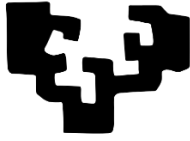


eman ta zabal zazu



Universidad
del País Vasco

Euskal Herriko
Unibertsitatea

A REDUCED REFERENCE VIDEO QUALITY ASSESSMENT METHOD FOR PROVISION AS A SERVICE OVER SDN/NFV-ENABLED NETWORKS

Author: Michail-Alexandros Kourtis

Supervisors: Fidel Liberal, Charilaos Koumaras

2018

(c)2018 MICHAEL ALEXANDROS KOURTIS

1	INTRODUCTION	26
1.1	CONTEXT AND MOTIVATION	26
1.2	CHALLENGES AND MAIN OBJECTIVES.....	28
1.3	STRUCTURE OF THE DOCUMENT.....	30
2	VIDEO QUALITY ASSESSMENT METHODS	33
2.1	SUBJECTIVE VIDEO QUALITY ASSESSMENT METHODS	35
2.2	OBJECTIVE VIDEO QUALITY ASSESSMENT METHODS.....	38
2.3	APPROXIMATIONS OF FR METHODS BY RR METHODS.....	41
2.4	CONCLUSIONS	42
3	THE PROPOSED REDUCED REFERENCE VIDEO QUALITY ASSESSMENT METHOD	43
3.1	OVERVIEW OF THE PROPOSED METHOD	43
3.2	FEATURE EXTRACTION USING SSIM	44
3.3	QUALITATIVE INTERPRETATION OF THE PROPOSED VIDEO QUALITY METRIC	48
3.4	PERFORMANCE COMPARISON OF SRR TO SSIM.....	49
3.5	PERFORMANCE COMPARISON OF SRR TO SUBJECTIVE DMOS AND OTHER ASSESSMENT METHODS.....	54
3.6	PERFORMANCE EVALUATION WITH OTHER RELATIVE COLORS.....	59
3.7	CONCLUSIONS	60
4	APPLICABILITY OF THE PROPOSED SRR AS A SERVICE ON TOP OF NFV/SDN- ENABLED NETWORKS.....	62
4.1	ANALYSIS OF NFV, SDN AND SRR AS A SERVICE.....	63
4.2	DEPLOYMENT OF THE PROPOSED SRR METHOD AT AN NFV ENVIRONMENT.....	66
4.2.1	REQUIREMENTS OF THE PROPOSED SDN/NFV-ENABLED NETWORK	66
4.2.2	PROPOSED SDN/NFV SRRAAS ARCHITECTURE	66
4.3	PROPOSED ARCHITECTURE IMPLEMENTATION.....	72
4.4	PERFORMANCE EVALUATION OF THE PROPOSED IS-QoE METHOD	75
4.5	CONCLUSIONS	78
5	VALIDATING THE SRRAAS METHOD OVER A SMALL CELL MOBILE NETWORK .	79
5.1	VIDEO QUALITY ASSESMENT IN 5G SMALL CELL NETWORKS.....	79
5.2	DEPLOYMENT OF THE PROPOSED SRR METHOD AT AN SMALL CELL ENVIRONMENT	82
5.2.1	REQUIREMENTS OF THE PROPOSED NFV-ENABLED SMALL CELL ARCHITECTURE.	82
5.2.2	PROPOSED NFV-ENABLED SMALL CELL ARCHITECTURE.....	83
5.3	PROPOSED NFV-ENABLED SMALL CELL ARCHITECTURE IMPLEMENTATION	84
5.4	PERFORMANCE EVALUATION OF SRR IN SMALL CELL NETWORKS.....	88
5.5	CONCLUSIONS	90
6	CONCLUSIONS AND FUTURE LINES	91
6.1	MAIN CONTRIBUTIONS	91
6.2	FUTURE LINES.....	97
6.3	ACKNOWLEDGEMENTS	99

7	PUBLICATIONS OF THE AUTHOR	100
8	ANNEX.....	103
8.1	VIDEO PROCESSING BASICS.....	103
8.2	PRINCIPLES OF VIDEO CODING.....	104
8.3	VIDEO CODING STANDARDS EVOLUTION	115
9	TECHNICAL AND GENERAL PURPOSE REFERENCES	131

List of Figures

Figure 1.1 - Structure of the document.....	31
Figure 2.1 - Full Reference Methods with multiple channels	38
Figure 2.2 - Full Reference Methods with single channel	39
Figure 2.3 - Reduced Reference Methods.....	40
Figure 3.1 - Overview of the proposed VQA Metric.....	43
Figure 3.2 - Typical SSIM index evaluation with original and target video sequences available at the same site	45
Figure 3.3 - SRR evaluation method using a reference video pattern as reference at both the service provider and end-user sites.....	47
Figure 3.4 - Variation of : (a) SSIMot (b) ideal SSIMtr (c) real conditions SSIMtr vs QP	48
Figure 3.5 - Representative Frames of the 40 test signals of the evaluation process.....	50
Figure 3.6 - Qualitative comparison of SRR and SSIM index applied on Kristen&Sara video sequence with QP=32.....	51
Figure 3.7 - Variation of MAPD vs QP : (a - solid line) Experimental, (b - dotted) Theoretical	53
Figure 3.8 - Scatter plot of objective SRR scores vs. DMOS v with the best fitting logistic function.....	55
Figure 3.9 - Basketball Pass SSIM vs SRR for QPs 12-32.....	57
Figure 3.10 - BQSquare SSIM vs SRR for QPs 12-32.....	58
Figure 3.11 - BQMall SSIM vs SRR QPs 12-32.....	59
Figure 4.1 - NFV architecture according to ETSI	63
Figure 4.2 - Overview of the proposed reference SDN/NFV architecture	67
Figure 4.3 - An overview of the experimental topology	74
Figure 4.4 - Openflow-based traffic steering within Openstack platform	74
Figure 4.5 - Label used to show the SDN-based traffic steering commands per hop	75
Figure 4.6 - Video quality degradation due to packet loss.....	76
Figure 4.7 - Video Quality is reinstated.....	78
Figure 5.1 - Architectural framework for combined PNF and VNF systems	83
Figure 5.2 - Virtualization of a Small Cell and split of functions between PNF and VNF	84
Figure 5.3 - An overview of the experimental testbed	86
Figure 5.4 - SRR variation vs frame sequence number without source adaptation for videos (a) KristenandSara and (b) BasketballDrill for various values of the backhaul bandwidth	89
Figure 6.1 - SRR variation vs frame sequence number with source adaptation for videos (a) KristenandSara and (b) BasketballDrill for various values of the backhaul bandwidth	99
Figure 8.1 - GOP Structure.....	105
Figure 8.2 - Example of variable block size coding	105
Figure 8.3 - A frame where motion vectors appear denoting the position of the best matching region.....	106
Figure 8.4 - A residual frame (on the left) denoting the differences between	107

Figure 8.5 - Steps of the Intra-Coding Prediction	108
Figure 8.6 - H.264/AVC Intracoding Prediction Modes.....	109
Figure 8.7 - H.264/AVC prediction example	109
Figure 8.8 - Example of DCT efficiency. Figure a is the source image, while b is reconstructed using only 1.....	111
Figure 8.9 - Huffman Coding Algorithm.....	114
Figure 8.10 - Timeline of video coding standards evolution.....	116
Figure 8.11 - H.261 Coding and Decoding steps.....	116
Figure 8.12 - Recursive quadtree representation of CU.....	128

List of Tables

Table 2.1 - Overview of the Subjective Quality Assessment Methods.....	36
Table 3.1 - MAPD for SRR and SSIM index for a white-video reference pattern.....	51
Table 3.2 - Comparison of the performance of VQA algorithms – LCC and SROCC.....	55
Table 3.3 - MAPD for the 40 Test Signals with Different Relative Colors.....	60
Table 7.1 - Resulting publications and relation to the PhD work	100
Table 8.1 - Quantization Examples	112

List of Abbreviations

Abbreviation	Meaning
3GPP	3rd Generation Partnership Project
AVC	Advanced Video Coding
CABAC	Context Based Arithmetic Coding
CAPEX	Capital Expenditure
CAVLC	Context Adaptive Variable Length Codes
CESC	Cloud Enabled Small Cell
COTS	Commercial off the Shelf
C-RAN	Cloud Radio Access Network
CU	Coding Unit
DCT	Discrete Cosine Transform
DMOS	Difference Mean Opinion Score
DPDK	Data Plane Development Kit
DSCQS	Double Stimulus Continuous Quality Scale
DSIS	Double Stimulus Impairment Scale
eNB (or eNodeB)	Evolved Node B
EPA	Enhanced Platform Awareness
EPC	Evolved Packet Core
FAT	Factory Acceptance Test
FCAPS	Fault Configuration Accounting Performance Security
FPGA	Field Programmable Gate Arrays
FR	Full Reference
GOP	Group of Pictures
GTP	GPRS Tunnelling Protocol
HEVC	High Efficiency Video Coding
HOT	HEAT Orchestration Template
HTTP	Hypertext Transfer Protocol
HVS	Human Visual System
IP	Internet Protocol

IPTV	Internet Protocol Television
ITU-R	International Telecommunication Union - Radiocommunication Sector
JCT-VC	Joint Collaborative Team on Video Coding
KPI	Key Performance Indicator
LCC	Pearson's Linear Correlation Scale
LTE	Long-Term Evolution
MAC	Media Access Control
MANO	Management and Orchestration System
MAPD	Mean Absolute Percentage Deviation
MEC	Mobile Edge Computing
ML2	Modular Layer 2
MME	Mobility Management Entity
MOCN	Multi-Operator Core Network
MOS	Mean Opinion Score
MPEG	Moving Pictures Experts Group
MSE	Mean Square Error
NFP	Network Forwarding Path
NFV	Network Function Virtualization
NFVI	NFV Infrastructure
NIC	Network Interface Card
NMS	Network Management System
NR	No Reference
NS	Network Service
NSD	Network Service Descriptor
NUMA	Non-Uniform Memory Access
ODL	OpenDaylight
OPEX	Operational Expenditure
OVS	Open vSwitch
PES	Packetized Elementary Stream
P-GW (or PDN-GW)	Packet Data Network Gateway
PNF	Physical Network Function
PSNR	Peak Signal to Noise Ratio
PU	Prediction Unit

QoE	Quality of Experience
QoS	Quality of Service
RAN	Radio Access Network
REST	Representational State Transfer
RR	Reduced Reference
SAT	Site Acceptance Test
SDN	Software Defined Networking
SFC	Service Function Chain
S-GW (or Serving-GW)	Serving Gateway
SLA	Service Level Agreement
SNR	Signal to Noise Ratio
SR-IOV	Single Root I/O Virtualization
SROCC	Spearman Rank Order Correlation Coefficient
SSCQE	Single Stimulus Continuous Quality Evaluation
SSIM	Structural Similarity Index Metric
TCP	Transmission Control Protocol
TR	Technical Report (by 3GPP)
TS	Technical Specification (by 3GPP)
TS	Transport Stream
TU	Transformation Unit
UE	User Equipment
VCEG	Video Coding Experts Group
VIM	Virtualised Infrastructure Manager
VLD	Virtual Link Descriptor
VNFD	Virtual Network Function Descriptor
VNFFG	Virtual Network Function Forwarding Graph
VNFM	Virtual Network Function Manager
VoD	Video on Demand
VQA	Video Quality Assessment
WAN	Wide Area Network

Abstract

The proliferation of multimedia applications and services has generated a noteworthy upsurge in network traffic regarding video content and has created the need for trustworthy service quality assessment methods. The recent market success of media services and the rising consumer demand, pushed innovation in the fields of video encoding techniques and adaptation and eventually the requirement for in service quality evaluation.

Facing new challenges in the multimedia and mobile network field, telecom operators and content providers have focused their interest towards novel technology approaches which can improve their customers' satisfaction and be easily deployable in the modern telecommunication networks. Currently, predominant position among the technological trends in telecommunication networks are Network Function Virtualization (NFV), Software Defined Networking (SDN) and 5G mobile networks equipped with small cells. In particular, the paradigm of SDN/NFV is based upon decoupling physical appliances from the network functions that operate on them and has gained significant attention due to its potential to reduce Operating Expenses (OPEX) and Capital Expenses (CAPEX), as well as extending the flexibility of network services. Similarly, the envision to build a diverse and "hyper-connected" network environment has led to the evolution of 5G. Notably, 5G proposes a joint radio-cloud architecture in an effort to place intelligence at the network edge by utilizing virtualization technologies.

Video Quality Assessment (VQA) methods are a very useful tool for both content providers and network operators, to understand of how users perceive quality and thus study the feasibility of potential services and adapt the network available resources to satisfy the user requirements. Full Reference (FR) VQA methods are accurate in measuring the end user's satisfaction, but they are not appropriate for implementation in commercial video distribution networks, because they need the original video at the customers' site, in order to compare it to the received video. Reduced Reference (RR) methods are less accurate, but they are appropriate for implementation in commercial telecommunication networks. Furthermore, to be successful, VQA methods must be able to be integrated in modern networks based on SDN/NFV and 5G technologies.

The thesis proposes a new RR method for Video Quality Assessment for as a Service provision, which combines the accuracy of a full reference metric, such as Structural Similarity Index (SSIM) Metric, with the advantage of being appropriate for deployment in modern telecommunication networks, based on Network Function Virtualization and SDN technologies, such as the forthcoming 5G networks. Initially, the novel VQA method is described as a standalone metric (i.e. SRR), analysed in detail and its performance is evaluated in comparison to other related VQA metrics. Then, it is proposed the advancement of the metric with additional features in order to be provided as a Service (i.e. SRRaaS), taking advantage of its integration within an SDN/NFV environment. Lastly, the further adoption of the SRRaaS method in 5G mobile networks equipped with

small SDN/NFV-enabled small cells is presented, along with the study of the challenges tackled. The deployment environment evolves as new technologies emerge, but the primary goal of the thesis is to provide an agile, efficient and appropriate video quality method for provision as a service over SDN/NFV networks. The thesis not only offers an innovative VQA algorithm, but also offers a thorough insight on current technologies which pave the future of telecommunications.

