International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Impact Factor (2012): 3.358

Design & Implementation of Saliency Detection Model in H.264 Standard

Sheeba Shabnam[1], H.N. Latha[2]

Student, Department of Electronics & Communication, BMS College of Engineering, Bangalore, India

Assistant Professor, Dept. of Electronics & Communication, BMS College of Engineering, Bangalore, India

Abstract: ***Saliency detection is defined as a region of interest which is focused beyond the background. Nowadays saliency detection is widely used in different fields and have many scopes in image and video processing applications. Existing saliency detection techniques are implemented in uncompressed domain. In this paper we proposed an algorithm for visual saliency detection model for videos by using H.264 standard. Since H.264 is in compressed domain, the storage space is reduced and processing speed is high to deliver data to the end users. Therefore H.264 bit streams is required in various internet based multimedia applications. First the features of H.264 bit streams are extracted then saliency computation is done by calculating its static & motion saliency map. Before extracting the features macro blocks are generated and the frames of input video bit stream are coded in each macro block and it splits into unpredicted DCT coefficients and predicted motion vectors. The DCT coefficients are used to extract the features of luminance, color & texture of the unpredicted/current video frame to obtain the static saliency map whereas the motion saliency map of predicted frame is computed by motion features. Combining these features of static & motion saliency map, the final saliency map is extracted which focusing only on the particular region of interest which we want in the scene by drawing a rectangle on that region of interest (ROI) along with the irrelevant background but the ROI part is focused beyond the background. Thus the proposed model for videos can predict the salient part accurately and efficiently in compressed domain while separating it with unfocused region. The experimental result indicates that our proposed model outperforms from the other existing saliency detection models.***

Keywords: Saliency detection, H.264, Visual attention, Static Saliency map, Motion Saliency map

1. Introduction

Saliency detection models are widely used nowadays in different fields of applications to extract the region of interest (ROIs) in images/frames for various image/video processing applications such as in classification, watermarking, Trans coding, and resizing. Video saliency detection algorithms is used to calculate the motion saliency map since motion is an important factor to attract human beings attention. During the past ten years, several models have been proposed for salient region detection for images saliency and video saliency detections, but all the existing models are implemented in uncompressed domain.

The Saliency detection provides a mechanism for selection of particular aspects of a visual scene that are most relevant to our ongoing behaviour while eliminating interference from irrelevant visual data in the background. Perhaps, one of the earliest definitions of attention was provided by William James in his text book “Principles of Psychology” [14] over the last decades, visual attention (VA) has been studied intensely, and research has been conducted to understand the deployment mechanisms of visual attention. According to the current knowledge, the deployment of visual attention called as “visual saliency” that is, the characteristics of visual patterns or stimuli, such as a twinkling bright star in the sky that makes them stand out from its surroundings dim stars. Bright star is dominant and draw our attention in an automatic and rapid manner. Various computational models of visual attention have been developed based on this belief for different applications such as robotics, navigation, image and video processing, and so on. Such computational models of human visual attention are commonly referred to as visual saliency models, and their goal is to predict where people are likely to look in a visual scene. Humans routinely judge the importance of image regions, and focus attention on important parts.

Compared with the image saliency detection model, currently, several studies have tried to detect salient regions in video [5], [8], [9]. But the Existing saliency detection models are implemented in uncompressed (pixel) domain. However, most video over Internet are typically stored in the compressed domain such as MPEG2, H.264, and MPEG4 Visual. The compressed videos are widely used in various Internet-based multimedia applications since they can reduce the storage space and greatly increase the delivering speed for Internet users. In the uncompressed video such as MPEG2/H.262, the visual saliency is detected by decompressing the compressed video and features are extracted in spatial domain. The full decompression process for video is not only time consuming but computation consuming as well. Therefore, the video saliency detection algorithm in H.264 is much desired for various Internet-based multimedia applications. Visual saliency, being closely related to how we perceive and process visual stimuli, is investigated by multiple disciplines including cognitive psychology neurobiology and computer vision. Recently, the authors proposed a saliency detection model for image retargeting in compressed domain [10], [15]. The saliency map is calculated based on the features extracted from the discrete cosine transform (DCT) coefficients of JPEG images. However, this model is designed for JPEG images and this will not include the motion feature extraction.

