ISSN:2320-0790

An International Journal of Advanced Computer Technology

Human Vision System's Region of Interest Based Video Coding

K. Asha[1], D. Anuradha[2], V.Saravanan[3], M.Rizvana[4]

[1,2]UG Scholar, IT Department, P.S.V. College of Engineering and Technology, Krishnagiri, India.
[3,4]Assistant Professor, IT Department, P.S.V. College of Engineering and Technology, Krishnagiri, India.
[1]kesavaluasha@gmail.com, [2]anurocks4ever@gmail.com, [3]v_saravanan18@yahoo.co.in, [4]rizvanait@gmail.com

Abstract: While watching a video human visual system gives more attention on the foreground objects than background objects. That is to say, human vision system pays more attention to region of interest, such as the human faces in the video content. Most of the video encoders compress video by considering every part of the video frames with equal importance. So the video size could not be reduced to maintain quality. The proposed system can detect the foreground and it can allocate different bit rates for different regions. By doing this the video quality can be maintained and the size can be reduced up to 40%.

Keywords: Video Compression, Human Vision System, Background, Foreground, Spatial, Temporal.

I. INTRODUCTION

Video compression algorithms may eliminate the redundancy in the video file to reduce the size [2, 5]. Most of the video codec uses audio compression to reduce the file size further. The video compression can be classified in to two types, they are, lossy and lossless compression. In lossy compression the file size can be reduced without any limitation and the original data cannot be reconstructed. In lossless compression the file size cannot be reduced beyond certain level and the original data can be reconstructed. The lossy video compression algorithms can reduce the frame rate of the video and/or it can reduce the bit rate of frame. In the lossless compression, the size of the video files can be reduced by identifying and eliminating the redundant data.

The video files should not be compressed beyond certain level, since the video clarity may be reduced abnormally. Every video is made up of frames and each frame is made up of pixels. Each pixel can be represented using definite number of bits in the computer memory. The video which contains more number of frames and more number of pixels may have high file size and video quality. The video with lesser number of frames and pixels may decrease the file size and quality of the video. That is, the frame rate and bit rate is directly proportional to the video quality.

Most of the video compression algorithms in the literature operate on neighboring pixels [9], called macroblocks. These macroblocks can be formed by dividing the entire video frames in to number of rows and columns. These macroblocks can be compared with the blocks of previous and/or next frames.

The sequence of frames may contain both spatial and temporal redundancy. These redundancies can be reduced by the video compression algorithms to make it in smaller size. Most video compression algorithms and codec uses both spatial image compression and temporal motion compensation techniques together [1].

Video compression techniques can be grouped in to two categories; they are interframe compression intraframe compression. In intraframe compression the system can consider the current frame only. It can reduce the file size by downgrading the bitrate of the pixel or number of pixel in each frame. Intraframe is suitable for devices with limited resources. Editing the video with intraframe compression is easy comparing the interframe compression. In interframe compression, more than one frame can be considered for compression. The previous frame and the next frames can be compared to identify the redundancy in the video data. If the current frame and the previous frame contain no change in a specific macroblock, then the system can create a command to copy that part from the previous frame instead of storing the same once again in the current frame. If more than one frame have small changes comparing each other, then the system simply create a command which instructs the decompressor to rotate, lighten, darken or shift the copy (These commands are shorter than the data spend by intraframe compression) [9]. The interframe compression is good for playback, but not suitable for editing work. Because, one frame may depends on other frame for getting the complete shape. If the original frame is missing, then construction of other frames may fail. Also, lot of buffers has to be maintained for smooth playback and editing [9].

Video format like MPEG2 uses interframe compression that contains a special frame called "I frame", which may not depend on other frame and requires more space than nearby frames. Increasing the number of I frames may increase the

file size and reducing the number of I frames may result in processing complexity. In the digital video, achieving the perceptual quality in the given bit rate is a challenging task. However, from human vision system point of view, a hard-to-predict area cannot catch as much attention as easily predictable area. In order to achieve constant visual quality across different part of the video psychophysical model must be considered during the bit allocation process. For the video codec design, the performance can be measured by peak signal-to-noise ratio. However the extra bits an encoder spent to increase peak signal-to-noise ratio does not cause an increase in visual quality. It is well known that the perceptual quality of visual sequence may not reflected by peak signal-to-noise ratio.

In this paper, we are focusing more on video rate control and foreground and background analysis. The capability for human vision to detect alteration in video sequences must be considered for proper bit allocation. The bit rate saving can be achieved by identifying the regions with various distortion level and by allocating low bit in the unwanted regions.

