Approaches to Video Quality Verification

Submitted By

Smit Bharada

15MCEC03

DEPARTMENT OF COMPUTER ENGINEERING

INSTITUTE OF TECHNOLOGY

NIRMA UNIVERSITY

AHMEDABAD-382481

May 2017

Approaches to Video Quality Verification

Major Project

Submitted in partial fulfillment of the requirements

for the degree of

Master of Technology in Computer Science and Engineering

Submitted By

Smit Bharada

(15MCEC03)

Guided By

Prof. Malaram Kumhar
Nirma University, Ahmedabad.

Mr. Manoj Gerald
Arris India Pvt. Ltd.

DEPARTMENT OF COMPUTER ENGINEERING

INSTITUTE OF TECHNOLOGY

NIRMA UNIVERSITY

AHMEDABAD-382481

May 2017

Certificate

This is to certify that the major project entitled **"Approaches to Video Quality Verification"** submitted by **Smit Bharada (Roll No: 15MCEC03)**, towards the partial fulfillment of the requirements for the award of degree of Master of Technology in Computer Engineering of Nirma University, Ahmedabad, is the record of work carried out by him under my supervision and guidance. In my opinion, the submitted work has reached a level required for being accepted for examination. The results embodied in this major project, to the best of my knowledge, haven't been submitted to any other university or institution for award of any degree or diploma.

Prof. Malaram Kumhar
Guide & Assistant Professor,
Information Technology Department,
Institute of Technology,
Nirma University, Ahmedabad.

Dr. Priyanka Sharma
PG Coordinator - CSE,
Associate Professor
Institute of Technology,
Nirma University, Ahmedabad

Dr. Sanjay Garg
Professor and Head,
Computer Engineering Department,
Institute of Technology,
Nirma University, Ahmedabad.

Dr Alka Mahajan
Director,
Institute of Technology,
Nirma University, Ahmedabad

Certificate

This is to certify that **Smit Bharada(Roll No: 15MCEC03)**, a student of M.Tech Computer Science and Engineering (CSE), Institute of Technology, Nirma University, Ahmedabad was working in this organization since 01/06/2016 and carried out his thesis work titled **"Approaches to Video Quality Verification"**. He was working in name of Software Intern under supervision of Mrs. Nethravathi Reddy (Mentor), and Mr. Manoj Gerald (Manager). He has successfully completed the assigned work and is allowed to submit his dissertation report. The results embodied in this project, to the best of our knowledge, haven't been submitted to any other university or institution for award of any degree or diploma. We wish him all the success in future.

Mrs. Nethravathi Reddy
External Guide & Project Manager,
Arris India Pvt. Ltd,
Bangalore

Mr. Gerald Manoj
Senior Manager,
Arris India Pvt. Ltd,
Bangalore

Statement of Originality

I, **Smit Bharada**, Roll. No. **15MCEC03**, give undertaking that the Major Project entitled "**Approaches to Video Quality Verification**" submitted by me, towards the partial fulfillment of the requirements for the degree of Master of Technology in **Computer Engineering** of Institute of Technology, Nirma University, Ahmedabad, contains no material that has been awarded for any degree or diploma in any university or school in any territory to the best of my knowledge. It is the original work carried out by me and I give assurance that no attempt of plagiarism has been made.It contains no material that is previously published or written, except where reference has been made. I understand that in the event of any similarity found subsequently with any published work or any dissertation work elsewhere; it will result in severe disciplinary action.

_____

Signature of Student

Date:

Place:

Endorsed by

Prof. Malaram Kumhar

(Signature of Guide)

Acknowledgements

It gives me immense pleasure in expressing thanks and profound gratitude to **Prof. Malaram Kumhar**, Assistant Professor, Computer Engineering Department, Institute of Technology, Nirma University, Ahmedabad for his valuable guidance and continual encouragement throughout this work. The appreciation and continual support he has imparted has been a great motivation to me in reaching a higher goal. His guidance has triggered and nourished my intellectual maturity that I will benefit from, for a long time to come.

It gives me an immense pleasure to thank **Dr. Sanjay Garg**, Hon'ble Head of Computer Engineering Department, Institute of Technology, Nirma University, Ahmedabad for his kind support and providing basic infrastructure and healthy research environment.

A special thank you is expressed wholeheartedly to **Dr Alka Mahajan**, Hon'ble Director, Institute of Technology, Nirma University, Ahmedabad for the unmentionable motivation she has extended throughout course of this work.

I would also thank the Institution, all faculty members of Computer Engineering Department, Nirma University, Ahmedabad for their special attention and suggestions towards the project work.

Smit Bharada
15MCEC03

Abstract

This report contains an overview of the tasks performed at Arris India Private Limited, Bengaluru within one year of internship.

Set-Top-Box (STB) is a leading equipment found in every home nowadays. To provide a continuous top quality stream to the user every day without any disturbance is a challenging task. Video Quality Assessment (VQA) of the video delivered to the user is a must. Without VQA one cannot know the perceptual quality of the video received at the end. With the advent of image and video in our daily lives, VQA has become a trending area of research. There are many techniques available to maintain and analyze the quality of a video. Out of them, Netflixs VMAF, RTM, timestamps and QR-Code based approach and some no-reference techniques are described in this report.

VQA is required because distortions occur in the video due to compression, scaling, encoding, packet loss, frame loss, etc. Various artifacts can be introduced in the video, which are to be identified and removed if possible. Brief description about the artifacts which may be introduced in the video is also provided.

Apart from VQA techniques, a study on some of the trending codecs was also carried out to have an understanding how actually compression works and also discover variety of codecs which could suit ever expanding video broadcasting industry within STB environment.

Abbreviations

STB	Set Top Box
MOVIE	Motion based Video Integrity Evaluation
SD	Standard Definition
HD	High Definition
VQA	Video Quality Assessment
VQM	Video Quality Metric
FR	Full Reference
RR	Reduced Reference
NR	No Reference
HVS	Human Visual System
MOS	Mean Opinion Score
TS	Timestamps
QoE	Quality of Experience
fps	Frames Per Second
SSIM	Structural Similarity Index
MS-SSIM	Multi-Scale Structural Similarity Index
VMAF	Video Multi-method Assessment Fusion
RTM	Real Time Monitor
MPEG	Motion Pictures Experts Group
PES	Packetized Elementary Stream
PID	Packet Identifier
DCT	Discrete Cosine Transform
RLC	Run Length Coding
VLC	Variable Length Coding
GOP	Group of Pictures
TAD	Targeted Advertisement

Contents

List of Figures

Chapter 1

Introduction

1.1 Overview

With the advancement in the Internet, videos and images have become a part of our daily lives. There are billions of videos and images uploaded, shared, downloaded and streamed across the world every day. These files in raw format can take up a lot of data, so it is not feasible to transfer data in raw format, compression is required for faster and easier transfer at low bandwidths. This compression is one of the reasons that degrades quality of a video, but it is an inevitable scenario. Without compression there can be no real time applications working. And as there will be no reduction in video and image data and keeping in mind the end user QoE, VQA is the only alternative. Quality of the data has to be examined and maintained over the network and improved if required before delivered to the end user.

Apart from the compression, video quality can be degraded by lossy networks and scaling also. These degradations in the video are known to be artifacts, thus VQA techniques focuses on identifying such artifacts and remove them if possible or just showcase the quality of video reduced because of such artifacts. The best way to assess a quality of video is through a human eye. There is no better alternative than a group of people viewing a video and rating its quality compared to the reference video. But in a fast paced world of Internet it is not possible everytime to gather a bunch of people and take their opinion. These approach is tedious and time consuming, so as a substitute certain video quality metrics are developed which tries to predict the quality of a video objectively. To

measure the effectiveness of these techniques usually a comparison is made to the subjective scores. One of the most used technique for subjective quality assessment is MOS - Mean Opinion Score. A group of people, accommodating from experts to layman are made to view the reference video in ambient light and surroundings, which are suitable for video observation. The score of reference video, whether it may be an SD or HD, or even 4K video, its score is always considered to be 100. Then the viewers are made to view the distorted in the same viewing conditions and ratings of the examiners are noted down. If the reference video was encoded at a different resolution the test video is also scaled to that resolution and then taken for examination. Mean of this opinion scores are taken as target to be achieved by the objective VQA, to know the effectiveness of the algorithm. More the results match, the better is the algorithm.

There are tons of algorithms available for VQA, but to find an optimum one depends on what kind of analysis a user wants. If quality is the ultimate goal, MOVIE or VQM, or some of the MMF algorithms can be useful. If speed at which VQA is carried out is important and most parts of the video consists of low motion activities, PSNR and MSE can be good alternatives. SSIM and edge extraction is like a midway between time and quality constraints. Ultimately all the VQA techniques tries to predict the perceived video quality objectively, an objective metric which predicts similar to the human eye, would be considered as the best. One of the most used technique nowadays to make predictions similar to the human eye, is the HVS - Human Visual System, which takes into account the high level and low level features of the video while analysis.

As already mentioned VQA is an integral part of video industry, but VQA is only needed because artifacts get introduced or loss of data occurs in the video due to compression or network anomalies. Thus this report also focuses on basic understanding of how compression works which is mentioned in the latter parts of the report. Another purpose of this study on codecs was to identify an appropriate codec which could cater to the needs of newly growing video content with respect to STB implementation. Next section briefly describes some of the artifacts which may be introduced in a video due to loss of useful data. Next part of the report focuses on some of the alternatives which were taken into consideration for VQA as part of the research work carried out at Arris

India. And the latter part briefs about a project End-to-End Automation Rack carried out alongwith the research.