The encoding standards for video and images are different. In this paper, we are using H.264 standard for saliency detection of Videos in compressed domain. In H.264 standard, The Y, C_r, C_b color space is used. Where Y represents the luminance component and (C_r and C_b) are used to represents the chroma components. In H.264 standard, the video frames are generated into 16 x 16 macro blocks are used for luminance and 8 × 8 macro-blocks for chroma components in 4:2:0 chroma sub-sampling. Each coded unit block consists of six 8 × 8 blocks: four 8 × 8 luminance channel blocks and two 8 × 8 chrominance blocks (one is the C_r channel block while the other is the C_b channel block). The data in video-bit stream are always processed by 8 × 8 blocks and, thus it performs here the saliency detection in 8 × 8 block level. The features of luminance, color, and texture are extracted from the DCT coefficients of the unpredicted frames (I frames) and these features are used for the static saliency map calculation for the unpredicted frames.

Whereas the motion feature is extracted from the motion vectors of the predicted frames (P frames) and this motion feature is used for the motion saliency map calculation of these predicted frames using the features of static saliency map and motion saliency map of video frames. The proposed model can predict the enhanced video frames efficiently in compressed domain which track the particular ROI part that we want. In addition to this extracting the motion saliency (ROI) features of two adjacent video frames blocks into a single frame is another advantage in H.264.The VOP (video object planes) are coded using macro-blocks. Usually, a VOP consist of one or many video packets/slices and each video packet is composed of integer number of consecutive macro blocks. The Experimental result show that the proposed model of saliency detection in H.264 is outperforms then the other existing ones. The resultant frame of enhanced saliency map of the two consecutive frame blocks into a single frame have some advantages such as in determining the saliency map location changes in each second and also become easy to compare the salient part in a single frame instead of using two frames and tracking the particular salient part that focused by drawing a rectangle on it.

Part I introduction described the features and importance of saliency detection and how it is processed in H.264 domain and the scope of the proposed model while Part II focuses on the research contribution which take place in the field of saliency detection is discussed .The rest of the paper is organized as follows. Proposed model shows how the features are extracted from H.264 bit streams followed by saliency computation, static and motion saliency map calculation and final map calculation are explained in section III. Design & implementation result is shown in section IV Concluding remarks are given in section V.

2. Related Work

There are several approaches to introduce saliency detection. There are two categories of approaches to automatically estimate saliency: bottom-up methods and top-down methods. A popular approach for computing bottom-up saliency was proposed by Itti et al. [3]. It is inspired by the human visual system and is based on low-level features: color, intensity, and orientation. The saliency map is obtained through the calculation of multi-scale center-surround differences by using these three features. A multi-resolution pyramid of the image is built, and significant changes in the features are searched for and combined into a single high-resolution map [2]. In the 1980s, Treisman *et al* [1] proposed the famous feature Integration Theory(FIT).According to this the early selective attention mechanism leads some image regions to be salient for their different features (including color, intensity, orientation, motion, so on) from their surroundings [1], [14]. In 1985, Koch *et al.* proposed a neuro-physiological model of visual attention in [13]. Later Hou et al [4] defined the concept of spectral residual to design a visual attention model. The spectral residual is computed based on the Fourier transform. Harel *et al.* [11] proposed a graph-based saliency detection model based on the study of visual attention for rapid scene. In this model, the saliency map is calculated based on two steps: forming activation maps on several features and the normalization of these feature maps. Achanta *et al.* [6] tried to obtain more frequency information to get a better saliency measure. The difference of Gaussian (DoG) is used to extract the frequency information in this model. Goferman *et al* [7] inform a context-aware saliency detection model by including more context information in the final saliency map. The center-surround differences of patches are used for saliency detection.

All above proposed model is based on image saliency detection. Some video saliency detection models have also been proposed Guo *et al* [5] proposed a phase-based saliency detection model for video of multi-resolution spatiotemporal saliency detection model and its application in image and video compression. This model obtains the saliency map through inverse Fourier transform on constant amplitude and the original phase spectrum of input video frames based on the following features: intensity, color, and motion. Then Itti [8] developed a model to detect the low-level surprising events in video; the surprising events are defined as the important information attracting human beings' attention in video Zhai *et al* [9] built a video saliency detection model by combining the spatial and temporal saliency maps. The color histograms of images are used for the spatial saliency detection, while the planar motion between images is adopted for the temporal saliency detection. In [12], the authors designed a dynamic visual attention model based on the rarity of features. The incremental coding length (ICL) is defined to measure the entropy gain of each feature for saliency calculation.

All these saliency detection models mentioned above are implemented in uncompressed domain. As to these saliency detection models, coded videos have to be decompressed into spatial domain to extract features for saliency detection. In this paper, we propose a saliency detection model in H.264 which is in compressed domain.