This paper is organized as follows. In section II, the foreground objects are identified by the motion attention index. The variation ranges can be calculated in section III, The proposed bit allocation scheme is explained in section IV. Finally a conclusion is given in section V.

II. Motion attention Index

Normally human attention on an object is directed by brain to complete a task. But in some situation the object can attract the humans and make them to listen, statically or dynamically. The static attraction is suitable for images and the dynamic attraction is more suitable for videos [3, 4, 6]. So we can use the dynamic model discussed in [1]. This is a simple model (many other models are also available [7, 8]) which is used to detect the moving object in a video, with considering the overall motion. This model uses three different values namely intensity inductor (Used to detect moving object in a video without considering global motion), spatial coherence inductor and temporal coherence inductor (Considers camera motion). The intensity inductor value (IIV) for a macro block (i,j) of n^{th} frame is calculated as

$$IIV_{nij} = \frac{\sqrt{motvecx^{2}_{nij} + motvecy^{2}_{nij}}}{max\, l_n}$$

Where maxIn is the maximum motion vector intensity in n^{th} frame, $motvecx^2_{nij}$ and $motvecy^2_{nij}$ are the motion vectors in the n^{th} frame. This IIV_{nij} is not sufficient because the camera movement can cause large intensities that cannot be calculated by this. To avoid such negative effect we must go for spatial inductor and temporal coherence inductor. The spatial coherence index value (SCIV) can be calculated for a macroblock (i,j) of n^{th} frame as

$$SCIV_{nij} = -\sum_{b=1}^{n_s} pd_n(b) Log(pd_n(b))$$

Where $pd_n(b)$ is a probability function, n_s is the number of histogram bin. Similarly the temporal coherence inductor value (TCIV) can be calculated by

$$TCIV_{nij} = -\sum_{b=1}^{n_s} pd_n(b) Log(pd_n(b))$$

Where $pd_n(b)$ is a probability distribution function and n_t is the number of bin for motion direction histogram. Now we have both the values which is required to compute the motion attention index value (MAIV) for the macro block (i,j). The MAIV values are used for detecting foreground and background in a video.

$MAIV_{nij} = IIV_{nij}$ X $TCIV_{nij}$ X (1 - IIV_{nij} X $SCIV_{nij}$)

The $MAIV_{nij}$ can be in between 0 and 1.

III. Regions With Variations

In section II, The detection of foreground and background in a video has done. Based on this detection we can allocate more bits for the macro blocks in foreground than background. By doing this lot of Bits can be saved. But, still more bits can be saved by adding our proposed idea with this.The proposed idea is, the blocks in the frames of the video can be examined. Each block can be compared with neighbor pixels for the variations. If the variation is too high, then the block can be marked to allocate more bits. Similarly if the variation is too low then the block can be marked to allocate fewer bits. This variation can be calculated in such a way that the maximum value should be 1 and minimum value should be 0. The pixels in the macroblock (i,j) can be compared with each other as follows to calculate the variations.

$VR_{nij} = VR1_{nij} + VR2_{nij}$

Where VR_{nij} is the variation range value, $VR1_{nij}$ and $VR2_{nij}$ can be calculated using the following code:

```
for(int i=1;i<maxy;i++)
{
    for(int j=1;j<maxx-1;j++)
    {
        for(int k=j+1;k<maxx;k++)
        {
            VR1nij+=Pnij(j,i)-Pnij(k,i)
        }
    }
}
for(int i=1;i<maxy;i++)
{
    for(int j=1;j<maxx-1;j++)
    {
        for(int k=j+1;k<maxx;k++)
        {
            VR2nij+=Pnij(i,j)-Pnij(i,k)
        }
    }
}
```

Fig. 1 Code for Calculating VR1 and VR2

Where maxx is the maximum number of pixels in x-axis and maxy is the maximum number of pixels in y-axis of a macroblock respectively. $P_{nij}(x,y)$ is the pixel value at location (x,y). Using the VR_{nij} value the smoothness or randomness of the macroblock can be detected. For smooth macroblock the VR_{nij} value can be less and for highly variation macroblock the VR_{nij} value can be large. This range of values can be converted to VR'_{nij} , the range between 0 and 1.