1.2 What are artifacts?

There are many causes due to which a video may seem distorted to end user. These distortions in the video are called artifacts. A brief description about various artifacts which can be present in a video are given below:

- Banding:
  Artifact is a problem of inaccurate colour presentation, usually due to insufficient bits being supplied to display them; which occurs in banding.

- Dropped Frames:
  Artifact looks similar to judder however it does not only occur with 24fps content, and may be random. It is also more noticeable and can cause the playback picture to almost pause/stop during playback.

- Ghosting:
  An artifact which will retain a previous image displayed on-screen or will display an off centered, usually transparent, image over the currently rendered image.

- Halo/Ringing effect:
  Artifact occurs around sharp transition points or lines, and appears like doubling of these points or lines. Usually occurs when mixing fast moving video with static OSD layers.

- Jaggies:
  Stairstepping artifact that appears where there should be smooth straight lines or curves; seen usually on edges of images during playback.

- Judder:
  Dropped frames during playback of 24fps content in pan and scan scenes, resulting in non-smooth motion; usually caused by bad 3:2 pulldown.

- Macroblocking:

  Pixelation that occurs in blocks (i.e. 16x16, etc.), and usually stands out from a smoothed out or normally displayed picture. Also known as blocking.

- Mosquito noise:

  Video artifact that appears near crisp edges of objects in video frames that are compressed. It occurs at decompression when the decoding engine has to approximate the discarded data.

1.3 Objective of study

There are many existing techniques to analyse the quality of video at hand, but certain approaches at some point can sound to be inappropriate to implement in every scenario. In the world of VQA, there is a tradeoff between accuracy and time. Finest algorithms like MOVIE and VQM give the most analogous results to the human perception of a video, but can seem to be complex and take up a lot of time to analyse. So this study aims to find a generic approach or a solution to counter the challenges faced while VQA and come up with a solution which is easy to implement and should give fair enough results, if not the best but in acceptable amount of time. Supplementary aim of this study is also to understand basics of compression and a research on codecs, for its feasibility and scalability with the STB environment and diverse video industry.

1.4 Scope of work

Video is an integral part of Arris, and the absolute goal is to let users have the best viewing experience possible.This report will tend on finding a good alternative for assessing a video quality in as much less time as possible. The most recent techniques and complex algorithms are not taken up as an alternative yet, because first a simple and effective solution was required and the direction of research is based on that. As a result of the research, an FR technique is developed which gives a good trade off between time and accuracy. Accuracy of this technique depends based on the algorithm which the user wants to implement. This technique will prove to be highly efficient if the same stream or content is to be analyzed regularly.

Chapter 2

Literature survey

2.1 Motivation

Since the time of cable connection, home television has become an integral part of everybodys life. From the time of analog signal to digitization, home television broadcasting and video quality have come a long way. In the earlier days, quality of video delivered was not considered as important as reliable delivery of content. But in this modern era, more importance lies in QoE of the user. Arris is a customer oriented company, so end user experience is of utmost concern. To maintain quality of the content till end of the chain is not an easy task, thus VQA is needed at the end delivery. Thus VQA helps to assess the problems due to which a video may get distorted, and once the root cause is found, actions can be taken to limit the damage as much as possible.

2.2 Significance

This research is dedicated to find a common solution for VQA, which could be applied to any type of video or image, despite of network dependency or the vividness of artifacts which may be present in the video. Also the research is not directed towards finding the best VQA method or metric, but rather on finding a mid-way solution which would take less time for assessment with good results analogous to human eye perception. At the end of the research it is expected to have a universal solution to VQA, which would be easy to perform.

2.3 Study

In the world of videos and images, VQA is the last thing to be carried out. Before progressing with the VQA, a video may come across many elements capturing, rendering, encoding, decoding, transmission, compressing and decompressing. These elements hugely affect how a video be received and perceived at the end. Studying all the elements thoroughly is beyond the scope of this research, but a brief understanding about all the elements could help in choosing the right VQA technique.

VQA can be mainly divided in two parts objective and subjective quality assessment. For obvious reasons, subjective quality assessment is the best because at the end it is a humans perception is the best metric we have. But subjective assessment is not possible at all times, thus we have to rely on objective assessment. Objective assessment can depend on many factors, VQA can be carried out based on pixel differences, bitstream anomalies or hybrid of both [6]. Going further into these topics, we can also make out the artifacts which may be introduced due to scaling of the video or while encoding and decoding and directly focus on those artifacts. These methodology can help in assessment of a video in a better way. Apart from these we can also focus on how the video is being transmitted across the network. The reliable the source the lesser the chances of lossy video. Also, nowadays videos and images are transferred across with a lot of compression, and at the clients end the video or image is reconstructed from the previous and the future frames of the video, thus network losses should be kept as minimum as possible, so that proper reconstruction can happen and chances of artifacts while decoding can be removed. Some streaming techniques does not allow upcoming frames to be transmitted prior to their turn, in such cases complete reliability is on the previous frames and thus bitstream errors are to be handled cautiously.

Considering the above mentioned elements, many VQA algorithms have come up. PSNR, Visual Greyscale, SSIM, Edge Extraction, VQM, MOVIE, Information Fidelity based approaches, cognitive approaches, etc are some of the well known metrics and approaches for VQA. Thus to choose an optimal approach from the existing techniques or combining the techinques based on its effectiveness in particular scenarios, the motive

of this research is to find a simple and effective universal solution. These solution can be presented in the form of a tool, which could assess the video quality depending on the conduct of the video.

Chapter 3

Approaches to video quality verification

3.1 FR Techniques

In this technique, the distorted or the test video is compared with the reference video as a whole. That is pixel by pixel comparison is done on respective frames of both videos under scrutiny. FR techniques gives the promising results of all the techniques because here the complete video is available for comparison with the test video. More the data to compare with, better the results. But these approaches can sometimes prove to be tedious when implemented in the real world scenario, as the chances of availability of the reference video at the users end is next to impossible. Real time applications like video conferencing, streaming applications cannot use this to achieve assessment results on the fly. Thus despite of being the best technique for VQA, RR and NR can come in handy in fast paced scenarios. Some of the tools and techniques used for FR VQA are mentioned below.

3.1.1 Netflix

Netflix is an American multinational entertainment company provides streaming media and video on demand both online and physically. As Netflix content distribution was to be integrated in STBs, the approach used by Netflix for video quality verification can prove useful for testing the end user viewing experience.

- Basic understanding:

  Netflix has open sourced its method of video verification by the name VMAF (Video Multi-method Assessment Fusion). Netflix provides a variety of content to the end user ranging from animation to high quality viewing (HD, 4K, etc.). Depending on the users capacity, appropriate content with best quality is provided. Netflix follows Adaptive Bit-rate Streaming to achieve this. Netflix works under TCP, so there are no chances of frame losses and bit errors while content delivery. So the only points taken into consideration while video quality measurement are the artifacts caused by compression and scaling of the content depending on the end device. HEVC, H.264/AVC and VP9 are the normally used codecs for compression of the video and testing purposes.

- VMAF:

  To test the surplus diverse content provided by Netflix, first a database is created which covers all the different kinds of streams which can be delivered to the end user. Netflix has provided a database consisting of 34 reference videos and about 300 distorted videos. How Netflix simulated the network or which tools were used for them is not mentioned. VMAF is a comparison algorithm, which compares two videos frame by frame by considering the overall perceptual quality of the video. A non-linear regressor is trained based on three major metrics:

  1. Visual Information Fidelity (VIF)
  2. Detail Loss Metric (DLM)
  3. Motion

  VMAF fuses these elementary metrics and tries to predict the subjective quality of the video objectively. Subjective quality of the video is determined by DMOS (Difference of Mean Opinion Score), which is the most widely used metric for subjective quality comparison. Depending on the training set, weights are assigned to the elementary metrics and then the SVM regressor is tested on other dataset to predict VMAF score for video verification. Detailed understanding of the same can be obtained from [4]. Also the code and link for the database can be obtained from [5]. According to the results shown in this blog, VMAF provides a good estimate

to the DMOS value. But there are certain issues while incorporating it in Arris's lab.

- Issues:
  - VMAF algorithm works, considering that there will be no frame loss or data loss in transmission. But in a generic case there might be losses in frames or data while capturing the video while scaling it, encoding it or splitting it which resembles to the actual scenario. Thus we cannot directly apply VMAF to these environment.
  - All the video comparison carried out in VMAF is done in raw format (YUV format). So if we want to check for some new videos other than those provided by Netflix or open databases, we have to re-convert the videos to YUV after decoding.
  - VMAF requires DMOS values to compare it with the objective quality metrics. Thus for new set of videos, DMOS values will be required using which the regressor could be trained and then can be used on test data. Calculating DMOS values for the videos is an extensive task, as it requires workforce to evaluate each and every video.
  - Usually a video capture card is used to record the test video, so this capturing can start from any frame. Thus syncing of the test video with the reference is required before frame by frame comparison starts. But in VMAF no such mechanism is given, we can only start comparison if both the frames are synced and guaranteed that there is no frame loss in between.
- Features:
  - VMAF Development Kit (VDK) provides means to create our own database. Thus we can test this algorithm on our streams provided that both the reference and distorted videos are available.
  - SVM model can be trained for our database by assigning weights to the elementary metrics according to our needs. Thus the model will work as we require by providing all the metrics and parameters necessary for it.