Resumen

La proliferación de servicios multimedia ha generado un incremento en el tráfico de red asociado al video y suscitado la necesidad de mecanismos confiables de evaluación de la calidad del servicio. La mayor demanda por parte del cliente y el reciente éxito comercial de los servicios multimedia ha promovido la innovación en las técnicas de codificación y transmisión adaptativa así como en los mecanismos de evaluación de la calidad.

Al mismo tiempo, de cara a afrontar los retos del mercado multimedia móvil los operadores de telecomunicaciones y los proveedores de contenidos han centrado su interés en aquellas tecnologías que puedan mejorar la satisfacción de los usuarios y, al mismo tiempo, garanticen un despliegue sencillo en las redes modernas. En la actualidad, las tendencias predominantes en dichas redes de telecomunicación consisten en la virtualización de funciones de red (*Network Function Virtualization NFV*) y en el uso de redes móviles 5G.

El paradigma NFV consiste en desacoplar los elementos hardware de las funciones de red que operan en ellos y ha ido ganando adeptos conforme se confirma su flexibilidad y capacidad para reducir el coste de adquisición (CAPEX) y de operación (OPEX).

Del mismo modo, la tendencia de crear un entorno de red diversa e “hiper conectada” ha llevado a la evolución hacia las tecnologías 5G, que propone también una arquitectura radio-cloud y desplegar inteligencia en los extremos de la red utilizando técnicas de virtualización.

Los métodos de evaluación de la calidad de video (*Video Quality Assessment (VQA)*) resultan una herramienta muy útil para proveedores de contenido y operadores de red de cara a entender cómo perciben la calidad los usuarios finales y, a partir de esa comprensión, analizar las posibilidades de hipotéticos nuevos servicios y adaptar los recursos disponibles para satisfacer los requisitos de esos usuarios.

Los métodos de evaluación de referencia completa (*Full Reference –VQA-*) generalmente proporcionan una medida acertada de la satisfacción de los usuarios pero no se pueden implementar de forma razonable sobre las redes comerciales de distribución de video al requerir de la fuente de video original en el equipamiento de usuario de cara a compararlo con el contenido recibido. Los métodos de referencia reducida (*Reduced Reference –RR-*), en cambio, son menos precisos pero su implementación en entornos comerciales resulta más sencilla. En cualquier caso, esos métodos de evaluación deberían poderse integrar en redes NFV y 5G.

Esta tesis propone un nuevo método de evaluación de la calidad de video de referencia reducida que trata de combinar la precisión de las métricas de referencia completa como SSIM (*Structural Similarity Index*) con la capacidad de poder desplegarse en redes modernas basadas en NFV o/y 5G. Para ello, inicialmente se presenta un nuevo método RR de evaluación de la calidad de video. Tras su descripción se compara su rendimiento

con otros métodos de evaluación. A continuación se plantea su integración en un entorno NFV y se describe cómo se puede aprovechar de dicho enfoque. Finalmente se analizan los retos de cara a su adopción en las futuras redes 5G. Puesto que el entorno tecnológico continúa en permanente evolución este trabajo de tesis se centra en desarrollar un método ágil y eficiente de evaluación de la QoE de los usuarios adaptable a esos cambios. En definitiva, esta tesis no sólo propone un método de evaluación de la calidad de video sino una revisión profunda de posibles mecanismos de despliegue de cara a garantizar que su evolución va emparejada a esos cambios.

Resumen Ejecutivo

El creciente interés de los consumidores por el contenido de video ha resultado a su vez en un universo cada vez mayor de servicios de provisión de video digital. Hoy, los proveedores de contenido pueden ofrecer su contenido a través de varios tipos de redes (televisión digital terrestre, satélite, Internet, redes móviles, etc.) y los operadores de red pueden desplegar sistemas de telecomunicaciones más eficientes y ubicuos, para hacer frente al enorme volumen de tráfico transportado a través de sus redes. El objetivo común tanto de los proveedores de contenido como de los operadores de red es mantener la satisfacción de los usuarios finales lo más alta posible.

Una evaluación confiable de la calidad del video juega un papel indispensable en el cumplimiento de la calidad de servicio comprometida (*Quality of Service* o QoS) y en la mejora de la calidad de la experiencia (*Quality of Experience* o QoE) del usuario final. Más específicamente, en un sistema de transporte de video, es importante monitorizar la QoS de transporte de red tanto a través de los parámetros de QoS de la red (como el retardo de paquetes y las tasas de pérdida de paquetes), como la QoS específica del servicio de video a través de parámetros más directamente relacionados con video (como el retardo en la reproducción de video y la calidad de codificación del mismo) que en última instancia contribuyen a la QoE del usuario.

La evaluación de calidad de video (*Video Quality Assessment* o VQA) permite comprender cómo perciben los usuarios la calidad y, por lo tanto, estudiar la viabilidad de posibles servicios y adaptar los recursos disponibles de la red para satisfacer los requisitos del usuario. La evaluación subjetiva de la calidad se considera generalmente como la técnica más confiable para evaluar un servicio multimedia ya que se basa en datos recopilados a partir de pruebas de opinión de observadores humanos. Desgraciadamente, estos métodos implican una cantidad considerable de recursos (principalmente tiempo y personas) y se consideran costosos y difícilmente escalables. Por lo tanto, los métodos subjetivos no resultan útiles por sí solos, para los operadores de redes ya que son imposibles de implementar y explotar en tiempo real en redes comerciales de distribución multimedia.

En paralelo a la evolución de los sistemas de distribución de contenido multimedia, en el campo de la creación de redes han surgido nuevos paradigmas que han transformado los nodos de red desde plataformas hardware de propósito específico a entornos virtualizados basados en tecnologías de redes definidas por software (*Software Defined Networks* o SDN) y virtualización de funciones de red (*Network Function Virtualization* o NFV) sobre hardware de propósito general. Además, las redes de acceso se están moviendo lentamente hacia la tecnología 5G. Uno de los elementos clave previstos del marco tecnológico de 5G es la capacidad de entregar inteligencia directamente a la red, en forma de dispositivos de red virtualizados, explotando conjuntamente esos paradigmas emergentes de NFV y Edge Cloud Computing. Siguiendo este cambio de paradigma, cada vez surgen más casos de uso de alto valor añadido a través de infraestructura o aplicaciones que tengan el potencial de ofrecerse como 'Servicio'.

De lo anterior, se puede concluir que un requisito crucial para que un método de VQA tenga éxito en la nueva era de las telecomunicaciones no solo es preciso que sea fiable, sino que también debe integrarse adecuadamente dentro de las infraestructuras de virtualización y redes móviles 5G y desplegarse fácilmente en tales infraestructuras. Esta tesis propone un nuevo método de VQA que cumpla con los requisitos anteriores. Más

precisamente, define, implementa y evalúa el rendimiento de un VQA denominado de referencia reducida (*Reduced Reference* o RR). Dicha métrica se calcula a partir de la ampliamente conocida Métrica de Similitud Estructural (SSIM). Además, se propone una arquitectura que define cómo la métrica de VQA puede integrarse dentro de infraestructuras virtualizadas basadas en tecnologías SDN y NFV e implementa el método como VNF sobre un banco de pruebas experimental y evalúa su desempeño.

Capítulo 1 introduce el alcance de la tesis, las motivaciones principales de la investigación y presenta una lista de los retos a los que se dirigen, junto con los objetivos de la tesis.

El Capítulo 2 se centra en la revisión de la literatura en el campo relacionado de Video Quality Assessment. Presenta las principales categorías de métodos VQA y analiza las ventajas y desventajas de cada una, de acuerdo con el tipo de aplicación en el que van a ser utilizados. La primera clasificación es entre métodos subjetivos y objetivos. Los métodos subjetivos se basan en la opinión de un grupo de espectadores con respecto a la degradación visual de una secuencia de video codificada en comparación con la secuencia original en bruto (no comprimida). En este capítulo, se describen varios métodos (según sus denominaciones en inglés Double Stimulus Impairment Scale, the Double Stimulus Continuous Quality Scale, the Absolute Category Rating, the Stimulus Comparison, the Single Stimulus Continuous Quality Evaluation, the Simultaneous Double Stimulus for Continuous Evaluation and the Subjective Assessment Methodology for Video Quality), junto con las escalas respectivas de calidad y el proceso seguido en cada uno. Este capítulo destaca que los métodos subjetivos son muy precisos para expresar la satisfacción de los usuarios finales, pero son caros (en tiempo y personas) y no apropiados para redes de distribución de video.

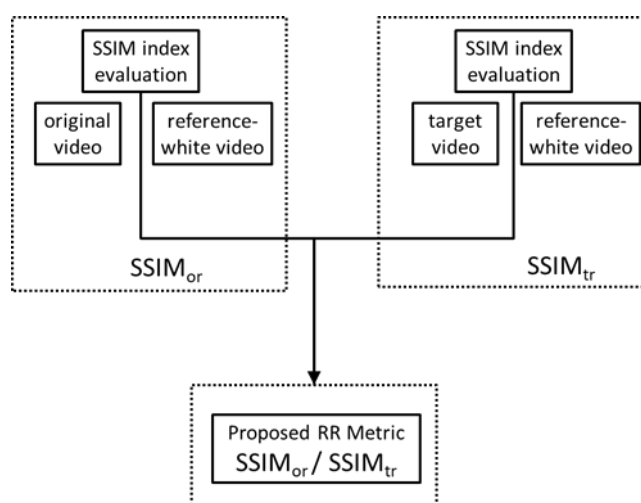
Los métodos objetivos de VQA son modelos computacionales matemáticos que utilizan varias características de las imágenes con el fin de aproximarse lo mejor posible a los resultados de pruebas subjetivas de una manera eficiente y rentable. Los métodos objetivos se clasifican en tres grandes grupos determinados por su enfoque y la métrica utilizada para la evaluación de la calidad: los de referencia completa (FR), los de referencia reducida (RR) y los de no referencia (NR). En la tesis, se describen las tres categorías anteriores, junto con las métricas correspondientes (por ejemplo, PSNR, SSIM, etc.) y las diversas características de las señales de video que se utilizan en cada procedimiento de evaluación. El análisis muestra que los métodos FR proporcionan una mayor precisión en comparación con RR y NR, pero no son apropiados para la integración en la cadena de provisión del servicio, dado que generalmente la señal original no está disponible en el sitio del usuario final. Por otro lado, los métodos RR son menos precisos, pero son apropiados para las redes de distribución de video, ya que solo requieren unas pocas características extraídas del video original en el sitio del usuario final. Los métodos NR son incluso algo menos precisos que los anteriores, pero en contrapartida no requieren ninguna información de la señal de video original. A este respecto, existe un compromiso entre la precisión y la idoneidad para las redes de distribución de video comerciales.

Esta tesis trata de hacer frente a este compromiso mediante un nuevo método de RR basado en el cálculo de SSIM, (que es una métrica FR), pero utilizando sólo unas pocas características extraídas del video original, por lo que puede ser categorizado como RR. El método propuesto ofrece por tanto una respuesta equilibrada entre la precisión de una métrica FR y la flexibilidad de una métrica RR (en términos de ser apropiada para la implementación en una red comercial de distribución de video).

En el Capítulo 3, se describe en detalle el método RR y se evalúa su rendimiento mediante un conjunto experimental de referencia. El método propuesto comprende el uso de características basadas en la evaluación del índice SSIM extraído tanto del fotograma de video original como del objetivo. Más específicamente, el índice SSIM se calcula para cada fotograma del video original utilizando como referencia un patrón de video estático, es decir, un video cuyos fotogramas son todos iguales. Por razones explicadas más adelante en la Sección 3 inicialmente se utiliza un patrón blanco. El índice SSIM ($SSIM_{or}$) para cada cuadro se transmite luego al usuario final. En el terminal del usuario final, se calcula el índice SSIM entre cada cuadro del video recibido (objetivo) y el mismo video estático patrón ($SSIM_{tr}$). La métrica RR propuesta (denominada SRR) es la relación de los dos índices SSIM, es decir:

$$SRR = SSIM_{or}/SSIM_{tr}$$

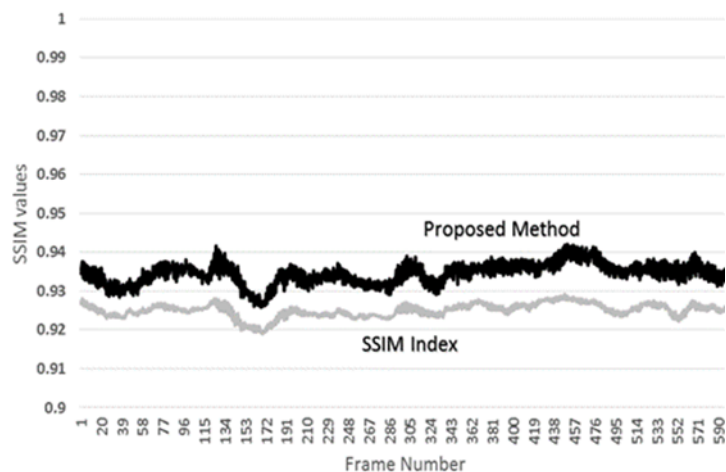
El proceso se muestra en la figura a continuación.



Descripción general de la métrica de VQA propuesta.

Como se describe en ese capítulo, el rendimiento del método propuesto se evaluó comparando la SRR con el índice SSIM original ($SSIM_{ot}$) para una gran cantidad de cuadros de video. Se seleccionó una amplia gama de videos, que incluyen cuarenta secuencias de video de diversa duración, resolución y contenido. Las secuencias de video seleccionados, incluyen once secuencias de video de referencia clásicas, dos secuencias de larga duración y veinte y siete secuencias de video no de referencia obtenidos de tráilers de películas. El número total de cuadros únicos que se usaron para evaluar el método propuesto es 60,866.