3. Proposed Model

In this section, we describe the proposed framework in details. The block-diagram of the proposed approach is depicted in Fig.1 which represents the static saliency map is obtained from the unpredicted frames and motion saliency is obtained from the predicted frames of H.264 input video bit stream. Initially the frames are generated from the given input video bit streams, then macro-blocks are generated, the macro-blocks are splits into unpredicted frames & predicted frames. For the given input video bit streamH.264, First we calculate the static saliency map by using the three features luminance, color and texture which is obtained from the DCT coefficient of the unpredicted frame (I frame). Meanwhile, the motion saliency map is calculated by using the motion features from the corresponding motion vectors of predicted frame (p frame) of the input video bit stream. By using this two saliency map of H.264, the enhanced final saliency map is obtained for each frame; the enhanced saliency can predict the salient part of each video frame efficiently in compressed domain. The proposed model computational complexity (time cost) is modest as compared to the other model.

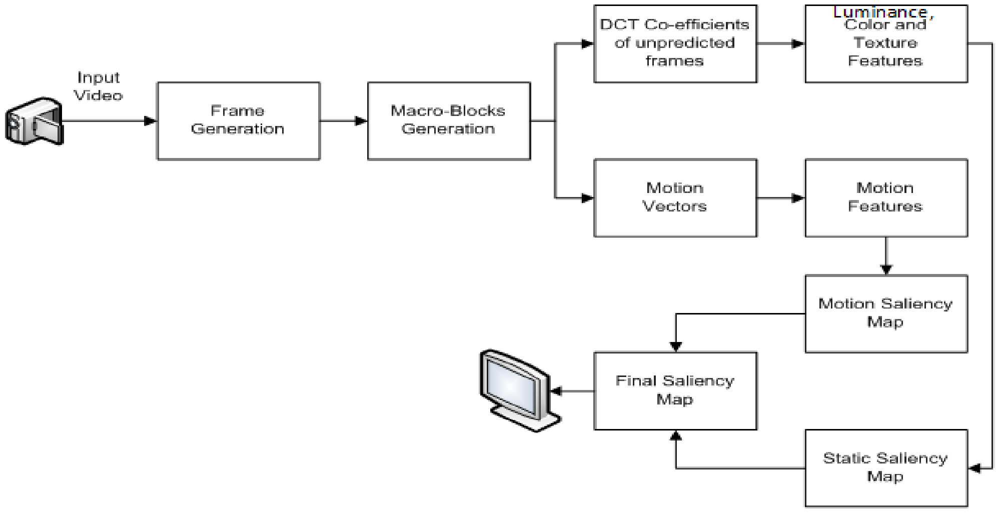

Figure 2.1: Block-diagram of the Proposed Model

A. Feature Extraction from H.264

The proposed model uses DCT coefficients of unpredicted frames (I frames) to obtain the static saliency map features and uses predicted frame (p frame) to obtain the motion saliency map features. In this section, we describe how the features are extracted and calculated from the DCT coefficient and motion vectors, which are used for extracting the static and motion saliency map. A natural video object is composed of a sequence of 2-D representations, which is referred to as video object planes (VOPs). The VOPs are coded using macro-blocks by exploiting both temporal redundancies and spatial redundancies. Usually, a VOP consists of one or several video packets (slices) and each video packet is composed of an integer number of consecutive macro blocks. In each macro-block, the motion vectors and DCT coefficients are coded.

The coded motion vector data are motion vector differences (predictively coded with respect to the neighboring motion vectors) after the variable length coding (VLC). The coded DCT coefficients are the 64 DCT coefficients encoded by zigzag scanning, run-length-encoded and the VLC. The differential motion vector can be extracted from the coded motion vector data based on the VLC table. Then it is added to a motion vector predictor component to form the real motion vector (MV) for predicted frames. In a similar way, VLC tables of DCT coefficients are used to decode the coded DCT coefficients. The fixed length decoding is used to obtain the real DCT coefficients for video frames. In H.264 video, DCT coefficients in one 8 × 8 block are composed of one DC coefficient and 63 AC coefficients. In each block, the DC coefficient is a measure of the average energy for the 8 × 8 block, while other 63 AC coefficients represent detailed frequency properties of this block. As mentioned above, YC_rC_b color space is used in H.264 video bit stream. In the YC_rC_b color space, the Y channel represents the luminance component, while C_r and C_b represent the Chroma components. Thus, the DC coefficients in DCT blocks from the Y, C_r, and C_b channels are used to represent one luminance feature and two color features for 8 × 8 blocks as follows:

$$L = DC_y \tag{1}$$
$$C_1 = DC_{C_r} \tag{2}$$
$$C_2 = DC_{C_b} \tag{3}$$

In (1) (2) (3) L, C_1 and C_2 represent one luminance and two Color features in each 8×8 DCT block respectively. DC_y, DC_{C_r} and DC_{C_b} are DC coefficients from the Y, C_r & C_b components in each DCT block, respectively.