- Technical requirements:

  Software requirements as per:[5]

software/package	version
Ubuntu/Mac	14.04 LTS/OS 10.10.5
gcc/g++	>=4.8
numpy	>=1.10.4
scipy	>=0.17.0
matplotlib	>=1.5.1
pandas	>=0.17.1
scikit-learn	>=0.18
h5py	>=2.2.1
VMAF	0.3.1
VDK	1.0.0

Table 3.1: VMAF software requirements

- Implementation:

  Required libraries and packages were installed in Mac OS, to test VMAF. Screenshots of comparison between two videos provided in the Netflix dataset are shown. The results are displayed in the form of VMAF_score which is the final score and the others are the scores for elementary metrics. adm2, vif_scalex scores range from 0 (worst) to 1 (best), and motion score typically ranges from 0 (static) to 20 (high-motion). For further experimentation and testing of the algorithm with specific parameters, [5] can be referred.

- Results:

  Here the comparison is done between two videos of 6 seconds, one being reference and other being distorted one. In total 47 frames are compared, each frame showing its respective results.

3.1.2 Real-Time Monitor

RTM is an Automated, Quality-of-Experience (QoE) Monitor, which measures the quality between two points. It does this after the decoder (STB) has had a chance to hide (or fix) the problems, which is exactly what the audience sees. RTM measures the video quality, audio quality, audio or video delay (offset), audio program loudness and VANC data accuracy. If any of these drop below a preset threshold, then an alarm is triggered and the sequences are saved. RTM saves valuable man-hours by alarming on all errors

```
apushpanathan:Desktop general$ cd vmaf-master/
apushpanathan:vmaf-master general$ export PYTHONPATH=/Users/general/Desktop/vmaf-master/python:$PYTHONPATH
apushpanathan:vmaf-master general$ ./run_vmaf yuv420p 576 324 resource/yuv/src01_hrc00_576x324.yuv resource/yuv/src01_hrc01_576x324.yuv

Asset: {"asset_dict": {"height": 324, "use_path_as_workpath": 1, "width": 576, "yuv_type": "yuv420p"}, "asset_id": 5162564171574674, "content_id": 5162564171574674, "dataset": "run_
vmaf", "dis_path": "resource/yuv/src01_hrc01_576x324.yuv", "ref_path": "resource/yuv/src01_hrc00_576x324.yuv", "workdir": "/Users/general/Desktop/vmaf-master/workspace/workdir/aaf95
30a-d52d-400b-ba44-79e3146a3c15"}
Executor: VMAF_V0.3.1
Result:
Frame 0: VMAF_feature_adm2_score:0.957, VMAF_feature_motion_score:0.000, VMAF_feature_vif_scale0_score:0.503, VMAF_feature_vif_scale1_score:0.877, VMAF_feature_vif_scale2_score:0.93
6, VMAF_feature_vif_scale3_score:0.964, VMAF_score:80.743
Frame 1: VMAF_feature_adm2_score:0.942, VMAF_feature_motion_score:4.105, VMAF_feature_vif_scale0_score:0.414, VMAF_feature_vif_scale1_score:0.820, VMAF_feature_vif_scale2_score:0.90
2, VMAF_feature_vif_scale3_score:0.945, VMAF_score:75.495
Frame 2: VMAF_feature_adm2_score:0.935, VMAF_feature_motion_score:3.740, VMAF_feature_vif_scale0_score:0.385, VMAF_feature_vif_scale1_score:0.805, VMAF_feature_vif_scale2_score:0.89
5, VMAF_feature_vif_scale3_score:0.943, VMAF_score:72.413
Frame 3: VMAF_feature_adm2_score:0.940, VMAF_feature_motion_score:3.597, VMAF_feature_vif_scale0_score:0.392, VMAF_feature_vif_scale1_score:0.806, VMAF_feature_vif_scale2_score:0.89
6, VMAF_feature_vif_scale3_score:0.942, VMAF_score:74.037
Frame 4: VMAF_feature_adm2_score:0.927, VMAF_feature_motion_score:3.354, VMAF_feature_vif_scale0_score:0.352, VMAF_feature_vif_scale1_score:0.768, VMAF_feature_vif_scale2_score:0.86
9, VMAF_feature_vif_scale3_score:0.926, VMAF_score:67.374
Frame 5: VMAF_feature_adm2_score:0.925, VMAF_feature_motion_score:3.466, VMAF_feature_vif_scale0_score:0.351, VMAF_feature_vif_scale1_score:0.767, VMAF_feature_vif_scale2_score:0.86
7, VMAF_feature_vif_scale3_score:0.922, VMAF_score:66.541
Frame 6: VMAF_feature_adm2_score:0.936, VMAF_feature_motion_score:3.168, VMAF_feature_vif_scale0_score:0.374, VMAF_feature_vif_scale1_score:0.780, VMAF_feature_vif_scale2_score:0.87
3, VMAF_feature_vif_scale3_score:0.922, VMAF_score:70.142
Frame 7: VMAF_feature_adm2_score:0.920, VMAF_feature_motion_score:3.531, VMAF_feature_vif_scale0_score:0.341, VMAF_feature_vif_scale1_score:0.748, VMAF_feature_vif_scale2_score:0.85
1, VMAF_feature_vif_scale3_score:0.907, VMAF_score:63.342
Frame 8: VMAF_feature_adm2_score:0.924, VMAF_feature_motion_score:2.883, VMAF_feature_vif_scale0_score:0.338, VMAF_feature_vif_scale1_score:0.754, VMAF_feature_vif_scale2_score:0.85
6, VMAF_feature_vif_scale3_score:0.911, VMAF_score:64.578
Frame 9: VMAF_feature_adm2_score:0.925, VMAF_feature_motion_score:3.348, VMAF_feature_vif_scale0_score:0.358, VMAF_feature_vif_scale1_score:0.765, VMAF_feature_vif_scale2_score:0.86
2, VMAF_feature_vif_scale3_score:0.913, VMAF_score:65.646
Frame 10: VMAF_feature_adm2_score:0.919, VMAF_feature_motion_score:3.224, VMAF_feature_vif_scale0_score:0.341, VMAF_feature_vif_scale1_score:0.748, VMAF_feature_vif_scale2_score:0.8
50, VMAF_feature_vif_scale3_score:0.908, VMAF_score:62.638
Frame 11: VMAF_feature_adm2_score:0.916, VMAF_feature_motion_score:2.751, VMAF_feature_vif_scale0_score:0.345, VMAF_feature_vif_scale1_score:0.746, VMAF_feature_vif_scale2_score:0.8
46, VMAF_feature_vif_scale3_score:0.902, VMAF_score:60.717
Frame 12: VMAF_feature_adm2_score:0.932, VMAF_feature_motion_score:3.097, VMAF_feature_vif_scale0_score:0.384, VMAF_feature_vif_scale1_score:0.774, VMAF_feature_vif_scale2_score:0.8
66, VMAF_feature_vif_scale3_score:0.917, VMAF_score:68.013
Frame 13: VMAF_feature_adm2_score:0.919, VMAF_feature_motion_score:3.499, VMAF_feature_vif_scale0_score:0.345, VMAF_feature_vif_scale1_score:0.748, VMAF_feature_vif_scale2_score:0.8
49, VMAF_feature_vif_scale3_score:0.908, VMAF_score:62.912
Frame 14: VMAF_feature_adm2_score:0.915, VMAF_feature_motion_score:2.953, VMAF_feature_vif_scale0_score:0.331, VMAF_feature_vif_scale1_score:0.739, VMAF_feature_vif_scale2_score:0.8
43, VMAF_feature_vif_scale3_score:0.904, VMAF_score:60.818
Frame 15: VMAF_feature_adm2_score:0.924, VMAF_feature_motion_score:2.726, VMAF_feature_vif_scale0_score:0.353, VMAF_feature_vif_scale1_score:0.760, VMAF_feature_vif_scale2_score:0.8
57, VMAF_feature_vif_scale3_score:0.913, VMAF_score:64.563
Frame 16: VMAF_feature_adm2_score:0.910, VMAF_feature_motion_score:2.977, VMAF_feature_vif_scale0_score:0.322, VMAF_feature_vif_scale1_score:0.733, VMAF_feature_vif_scale2_score:0.8
39, VMAF_feature_vif_scale3_score:0.903, VMAF_score:58.788
Frame 17: VMAF_feature_adm2_score:0.913, VMAF_feature_motion_score:3.283, VMAF_feature_vif_scale0_score:0.334, VMAF_feature_vif_scale1_score:0.743, VMAF_feature_vif_scale2_score:0.8
47, VMAF_feature_vif_scale3_score:0.907, VMAF_score:60.322
Frame 18: VMAF_feature_adm2_score:0.927, VMAF_feature_motion_score:3.185, VMAF_feature_vif_scale0_score:0.368, VMAF_feature_vif_scale1_score:0.767, VMAF_feature_vif_scale2_score:0.8
62, VMAF_feature_vif_scale3_score:0.914, VMAF_score:66.070
Frame 19: VMAF_feature_adm2_score:0.912, VMAF_feature_motion_score:3.321, VMAF_feature_vif_scale0_score:0.330, VMAF_feature_vif_scale1_score:0.734, VMAF_feature_vif_scale2_score:0.8
40, VMAF_feature_vif_scale3_score:0.901, VMAF_score:59.655
Frame 20: VMAF_feature_adm2_score:0.912, VMAF_feature_motion_score:3.839, VMAF_feature_vif_scale0_score:0.333, VMAF_feature_vif_scale1_score:0.746, VMAF_feature_vif_scale2_score:0.8
51, VMAF_feature_vif_scale3_score:0.906, VMAF_score:60.673
Frame 21: VMAF_feature_adm2_score:0.927, VMAF_feature_motion_score:4.249, VMAF_feature_vif_scale0_score:0.375, VMAF_feature_vif_scale1_score:0.781, VMAF_feature_vif_scale2_score:0.8
74, VMAF_feature_vif_scale3_score:0.922, VMAF_score:67.877
Frame 22: VMAF_feature_adm2_score:0.922, VMAF_feature_motion_score:4.385, VMAF_feature_vif_scale0_score:0.364, VMAF_feature_vif_scale1_score:0.770, VMAF_feature_vif_scale2_score:0.8
65, VMAF_feature_vif_scale3_score:0.916, VMAF_score:65.639
Frame 23: VMAF_feature_adm2_score:0.929, VMAF_feature_motion_score:4.835, VMAF_feature_vif_scale0_score:0.388, VMAF_feature_vif_scale1_score:0.786, VMAF_feature_vif_scale2_score:0.8
```

Figure 3.1: VMAF Results - 1

which affirms what the operator already has seen or alerts them to a potential problem which was missed. This system is being used in Arris India Private Limited for video quality assessment. For detailed understanding of RTM, [7] can be viewed.