Las secuencias de video originales sin comprimir se codificaron en tres valores de QP: 12, 22 y 32, que cubren satisfactoriamente el rango de calidad de video típicamente alcanzados en las señales de video codificadas / comprimidas. Los valores más altos de QP no se examinaron, ya que conducen a una calidad de video inaceptable. En la Figura siguiente se muestra una comparación cualitativa inicial entre el índice SRR y SSIM, que muestra la variación del índice SRR y SSIM para cada cuadro de una secuencia de video (Kristen y Sara) con QP = 32. De la figura a continuación es obvia la similitud cualitativa del índice SRR propuesto y SSIM.



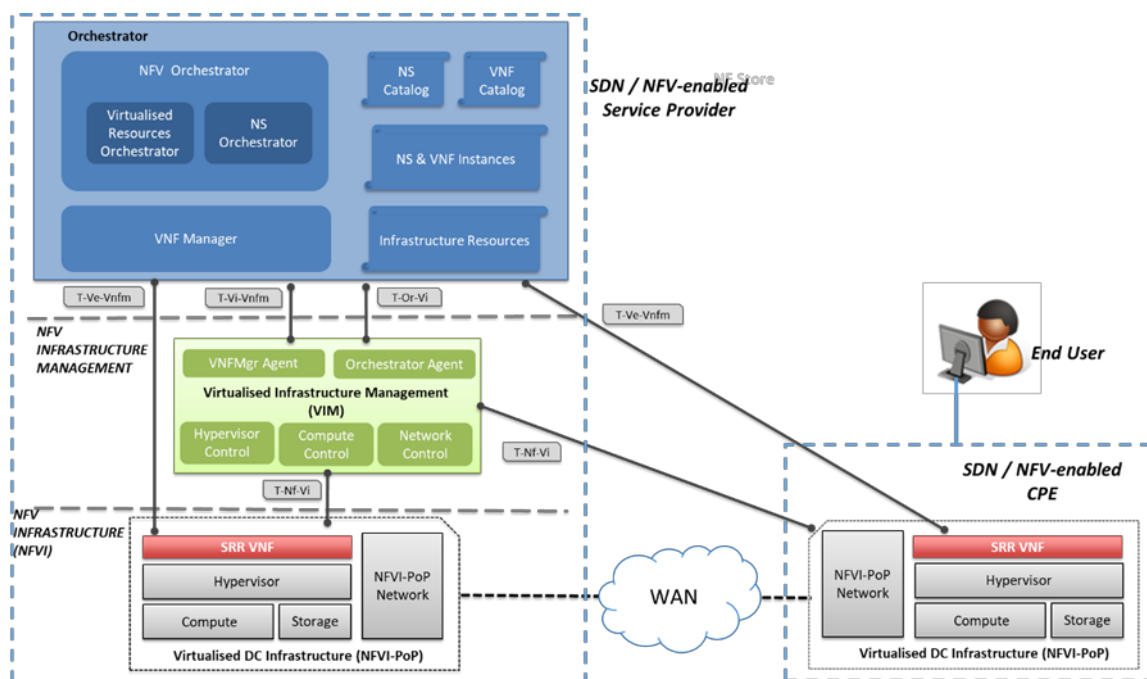
Comparación cualitativa del índice SRR y SSIM aplicado en la secuencia de video de Kristen & Sara con QP = 32.

Para la medición cuantitativa de la actuación del método propuesto, se calculó la desviación promedio absoluta porcentual (MAPD) para cada trama i entre SRR y el índice de SSIM. Los resultados experimentales muestran que la precisión del método propuesto varía de 0.62% (QP = 12) a 2.56% (QP = 32), lo que representa el peor de los casos. Los resultados también muestran que el método de SRR propuesto mantiene un rendimiento satisfactorio en todo el rango potencial de valores de QP, aunque se logra una mayor precisión en valores de QP más bajos. La comparación entre la correlación de QP y la verdad del terreno (es decir, MOS) versus la correlación de los puntajes de SRR y la verdad del terreno proporciona el resultado de 28.65, que muestra la ventaja y el mejor rendimiento de la métrica propuesta.

El rendimiento del método SRR se comparó con el DMOS subjetivo y otros métodos de evaluación, utilizando el Coeficiente de correlación lineal (LCC) de Pearson, el coeficiente de correlación de Spearman (SROCC). Los resultados experimentales muestran que la precisión del método SRR propuesto resulta superior a los métodos RR VQA RR-LHS, J.246 y método de Yang tanto en términos de LCC como de monotonicidad, excepto en la métrica RR, que proporciona mejores resultados. De manera similar, la precisión del método de SRR propuesto es mejor que los métodos de VQA de referencia completa de PSNR y VSNR en términos de LCC, mientras que es ligeramente menor en rendimiento respecto VQM y el índice de SSIM, como se esperaba, debido a la naturaleza de referencia reducida del metodología propuesta. En términos de monotonicidad (SROCC), el método propuesto funciona mejor que el PSNR, pero inferior al resto de los métodos de VQA, aunque sin desviarse significativamente de su rango de rendimiento.

Finalmente, una ventaja importante del método propuesto reside en el bajo bitrate de la señal de información de referencia que debe enviarse al usuario final. Éste oscila entre 400-600 bps, significativamente menor que la tradicionalmente necesaria para otros métodos de RR, de 15 kbps a 256 kbps.

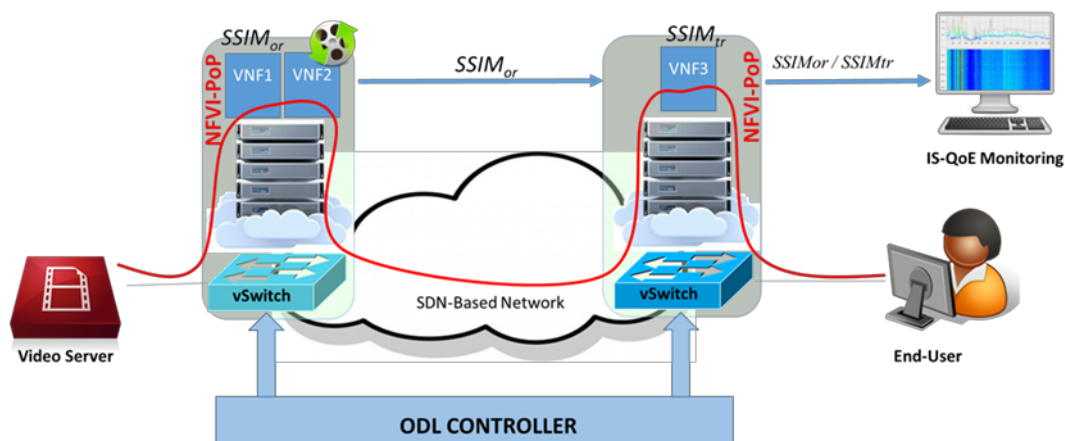
El Capítulo 4 se encarga de examinar cómo se puede aplicar el método SRR propuesto a un entorno de provisión de servicio, combinando la agilidad que ofrecen la virtualización de funciones de red (NFV) y las redes definidas por software (SDN), con el fin de proporcionar "métrica SRR como servicio (SRRaaS)". Inicialmente, se analizan los requisitos específicos, a fin de permitir el despliegue de la SRR propuesta en forma de VNFaaS, a partir de los cuales se propone la siguiente arquitectura de referencia.



Descripción general de la arquitectura de referencia SDN / NFV de referencia

La implementación de SRRaS se realiza de acuerdo con las especificaciones de ETSI NFV, a fin de proporcionar un servicio de red operativo compatible a un sistema habilitado para NFV.

Basado en la arquitectura de referencia anterior se implementó un banco de pruebas de prueba de concepto para realizar Pruebas de Aceptación en Fábrica (FAT), utilizando Openstack, Opendaylight y OVS para desplegar las tecnologías SDN / NFV. El banco de pruebas también debía cumplir los requisitos de la implementación de tres fases del método SRR de la siguiente manera: En la fase I, se evalúa un valor SSIMor inicial en el sitio del proveedor del servicio y se puede implementar como un VNF en ese sitio. En la fase II, se evalúa un valor SSIMtr y en la fase III, la combinación de SSIMor y SSIMtr proporciona la métrica SSR. La Fase II y III pueden implementarse dentro de una VNF en el sitio del consumidor. Por otra parte, la puesta en práctica en el servicio de SRRaaS, requiere que el tráfico debe ser dirigido adecuadamente a través de las VNFs. La arquitectura del banco de pruebas se representa en la siguiente figura. Se compone de dos puntos de presencia (PoPs) NFVI-, uno en la entrada (proveedor de servicios) y el otro en los puntos de salida (usuario final). Cada uno de estos dominios de red integra un switch virtual compatible con SDN (OVS), que está bajo la administración y el control del controlador SDN de OpenDaylight. En ambos puntos de entrada y salida del dominio, una plataforma cloud Openstack instancia las VNFs que soportan los PoPs NFVI.



Una visión general de la topología experimental

La Fase I se implementa en la VNF1, mientras que las Fases II y III en la VNF3. La VNF2 es un transcodificador de video para limitar la tasa de bits del video dentro del ancho de banda del canal de transmisión.

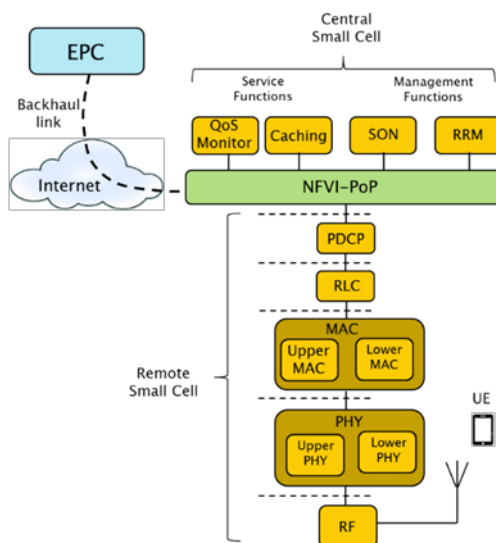
Al introducir tráfico de fondo para provocar la degradación de la calidad del video, se demostró en varios experimentos que el método de SRR propuesto se puede implementar de manera distribuida como un grupo de VNFs, sin que esta división afecte al rendimiento o la capacidad de respuesta del método.

Al ofrecer el método de SRR propuesto como una métrica QoE en servicio (IS- QoE), que se puede implementar como una VNF (VNFaaS) se mejora su aplicabilidad en redes de distribución de video comerciales. Así, SRRaaS se puede instanciar en cualquier punto dentro de la cadena de distribución de video, que es una característica muy importante para los operadores de red ya que les permite encontrar la ubicación dentro de su red, donde se produce una degradación de calidad de video y aplicar acciones correctivas específicas para solucionar el problema.

El Capítulo 5 valida el método SRRaaS , que se probó en el capítulo anterior en un entorno FAT, en una red realista "5G" basada en una arquitectura de small cells para expandir el proceso de validación en las Pruebas de aceptación en Sitio (SAT). Más específicamente, se muestra que el método SRRaaS es adecuado para su implementación como una VNF dentro de una infraestructura de TI ubicada en la proximidad de la small cell. El enfoque propuesto permite la supervisión en servicio de la calidad de video entregada, lo que resulta una herramienta muy útil para los operadores de redes móviles, para monitorizar la satisfacción de sus clientes. Una ventaja del método propuesto, cuando se aplica en una red de este tipo, es que el proceso complejo de evaluación de calidad de video y que consume energía se realiza en el borde de la red, y no en el propio UE, reduciendo significativamente su consumo de energía.

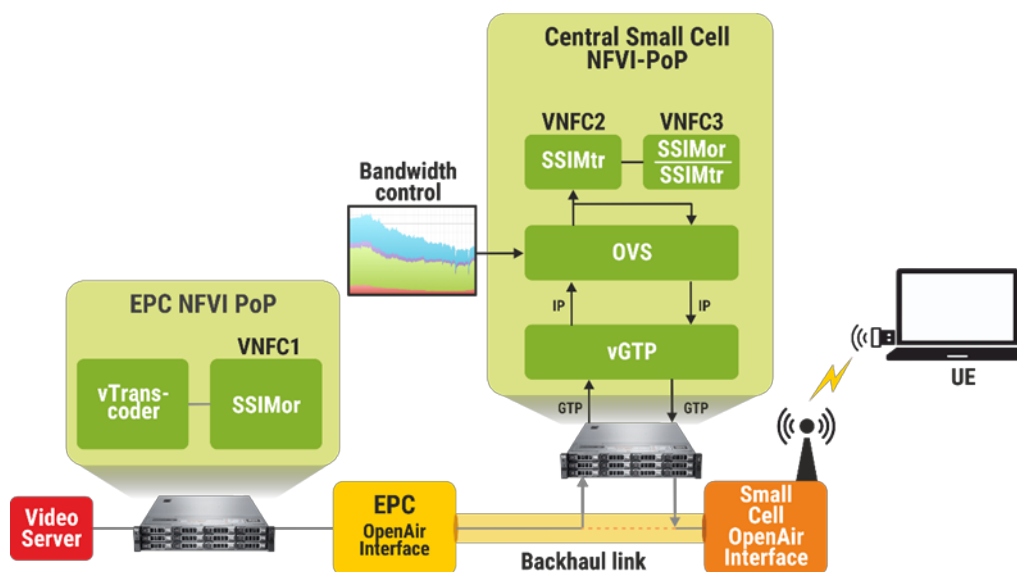
Para la experimentación con SAT, se ha desplegado una plataforma experimental LTE, basada en productos comerciales de Evolved Packet Core (EPC) y Small Cell enfocado a probar y evaluar el rendimiento del método propuesto. Al entorno se han incorporado dos PoPs NFVI-, que alojan a las VNFs, de los cuales se compone el SRR. En el trabajo realizado se analiza también el problema del proceso de (des)encapsulado GTP de los paquetes de datos en el enlace S1 y se plantea una solución.