The luminance blocks share two 8 × 8 chrominance blocks in the 4:2:0 chrominance format. The AC coefficients include the detailed frequency information and existing studies have shown that AC coefficients can represent texture information for image blocks. In YC_rC_b color space, C_r and C_b components mainly include color information and little texture information is included in these two channels. Thus, only the AC coefficients in Y component are used to represent the texture information for images. In one DCT block, most of the energy is included in the first several low-frequency coefficients, which are in the left-upper corner of the block. The AC coefficients in the right-bottom corner of DCT blocks are equal to or close to zero and they are neglected during the quantization in coding process. In the progressive coding, the AC coefficients in one DCT block are ordered by zigzag scanning. As the high-frequency AC coefficients include little energy for each DCT block, we use the first several AC coefficients to represent the texture feature of the DCT block. The existing study has shown that the first nine AC coefficients can represent most energy in each DCT block. Therefore, here, we use the first nine AC coefficients in each DCT block to represent the texture features follows:

$$T = AC_{01}, AC_{10}, AC_{20}, AC_{11}, \ldots\ldots, AC_{30} \tag{4}$$

The extracted motion vectors from the video bit stream are used to calculate the motion feature for predicted frames. In H.264 ASP video, there are two types of predicted frames: P frames use motion compensated prediction from a past reference frame, while B frames are bidirectional predictive-coded by using motion compensated prediction from a past and/or a future reference frame. As there is just one prediction direction (predicted from a past reference frame) for P frames, the original motion vector (MV) are used to represent the motion feature for P frames. Therefore we are using only p frame to extract the motion feature.

$$V=MV \tag{5}$$

While the original motion vector is used to represent the motion feature for each DCT block in P frames. Now, we can get luminance, color, and texture features for non-predicted frames as *L*, *C*1, *C*2, and *T* and the motion feature of predicted frames can be obtained as *V*. It will describe how to use these features in compressed domain to calculate the saliency map for video frames.

B. Saliency Computation

The Above description is based on the extraction of the features of unpredicted and predicted frames. Using these features, in this paper we calculate the enhanced saliency map for H.264 standard.

1) Static and Motion Saliency map Calculation:

The viewer always attracted to the region which have different feature from its surrounding. The features which can used to isolate the image region from its background include color, intensity and motion and so on. In this paper we use the features of luminance, color, texture and motion extracted from the DCT coefficients and motion vectors are used to calculate the DCT block differences for saliency detection. On the retina of human eye, the fovea has the highest density of cone photoreceptor cells and thus the focus area is perceived at the highest resolution. The visual acuity decreases with the increasing eccentricity from the fixation areas using a Gaussian model to simulate this mechanism for weighting the center-surround differences among image blocks for saliency detection. The feature map of each video frame is calculated as follows:

$$S_i^k = \sum_{j \neq i} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{d_{ij}^2}{2\sigma^2}} D_{ij}^k \tag{6}$$

In (6) S_i^k indicates saliency value of the i^{th} DCT block in the feature map with feature *k*, *k* $\in$\{*L*,*C*1,*C*2, *T*, *V*\}, σ is a parameter of the Gaussian distribution, $d_{ij}$is the Euclidean distance between DCT blocks *i* and *j*, $D_{ij}^k$is the feature differences between DCT blocks *i* and *j* with feature *k*.

The static saliency map S_s for unpredicted frames is calculated as linear combination of four feature maps from the luminance, color, and texture features (*L*,C_1,C_2&*T*) as follows:

$$S_s=\sum \beta_\theta N_\theta \tag{7}$$

In (7) *N* is normalization operation, $\theta \in \{S^k\}$, β_θ is the parameter determining the weight for each feature map. The motion feature *V* obtained from (5) is used to calculate the motion feature differences between DCT blocks. Here, the motion feature differences are computed as the Euclidean distance between motion features of DCT blocks. Therefore, the motion saliency map S_m can be obtained based on (6) for predicted frames. The final saliency map of video frames is calculated by combining the static saliency map S_s and motion saliency map S_m. We will describe how to compute the final saliency map as follows.