- Issues:
  - RTM compares the two videos only on the basis of PSNR or DMOS (with MS-SSIM algorithm). These algorithms are not that effective as fused metric of VMAF, which provides a better estimate to the visual perception of human eye.
  - RTM runs slower than VMAF (provided that the SVM model is trained). This happens because RTM searches for the sync frame everytime there is a frame loss or data loss among the two videos.

```
65, VMAF_feature_vif_scale3_score:0.916, VMAF_score:65.639
Frame 23: VMAF_feature_adm2_score:0.929, VMAF_feature_motion_score:4.835, VMAF_feature_vif_scale0_score:0.388, VMAF_feature_vif_scale1_score:0.786, VMAF_feature_vif_scale2_score:0.8
74, VMAF_feature_vif_scale3_score:0.920, VMAF_score:69.050
Frame 24: VMAF_feature_adm2_score:0.953, VMAF_feature_motion_score:5.013, VMAF_feature_vif_scale0_score:0.468, VMAF_feature_vif_scale1_score:0.853, VMAF_feature_vif_scale2_score:0.9
19, VMAF_feature_vif_scale3_score:0.951, VMAF_score:80.845
Frame 25: VMAF_feature_adm2_score:0.934, VMAF_feature_motion_score:4.501, VMAF_feature_vif_scale0_score:0.395, VMAF_feature_vif_scale1_score:0.804, VMAF_feature_vif_scale2_score:0.8
91, VMAF_feature_vif_scale3_score:0.936, VMAF_score:72.061
Frame 26: VMAF_feature_adm2_score:0.930, VMAF_feature_motion_score:3.884, VMAF_feature_vif_scale0_score:0.372, VMAF_feature_vif_scale1_score:0.791, VMAF_feature_vif_scale2_score:0.8
83, VMAF_feature_vif_scale3_score:0.932, VMAF_score:69.767
Frame 27: VMAF_feature_adm2_score:0.932, VMAF_feature_motion_score:3.725, VMAF_feature_vif_scale0_score:0.377, VMAF_feature_vif_scale1_score:0.793, VMAF_feature_vif_scale2_score:0.8
85, VMAF_feature_vif_scale3_score:0.933, VMAF_score:70.404
Frame 28: VMAF_feature_adm2_score:0.920, VMAF_feature_motion_score:3.379, VMAF_feature_vif_scale0_score:0.345, VMAF_feature_vif_scale1_score:0.762, VMAF_feature_vif_scale2_score:0.8
63, VMAF_feature_vif_scale3_score:0.919, VMAF_score:64.320
Frame 29: VMAF_feature_adm2_score:0.916, VMAF_feature_motion_score:3.296, VMAF_feature_vif_scale0_score:0.344, VMAF_feature_vif_scale1_score:0.758, VMAF_feature_vif_scale2_score:0.8
59, VMAF_feature_vif_scale3_score:0.918, VMAF_score:62.687
Frame 30: VMAF_feature_adm2_score:0.929, VMAF_feature_motion_score:3.811, VMAF_feature_vif_scale0_score:0.371, VMAF_feature_vif_scale1_score:0.778, VMAF_feature_vif_scale2_score:0.8
73, VMAF_feature_vif_scale3_score:0.924, VMAF_score:68.242
Frame 31: VMAF_feature_adm2_score:0.917, VMAF_feature_motion_score:3.393, VMAF_feature_vif_scale0_score:0.341, VMAF_feature_vif_scale1_score:0.750, VMAF_feature_vif_scale2_score:0.8
55, VMAF_feature_vif_scale3_score:0.913, VMAF_score:62.428
Frame 32: VMAF_feature_adm2_score:0.914, VMAF_feature_motion_score:3.632, VMAF_feature_vif_scale0_score:0.334, VMAF_feature_vif_scale1_score:0.742, VMAF_feature_vif_scale2_score:0.8
48, VMAF_feature_vif_scale3_score:0.909, VMAF_score:61.149
Frame 33: VMAF_feature_adm2_score:0.923, VMAF_feature_motion_score:3.818, VMAF_feature_vif_scale0_score:0.356, VMAF_feature_vif_scale1_score:0.761, VMAF_feature_vif_scale2_score:0.8
62, VMAF_feature_vif_scale3_score:0.917, VMAF_score:65.367
Frame 34: VMAF_feature_adm2_score:0.913, VMAF_feature_motion_score:3.195, VMAF_feature_vif_scale0_score:0.342, VMAF_feature_vif_scale1_score:0.746, VMAF_feature_vif_scale2_score:0.8
50, VMAF_feature_vif_scale3_score:0.909, VMAF_score:60.684
Frame 35: VMAF_feature_adm2_score:0.916, VMAF_feature_motion_score:3.429, VMAF_feature_vif_scale0_score:0.351, VMAF_feature_vif_scale1_score:0.753, VMAF_feature_vif_scale2_score:0.8
53, VMAF_feature_vif_scale3_score:0.911, VMAF_score:61.964
Frame 36: VMAF_feature_adm2_score:0.933, VMAF_feature_motion_score:3.295, VMAF_feature_vif_scale0_score:0.385, VMAF_feature_vif_scale1_score:0.783, VMAF_feature_vif_scale2_score:0.8
75, VMAF_feature_vif_scale3_score:0.926, VMAF_score:69.510
Frame 37: VMAF_feature_adm2_score:0.916, VMAF_feature_motion_score:3.347, VMAF_feature_vif_scale0_score:0.348, VMAF_feature_vif_scale1_score:0.754, VMAF_feature_vif_scale2_score:0.8
55, VMAF_feature_vif_scale3_score:0.913, VMAF_score:62.245
Frame 38: VMAF_feature_adm2_score:0.915, VMAF_feature_motion_score:3.975, VMAF_feature_vif_scale0_score:0.344, VMAF_feature_vif_scale1_score:0.752, VMAF_feature_vif_scale2_score:0.8
56, VMAF_feature_vif_scale3_score:0.914, VMAF_score:62.291
Frame 39: VMAF_feature_adm2_score:0.926, VMAF_feature_motion_score:3.266, VMAF_feature_vif_scale0_score:0.360, VMAF_feature_vif_scale1_score:0.769, VMAF_feature_vif_scale2_score:0.8
68, VMAF_feature_vif_scale3_score:0.920, VMAF_score:66.441
Frame 40: VMAF_feature_adm2_score:0.920, VMAF_feature_motion_score:3.917, VMAF_feature_vif_scale0_score:0.341, VMAF_feature_vif_scale1_score:0.752, VMAF_feature_vif_scale2_score:0.8
57, VMAF_feature_vif_scale3_score:0.915, VMAF_score:64.250
Frame 41: VMAF_feature_adm2_score:0.919, VMAF_feature_motion_score:3.796, VMAF_feature_vif_scale0_score:0.347, VMAF_feature_vif_scale1_score:0.758, VMAF_feature_vif_scale2_score:0.8
60, VMAF_feature_vif_scale3_score:0.918, VMAF_score:63.866
Frame 42: VMAF_feature_adm2_score:0.935, VMAF_feature_motion_score:4.184, VMAF_feature_vif_scale0_score:0.380, VMAF_feature_vif_scale1_score:0.786, VMAF_feature_vif_scale2_score:0.8
80, VMAF_feature_vif_scale3_score:0.931, VMAF_score:71.385
Frame 43: VMAF_feature_adm2_score:0.921, VMAF_feature_motion_score:3.748, VMAF_feature_vif_scale0_score:0.361, VMAF_feature_vif_scale1_score:0.770, VMAF_feature_vif_scale2_score:0.8
68, VMAF_feature_vif_scale3_score:0.922, VMAF_score:65.299
Frame 44: VMAF_feature_adm2_score:0.925, VMAF_feature_motion_score:4.216, VMAF_feature_vif_scale0_score:0.359, VMAF_feature_vif_scale1_score:0.770, VMAF_feature_vif_scale2_score:0.8
67, VMAF_feature_vif_scale3_score:0.921, VMAF_score:66.862
Frame 45: VMAF_feature_adm2_score:0.937, VMAF_feature_motion_score:4.213, VMAF_feature_vif_scale0_score:0.393, VMAF_feature_vif_scale1_score:0.801, VMAF_feature_vif_scale2_score:0.8
89, VMAF_feature_vif_scale3_score:0.936, VMAF_score:72.799
Frame 46: VMAF_feature_adm2_score:0.931, VMAF_feature_motion_score:4.877, VMAF_feature_vif_scale0_score:0.392, VMAF_feature_vif_scale1_score:0.798, VMAF_feature_vif_scale2_score:0.8
85, VMAF_feature_vif_scale3_score:0.931, VMAF_score:70.839
Frame 47: VMAF_feature_adm2_score:0.939, VMAF_feature_motion_score:4.970, VMAF_feature_vif_scale0_score:0.419, VMAF_feature_vif_scale1_score:0.817, VMAF_feature_vif_scale2_score:0.8
98, VMAF_feature_vif_scale3_score:0.941, VMAF_score:74.404
Aggregate: VMAF_feature_adm2_score:0.925, VMAF_feature_motion_score:3.592, VMAF_feature_vif_scale0_score:0.366, VMAF_feature_vif_scale1_score:0.772, VMAF_feature_vif_scale2_score:0.
868, VMAF_feature_vif_scale3_score:0.921, VMAF_score:66.628
Done.
apushpanathan:vmaf-master general$
apushpanathan:vmaf-master general$
```

