based compression was proposed. A dynamic hybrid color space model was proposed for ROI-separation, which was used to identify face skin region and other regions in a video frame. The eye-mouth features were used to detect the face or ROI region. A lossy compression algorithm was then used to compress non-ROI region, while a lossless algorithm was used to compress the ROI-region.

Experimental results proved that the accuracy of ROI-detection was high when compared to algorithm that used conventional and existing color space models and also helped to improve compression ratio, while maintaining the visual quality of the video. In future, appropriate techniques to enhance the lossless and lossy compression algorithms will be probed.

REFERENCES

1. Abdelali, H.A., Essannouni, F., Essannouni, L. and Aboutajdine, D. (2016) An adaptive objective tracking using Kalman filter and probability product kernel, Modelling and Simulation in Engineering, Hindawi Publishing Corporation, Vol.2016, Article ID 2592368, Pp.1-8.
2. Ahmed, S., Osama, A. and El-Sayed, E. (2018) Comparative study of video compression techniques.Menoufia J. of Electronic Engineering Research, Vol. 27, No. 1, Pp. 1-33.
3. Antony, A. and Ganapathy, S. (2015) Highly efficient near lossless video compression using selective intra prediction for HEVC lossless mode, AEU - International Journal of Electronics and Communications, Vol. 69, Issue 11, Pp. 1650-1658.
4. Campadelli, P. and Lanzarotti, R. (2002) Localization of facial features and fiducial points, Proceedings of International Conference on Visualization, Imaging and Image Processing, Pp. 491–495.
5. Gowda, S.N. and Yuan, C. (2019) ColorNet: Investigating the importance of color spaces for image classification, Asian Conference on Computer Vision, Pp. 1-17.
6. Hsu, R.L., Mottalen, M.A. and Anil, K. J. (2002) Face detection in color images, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 24, Issue 5, Pp. 696-706.
7. https://en.wikipedia.org/wiki/Foreground_detection#Using_frame_differencing, Last Accessed During April, 2019.
8. https://en.wikipedia.org/wiki/Random_forest, Last Accessed During April, 2019.
9. https://media.xiph.org/video/derf, Last Accessed During April, 2019.
10. https://www.nist.gov/itl/iad/image-group/color-feret-database, Last Accessed During April, 2019.
11. Hussain, A.J. and Ahmed, Z. (2019) A survey on video compression fast block matching algorithms, Neurocomputing, Vol. 335, Pp. 215-237.
12. Jing, H., Yunyi, S. and Hai, C. (2019) Improved Viola-Jones face detection algorithm based on HoloLens, EURASIP Journal on Image and Video Processing, Article No. 41, Vol. 2019, Issue 1, Pp. 1-11.
13. Kung, S.Y. (2014) Kernel Methods and Machine Learning, Part II - Dimension-reduction: PCA/KPCA and feature selection, Cambridge University Press, New Jersey, PP. 77-78.
14. Manosij, G., Ritam, G., Ram, S. and Ajith, A. (2019) A wrapper-filter feature selection technique based on ant colony optimization, Neural Computing and Applications, Pp.1-19.
15. Oghaz, M.M., Maarof, M.A., Rohani, M.F. and Yaghoubyan, S.H. (2015) A hybrid color space for skin detection using genetic algorithm heuristic search and principal component analysis technique. PLoS One, Vol. 10, No. 8, Pp. 1-21.
16. Oghaz, M.M., Maarof, M.A., Rohani, M.F. and Zainal, A. (2017) A hybrid color space for skin recognition for real-time applications, Journal of Computational and Theoretical Nanoscience, Vol. 14, No. 3, Pp. 1-10.
17. Pansare, J.R. and Jadhav, K.R. (2018) Novel Technique for Lossless Compression of Video Using RCRC, International Journal of Innovative Research in Computer and Communication Engineering, Vol. 6, Issue 3, Pp. 2427-2435.
18. Pawłowski, P., Piniarski, K. and Dąbrowski, A. (2018) Selection and tests of lossless and lossy video codecs for advanced driver-assistance systems,Signal Processing: Algorithms, Architectures, Arrangements, and Applications, Pp. 344-349.
19. Wang, Z., Hou, Z., Hu, R.and Xiao, J. (2019) A Lossless Recompression Approach for Video Streaming Transmission,IEEE Access,Vol. 7, Pp. 35162-35172.
20. Wei, S.T., Shen, S.R., Liu, B.D. and Yang, J.F. (2009) Lossless image and video coding based on H.264/AVC intra predictions with simplified interpolations,16th IEEE International Conference on Image Processing, Pp. 633-636.
21. Xue, H., Zhang, Y. and Wei, Y. (2016) Fast ROI-based HEVC coding for surveillance videos,19th International Symposium on Wireless Personal Multimedia Communications, Pp. 299-304.