Inicialmente, se propone una arquitectura de small cells con capacidades NFV (ver figura a continuación), que se basa en el marco general elaborado en el Small Cell Forum (SCF).



Descripción general de la arquitectura general propuesta , basada en el marco SCF

En base a la arquitectura general anterior, se implementó una prueba de concepto para realizar pruebas SAT, utilizando Openstack y Open Virtual Switches, así como una EPC y Small Cells comerciales. El banco de pruebas también debía cumplir con los requisitos de la implementación de tres fases del método SRR como se describió anteriormente y un rutado de tráfico apropiado entre los diversos componentes. En la siguiente figura, se representa la arquitectura implementada. El primer NFVI-PoP estaba ubicado entre el servidor de video y el EPC y el otro era el Central Small Cell NFVI- PoP , ubicado entre el EPC y la small cell remota.



Una visión general del banco de pruebas experimental.

Las fases SRR de SSIMor , SSIMtr y su relación se representan como VNFC1, VNFC2 y VNFC3, respectivamente. También se creó una instancia de OVS en la small cell central NFVI- PoP

El rendimiento de SRR se ha evaluado en una situación de reducción del ancho de banda del backhaul. Los resultados experimentales muestran que el método propuesto es capaz de detectar con éxito la reducción de la calidad del video cuando se degrada el enlace. Otra conclusión es que la reducción de la calidad de video está asociada a la variación de ancho de banda para videos de baja actividad, mientras que no está vinculada a la variación del ancho de banda, para videos de alta actividad. El método de SRR propuesto se puede ofrecer como un servicio a los operadores de redes móviles y les proporciona una herramienta para monitorizar la satisfacción de sus clientes.

El Capítulo 6 presenta las conclusiones de la tesis y las líneas futuras. A este respecto, la tesis presentada muestra cómo el método SRR puede ser instanciado en una infraestructura virtualizada pero no analiza en detalle la interacción de las correspondientes VNFs con las capas superiores de gestión y control del entorno virtualizado. Un trabajo futuro, más allá del alcance de esta tesis, es el desarrollo de los VNF que comprenden SRR para que sean compatibles con la arquitectura MANO.

Otra línea futura es incluir la métrica SRR en un marco de monitorización de NFV más general, que proporcione métricas de VQA a los sistemas de gestión definidos en la arquitectura MANO, que a su vez aplicará las acciones correctivas. Especialmente en redes 5G, las métricas recopiladas de los elementos de radio / nube / software, junto con las métricas de SRR, permitirán una administración más eficiente de los recursos de red 5G y una QoE más alta para los usuarios finales.

Una dirección futura adicional del método SRR propuesto en un despliegue comercial es la provisión en un mecanismo de adaptación de lazo cerrado. Más específicamente, en este enfoque futuro no se permitiría que la calidad del video disminuyera significativamente, a fin de retener la QoE de los clientes. Se aplicaría un mecanismo de adaptación de la tasa de origen, protección contra pérdidas (por ejemplo, FEC) o técnicas de ocultación de errores, para mantener la degradación de la calidad del video lo más pequeña posible. Para adaptarse a tal evento, dentro del propio trabajo del doctorando se realizaron experimentos adicionales, en presencia de un mecanismo de adaptación de velocidad de fuente a través de un transcodificador. Los resultados experimentales preliminares muestran que el método propuesto, cuando se aplica en una arquitectura de circuito cerrado, proporciona mediciones precisas, que logran una degradación de la calidad del video muy pequeña, cuando se aplican los mecanismos de adaptación de la tasa de origen.

Capítulo 7 , se presenta las publicaciones resultados y relación con el trabajo de doctorado, como se muestra en la siguiente tabla.

Capítulo relacionado	Publicación científica
Capítulo 3	Kourtis, M.-A.; Koumaras, H. & Liberal, F. Reduced-reference video quality assessment using a static video pattern, Journal of Electronic Imaging, SPIE-Intl Soc Optical Eng, 2016, 25, 043011
Capítulo 4	Koumaras, H.; Kourtis, M.-A.; et al. In-service Video Quality assessment based on SDN/NFV techniques,

	2016 23rd International Conference on Telecommunications (ICT), IEEE, 2016
Capítulo 4	Kourtis, M.-A.; McGrath, M. J.; et al. T-NOVA: An Open-Source MANO Stack for NFV Infrastructures, IEEE Transactions on Network and Service Management, Institute of Electrical and Electronics Engineers (IEEE), 2017, 14, 586-602
Capítulo 5	Kourtis, M.-A.; Koumaras, H.; Xilouris, G. & Liberal, F. An NFV-Based Video Quality Assessment Method over 5G Small Cell Networks IEEE MultiMedia, Institute of Electrical and Electronics Engineers (IEEE), 2017, 24, 68-78
Capítulo 5	Khodashenas, P. S.; Blanco, B.; Kourtis, M.-A.; Taboada, I.; Xilouris, G.; et al., Service Mapping and Orchestration Over Multi-Tenant Cloud-Enabled RAN, IEEE Transactions on Network and Service Management, Institute of Electrical and Electronics Engineers (IEEE), 2017, 14, 904-919
Capítulo 4	Trajkovska, I.; Kourtis, M.-A.; et al. SDN-based service function chaining mechanism and service prototype implementation in NFV scenario Computer Standards & Interfaces, Elsevier BV, 2017, 54, 247-265

Finalmente , el Capítulo 8 proporciona a modo de anexo una visión general de los conceptos básicos de los métodos de codificación de video que se necesitan para comprender completamente los aspectos tratados en esta tesis.

Laburpena

Multimedia zerbitzuen ugartzeak bideokonferentzia bidezko trafikoa handitu du eta, ondorioz, zerbitzuaren kalitatea ebaluatzeko mekanismo fidagarrien beharra agertu da. Bezeroen eskariaren handitzeak eta azkenaldiko multimedia zerbitzuen arrakasta komertzialak berrikuntza nabarmenak sustatu ditu kodifikazio eta transmisio moldatzaile teknikan eta baita kalitatea ebaluatzeko mekanismoetan ere.

Aldi berean, sakelako multimedia merkatuaren erronkei begira, telekomunikazio operadoreek eta eduki hornitzaileek erabiltzaileen asebetetzea hobetu dezaketen eta, aldi berean, sare modernoetan hedatze erraza bermatzen duten teknologietan zentratu dute beraien interesa. Gaur egun, telekomunikazio sare horietan ematen ari diren joera nagusienak sare funtzio birtualizazioa (NFV Network Function Virtualization) eta 5G sare mugikorren erabilera dira.

NFV paradigma hardware elementuak eta beraien gainean dabilzaten sare funtzioak banatzean datza, eta bere malgutasuna eta eskuratze kostua (CAPEX) eta eragiketa kostua (OPEX) murrizteko ahalmena baieztatzen joan den gisa, jarraitzaileak lortzen joan da

Era berean, sare ingurugiro aniztuna eta "hiper-konektatua" sortzeko joera 5G teknologiarantz bideratzen duen eboluzioa bultzatu du, birtualizazio teknikak erabiliz cloud-irratia arkitektura eta baita sare muturrean intelijentzia hedatzea ere proposatzen duena.

Bideoaren kalitatea ebaluatzeko metodoak (Video Quality Assessment (VQA)) oso tresna baliagarriak dira edukien hornitzaileentzat eta sare operadoreentzat, azken erabiltzaileek kalitatea nola hautematen duten ulertu ahal izateko eta, ulertze horretatik aurrera, zerbitzu berri hipotetikoek aukerak eta erabiltzaile horien eskakizunak betetzeko erabilgarri dauden baliabideen egokitzea aztertzeko.

Erreferentzia osoko ebaluaketa metodoak (Full Reference -VQA-), oro har, erabiltzaile asebetetze neurri ona ematen dute baina ezin dira bideo banaketa sare komertzialetan ezarri arrazoizko modu baten jatorrizko bideo-iturria erabiltzailearen ekipamenduan behar dutelako jasotako edukiekin alderatzeko. Erreferentzia-metodo murriztuak (Reduced Reference -RR-), ordea, ez dira hain zehatzak, baina ingurune komertzialetan ezartzea errazagoa da. Edonola ere, ebaluazio-metodo horiek NFV eta 5G sareetan integratu ahal izango lirarteke.

Tesi honek bideo-kalitatea ebaluatzeko erreferentzia-metodo murriztu berri bat proposatzen du, SSIM (Structural Similarity Index) bezalako erreferentzia osoko metriken prezisioa eta NFV eta/edo 5G sare modernoetan hedatzeko gaitasuna konbinatzen dituena. Horretarako, hasieran bideoaren kalitatea ebaluatzeko RR metodo berri bat aurkezten da. Bere deskribapenaren ondoren, errendimendua beste ebaluazio metodo batzuekin alderatzen da. Hurrengo urratsa NFV ingurune batean integratzea da eta hurbilketa hori nola erabili daitekeen azaltzea. Azkenean, etorkizunean 5G sareetan ezartzeko erronkak aztertuko dira. Ingurune teknologikoak eboluzio jarraian dagoenez, tesi lan hau aldaketa horietara moldatzen den erabiltzaileen QoE ebaluaketa metodo malgu eta eraginkorraren garapenean oinarritzen da. Laburbilduz, tesi honek bideo-kalitatea ebaluatzeko metodo bat ezezik, hedatze

mekanismo autagarrien berrikusketa sakona ere proposatzen du, bere eboluzioa aldaketa horiekin bat datorrela bermatzeko.

1 INTRODUCTION

1.1 CONTEXT AND MOTIVATION

The thriving interest of consumers for video content has brought the world to the universe of digital video provision closer than ever before. Today, content providers are able to offer their content over various types of networks (digital terrestrial TV, satellite, Internet, mobile networks etc) and network operators are able to deploy more efficient and ubiquitous telecommunication systems, in order to confront with the ever increasing traffic volume through their networks. The common target of both content providers and network operators is to keep the end users' satisfaction as high as possible.

The reliable assessment of video quality plays an important role in meeting the promised quality of service (QoS) and in improving the end user's quality of experience (QoE). More specifically, in a video transport system, it is important to monitor the network transport QoS through network QoS parameters, such as packet delay and packet loss rates, as well as the QoS of the video service through video related parameters, including start-up delay of the video playback and video quality, which ultimately contribute to the user's QoE.

Video Quality Assessment (VQA) enables the understanding of how users perceive quality and thus study the feasibility of potential services and adapt the network available resources to satisfy the user requirements. Subjective quality assessment is usually regarded as the most reliable technique to evaluate a multimedia service since it is based on data gathered from opinion tests with human viewers. Subjective VQA methods are very accurate in expressing the end users' satisfaction and they are very useful tools for the video content providers and video encoding vendors, in their effort to provide higher and higher video quality. However, these methods usually involve a considerable amount of resources (mainly time and people) and are considered expensive. Thus, subjective methods are not useful for networks operators as they are impossible to be implemented and exploited in commercial multimedia distribution networks.

Considering the above restriction and the relevance of QoE for video platform services (like Youtube or Netflix - as it can decide the success or fail of these platforms), the scientific community has devoted particular attention to this area, developing objective metrics that automatically evaluate the quality of the video, trying to model the subjective analysis of humans. Furthermore, during the service provision the video stream may need to be dynamically transcoded at different formats/profiles. This may be required because multimedia services need to fit in the current network conditions and the terminal device specifications. However, this in-service transcoding process introduces to the multimedia service a wide variety of encoding impairments that degrade the quality level of the consumed multimedia service. So, it is desired by both

the content providers and the network operators to be able to assess the quality not only during the initial coding process of the source signal, but also across the media service delivery path till the end-user, providing useful feedback for service adaptation actions and also optimal traffic steering decisions.

Objective quality assessment methods are based on mathematical estimators of the quality. They usually require a set of subjective scores to build the estimation model, but then they can compute quality without further human intervention. Objective methods also yield the same quality estimation every time a certain video sample is given as input. This is not true for subjective methods due to the inherent nature of human judgements. Furthermore, these methods are cheaper, faster and easier to be applied on commercial multimedia networks, offering in-service monitoring of the video content.

From the above, it is evident that there is a need for objective VQA methods, which can fulfil the content providers' as well as the network operators' needs for offering high levels of QoE to their customers. Furthermore, with the advent of new technologies like virtualisation, SDN (Software Defined Networking), Network Function Virtualisation (NFV) and new generation mobile networks (5G), the VQA methods have to fulfil additional requirements. They have to be easily deployed in current and future telecommunication networks. Up to now, the telecom operators' infrastructures are based on fixed hardware platforms in their core networks and 3g/ 4G systems in their access networks. Their main concern, until now, was to offer higher bandwidth, better management of network resources and improved coverage and capacity.

However, technology advent has transformed the core networks from hardware platforms to virtualised environments based on SDN and NFV technologies. Also, the access networks are slowly moving towards 5G technology. One of the envisaged key elements of the 5G technological framework is the capability to deliver intelligence directly to network's edge, in the form of virtual network appliances, jointly exploiting the emerging paradigms of NFV and Edge Cloud Computing. 5G network infrastructures need to offer rich virtualisation capabilities and support dynamic processing capabilities on-demand, optimally deployed close to the user. The potential benefits from such an approach trigger the interest of evolving business entities like Communications Service Providers (CSPs), Mobile Network Operators (MNO), Mobile Virtual Network Operators (MVNO) and Over-The-Top (OTT) content and service providers, allowing them to gain an extra share in the network market by pursuing emerging business models. Following this direction, novel business cases will produce added value from any kind of infrastructure or application that has the potential to be offered 'as a Service'.

From the above, it can be concluded that a crucial requirement for a VQA method to become successful in the new telecommunications era, is not only to be accurate, but also to be integrated within virtualisation infrastructures and 5G mobile networks and to be easily deployed in such infrastructures. This thesis proposes a new VQA method that is compliant with the above requirements.