Figure 3.2: VMAF Results - 2

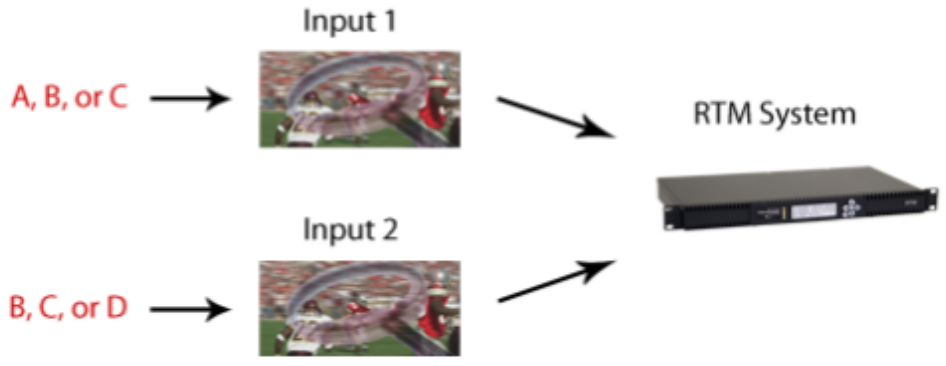

Figure 3.3: RTM

– After the impairments are found using RTM Monitor, the two videos have to be fed manually in RTM Player for further analysis. There is no direct

integration between the two systems and only one can be launched at a time.

- Features:
  - Even though RTM runs slower than VMAF due to lip-sync, this feature happens to be useful in the environment of STBs, because frame losses and capture of black frames can occur in this setup. Where VMAF would fail, RTM can continue its analysis.
  - RTM provides a GUI for both RTM Monitor and RTM Player, which is easier to operate.
  - RTM supports wide range of video streams and video formats, which can be viewed side by side for assessment. Also the impaired video sequences are recorded for further analysis.

3.1.3 Other techniques

As the above mentioned techniques showed some issues which are not suitable for generic scenarios which one can come across during VQA, a simple approach was developed which could give optimum results within a considerable amount of time, but with one leverage (Timestamps or QR-Code to be added in the reference and test videos). In these approach, the reference video would be framed and stored using ffmpeg and named according to the frame number present in the timestamp or the qr-code within the video. Then the distorted video would be scanned frame by frame, extract the information held in the timestamp or the qr-code and a direct comparison would be made with reciprocating name of the frame obtained from reference video. These solution would solve majority of the issues faced in the previous techniques. The major overhead is the the addition of timestamp or qr-code in both streams for frame syncing and a direct comparison. But the execution speed increases five holds as compared to the previous technique.

- Timestamps:

  As mentioned above here frame recognition is based upon the timestamp held within the frame. Timestamps are recognized using template matching. We need to have a dataset of general timestamps which are used most often (for accurate results). Algorithm used for frame by frame comparison is SSIM or MS-SSIM (available

in the scikit image library). Operations on video like frame extraction, resizing, seeking are achieved by open source library opencv.

- QR-Code:

  As compared to timestamps, the major difference lies in the frame recognition via qr-codes within the video. QR-Code recognition is carried using another open source library Z-Xing, which detects the qr-code lying in the frame and also extracts the information from it. After frame identification, algorithm used for image comparison was edge extraction.

- Technical requirements:

software/package	**version**
g++	>=4.8
ffmpeg	>=3.2.2
Python2	>=2.7
scipy	>=0.18.1
numpy+mkl	>=1.11.2
scikit-image	>=0.12.3
Zxing	>=3.3.0

Table 3.2: TS and QR-Code software requirements

- Implementation:

  At first, the reference video is framed and stored according to the timestamps or qr-codes present inside the video frames using ffmpeg. Python2 has to be installed on the machine with the required libraries as mentioned in the technical requirements. Opencv also has to be installed and add its ffmpeg.dll and cv2.pyd in python installation directory. Ffmpeg.dll is required so that all the file formats could be supported for processing. After the pre-requisites are done, using opencv and python libraries, frame by frame comparison is carried out based on timestamp and qr-code match. Scikit-image provides SSIM and MS-SSIM algorithm for image comparison and the results of exch frame are stored in a file.

- Results:

Mask	Approach	Algorithm	Length	Resolution	Fps	frames	Run time
with	TS	SSIM	1 min	1280 x 720	59.94	3590	31 min 7 sec
	TS (Parallel)	SSIM	1 min	1280 x 720	59.94	3575	16 min 24 sec
	QR-Code	Edge Extraction	1 min	1280 x 720	50	3001	39 min 8 sec
without	TS	MS-SSIM	1 min	1280 x 720	59.94	3590	24 min 7 sec
	TS (Parallel)	MS-SSIM	1 min	1280 x 720	59.94	3590	12 min 35 sec

Table 3.3: TS and QR-Code results

3.2 RR Techniques

RR techniques are less fruitful in terms of VQA than FR techniques, but are faster and can be used for real time applications. In this, certain features from the reference video are extracted and passed along with the test video. Respective features can be extracted from the test video and correlated with the features of the reference video on the fly. Hence VQA can be performed and examined at any point of time with less data required for comparison. Further research in this area was not carried out because prior focus was required with the FR and NR approach. In future, research in this area would be undertaken.

3.3 NR Techniques

Other than FR and RR techniques which require a reference video or certain features of it for comparison, there are NR techniques which tries to assess video quality only through the distorted video alone. These can prove to be most useful in terms of bandwidth requirements because the test video is all we need for VQA. The downs of this approach are that the results obtained from NR-VQA are menial as compared to above two approaches, especially in lossy network conditions. But still its less data requirement and being computationally light are good enough facts for it to find a place in market. Some promising techniques relying on NR-VQA which were taken into consideration are mentioned below.

3.3.1 HVS based technique

This technique proposes a non-reference metric that works on the approach of Human Visual System (HVS), which tries to identify the salient regions of the frame and focus more on those parts of the image. This metric estimates the degree of blur and blockiness in each video frame from the impaired video only, and uses it with the saliency map to derive a weighting function. The metric is obtained after taking into account Stationary Saliency and Motion Saliency of the video. These models helps us to find the important and non-important regions of the video. According to these saliency maps, a fused saliency map is formed which has retains the features of both the maps according to the weights assigned to them. Based on the saliency map, metric can be proposed depending on the blurriness and blockiness in the video frames. The final metric obtained is named Quality Prediction Metric (QPM). Equations for the same are given in [1]. The QPM scores are compared to the subjective scores for video quality assessment. Detailed understanding of the metric is given in [1].

3.3.2 Technique based on encoding scheme

This technique is based on the general process involved in the compressions of a video. Modern day codecs compress the video through many steps transformation, quantization, motion estimation, motion compensation and entropy coding. Out of these quantization and entropy coding are the major steps for compression. The no-reference quality metric presented in this paper is based on three factors:

1. Quantization Parameter Factor Quantization induces majority of distortion and compression in a video, thus assess the quality of a video Quantization Parameter (QP) is an important factor.

2. Motion Factor: Video signals do suffer from spatial distortion but majority of artifacts are introduced by the temporal effects in a video sequence. Thus this metric is used to calculate the local and global motion consistency of the video based on P-frames and the object taken into consideration.

3. Bit Allocation factor: HVS is less sensitive to the loss of high frequency information than that of low frequency information. This phenomenon inspires the idea of Region of Interest (ROI) based video coding method in transform domain.

The final metric is obtained by combining the features of all the three parameters for a sequence of n-frames into consideration. Paper also gives the results obtained by testing the metric on LIVE Video Quality Database. Detailed understanding of the metric can be obtained from [3].

3.3.3 Machine learning based technique

The method described in this paper analyses the received video stream in terms of eight NR features which are dependent on both the bitstream and the pixel levels. Also in addition to that the network is sensed to obtain two network measurements, namely - nominal bitrate and estimated level of packet loss. These ten features in all will serve as input to a Supervised Learning (SL) algorithm, which will be trained on the basis of previously learned samples of video quality. Then this trained model will be deployed on the client side to perform a predictive assessment of the quality of the video under scrutiny. Different SL models are trained and tested on the LIVE video quality database, their comparative results with the subjective measurement and the computation complexity are also mentioned in the paper. Detailed understanding of the technique can be obtained from [2].

Chapter 4

Codecs

In the world multimedia, audio, video and image data are inadvertent. Hence the size of data to be transferred is very huge compared to the bandwidth available. Codecs are a solution to this dilemma, which compresses data before transfer and reforms it before presenting to the target, hence meeting with bandwidth availability. Codecs have been part of video industry since the advent of digital content delivery mechanism as compared to analog distribution. Codec comprises of two terminologies - encoder and decoder. Encoder compresses the data while decoder decompresses it.

