Spatial and Temporal Complexity within Extracted Features of H.264 Bit Stream Information

Amitesh Kumar Singam, Muhammad Shahid

Abstract

These days, Need of Application developers towards developing back end visual productivity applications gave rise to specific recommendations while maintaining Standards forced towards huge competition between quality of service and User Experience. Moreover, our interest lies in quantifying the Spatial (Observations) and Temporal (observers) Information based on decisions not consistency by Quality of Experience (experts based). Finally, we intend to validate our data acquired through subjective experiments by formulating those effective features into one metric based on human visualizing characteristics. This research concludes with the trade offs between quality prediction and complexity of Visual content within h.264 bit stream information extracted at macro block layer are expected to have strong spatial and temporal complexity with the subjective scores and We concluded that our newly developed algorithm which is used to calculate spatial and temporal distortions for successive frames or Group of Pictures at frame by frame level has achieved expected outcome towards benefiting research community.

Spatial and Temporal Complexity, Video Quality Assessment

INTRODUCTION

Subjective analysis is conducted based on human perception since it is concerned with how video is perceived by a viewer or subject and expresses subjects opinion on a particular video sequence in comparison with its original video sequence. The subject has to vote for the video sequences under certain test environment conditions for example the ITU-Recommendations [1]. Human perception is considered as the true judgment of video quality and precise measurement of video quality but it is quite expensive and tedious in terms of time such as preparation, running and human resources. Objective quality assessment is therefore essential.

Objective Video Quality Metric should be designed based on HVS (Human Visualizing System) characteristics. Some aspects of HVS like contrast, orientation sensitivity, spatial and temporal masking effects, frequency selectivity and color perception are incorporated in the design of objective quality metrics. Even though it is computationally very expensive and complex to design a quality metric with above aspects. It is useful for a wide range of applications if it correlates well with human perception.

The impairments visibility which is related to video processing system is subjected to spatial and temporal properties of video content, since subjective analysis is quite expensive and time conservative method, objective metric has been developed considering HVS. In our thesis work we have done experiment on both subjective and objective analysis.

SPATIAL AND TEMPORAL INFORMATION

Even though this research is assumed as an Hypothesis work, i.e, in order to combine audio with Visual Characteristics together, we initially considered the typical 3 co-ordinates plane within two-dimensional projected space, in technical terms it should be Chrominance, Luminance and colour domain and to quantify spatial artifacts and temporal impairments, we deployed some mathematical operations for spatial information within chrominance and temporal information within luminance, finally colour domain is projected out of 2 dimensional space by default which means 3rd coordinate is not colour domain within two dimensional space.

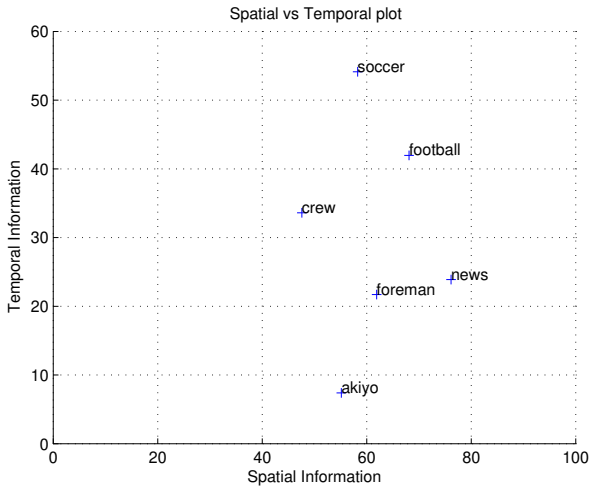

Figure 1: Spatial and Temporal Information

Spatial and Temporal Complexity

The spatial-temporal information is essential for some unspecified reasons and due to quality of transmitted video sequence is highly dependent on this whereabouts. The formulations for quantifying spatial and temporal information for test sequences are shown in following sections [4].

Implementation

$Moodel_{Metric}$ is newly developed algorithm which is used to calculate spatial and temporal distortions for successive frames or Group of Pictures at frame by frame level or macro block level which is expressed as

$$Moodel_{Metric} = max_{time}[std_{GoP}[sobel^*(f_n)]] \quad (1)$$

- $sobel^*(f_n)$ is used for quantifying distortions by self defined function for n frames at time(t).
- std_{GoP} is standard deviation over pixels between successive frames.
- max_{time} is maximum value within frequency domain.

$Sobel^*$**function** is calculated for amount of change in Spatial artifacts and temporal impairments within a video sequence and it is computed as

$$sobel*(f_n) = max_{time}[std_{GoP}[M_n(i,j,k)] \quad (2)$$

M(i,j,k) is changes in motion information between pixel values for sequential frames. M_n(i, j,k) as a function of kth frame and it is expressed as

$$M_n(i,j,k) = f_n(i,j,k) - f *_n (i,j,k)$$

f_n(i,j,k) is pixel at i^{th} row and j^{th} column for k^{th} frame in a video sequence and $f*_n$(i, j,k) is pixel at i^{th} row and j^{th} column of a kth frame. The higher motion content in successive frames of a video sequence leads to higher Spatial content is a misconception and more over dependency of spatial and temporal distortions are based on decisions not consistency of a subject in our case. below plots illustrates observation of observers based on decisions.