1.2 CHALLENGES AND MAIN OBJECTIVES

The main challenge addressed in this PhD deals with the problem of measuring the quality of encoded video signals over modern network infrastructures in an accurate and efficient manner, while minimizing the required information from the original video signal. It also addresses the problem of implementing an innovative video quality assessment method in a virtualised environment and offer it as a Virtual Network Function (VNF). Finally the thesis addresses the problem of integrating the proposed method over 5G networks, while evaluating its performance in each of the above cases.

More specifically the thesis addresses the following challenges:

- C1. Define an efficient yet accurate objective VQA method:** Most widely used VQA methods are either Full Reference ones, i.e. need to have the original video signal on site to perform the assessment, or Reduced Reference ones, i.e. require a set of features extracted from the original video signal. FR methods are more accurate, but are not appropriate for video distribution networks, since the original signal is not available at the user's site. This thesis will address this challenge by proposing a VQA method that is based on the evaluation of a FR metric (Structural Similarity Index Metric-SSIM), but it is also suitable to be implemented in video distribution networks, since it does not require the original video signal as the user's site, but only a single feature extracted from it.
- C2. Ensure the applicability of the proposed method in NFV/SDN environments.** Most VQA methods do not take under consideration their applicability in virtualised environments, based on NFV/SDN technologies.
- C3. Analyse how multimedia applications should be coupled and interwork with the 5G network components and how to deliver a fully operational 5G-ready NFV enabled system with QoS/QoE capabilities.** The research efforts in currently evolving 5G networks are mainly focused on the required advances in network technologies, such as spectrum, radio access, SDN, NFV, flexible management etc. Less effort has been allocated on the multimedia applications and services that will make use of and exploit advanced 5G network capabilities.

Based on the above challenges, the objectives of this thesis are the following:

Primary objective	Secondary objectives
Objective 1 To define, implement and evaluate an accurate and efficient Reduced Reference video quality evaluation	<ul style="list-style-type: none"> • To define the proposed video quality metric. • To define its fundamental relation to SSIM. • To identify the evaluation metrics and assess its performance. • To compare its performance with relative work in the VQA field. • To perform all the necessary evaluation tests.

<p>metric, based on the calculation of the Structural Similarity Index Metric (SSIM), which is a par excellence FR metric.</p>	
<p><u>Objective 2</u> To implement the proposed SSIM Reduced Reference (SRR) metric as a VNF, suitable to be instantiated in an NFV infrastructure (NFVI), and assess its performance.</p>	<ul style="list-style-type: none"> • To define conversion aspects of the proposed VQA method into a VNF. • To define the functionalities of the required VNFs to implement the proposed method in a virtualized environment. • To identify traffic steering problems between the various VNFs and propose a solution, based on SDN techniques. • To implement the proposed SRR method over an experimental testbed, employing two NFVI Points of Presence (NFVI-PoPs). • To evaluate its performance over the NFV/SDN enabled testbed.
<p><u>Objective 3</u> To implement and evaluate the performance of the proposed SRR method in a 5G network, based on small cells topology, enhanced with virtualisation capabilities, in order to deliver a fully operational 5G NFV enabled system with QoE capabilities .</p>	<ul style="list-style-type: none"> • To propose an energy efficient 5G network architecture that supports VQA at the edge of the network, thus reducing the power consumption of the UE. • To investigate virtualization aspects in 5G networks following the small cell deployment architecture. • To implement the proposed SRR method as a VNF over an experimental testbed, deployed with small cells, which are enhanced with virtualization capabilities. • To evaluate its performance over the above 5G enabled testbed.. • To evaluate its performance in a closed loop architecture, where the SRR metric is used to trigger a source video rate adaptation mechanism.

Thus, Challenge 1 (C1) will be addressed by proposing a VQA method that is based on the evaluation of a FR metric (Structural Similarity Index Metric-SSIM), but it is also suitable to be implemented in video distribution networks, since it does not require the original video signal as the user's site, but only a single feature extracted from it.

Later, Challenge C2 (efficient deployment) will be faced by implementing the proposed method as a VNF, which can be instantiated in a virtualised Network Function Virtualisation Infrastructure (NFVI). Furthermore, being implemented as a VNF the

proposed method addresses the challenge of monitoring the video quality in real-time and across the service provision chain.

Regarding Challenge C3, among the various 5G network architectures that are candidates for the commercial deployment of 5G, this thesis addresses challenges related to how a VQA method, such as the one proposed, can be integrated in a 5G network architecture based on small cells, supported by a virtualised environment. The main reason for this selection is that any VQA method is a complex and power consuming process, so it is better to be performed at the edge of the network, rather than at the UE itself, because it will significantly reduce the impact on UE's battery life. As explained later in section 5, it is equivalent to the corresponding VQA at the UE, because, the main reason for video quality degradation in 5G small cell architectures is due to the unreliable backhaul link over the Internet, and not on the RF link between the UE and the small cell.

1.3 STRUCTURE OF THE DOCUMENT

The document is structured in 6 main sections, organized and related among them as depicted in Figure 1.1.

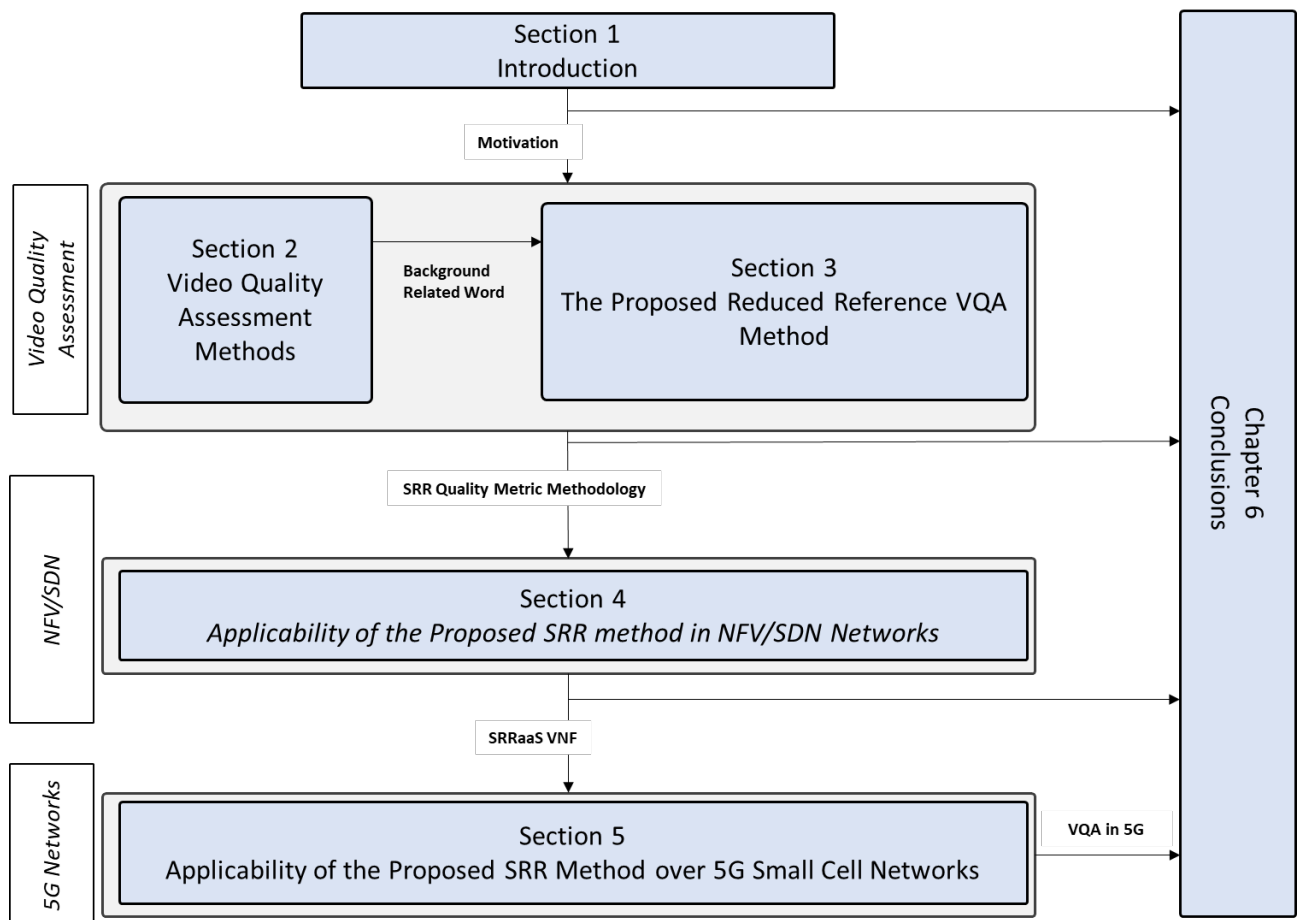


Figure 1.1 - Structure of the document

- Section 1: Introduction

Section 1 states the main motivations of the PhD work, identifies a series of open issues and challenges that need to be addressed and closes with a description of the objectives of the thesis.

- Section 2: Video Quality Assessment Methods

Section 2 provides the literature review of video quality assessment mechanisms. It also presents the categorization of video quality assessment methods and describes the motivation of introducing the proposed algorithm.

- Section 3: The Proposed Reduced Reference Video Quality Assessment Method

Section 3 describes the novel Reduced Reference Video Quality Assessment Method proposed, which is based upon the widely-used Image Quality Metric SSIM. The section also evaluates the performance of the proposed metric and compares it to a set of other VQA algorithms. The experimental process followed is explained in detail. Finally, the bit rate of the reference signal is calculated and its advantage over other Reduced Reference methods is discussed.

- Section 4: Applicability of the proposed SRR method in NFV/SDN networks

This section investigates the applicability of the proposed VQA method in modern telecommunication networks and more specifically, it examines the implementation of the proposed method in an NFV/SDN environment. The experimental testbed is described in detail and specific traffic steering issues are discussed and a solution is proposed. Additionally, the implementation of the proposed SRR method in a virtualized infrastructure gives the ability to perform Factory Acceptance Tests (FATs) for the proposed metric in the designed testbed. Section 4 also evaluates the performance of SRR as an in-service QoE metric, which can be instantiated as a VNFaaS in any point within the video distribution chain and discusses its importance for network operators, for identifying the location within their networks, where video quality is degraded.

- Section 5: Applicability of the proposed SRR method over 5G small cell networks

This section investigates the implementation and performance evaluation of the SRR method in 5G networks, supporting small cells and virtualisation capabilities. The experimental testbed is described in detail and performance of SRR is evaluated for different types of video signals and as a next step in the validation process of the proposed method Site Acceptance Tests (SAT) are performed. The advantage of the

proposed architecture as a monitoring mechanism in Small Cell NFV enabled architectures is discussed in detail.

- Section 6: Conclusions and Future Lines

Finally, Section 6 summarizes the main conclusions of the research activities in this thesis and identifies future research lines. Among, the future steps a set of initial experimental results are presented in the scope of a Forward Error Correction (FEC) mechanism developed on top of the proposed VQA method.

2 VIDEO QUALITY ASSESSMENT METHODS

Currently the available video quality assessment methods are divided into two major categories, the subjective and the objective ones. More specifically, the subjective video quality assessment methods [Seshadrinathan-2010] are based upon the opinion score of a group of viewers regarding the visual degradation of an encoded video sequence compared to the original uncompressed sequence, establishing them as the primary choice for video quality evaluation tests in terms of reliability. The subjective video quality evaluation methods are expensive and time-consuming mainly due to their demanding setup within a controlled room/environment with sophisticated apparatus, which leads to the fact that they cannot be commercially exploited and especially within the service provision chain for monitoring purposes of the delivered video service.

Correspondingly, the objective video quality assessment techniques are mathematical computational models that utilize various image characteristics (e.g. luma, chroma) [Wang-2011; Wang-2006; Wang-2005; Li-2009; Cover-1991; Ma-2011; Soundararajan-2011; Redi-2010; Zeng-2010_a; Zeng-2010_b], or other image statistics (e.g. blockiness [Le-Callet-2010]), in order to approximate as best as possible the subjective test results in an efficient and cost effective way. The objective methods are categorized into three groups determined by their approach and the metric used for the quality assessment: the Full Reference ones (FR), the Reduced Reference ones (RR), and the No-Reference ones (NR).

The FR methods evaluate the video quality by comparing the frames of the original video and the target video. The methods perform multiple channel decomposition of the video signal, where the objective method is applied on each channel, which features a different weigh factor according to the characteristics of the Human Visual System, using Contrast Sensitivity Functions (CSF), Channel Decomposition, Error Normalization, Weighting and finally Minkowski error pooling for combining the error measurements into single perceived quality estimation [Wang-2003]. Also, in the bibliography it has been proposed full reference methods of single channel, where the proposed objective metric is applied on the video signal, without considering varying weight functions. Some full reference metrics that are based on the video structural distortion have been proposed [Wang-2004_b], among which the widely known Structural SIMilarity or SSIM index, which has a very wide range of applicability across many different fields [Gunawan-2003; Gunawan-2008]. All FR methods, including SSIM, provide higher accuracy and credibility in comparison to the rest categories (RR and NR), but in the evaluation process they require both the original and the encoded video sequences at the same site, making them inappropriate for integration in the service provision chain, where the original signal is not available at the end user site.

The RR methods are able to evaluate the video quality level based on metrics, which use only some extracted structural features from the original signal [Wolf-1999; Wolf-2005].

The concept of the RR metrics was introduced by [Wang-2005; Kusuma-2003], where the RR metric was based upon the extraction of various spatial and temporal features of the reference video, which are easily exposed to distortions added by the standard video compression process. RR metrics can be roughly categorized into three categories:

- The first category includes all methods based upon models of low level statistical properties of the original natural image. An RR metric that belongs to this category, is described in [Gunawan-2003], which provides a condensed amount of RR information obtained by the comparison of the marginal probability distribution of wavelet coefficients in different wavelet sub bands with the probability density function of the wavelet coefficients of the decoded signal, while using the Kullback-Leiber divergence as a distance between distributions.
- The second category of RR metrics includes methods that capture visual distortions, in order to quantify the decoded signal's quality [Chono-2008; Carnec-2003; Carnec-2005; Barlow-1961]. However, this type of metrics performs well only, when there is sufficient knowledge about the degradation process the signal underwent. It is not efficient to apply these techniques on general cases, whose distortion has not been previously assumed.
- The third and last category of RR metrics is based upon models of the viewer's perception, for example the HVS [Simoncelli-2001; ITU-T2008; Pinson-2004; Ma-2012]. These models exploit and apply different psychological and psychological vision studies on the end users, in an attempt to imitate the behaviour of subjective test groups.