The purpose of this study was to find a suitable codec which could prove future compatible with the variety of data getting generated everyday. A codec which fits in conventional STB environment alongwith support for VOD, OTT and online streaming services. Some of the trending codecs like MPEG, VP8, VP9 and HEVC are studied to gain deeper understanding about each one and find suitable answer. Generic explanation about working of a codec in video compression is mentioned below with respect to MPEG as it is one of the most widely used codec by content distributors worldwide (including the content provided in ARRIS).

4.1 MPEG

MPEG stands for 'Moving Pictures Experts Group'. First MPEG standard was released in 1993, since then many standards have been launched under MPEG specification, the latest of them being MPEG-DASH. MPEG is not a single format, but a combination of various standardized tools which can be suitable for a range of applications [9].

MPEG comprises of different elementary streams which together form a program stream

or a transport stream which is viewed to the viewer. Elementary streams can be one video stream and one or more audio streams which are multiplexed to form a program stream. For convenience, elementary streams are packetized to form Packetized Elementary Stream (PES). These PES packets contain timestamps to maintain audio-video lip-sync. A program stream contains a single program while a transport stream contains multiple programs in the same stream. To maintain decorum among individual programs, transport streams are further divided into fixed sized packets are given PID (packet identifier) depending on the program to which the elementary stream belongs.

MPEG supports many profiles and levels to achieve a compression factor depending

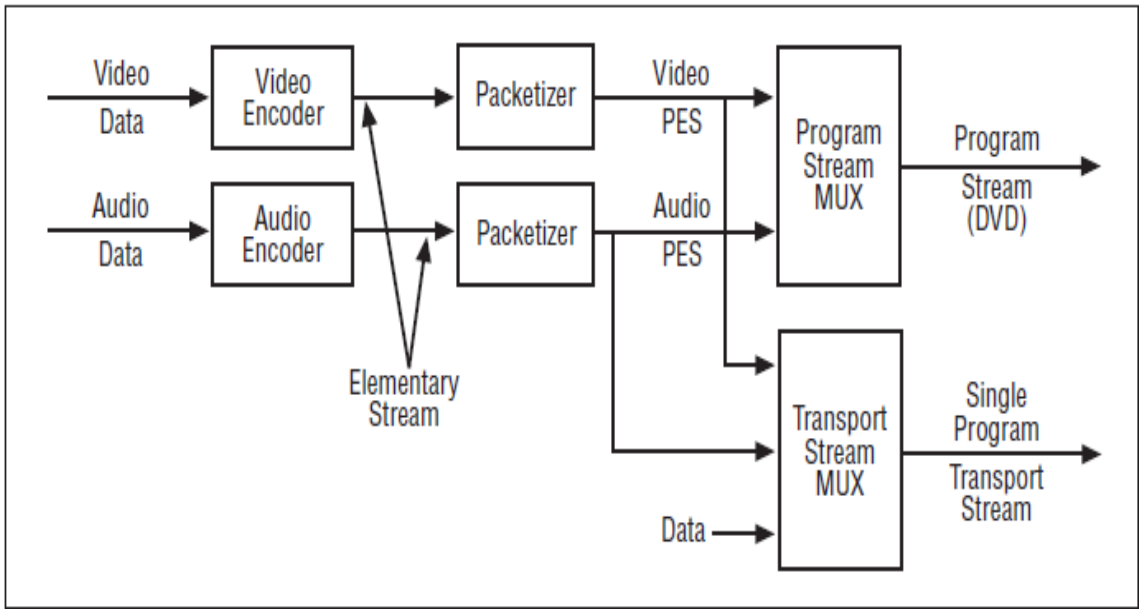

Figure 4.1: MPEG streams

on the need of application. Like bitrates required for HD content will be much higher than that for SD. Also MPEG is a lossy codec, which means the data presented is different from the original data. An ideal coder has to present all the entropy within the video and redundant data has to be reused to achieve compression. So MPEG performs depending on the application requirement and bandwidth availability, to achieve a certain compression factor and quality of data with profiles and levels. MPEG achieves its compression exploring spatial and temporal redundancy within the video. Firstly, temporal redundancy is explored, which is redundancy between successive frames. After that spatial redundancy is to find similar regions or macroblocks within the frame itself. Further compression is achieved through entropy coding. Process of compression and steps involved in it are briefly explained below:

4.1.1 Transformation

Transformation is the process in which data is converted from time domain to frequency domain i.e. any sine-cosine waveform can be expressed in terms of amplitude and phase modulation if frequency is known. Thus the signals received can be expressed as an integration of multiple cosine waves having particular amplitude, phase modulation and frequency. For this, a set standard frequencies are decided, which is called a basis function. The transform finds each frequency in the input signal by multiplying different frequencies within basis function and integrating them. Thus we end up with coefficients representing multiplication factor for all the basis frequencies. In MPEG, DCT is used as a transform, which takes 64 frequencies as part of basis function. So after transform is applied 8x8 block can be represented with 64 coefficients representing the contribution of individual waveform in the construction of the input signal. As shown in figure, horizontal spatial frequencies increase from left-to-right while vertical spatial frequencies increase from top-to-bottom. Codecs take advantage of the fact that not all spatial frequencies are simultaneously present, thus majority of coefficients would be zero. Hence compression can be achieved with this redundant data and entropy can be transferred.

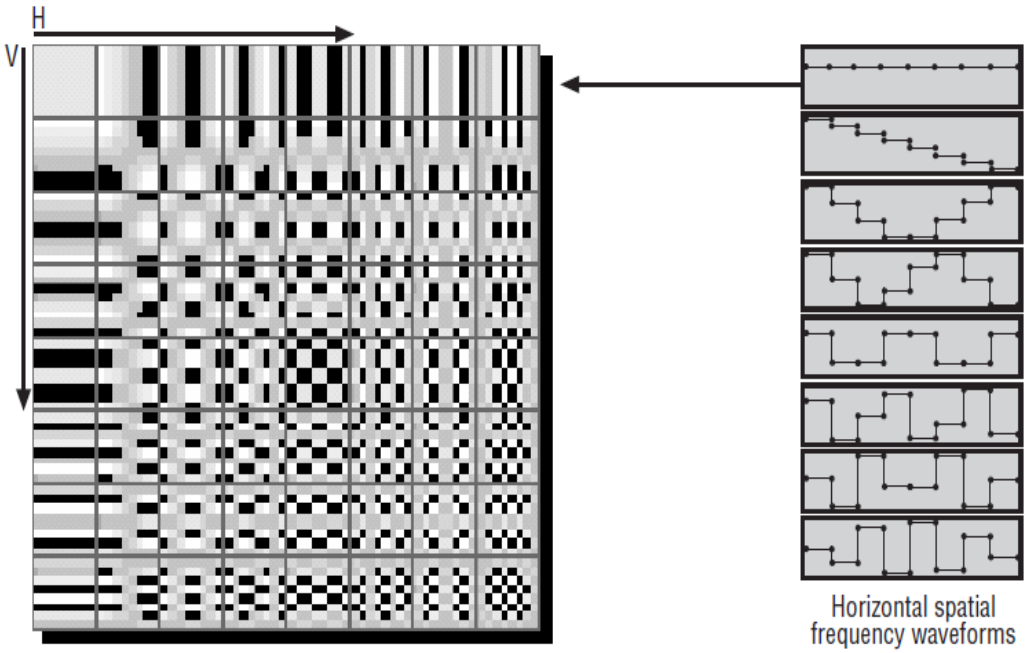

Figure 4.2: spatial frequency (basis function) [8]

4.1.2 Quantization

Transformation does not compress the data, in fact it might increase the data. But DCT transforms data such that only entropy data is left into consideration while rest can be removed. Quantization introduces majority of compression among all the steps involved. After quantization original data cannot be reformed again, some data loss always occurs. Quantization is nothing but dividing the coefficients obtained from DCT by certain numbers so that range of values confines to a certain scope. So for example all the coefficients might be having a value between 0-128 after quantization.

4.1.3 Weighting

Weighting is smart way of quantizing data. As already mentioned, not all spatial frequencies are simultaneously present in a frame. Also most of the frequencies contributing to generate input signals are lower frequencies. Higher frequencies are rarely used. So through weighting, more quantization factor can be kept for higher frequency while lesses for lower frequencies. As human eye is less susceptible to high frequency data, no change would be observed even if that data is changed to considerable amount.

4.1.4 Scanning

After transformation and quantization, majority of the coefficients would be zero. Scanning is the process of traversing data within frame so that most number of zeroes can be encountered. Widely used scanning methods are raster scan and zigzag scan. Both of them start from top-left and ends at bottom-right. MPEG usually uses zigzag scan as there is a higher possibility of finding chain of zeros via this scan. Zigzag scan is shown in the figure.