2) The Final Saliency map Calculation

Based on the above description, we can obtain the static saliency map for unpredicted video frames (I frames) and the motion saliency map for predicted video frames (P frames). The motion saliency map of unpredicted frames cannot be calculated since there are no motion vectors for unpredicted frames in video bit stream. However, there may be motion in unpredicted frames. Generally, the content in the consecutive video frames will not change greatly and thus the saliency maps (static or motion) of the consecutive video frames are very similar in video. Therefore, we can use the static or motion saliency map of the previous video frames to represent that of the current ones.

As there is no motion saliency map for unpredicted frames, the motion saliency map of the previous predicted frame is adopted to represent that of the current unpredicted frame. Thus, the final saliency map for unpredicted frames (I frames) is calculated as follows:

$$S = f(S_s, S_{mp}) \tag{8}$$

In (8) *S* is the final saliency map of the current unpredicted frame, S_s is the static saliency map of the current unpredicted frame, S_{mp} is the motion saliency map of the previous predicted frame, *f* (S_1,S_2) is the fusion function to get the final saliency map from the saliency maps of S_1 and S_2. Similarly, the final saliency map of the predicted frames (P frames) is computed as follows:

$$S = f(S_{sp}, S_m) \tag{9}$$

In (9) *S* is the final saliency map of the current predicted frame,S_{sp} is the static saliency map of the previous unpredicted frame, S_m is the motion saliency map of the current predicted frame. Using (8) and (9), we can calculate the final saliency map for video frames based on the static saliency and motion saliency maps.

The spatial variance is adopted to measure the feature contrast in the static and motion saliency map. Spatial variance is calculated as follows:

$$v_k = \frac{\sum_{(i,j)} \sqrt{(i-s_{i,k})^2 + (j-s_{(j,k)})^2} * s_k(i,j)}{\sum_{(i,j)} s_k(i,j)} \tag{10}$$

In (10) $v_k$is the spatial variance of the saliency map S_k, $S_k(i,j)$ is the saliency value at the location (i, j) in the saliency mapS_k, represents the spatial expectation values of salient region in horizontal and vertical directions respectively.

4. Proposed Model Design and Implementation Result

This section describes the design of the proposed model, implementation and performance evaluation.

A. Design of the proposed model:

Matlab is used for the implementation of this proposed model. The entire model can be divided into four parts:

- I-frame feature
- P-frame feature
- Frame generation details
- Main toolbox

1) I-frame features:

I-frame indicates the intra-predicted frames or current frame of the video. This frame is used to determine static saliency map by applying DCT coefficient on the YC_rC_b color space. The flowchart is shown below in Fig.4.1