RR methods are more flexible for in-service integration, since they require only partial information of the original video signal, but they have reduced accuracy and credibility in comparison to the FR metrics.

Finally, the NR methods evaluate the video quality on the basis of processing the frames of the target video, alone. As they do not require any information from the original video sequence, they can be easily integrated within the service provision chain. However, their performance is limited to specific visual artifacts (e.g. tiling), restricting its range of applicability to special cases only. Thus, from the family of the objective methods, the RR and the NR metrics are more suitable for in-service integration, but they suffer from limited efficiency in comparison to the FR, which offer high accuracy, but applicability limitations, since they require the original video sequence for assessing the video quality.

It is evident that RR methods require an additional communication channel within the network architecture, to transmit the extracted features from the video provider site to the end-user terminals. It is obvious that the required bandwidth depends of the specific RR method. A bandwidth in the range of 1kbps to 150kbps is usually required, depending of the RR method and the feature extraction type [Wolf-1999; Razaak-2014]. However, the method described in [Razaak-2014], which requires under 1kbps bandwidth, performs very poorly in terms of absolute difference compared to subjective tests.

An alternative technique is that the extracted features (or in general the reference information) will be encapsulated inside the forward link, along with the video transmission. The more features are extracted and transmitted to the end-user, the more accurate is the objective video quality. However, more features require higher bandwidth of the communication network. So, in RR methods there is a trade-off between the accuracy of the video quality assessment and the constraints in the network bandwidth.

Currently, the evaluation of the video quality is a matter of objective and subjective evaluation procedures, each time taking place after the encoding process. Subjective video quality evaluation processes require large amount of human resources, establishing it as a time-consuming process (e.g. large audiences evaluating video/audio sequences). Objective evaluation methods, on the other hand, can provide video quality evaluation results faster, but require large amount of machine resources and sophisticated apparatus configurations. Towards this, objective evaluation methods are based and make use of multiple metrics, which are related to the content's artifacts (i.e. tiling, blurriness, error blocks, etc.) resulting during the encoding and transmission process. These two categories of video quality evaluation methods will be analyzed and discussed hereby briefly.

2.1 SUBJECTIVE VIDEO QUALITY ASSESSMENT METHODS

The subjective test methods, which have mainly been proposed by International Telecommunications Union (ITU) and Video Quality Experts Group (VQEG), involve an audience of people, who watch a video sequence and score its quality as perceived by them, under specific and controlled watching conditions. Afterwards, the statistical analysis of the collected data is used for the evaluation of the perceived quality. The Mean Opinion Score (MOS) is regarded as the most reliable method of quality measurement and has been applied on the most known subjective techniques.

Subjective test methods are described in ITU-R Rec. T.500-11 (2002) and ITU-T Rec. P.910 (1999), suggesting specific viewing conditions, criteria for observers and test material selection, assessment procedure description and statistical analysis methods. ITU-R Rec. BT.500-11 described subjective methods that are specialized for television applications, whereas ITU-T Rec. P.910 is intended for multimedia applications.

The most known and widely used subjective methods are:

- Double Stimulus Impairment Scale (DSIS) -- This method proposes that observers are shown multiple reference and degraded scene pairs. The reference scene is always first. Scoring is on an overall impression scale of impairment: imperceptible, perceptible but not annoying, slightly annoying, annoying, and very annoying. This scale is commonly known as the 5-point scale with 5 being imperceptible and 1 being very annoying.

- **Single Stimulus Methods** -- Multiple separate scenes are shown. There are two approaches: SS with no repetition of test scenes and SS where the test scenes are repeated multiple times. Three different scoring methods are used:
 - **Adjectival:** the aforementioned 5-grade impairment scale, however half-grades may be allowed.
 - **Numerical:** an 11-grade numerical scale, useful if a reference is not available.
 - **Non-categorical:** a continuous scale with no numbers or a large range, e.g. 0 100.
- **Stimulus Comparison Method:** Usually accomplished with two well matched monitors, where the differences between scene pairs are scored in one of two ways:
 - **Adjectival:** a 7-grade, +3 to -3 scale labeled: much better, better, slightly better, the same, slightly worse, worse, and much worse.
 - **Non-categorical:** a continuous scale with no numbers or a relation number either in absolute terms or related to a standard pair.
- **Single Stimulus Continuous Quality Evaluation (SSCQE)**

According to this method, the viewers watch a program of typically 20–30 minutes without the original reference to be shown. The test program has been processed by the system under test. The subjects/viewers using a slider continuously rate the instantaneously perceived quality on scale from 'bad' to 'excellent', which corresponds to an equivalent numerical scale from 0 to 100.

- **Double Stimulus Continuous Quality Scale (DSCQS)**

At DSCQS the viewers watch multiple pairs of quite short (i.e. 10 seconds) reference and test sequences. Each pair appears twice, with random order of the reference and the test sequence. The viewers/subjects are not aware of the reference/test order and they are asked to rate each of the two separately on a continuous quality scale namely ranging from 'bad' to 'excellent', which corresponds to an equivalent numerical scale from 0 to 100. This method is usually used for evaluating slight quality differences between the test and the reference sequence.

The aforementioned methods are described in the ITU-R Rec. T.500-11 document and are mainly intended for television signals. Based on slight modifications and adaptations of these methods, some other subjective evaluation methods (namely Absolute Category Rating (ACR), Degradation Category Rating (DCR) etc.) for multimedia services are described in ITU-T Rec. P.910. In Table 2.1 a detailed overview of Subjective VQA methods, along with their descriptions, can be found.

Table 2.1 - Overview of the Subjective Quality Assessment Methods

No.	Method	Description
1	Double Stimulus Impairment Scale (DSIS) [ITU-R Rec.BT.500-11]	The reference and test video are shown only once. The experts rate the amount of impairment in a discrete 5 level scale with a range from very annoying to imperceptible. Degradation Category Rating (DCR) recommended by ITU-T Rec.P.910 is a method similar to DSIS.
2	Double Stimulus Continuous Quality Scale (DSCQS) [ITU-R Rec.BT.500-11]	Pair of videos comprising reference video and test video is presented twice. A continuous quality scale of 0-100 ranging from bad to excellent is used.
3	Absolute Category Rating (ACR) [ITU-T Rec.P.910]	The observer will watch the test video without any reference. It is a single stimulus method that uses discrete 5 level scale (bad to excellent). ACR-HR (Hidden Reference) provides a variation of ACR.
4	Stimulus Comparison (SC) or Pair Comparison (PC) [ITU-T Rec.P.910]	The test videos from the same scene but different conditions are paired and the experts make judgement for each pairs.
5	Single Stimulus Continuous Quality Evaluation (SSCQE) [ITU-R Rec.BT.500-11]	The observers view video clip of small duration. Using a slider, the experts provide continuous judgement on perceived quality.
6	Simultaneous Double Stimulus for Continuous Evaluation (SDSCE) [ITU-R Rec.BT.500-11]	Two parallel screens are used by the observer and the quality testing is done by comparing reference and impaired video.
7	Subjective Assessment Methodology for Video Quality (SAMVIQ) [ITU-T Rec. BT.1788]	The assessment video is played back according to the subject's need and pace and the rating will be given instantaneously.

2.2 OBJECTIVE VIDEO QUALITY ASSESSMENT METHODS

The preparation and execution of subjective tests is costly and time consuming and its implementation today is limited to scientific purposes, especially at Video Quality Experts Group (VQEG) experiments.

For this reason, a lot of effort has recently been focused on developing cheaper, faster and easier applicable objective evaluation methods. These techniques successfully emulate the subjective quality assessment results, based on criteria and metrics that can be measured objectively. The objective methods are classified, according to the availability of the original video signal, which is considered to be in high quality.

Most of the proposed objective methods in the literature require the undistorted source video sequence as a reference entity in the quality evaluation process, and due to this, these methods are characterized as Full Reference Methods. The methods perform multiple channel decomposition of the video signal, where the proposed objective method is applied on each channel, which features a different weigh factor according to the characteristics of the Human Visual System. The basic block diagram of the full reference methods with multiple channels is depicted on Figure 2.1.

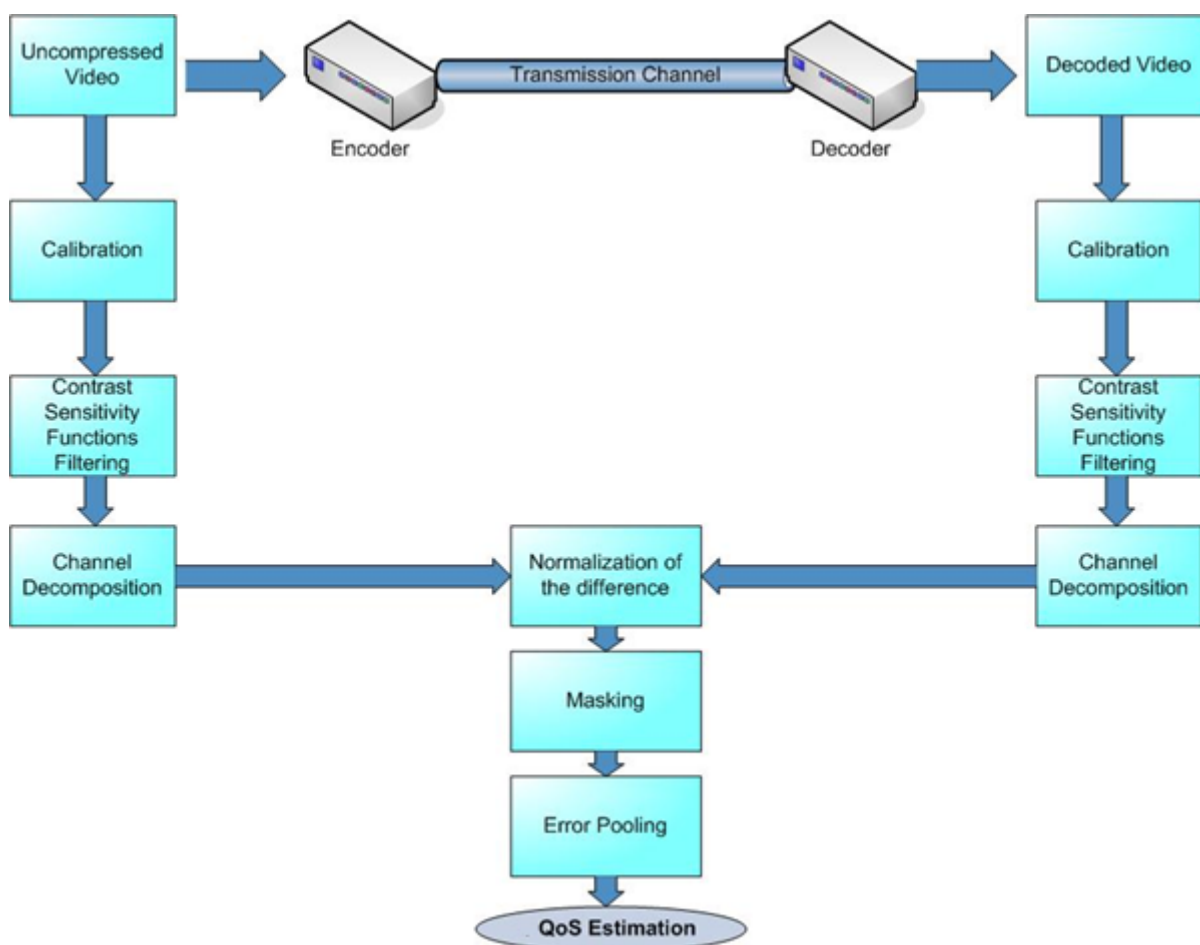


Figure 2.1 - Full Reference Methods with multiple channels

Similarly, in the bibliography it has been proposed full reference methods of single channel, where the proposed objective metric is applied on the video signal, without considering varying weight functions. The block diagram of these methods is depicted on Figure 2.2. However, it has been reported [VQEG-2000; Lu-2002] that these complicated methods do not provide more accurate results than the simple mathematical measures (such as PSNR). Due to this, some new full reference metrics that are based on the video structural distortion, and not on error measurement, have been proposed [Wang-2003].

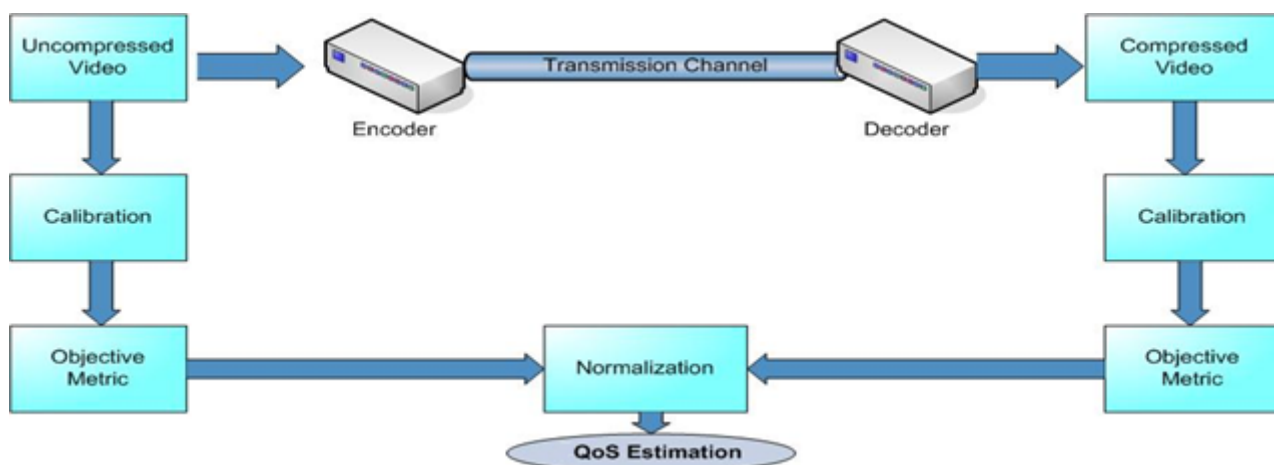


Figure 2.2 - Full Reference Methods with single channel

On the other hand, the fact that these methods require the original video signal as reference deprives their use in commercial video distribution services, where the initial undistorted video signal is not accessible at the end user's site, where the VQA is supposed to take place. This requirement makes the implementation of the Full Reference Methods impractical.