4.1.5 Entropy Coding

After the sequence of bits is obtained from scanning, entropy coding tries to compress data even more by reducing number of bits required to represent same data. This is lossless compression. Two alternatives are widely used for entropy coding, Run Length Coding (RLC) and Variable Length Coding (VLC). RLC achieves compression by sending the number of occurrences of 0's or 1's in a chain rather than individual bits. Longer the chain more is the compression. While VLC, assigns acronyms to numbers, for example

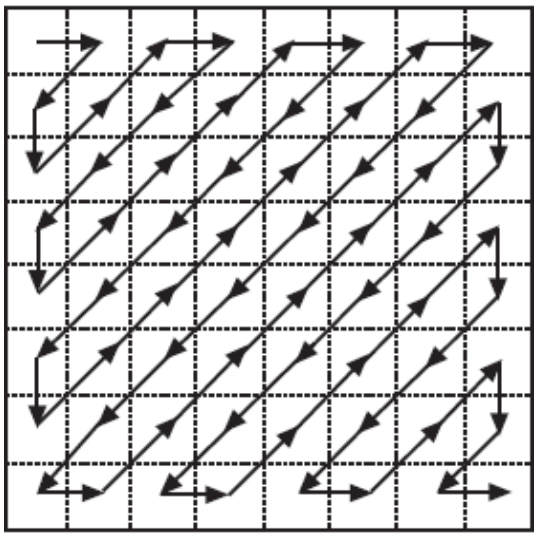

Figure 4.3: Zigzag scan [8]

the coefficient which is repeated the most will be assigned smallest acronym, and this acronym would be sent instead of the coefficient. This way MPEG decides on some of the recurring coefficients and replaces them with short acronyms, so that compression can be achieved. VLC is based on Huffman Coding.

4.1.6 Temporal Coding

Temporal Coding increases the possibility of redundancy by many folds in a video. As a video is usually of 30-50 fps, most of the successive frames will be 90% similar to the previous frame. Temporal coding takes advantage of this fact and sends only the difference between the frames rather than the whole frames. Thus a lot of redundant data is avoided and also size of difference data is very less than the actual frame. Temporal coding can be achieved two ways explained below:

- Motion Compensation:

  Motion compensation is used to increase similarities between successive frames. Motion reduces this similarity, for example an object moving across the screen in consecutive frames. The object changes its position but not its appearance. Hence same block of pixels can represent the object but at a different position. To achieve this following actions can be taken:

  - Compute motion vector for current frame with respect to the previously coded frame

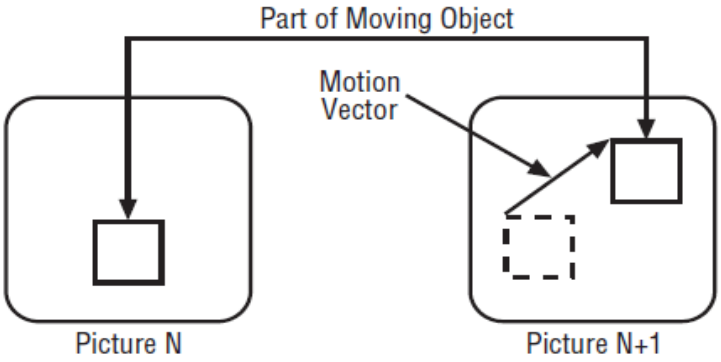

Figure 4.4: Motion Compensation

- Get the predicted frame by substituting motion vector blocks (Picture N+1)
- Find the difference between actual and predicted frame
- Send motion vector and difference picture which are enough to generate actual frame from previous frame

Using motion compensation size of data required to represent the frame may reduce to half of its original size.

- Bidirectional Coding:

  Bidirectional coding is same as motion compensation, the only addition being that the predicted frame is constructed referring to both previous frames and future frames in a video sequence. Using bidirectional coding, size of data required to represent a frame may reduce to quarter of a frame. Bidirectional coding can be further understood from types of frames explained below.

4.1.7 Types of frames

If a codec, to achieve high compression always uses previously coded frames then even loss of a single frame will disrupt the rest of the video. As all the frames could be relying on a single point of failure. To avoid this GOP - Group of Pictures is introduced. GOP is like a restart for a video sequence that means even if a frame is lost, video sequence only within that GOP would be affected and video could be started with any upcoming GOP. Generally GOP consists of about 12-15 frames depending on the codec in consideration. It forms a sequence of I-B-B-P-B-B-P-B-B-P-B-B-I. Each of the frame types is explained below:

- I-Frame: I frames are intra-coded frames, that means it only uses spatial coding. I frames do not use motion compensation, so that it could be re-constructed independently irrespective of other frames. This frame takes up maximum amount of data for representation as scope for compression is less.

- P-Frame: P frames are future frames which will come later in video sequence. These frames are constructed only from I frames. Data required to represent may reduce to half of the original data.

- B-Frame: B frames are constructed using I and P frames. These frames lie in between I and P frames within GOP and data required to represent these frames may reduce to quarter of its original size. Most of the frames within a video are compresses as B-frames

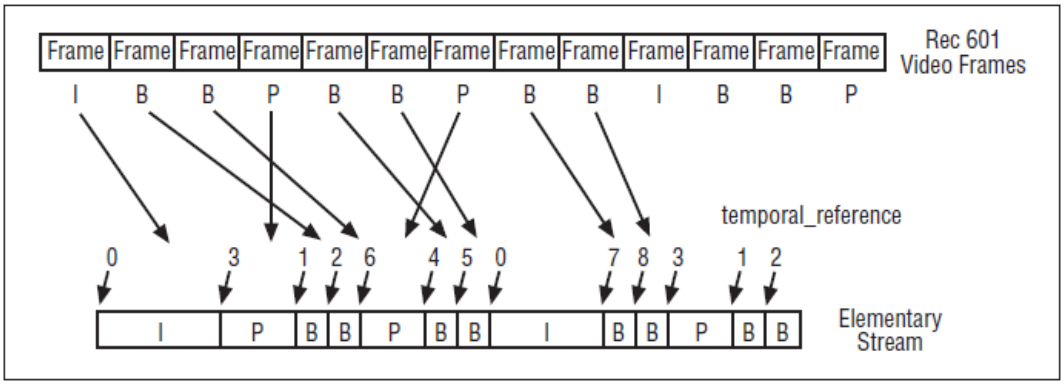

Figure 4.5: Types of frames in MPEG [8]

The order in which these frames are sent is different from the order in which they are displayed to the user. This is required for the decoder to decode B-frames which depends on P-frames. The oredering of frames is shown in the figure.

4.1.8 Pre-processing

The main purpose of a codec is to eliminate redundant data and transfer only entropy to the other side. More the redundancy, more is the compression factor. Certain phenomena like noise and film grain in a video may reduce these redundancy and hence possible compression factor decreases. Pre-processing is a stage which tries to avoid such problems. Pre-processing would remove the unnecessary grain and noise within the video before encoder starts encoding.

4.2 Other codecs

All the codecs, pretty much follow the above mentioned steps to encode a video. Similar steps are involved in VP8 and HEVC [10] also with differences in the size of macroblocks used, quantization tables, temporal coding used, types of frames. After a brief study on this codecs, it appears that some codec may provide efficient results in a particular scenario while some other codec may be better otherwise. Thus a codec should be chosen as per the application which is to be implemented. In general we can say HEVC is supposedly one of the best codec available out there as it contains different profiles for every application, thus it can provide best compression if chosen the main profile. But time of compression also matters in some applications, in such cases codecs like VP8, VP9 come into picture, which does not support all the profiles and levels as HEVC but outperforms it with low bandwidth and resolution contents. VP8, VP9 are mainly used for web viewing while HEVC can prove to be fruitful in other scenarios also.

Chapter 5

Health/Sanity checkup

End-to-End Automation Rack is a solution developed at Arris India, which provides end users a control over the test cases to schedule and run which could check sanity of the STB or can also perform other user specific tests. These tests can be as simple sa checking the UI of the box getting displayed or to check the IR commands resulting into specific actions getting performed onto the box or to complex tasks like audio, video analysis.

End user would select the build to be loaded into the box, select the test case which is to be ran, select a platform, initialize the parameters required for the task and run the test. The results would be returned to the user in the form of mail or a message or could be displayed on the desktop or mobile screen from which he/she will be operating. Overall process is explained in the form of a flowchart, and the individual units are explained in the coming sub-sections:

5.1 Builds

Builds consists of a set of iips which can be loaded into the STB, after which the box can start functioning to the instructions through APIs inside the box. Only iips are not enough for the functioning of the STB, a platform dependent bin file is also required

5.2 Campaign Creation and Deployment

The build - set of iips and platform (framework) - bin file, these when packaged together forms a campaign. Before loading these files into the box a campaign is created on the server and only this single campaign is deployed to the box rather than individual file. The whole process from creating campaign to deploying it when done manually took