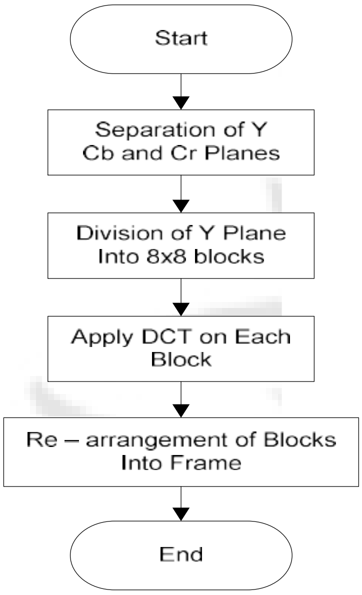

Figure 4.1: Flowchart of I-frame

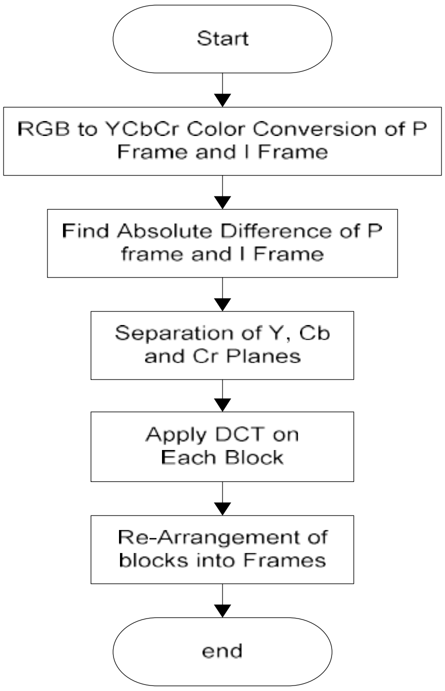

Figure 4.2: Flowchart of P-frame

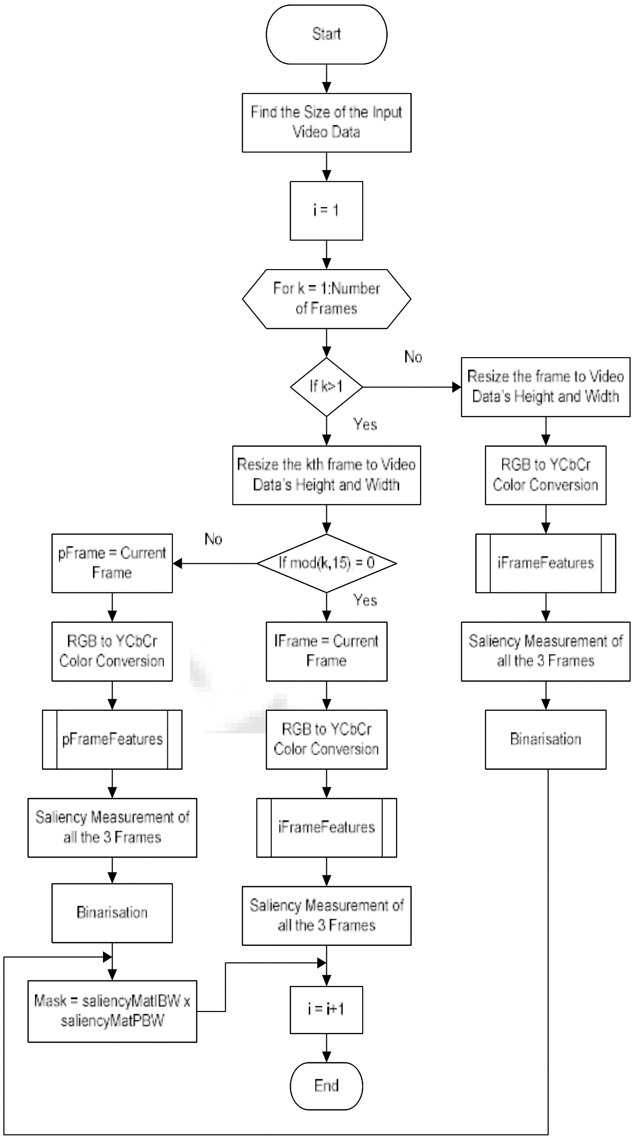

Figure 4.3: Flowchart of frame generation details

2) P-frame features:

P-frame is used for predicted frames to determine the motion saliency map of the input video. The flowchart of p-frame is given in Fig.4.2 that describes the RGB color conversion of input video into YC_rC_b of predicted and unpredicted frames. There is a motion in video in each second slight changes may occur. Absolute difference of adjacent frames is calculated. After the separation of YC_rC_b apply DCT on each blocks and rearrange into frames. In this way motion features are calculated for detection of motion saliency map.

3) Frame generation details:

The details of frame generation is shown in the flowchart Fig.4.3 that describe the whole process of frame generation conditions for feature extraction and saliency computation of I & P frame respectively.

4) Main toolbox:

The performance of the proposed model is implemented in the main toolbox, by giving input as any database video clips of H.264 bit stream in AVI format. The flowchart of the main toolbox is given below.

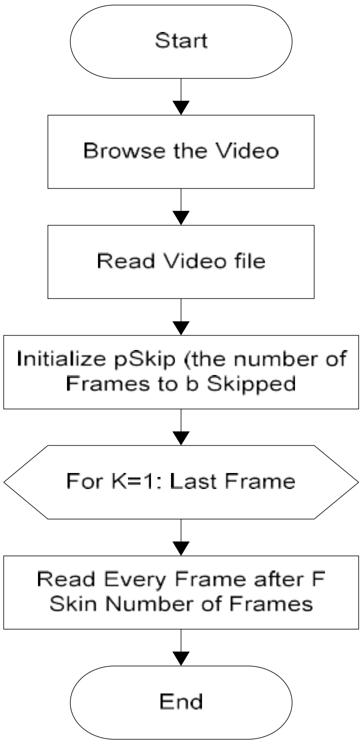

Fig.ure 4.4: Flowchart of Main toolbox

B. Result Implementation

The performance of the proposed model is done by using any public database of H.264 video bit streams. In this paper, we are using a database of 10 sec video clips of H.264 standard in AVI format. The performance of the video saliency detection model is determined by comparing the response values at saccade locations and random location in the saliency map. A high salient value at human saccade location indicates that saliency detection model can predict the location of saliency map exactly and efficiently.