Due to these reasons, later research has been focused on developing methods that can evaluate the video quality based on metrics, which use only some extracted structural features from the original signal (Reduced Reference Methods) [Gunawan-2003]. The block diagram of the reduced reference methods is depicted on Figure 2.3, which shows that the RR methods are designed to assess the perceptual quality of distorted images with only partial information about the reference images. Reduced-reference extracted features, which are used in the evaluation procedure, can take several different form, such as scalar, vector, or matrix. The attractiveness of quality assessment based on reduced-reference approach is the choice of the amount of information required that makes up the reduced- -reference overhead data. This amount can be dictated in practice by the accessible bandwidth of the ancillary channel to transmit the reduced-reference data or similarly by the available storage to cache them. For this purpose, it has stated that the bit rates of the reduced-reference channel could be either zero (no-reference), 15 kbps, 80 kbps, or 256 kbps [VQEG, 2007].

Finally, some methods and techniques have been proposed in the bibliography that do not require any reference video signal (No Reference Methods) [Lu-2002].

Nevertheless, due to the fact that the LTE vision is the provision of audiovisual content at various quality and price levels [Seeling-2004], there is great need for developing methods and tools that will help service providers to predict quickly and easily the video quality level of a media clip. These methods will enable the determination of the specific encoding parameters that will satisfy a certain quality level. All the previously mentioned post-encoding methods may require repeating tests in order to determine the encoding parameters that satisfy a specific level of user satisfaction. This procedure is time consuming, complex and impractical for implementation on the LTE multimedia mobile applications.

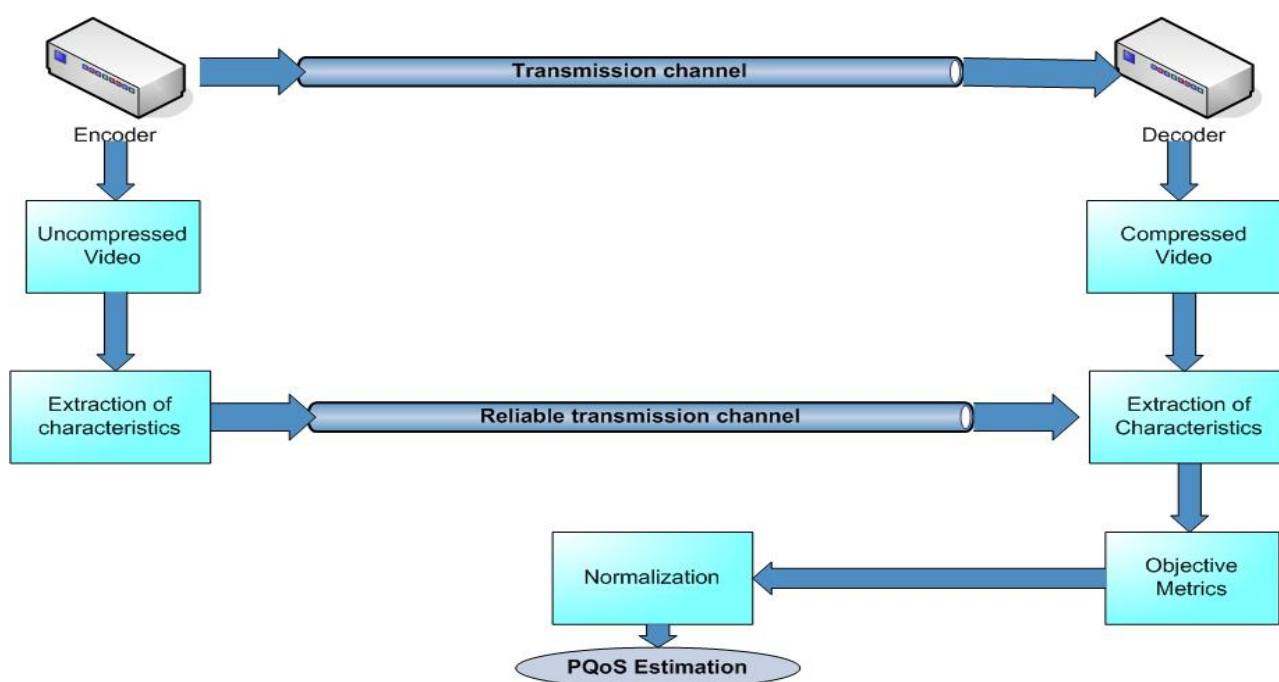


Figure 2.3 - Reduced Reference Methods

Towards this, recently it has been performed research in the field of pre-encoding estimation and prediction of the QoS (Quality of Service) level of a multimedia service as a function of the selected resolution and the encoding bit rate [Koumaras-2004]. These methods provide fast and quantified estimation of the QoS, taking into account the instant QoS variation due to the Spatial and Temporal (S-T) activity within a given encoded sequence. Quantifying this variation by the Mean QoS (MQoS) as a function of the video encoding rate and the picture resolution, it is finally used the MQoS as a metric for pre-encoding QoS assessment based on the fast estimation of the S-T activity level of a video signal.

2.3 APPROXIMATIONS OF FR METHODS BY RR METHODS

The FR metrics provide higher accuracy and credibility in comparison to the rest categories, but they require both the original and the encoded video signals in the evaluation process, making them inappropriate for integrating them in the service provision chain, where the original signal is not available. The RR metrics are more flexible for in-service integration, since they require only partial information of the original video signal, but they have reduced accuracy and credibility in comparison to the FR metrics, which are not generally preferred for in-service use. Respectively, the NR metrics, have limited applicability on special cases, for which the specific NR metric has been developed.

Thus, it has been created the need for developing a new generation of video quality assessment methods that will have a balanced trade-off between the FR-metric accuracy and the RR-metric flexibility. The new trend has initiated a novel research activity, which novel methods have been defined that extend the applicability of widely-used FR metrics (e.g. SSIM) to reduced reference environments.

Toward this direction, [Rehman-2012] is motivated by the success and performance of the FR SSIM metric and attempts to create an RR metric which approximates the SSIM index, without applying the SSIM directly, but by exploiting the Divisive Normalization Transform (DNT) – domain image statistical properties and the algorithm steps of the SSIM metric. This method can be exploited by and be applied on a general-purpose and distortion independent video quality system, as it is based on natural image statistical modelling, which also efficiently represents the signal's content. The efficient content description can reduce significantly the size of RR features transmitted. However, this RR estimate is built on the hypothesis that the image distortion type is fixed, therefore its applicability is limited.

Another similar work [Tagliasacchi-2010], which belongs to the aforementioned second category of RR metrics, and also tries to approximate the SSIM index, uses an entirely different approach by quantifying visual degradations generated by channel transmission errors on the transmitted signal, and then uses distributed source coding (DSC) techniques in order to compress the generated RR features, in order to minimize the required transmission bitrate. However, this technique, as the aforementioned one, cannot be applied on various cases, as it is dependent on the distortion error caused by the transmission.

Therefore, there still need for an efficient approximation of a FR metric (e.g. SSIM), which will facilitate its applicability under various conditions. In this context, the work in this thesis falls within the framework of the latter mentioned works about SSIM approximation, but providing a general-purpose method, which is not limited to the distortion types, as it assesses the required RR feature both content independent and distortion independent, by using the relativity of a monochromatic video signal as a point of reference.

2.4 CONCLUSIONS

This section refers to the classification of VQA methods and analyses their advantages and disadvantages, according to the type of application that they will be used. The first classification is in subjective and objective methods. Subjective methods are very accurate in expressing the end users' satisfaction and the Mean Opinion Score (MOS) is regarded as the most reliable metric of video quality measurement. However, they are expensive (in time and people) and not appropriate for video distribution networks.

The three large groups in objective methods are FR, RR and NR. All categories try to emulate as much as possible subjective ones. One of the most widely known metrics for FR is SSIM index. All FR methods, including SSIM, provide higher accuracy in comparison to RR and NR, but they are not appropriate for integration in the service provision chain, because the original signal is not available at the end user site. On the other hand, RR methods are less accurate, but they are appropriate for video distribution networks, as they require only a few features extracted from the original video, at the end user's site. NR methods are even less accurate than the previous ones, but they do not require any information from the original video signal.

In other words, there is a compromise between accuracy and appropriateness for commercial video distribution networks. This thesis will propose a new RR method based on the calculation of SSIM, (which is a FR metric), but using only a few features extracted from the original video, so it can be categorised as RR. In this way, the proposed method should offer the accuracy of a FR along with the appropriateness to be implemented in a commercial video distribution network.

3 THE PROPOSED REDUCED REFERENCE VIDEO QUALITY ASSESSMENT METHOD

3.1 OVERVIEW OF THE PROPOSED METHOD

From the above chapter it is evident that although FR methods are more accurate as compared to RR methods, they are not appropriate for the distribution of multimedia content over telecommunication systems, such as digital TV and 4G/5G mobile networks. On the other hand, the RR methods are more suitable for in-service integration, since they require only partial information of the original video signal. In this section, a novel RR method is proposed, which is suitable for in-service use.

The proposed method comprises using features based on the evaluation of the SSIM index extracted from both the original and the target video frames. More specifically, the SSIM index is calculated for each frame of the original video using as reference a static white pattern, i.e. a video whose frames are all white (SSIM_{or}) (see Section 3.2 for the rationale of this choice). The SSIM index for each frame is then transmitted to the end-user site. At the end-user terminal the SSIM index is calculated between each frame of the received (target) video and the same static white pattern (SSIM_{tr}). The proposed RR metric is the ratio of the two SSIM indexes, and is depicted in Figure 3.1 below.

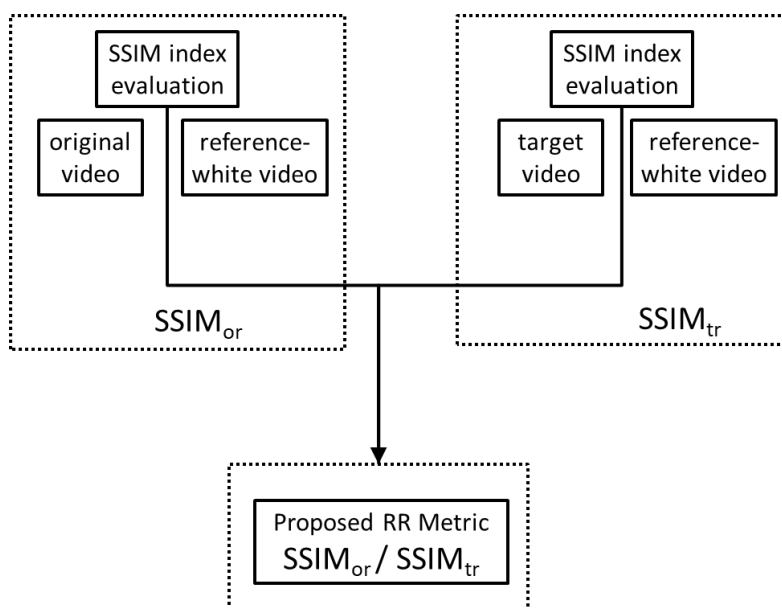


Figure 3.1 - Overview of the proposed VQA Metric

Through experimental measurements, it will be shown that the proposed metric has a value very close to the SSIM index as calculated from the comparison of original and the target video frames, resulting a Mean Absolute Percentage Deviation (MAPD) lower than 2.56 %. Another advantage of the proposed method is the low bit rate reference information signal that needs to be sent to the end-user, which ranges between 400-600 bps.

3.2 FEATURE EXTRACTION USING SSIM

Among the most reliable objective evaluation metrics is the SSIM [Wang-2004], which measures the structural similarity between two image sequences, exploiting the general principle that the main function of the human visual system is the extraction of structural information from the viewing field. If x and y are two video signals, then SSIM is defined as:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (1)$$

where μ_x, μ_y are the mean of x and y , $\sigma_x, \sigma_y, \sigma_{xy}$ are the variances of x, y and the covariance of x and y , respectively. The constants C_1 and C_2 are defined as:

$$C_1 = (K_1L)^2 \quad C_2 = (K_2L)^2 \quad (2)$$

where L is the dynamic pixel range and $K_1 = 0.01$ and $K_2 = 0.03$, respectively.

In the typical SSIM index evaluation process, it is assumed that both the original and the target video sequences are available at the same site, as shown in Figure 3.2. $SSIM(x,y)$ evaluation for every frame can be based on any software implementation of equation (1), where x is the original video sequence VSo in Figure 3.2, y is the target video sequence VSt and $SSIMot$ is their $SSIM(x,y)$ index.