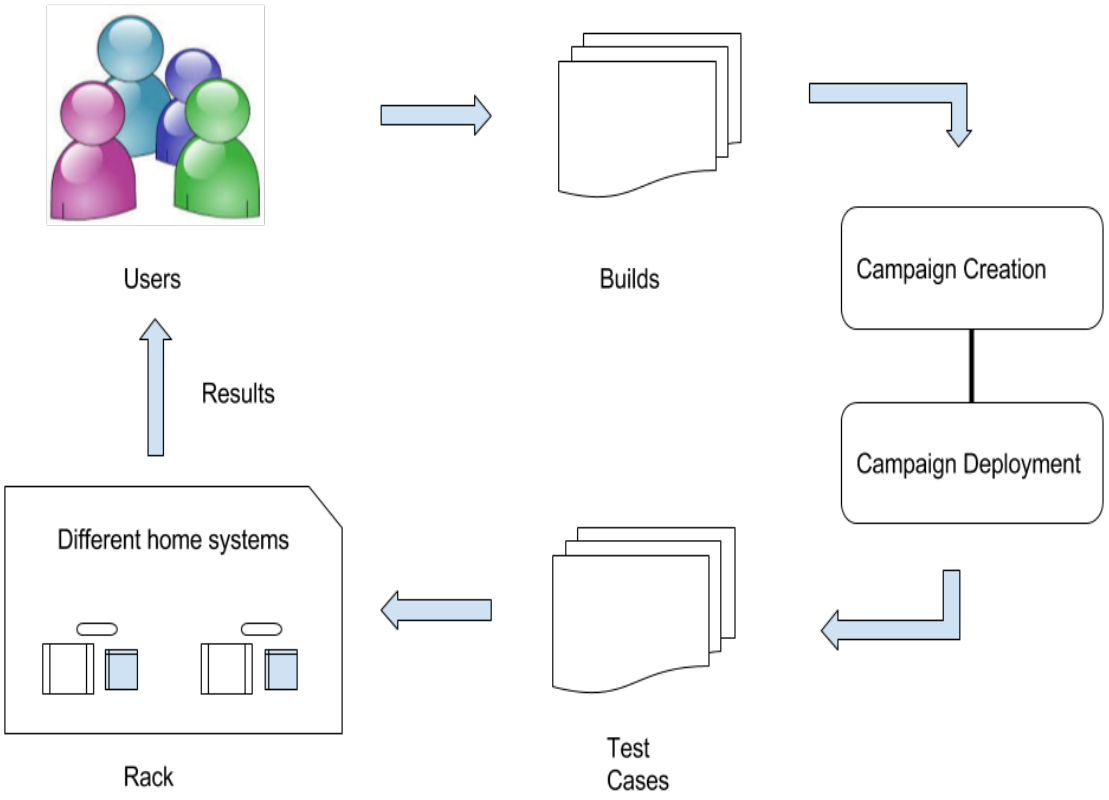

Figure 5.1: Health/Sanity checkup

about 45 min to 1 hour, and the frequency of campaign deployments would be at least one or two per day. To save considerable amount of man hours, the whole process of campaign creation and deployment is automated using sikuli 1.1.0. User initializes the parameters which are required and at the click of a button the required campaign would be created and deployed onto the server. Automated execution reduced the time to about 2.5 folds. I.e about 15-20 min are required. The whole automated process of campaign creation and deployment is summarized in the flowchart given below and individual units are also explained below.

- Build kit and platform (framework): The set of iips and bin which are to loaded are collected/downloaded into the PC and are uploaded to a server where the campaign is formed.

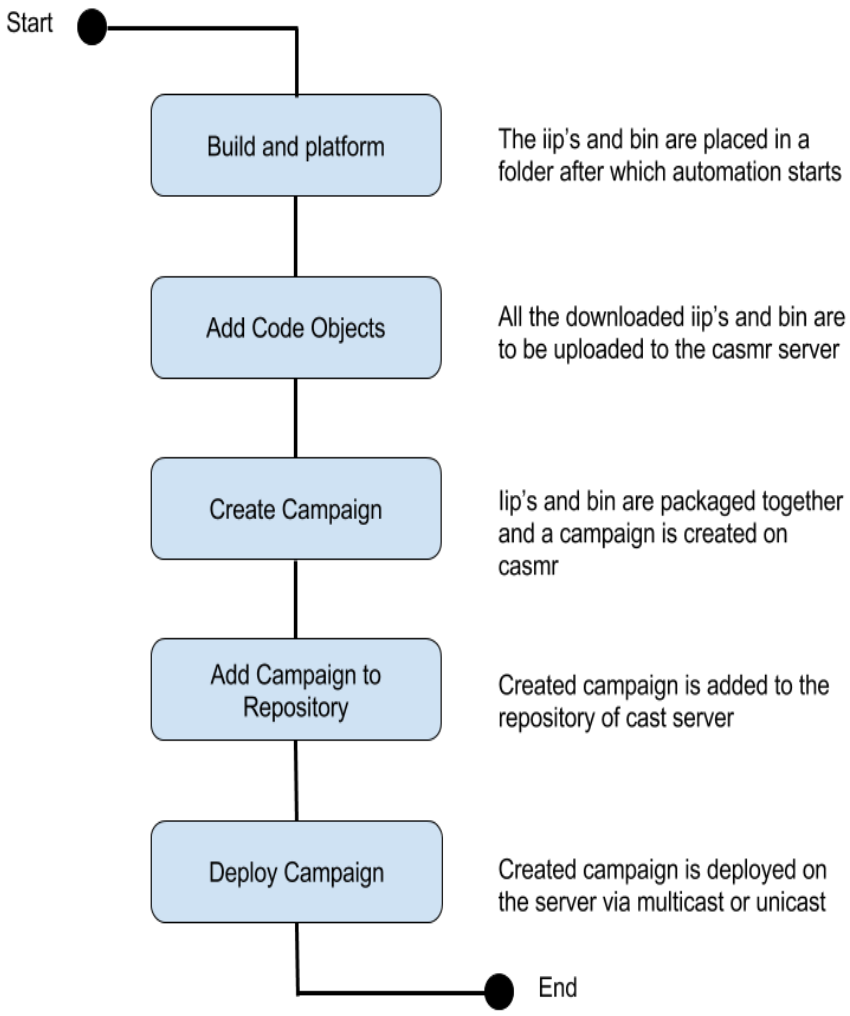

Figure 5.2: Campaign Creation and Deployment

- Adding Code Objects: The set of iips and the bin files are called code objects. These code objects are uploaded to the casmr server one by one and then the campaign is created on the server itself.

- Campaign Creation: Individual code objects loaded to casmr are packaged together and a campaign is formed on the server. Also the target model (STB model) of the box is specified here.

- Campaign Deployment: Campaign Deployment is done on a different server named cast. The campaigns created in the casmr are added to the cast repository. The these campaigns are deployed on the server via multicast or unicast.

Any of the deployed campaigns can be loaded on the box, and thus the STB can start functioning according to the loaded build and framework.

5.3 Test Cases

After the deployed campaign has been loaded into the box, a certain test case which is compatible with the build and platform can be scheduled and run in the box. These tests are scheduled and run in a proprietary automation tool of Arris India, which runs the scripts in a sequential manner along-with add-ons like capturing device,support for IR commands blasted through RedRat and command line tool which can be incorporated in the scripts. Two of the test cases are described below.

- Checking smooth transition of TAD: This test case checks whether the ad displayed in the channel playing in the STB is a network advertisement or TAD (targeted advertisement). There is a particular sequence in which transition occurs in case of TADs which is different from normal advertisement. The test case would start a video recording through a capture device and analyze the video to mark the presence of a TAD or no TAD. Also it accounts for smooth transition, means checks whether the number of black screens in the video are below a specific threshold to know that the transition wa smooth or not. The result would be collected in an excel sheet and can be examined later.

- VQA via timestamps/QR - Codes: As mentioned above in the FR techniques section, these VQA can be carried out through the automation tool as well. Despite of being an open source solution, automation tool supports command line tool, thus a call to the python scripts can be made via command line and all the specs of the machine on which automation tool is running can be used. The results of this script are in terms of number of black screens, freeze frames and VQA of every frame in the video, stored in an excel sheet.

5.4 Rack

Rack is nothing but the hardware setup which might be located anywhere in the world. These setup would include all the necessary equipments and tools which are necessary for a test to progress smoothly. Thus scheduling test cases on the case can be achieved via end-to-end automation rack.

Chapter 6

Conclusion

After going through different approaches for VQA, we can conclude that, each approach has its pros and cons. FR technique provides the best analogical results with the subjective results but are complex and compute intensive. Also reference video is required for assessment, which might not be available at the users end in every scenario.

On the other hand, NR techniques are not as accurate as the FR ones but are more user friendly. These techniques are easy to compute compared to FR techniques and can be very useful in cases of live streaming and real time monitoring applications where video quality maintenance is required on the go. Thus VQA would highly depend on the use case or the environment in which the video is to be tested, and a universal approach cannot be obtained with only one metric or technique but some kind of amalgamation of both would be required.

In terms of codecs, depending on the content to be distributed and on network condition, a codec should be selected. Because every codec has its own peculiarities and can prove to be very efficient with some particular applications. In general, VP8 and HEVC performs better for overall cases. Out of these two also, VP8 is better suited for low bandwidth and video quality upto HD content, while HEVC performs better with higher quality content.

Bibliography

[1] Guraya, Fahad Fazal Elahi, et al. "A non-reference perceptual quality metric based on visual attention model for videos." Information Sciences Signal Processing and their Applications (ISSPA), 2010 10th International Conference on. IEEE, 2010.

[2] Vega, Maria Torres, Decebal Constantin Mocanu, and Antonio Liotta. "Predictive No-Reference Assessment of Video Quality." arXiv preprint arXiv:1604.07322 (2016).

[3] Kulkarni, Yogini, and Charudatta V. Kulkarni. "No-Reference Video Quality Assessment." ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 4, Issue 6, June 2014

[4] http://techblog.netflix.com/2016/06/toward-practical-perceptual-video.html

[5] https://github.com/Netflix/vmaf

[6] Liu, Tsung-Jung, Weisi Lin, and C-C. Jay Kuo. "Recent developments and future trends in visual quality assessment." Proceedings of Asia-Pacific Signal and Information Processing Association Annual Submit and Conference. 2011.

[7] http://videoclarity.com/PDF/ClearViewSystemGuide.pdf

[8] MPEG http://www.img.lx.it.pt/ fp/cav/Additional_material/MPEG2_overview.pdf

[9] MPEG-2 Digital Video Technology Testing, Hewlett - Packard

[10] Feller, Christian, et al. "The VP8 video codec-overview and comparison to H. 264/AVC." Consumer Electronics-Berlin (ICCE-Berlin), 2011 IEEE International Conference on. IEEE, 2011.