To evaluate the performance of saliency detection models, Kullback-Leiber (KL) distance is used to measure the similarity between the distribution of human saccade location and random location as follows:

$$KL(H,R) = \frac{1}{2}\left(\sum_n h_n \log \frac{h_n}{r_n} + \sum_n r_n \log \frac{r_n}{h_n}\right) \quad (11)$$

In (11) H and R is saliency distributions at human saccade locations and random locations with probability density functions h_n and r_n respectively, n is the saliency value bins. A good video saliency is detected if high salient values occur at human saccade location (ROI) and low salient value at random locations (background).

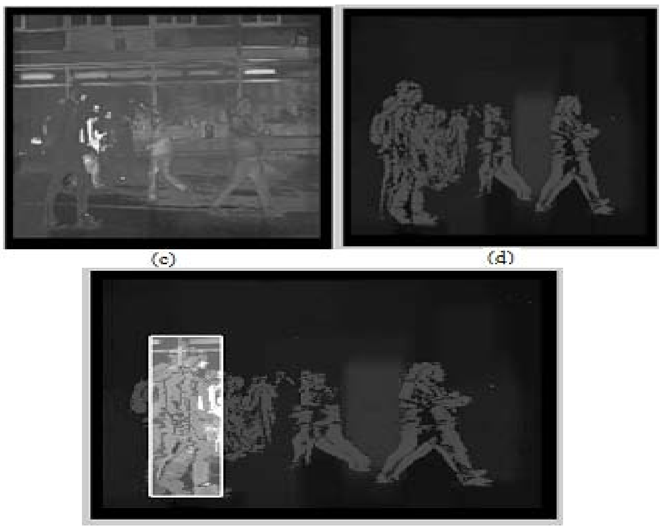

Figure 4.5: Result of Saliency Map (Video clips of people walking on the road here the current frame of a scene (a) and tracking one person of interest which we want in the scene by drawing rectangle on it (b) detecting the static saliency map of the frame in static (c) and motion saliency map (d), Using the static and motion saliency map of the frame the final saliency map of the frame is extracted which focused only on the tracking person (d)).

The result implementation of the proposed model is shown in Fig.4.5 which evaluate the performance of this saliency detection model by using any public database video clips of H.264 video stream in AVI format. The result in the Fig.4.5 depicts that the first frame is the original current video frame, The video clips of 10 frames/sec is taken in this experiment, that is k integer consecutive macro-blocks are generated (k=1,2,….10). The DCT coefficient and motion vectors are coded in each macro-block. From the DCT coefficients of non-predicted frames (I frame), the features of luminance(y frame), color (C_r & C_b frame) and texture features are extracted. The texture features are obtained from the luminance component is explained already in features extraction. From the coded motion vectors in each macro-block, the motion features are extracted from the predicted frames (p frames). The demonstrated result shows the result of static saliency map, motion saliency map, and final saliency map of a current video frame in compressed domain. The final saliency map is extracted by using the static and motion saliency map which focused only on particular person which we want in the scene by forming a mask on it.

CCTV cameras are used widely in many occasions, by using that video bit stream of H.264 we can extract that salient part (ROI) of the frame easily. Extracting the saliency map of each frame has many applications. Some of the applications are mentioned below as follows:

1) Automated video surveillance: In these applications computer vision system is designed to monitor the movements in an area, identify the moving objects and report any doubtful situation. The system needs to discriminate between natural entities and humans, which require a good object tracking system. Extracting the saliency map of the frame is benefited to catch that particular salient region and enhance it.

2) Robot vision: In robot navigation, the steering system needs to identify different obstacles in the path to avoid collision. If the obstacles themselves are other moving objects then it calls a real-time object tracking system and saliency detection technique will be useful.
3) Traffic monitoring: In some countries highway traffic is continuously monitored using cameras. Any vehicle that breaks the traffic rules or is involved in other illegal act can be tracked down easily by detecting and extracting its saliency map.