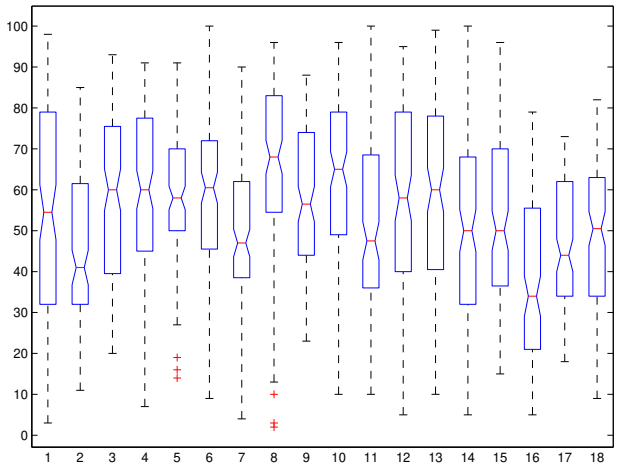

Figure 2: Illustrates observation of observers based on decisions

SCIENTIFIC ANALYSIS OF VISUAL INFORMATION BASED ON DATA SCREENING METHOD

An assumption of consistency(decisions) of observations by observers within two scenarios et al[3] for conditional decisions judged by decision making tree and moreover it exploits inconsistency within spatial(observations) and temporal (observers)information for second decision due to impact of I frame Effect during video encoding process and In Group Of Pictures, intra coded frame contains high quality but gradually reduced for following sequence of P-frames and the occurrence of I frame effect will have huge influence on subjective test which is the main cause for up sampling or feature mapping by a bi-linear interpolation algorithm in FFMPEG[2].

Amitesh et al.[5] mentioned about Scope of subjective quality assessment i.e, this kind of assessment is essential towards considering subjective scores as independent variables and also hypothetically validated those independent variables which are limited to few concepts, because after investigating Human Visual Characteristics author concluded that selectively subjective scores are considered as true values judged by humans.

FRAME LEVEL QUALITY EVALUATION

For instance, its not per video, we evaluated our proposed model at frame level, Moreover, these features were extracted out of H.264 bit stream data and also These features are expected to have high correlation with corresponding perceptual quality score of the selected video, mainly feature extraction of 120 videos has been processed using JM Reference software Version 16.0. The properties of encoded videos are acquired from bit stream data of H.264/AVC which has been generated as a trace file while encoding process. Rather than using completely decoded frame, our interest lies in reversing the entropy encoding of bit stream. By analyzing three successive Nal, Slice and Macro Block Layer[6].

EVALUATION OF PROPOSED MODEL

We considered distribution of test configuration towards illustration between Mean Opinion Scores and Test conditions for selected video sequences. Finally due to missing motion vectors, coding information can be quantified by error concealment i.e, only in case of default encoding settings[7]and in other case if rate distortion setting are not default and motion vector feature cannot be used instead spatial distortions should be quantified based on motion intensity by metric called Motion Dynamics[3].

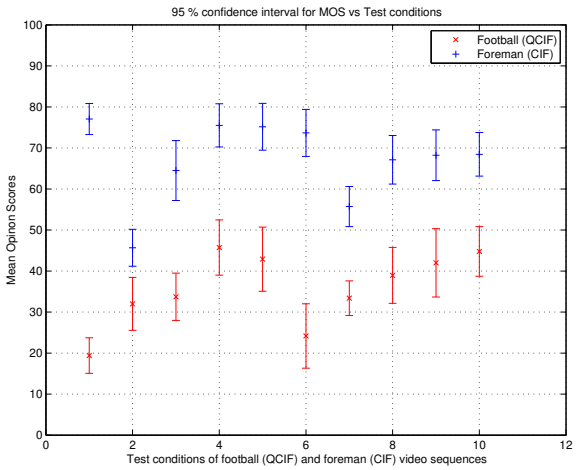

Figure 3: illustration between Mean Opinion Scores and Test conditions for selected video sequences

CONCLUSION

We conclude that Spatial and Temporal complexity quantifies distortions by comparing 95 % confidence interval of individual subject based on decisions not consistency.

ACKNOWLEDGEMENTS

The author wishes to thank User Experience community

REFERENCES

[1] ITU-R radio communication sector of itu, recommendation itu-r bt.500-12, 2009. http://www.itu.int/.

[2] FFMPEG documentation, 2011. http://ffmpeg.org/ffmpeg.html.

[3] Pashike, V. R. R., SINGAM, A. K., and Shahid, M. Human perceptions based on translations of recurrent neural networks principles for low latency applications. *Preprints* (August 2023).

[4] Shahid, M., Singam, A. K., Rossholm, A., and Lövström, B. Subjective quality assessment of h.264/avc encoded low resolution videos. In *2012 5th International Congress on Image and Signal Processing* (2012), pp. 63–67.

[5] singam, A. Computational Complexity of Human Visualization Characteristics based on Well Known Metric.

[6] Singam, A. K., and Pashike, V. R. R. Master thesis electrical engineering low complex blind video quality predictor based on support vector machines.

[7] Singam, A. K., Pashike, V. R. R., and Shahid, M. A. An error concealment based on translations of generalized neural networks. *SSRN Electronic Journal* (2023).