IV. Proposed Scheme

To develop a good video coder the ability of human eyes to detect coding distortion should be considered. The basic idea of the proposed algorithm is to allocate more number of bits for foreground objects and to allocate few bits for background objects. The video can be split into number macro blocks and then we have to find whether the macro block contains foreground object or a background object by motion attention index and variation range. The motion attention index can be calculated by section II and variation range can be calculated using section III. The human vision system's region of interest (HROI) value can be calculated by using both motion attention index and variation range (VR) values. Then based on this value the coder can take a decision how much bit to allocate for a particular block. The HROI value can be calculated using the below algorithm.

To develop a good video coder the ability of human eyes to detect coding distortion should be considered. The basic idea of the proposed algorithm is to allocate more number of bits for foreground objects and to allocate few bits for background objects. The video can be split into number macro blocks and then we have to find whether the macro block contains foreground object or a background object by motion attention index and variation range. The motion attention index can be calculated by section II and variation range can be calculated using section III. The human vision system's region of interest (HROI) value can be calculated by using both motion attention index and variation range (VR) values. Then based on this value the coder can take a decision how much bit to allocate for a particular block. The HROI value can be calculated using the below algorithm.

Step1. Calculate the $MAIV_{nij}$ for the specific block.
Step2. If the $MAIV_{nij}$ is greater than threshold then the maximum value can be set to HROI.
Step3. Else calculate VR'_{nij} for the specific block.
Step4. If VR'_{nij} is greater than certain level (Threshold) then assign VR'_{nij} value to the HROI.
Step5. If both $MAIV_{nij}$ and VR'_{nij} are not greater than expected threshold value then assign a fixed value to HROI.
Step6. Repeat the above steps for all the blocks.

Fig. 2 Algorithm for Calculating HROI.

For each macro block the HROI can be calculated using the above algorithm. These values of HROI can be given as an input for encoding algorithm so that it can decide the amount of bits spend for a particular macro block. The decided bit rate always less than original bit rate or it can be equal, never greater than the original bit rate. Hence the result of the encoding can reduce the size of the video file.

V. Conclusion

In this paper we have proposed a VDSI based bit allocation scheme. This scheme makes the coder to allocate more bits for foreground where the viewers can easily identify the distortion and few bits for background where the viewers cannot easily identify the distortion. Thus the video size is reduced without affecting the perceptual quality. In future the face detection algorithms can be added along with this to allocate more bits where the human faces are available in frames. This can further improve the quality of the video.

Acknowledgment

Our sincere thanks to our honorable Chairman **Dr.P.Selvam M.A., B.Ed., M.Phil., Ph.D., D.Litt.,** P.S.V. College of Engineering and Technology, Krishnagiri, for giving this opportunity. We express my profound gratefulness to **Dr.K.Rangasamy M.E., M.B.A., Ph.D.,** Principal, P.S.V. College of Engineering and Technology for his continuous encouragement in publishing research articles.

References

[1]. Y.-F.Ma and H.-J. Zhang, "A model of motion attention for video skimming," in Proc. ICIP, vol. 1, Sept. 2002, pp. I-129–I-132.

[2]. Y. Takahashi, N. Nitta, and N. Babaguchi, "Video summarization for large sports video archives," in Proceedings of the IEEE International Conference on Multimedia and Expo (ICME '05), pp. 1170–1173, Amsterdam, The Netherlands, July 2005.

[3]. Y.-F. Ma, L. Lu, H.-J. Zhang, and M. Li, "A user attention model for video summarization," in Proceedings of the 10th ACM International Multimedia Conference and Exhibition, pp. 533–542, Juan Les Pins, France, December 2002.

[4]. H.J Zhang, et al, "An integrated system for content-based video retrieval and browsing," Pattern Recognition, vol.30, no.4, pp.643-658, 1997.

[5]. C. Kim and J. N. Hwang, "An integrated scheme for objectbased video abstraction," Proc. Of ACM Multimedia 2000. Los Angeles, CA, 2000.

[6]. S. Z. Li, et al., "Statistical Learning of Multi-View Face Detection," Proc. of ECCV 2002.

[7]. Y. Tian, Max Lu, and A. Hampapur, "Robust and Efficient Foreground Analysis for Realtime Video Surveillance," IEEE CVPR, San Diego. June, 2005.

[8]. S. Cheung and C. Kamath, "Robust background subtraction with foreground validation for urban traffic video", EURASIP Journal of Applied Signal Processing, Special Issue on Advances in Intelligent Vision Systems, 2005.

[9]. V.Saravanan, Dr.A.Sumathi, S.Shanthana and M.Rizvana, "Dual Mode Mpeg Steganography Scheme For Mobile and Fixed Devices", International Journal of Engineering Research and Development, Volume 6, Issue 3 PP. 23-27, Mar 2013.