Therefore saliency detection techniques are usable and have many applications in image/video processing. In this paper, we extracted the final saliency map to track the particular region efficiently, by using the static saliency map of the unpredicted (current) frame and motion saliency map of the predicted frame, the static saliency map is obtained by taking the features of DCT coefficient and motion saliency map is obtained by motion vectors. Using the static and motion saliency map, the final saliency map is extracted which have the good feature contrast of each video frame of H.264 which is focused beyond the background. Therefore the result implemented in compressed domain which outperforms from the other existing models. The ROC curve of our proposed model is shown in Fig.4.6

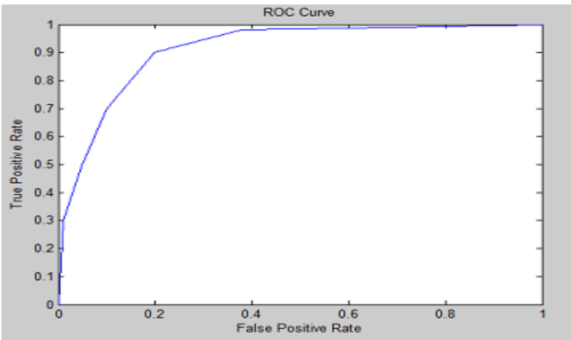

Figure 4.6: ROC curve of the proposed model for video saliency detection

5. Conclusion

In this paper, we are using H.264 standard to detect the visual saliency in compressed domain. The static saliency map is obtained by using the luminance, color and texture feature. While the motion features are extracted from the motion vectors for calculating its motion saliency map. Combining the static and motion saliency map, the final saliency map is obtained which is focused only on the particular salient part which we want in the scene by forming a mask on that salient part along with the unfocused background. The proposed model efficiently extracted the saliency map in compressed domain, since we are using H.264 video bit stream which can more conveniently used in various internet based multimedia applications and in research field. The earlier saliency detection models are based on uncompressed domain. Compared with existing model, our implemented result is significant and has vast scope

6. Acknowledgement

The Authors would like to thank VGST (Vision Group on Science and Technology), Government of Karnataka, India for providing infrastructure facilities to do this project.

References

[1] Treisman and G. Gelade, "A feature-integration theory of attention" *Cognitive Psychology*, vol. 12, no. 1, pp. 97–136, 1980.
[2] J. Wolfe, "Guided search 2.0: A revised model of guided search,"*Psychonomic Bull. Rev.*, vol. 1, no. 2, pp. 202–238, 1994.
[3] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Mach.Intell*, vol. 20, no. 11, pp. 1254–1259, Nov. 1998.
[4] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," in *Proc. IEEE Int. Conf. Comput. Vision Pattern Recognit*, pp. 1–8, Jun. 2007.
[5] Guo and L. Zhang, "A novel multi-resolution spatiotemporal saliency detection model and its applications in image and video compression," *IEEE Trans. Image Process.*, vol. 19, no. 1, pp. 185–198, Jan. 2010.
[6] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," in *Proc. IEEE Int. Conf. Comput. Vision PatternRecogni*, pp. 1597–1604, Jun. 2009.
[7] S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," in *Proc. IEEE Int. Conf. Comput. Vision Pattern Recognit*, pp. 2376–2383, Jun. 2010.
[8] L. Itti and P. Baldi, "Bayesian surprise attracts human attention," *Adv. Neural Inform. Process Syst*, vol. 46, nos. 8–9, pp. 1194–1209, Apr.2006.
[9] Y. Zhai and M. Shah, "Visual attention detection in video sequences using spatiotemporal cues," in *Proc. ACM Int. Conf. Multimedia*, 2006, pp. 815–824.
[10] Y. Fang, Z. Chen, W. Lin, and C. Lin, "Saliency-based image retargeting in the compressed domain," in *Proc. ACM MM*, pp. 1049–1052, 2011.
[11] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," *Adv. Neural Inform. Process. Syst.*, vol. 19, pp. 545–552, MIT Press, 2007.
[12] X. Hou and L. Zhang, "Dynamic visual attention: Searching for coding length increments," *Adv. Neural Inform. Process. Syst.*, vol. 21, pp. 681–688, MIT Press, 2008.
[13] Koch and S. Ullman, "Shifts in selective visual attention: Towards the underlying neural circuitry," *Human Neurobiology*, vol. 4, no. 4, pp. 219–227, 1985.
[14] Treisman and S. Gormican, "Feature analysis in early vision: Evidence from search asymmetries," *Psychology Rev.*, vol. 95, no. 1, pp. 15–48, 1988.
[15] Y. Fang, Z. Chen, W. Lin, and C.-W. Lin, "Saliency detection in the compressed domain for adaptive image retargeting," *IEEE Trans. Image Process.*, vol. 21, no. 9, pp. 3888–3901, Sep. 2012.

Author Profile

Sheeba Shabnum, Persuing M-tech (2012-2014) E&C Department, from BMS College of Engineering Bangalore Karnataka, India

H. N. Latha is working as Assistant Professor, Department of E&C BMS College of Engineering, Bangalore, Karnataka, India