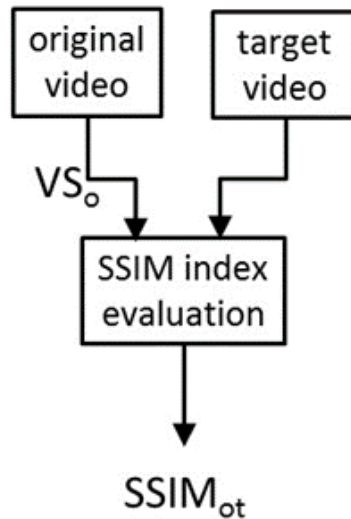


Figure 3.2 - Typical SSIM index evaluation with original and target video sequences available at the same site

According to [Wang-2004], where SSIM index is defined and introduced, SSIM comprises of three image characteristics components, the luminance, the contrast and the structure comparison. Their combination results in the widely used SSIM index. The three separate comparison components are:

$$l(x, y) = \frac{2(1+R)}{1+(1+R)^2 + \frac{c_1}{\mu_x^2}} \quad (3)$$

$$c(x, y) = \frac{2\sigma_x\sigma_y + c_2}{\sigma_x^2 + \sigma_y^2 + c_2} \quad (4)$$

$$s(x, y) = \frac{\sigma_{xy} + c_3}{\sigma_x\sigma_y + c_3} \quad (5)$$

The combination of (3) luminance, (4) contrast and (5) structure comparisons results in the SSIM index formula:

$$SSIM(x, y) = [l(x, y)]^\alpha * [c(x, y)]^\beta * [s(x, y)]^\gamma \quad (6)$$

According to the SSIM index formula, the parameters α , β and γ adjust the relative importance of each of the three components in the calculation of the SSIM, but for simplicity reasons the authors of [Wang-2005] have selected the case that $\alpha=\beta=\gamma=1$, which results to the well know expression of the SSIM index. However, the decision that

the three components participate equally in the final calculation of SSIM index is not sufficiently justified by the authors in [4] and it seems that it is a decision taken only for simplicity reasons.

This motivate us to further research on the sensitivity analysis of the SSIM accuracy under different α , β , γ weights. More specifically, considering the purpose of this chapter to develop a novel and flexible metric suitable for in-service applicability, we notice that parameter γ specifies the importance of the structural similarity between signals x and y , which is the most influential factor for the FR requirement in the applicability of the SSIM, since according to the following type, the σ_{xy} factor is needed to be calculated for the measurement of equation (5).

In the proposed method, in order to investigate the in-service applicability of the SSIM (i.e. in in-service cases that the calculation of equation (5) is not feasible due to the lack of the reference signal), we research in this chapter the case that $\gamma \rightarrow 0$, so the relevant importance of equation (5) in the SSIM calculation is limited and therefore the requirement for the reference signal to be available together with the encoded signal for the estimation of equation (5) stops to exist, allowing the decomposition of the SSIM index to exclusively the equation (3) and equation (4) parameters, where does not exist any parameters requiring the existence of both the reference and the encoded signal at the same place.

Based on this analysis, in the proposed method (see Figure 3.3), the SSIM index is used as a tool, in order to extract features from both the original and the target video sequences, using a reference video pattern. In this respect, an initial SSIMor value is evaluated for every frame at the service provider site, by comparing the original video sequence (VSo) with a video reference pattern (VSr), e.g. a video sequence of static single color video frames of the same resolution and frame rate, which is artificially generated.

The initial color choice was made based on the fact that SSIM index formula is comprised by 3 components, as shown in equation (6): the luma, the contrast and the structure of a video frame. Out of the three components, luma is the most influential one for the HVS [Lucas-2017] and also it has a discrete arithmetic representation scale (ranging from 0 to 255 in the YUV color space). Therefore, luma is the most appropriate candidate for producing the static reference pattern due to its simplicity in reproduction by an imaging software. Among the various luma values, the selection of 255 (white color) was the initial option for the experimental process, because it is the highest luma value and is equal to the maximum dynamic pixel range L as shown in equation (2).

The evaluation of SSIMor can be based on any software implementation of equation (1), where x is VSo and y is VSr and refers to each frame. SSIMor, can be considered as a feature of each frame of the original video and is sent to the end-user site by any means, i.e. either through the same communication channel as the video, or any other communication channel with a sufficient bandwidth, or by embedding it inside the transport stream of the video sequence. In any case, it is considered that the SSIMor value is recovered at the end-user site.

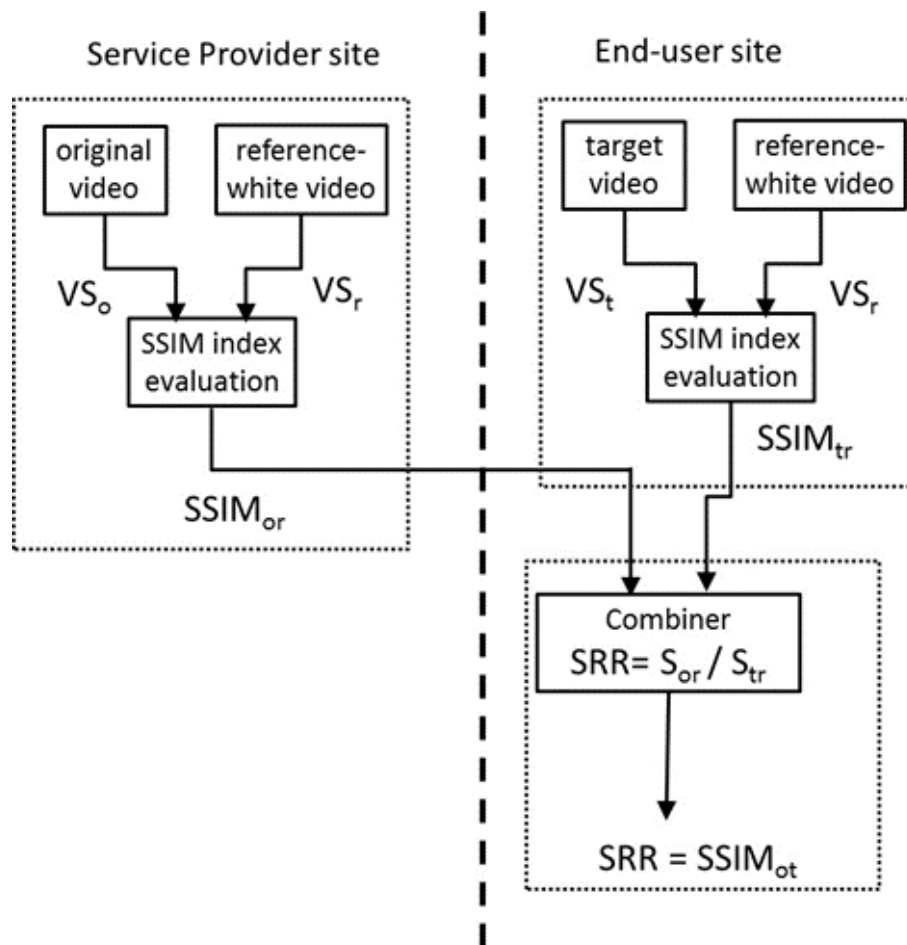


Figure 3.3 - SRR evaluation method using a reference video pattern as reference at both the service provider and end-user sites

Referring to Figure 3.3, an SSIM_{tr} value is evaluated at the end-user site, by comparing the received (target) video signal (VS_t) with a reference video pattern (VS_r), which is identical to the one used at the service provider site and is also artificially generated at site. The evaluation of SSIM_{tr} is based on the same software implementation of equation (1), used at the transmitter site and refers to each frame of the target video sequence. SSIM_{tr} can be considered as a feature of each frame of the target video.

The ratio of these two SSIM values can be considered as a new metric, based on SSIM, namely SSIM Reduced Reference (SRR), i.e.:

$$SRR = SSIM_{or} / SSIM_{tr} \quad (7)$$

Comparing SRR with the SSIM index between the original and the target video sequences, experimental results in section 3.3 show that SRR approximates efficiently SSIM, with a mean absolute percentage deviation less than 2.56% and satisfactory correlation coefficient values with subjective Mean Opinion Scores (MOS).

Concerning the channel requirements and the overhead of the proposed method, SSIM_{or} is a number less than 1 and it can be represented by 2 bytes per frame for an

accuracy of four decimal places (10^{-4}). In this case, the required bit rate to be transmitted to the end-user site is 400 bps per 25 frames/sec. For an increased accuracy per frame of six decimal places (10^{-6}), 3 bytes are required, resulting 600 bps for the SSIMor. Even 600 bps is significantly lower than the value of approximately 15kbps to 256kbps, required for other RR methods, as mentioned in section 2. [VQEG, 2007]

3.3 QUALITATIVE INTERPRETATION OF THE PROPOSED VIDEO QUALITY METRIC

In order to interpret the proposed quality metric and its relation to the SSIM index, we consider a video sequence, which is encoded at various QP values, resulting to different bit rates and quality levels. The SSIM index between the original and the target video sequence for each frame is denoted as $SSIMot(QP)$, which is a descending function of QP. Given that the maximum value of $SSIMot$ is equal to 1, the SSIM variation is an exponential function as shown in [Koumaras-2007_a]. Therefore:

$$SSIMot(QP) = e^{-\alpha * QP} \quad (8)$$

where α is a coefficient that depends on the content of the video signal [Koumaras-2007_b].

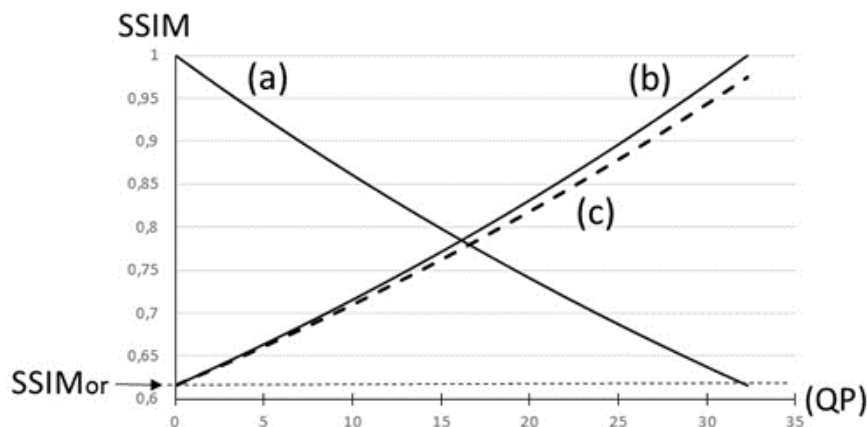


Figure 3.4 - Variation of : (a) SSIMot (b) ideal SSIMtr (c) real conditions SSIMtr vs QP

The plot of $SSIMot(QP)$ vs QP is shown in Figure 3.4, curve (a). As QP increases, more and more information from the original video frame is lost, i.e. the spatial information (i.e. color and intensity of each pixel) of VSt is lost. In the extreme case where the information is completely lost, all pixels are equal in color and density, i.e. the VSt is degraded to a sequence of uniform frames, as for example a white video pattern (similar to the used reference video pattern). In this extreme case, the lowest value of $SSIMot(QP)$ is the SSIM index between the original video sequence and the reference video pattern, which can be denoted as SSIMor

$$\min[SSIMot(QP)] = SSIMor \quad (9)$$

The previous analysis refers to the comparison between the original and the target video frames. Respectively, if we consider the comparison between the target video frames and the reference video (a white video pattern earlier explained), the $SSIM_{tr}(QP)$ index vs QP will be an ascending function. This is because at low QPs the target frame is very close to the original and therefore will greatly defer from the (white) reference video pattern, resulting a low value $SSIM_{tr}$. In the extreme case of very low QP the target video frame is identical to the original and therefore :

$$\min[SSIM_{tr}(QP)] = SSIM_{or} \quad (10)$$

Furthermore, as QP increases the target frame will more and more resemble to the reference white video pattern, resulting a higher value of $SSIM_{tr}$. In the ideal case, for very high QP the target frame is identical to the reference one and the maximum $SSIM_{tr}$ equals 1. Considering that the variation of $SSIM_{tr}$ is an exponential function of QP as above and also is symmetric to $SSIM_{ot}$ it can be deduced that :

$$SSIM_{tr}(QP) = SSIM_{or} * e^{a*QP} \quad (11)$$

The plot of $SSIM_{ot}(QP)$ vs QP is shown in Figure 3.3, curve (b).

From equations (4) and (5) it can be deduced that:

$$SSIM_{ot}(QP) * SSIM_{tr}(QP) = SSIM_{or}$$

Or

$$SSIM_{ot} = SSIM_{or}/SSIM_{tr} \quad (12)$$

Comparing equations (7) and (12) it is evident that :

$$SRR = SSIM_{ot} \quad (13)$$

i.e. in the ideal case, the proposed SRR equals the $SSIM$ index. However, in practical situations $SSIM_{tr}$ differs from equation (7). Although its minimum value is $SSIM_{or}$, its maximum value is not equal to 1 because even for high QP the degraded frame is not identical to the white reference one. As an example, the plot of $SSIM_{ot}$ in a practical situation is shown in curve (c) of Figure 3.4. This deviation will affect the performance of the proposed method, and SRR will differ from $SSIM$ index by an amount that depends on the difference between curves (b) and (c) of Figure 3.4.

3.4 PERFORMANCE COMPARISON OF SRR TO SSIM

The performance of the proposed method is evaluated by comparing SRR with the original $SSIM$ index ($SSIM_{ot}$) for a large number of video frames. In this respect, a wide range of video sets were selected, which include forty video sequences of various length,

resolution and content. The selected video sequences include eleven reference video sequences, two long duration sequences (BigBuckBunny, Elephant's Dream), and twenty-seven non-reference video sequences retrieved from movie trailers, representative frames can be seen in Figure 3.5. The total number of unique frames, which were used for evaluating the proposed method is 60,866.



Figure 3.5 - Representative Frames of the 40 test signals of the evaluation process

The original uncompressed video sequences were encoded at three QP values: 12, 22 and 32, which cover satisfactorily the achieved video quality range of the encoded/compressed video signals. Higher QP values were not examined, because they lead to unacceptable video quality [Razaak-2014]. An initial qualitative comparison between SRR and SSIM index is depicted in Figure 3.6, which shows the variation of SRR and SSIM index for each frame of a video sequence (Kristen&Sara) with QP=32. From Figure 3.6 it is obvious the qualitative similarity of SRR and SSIM index.

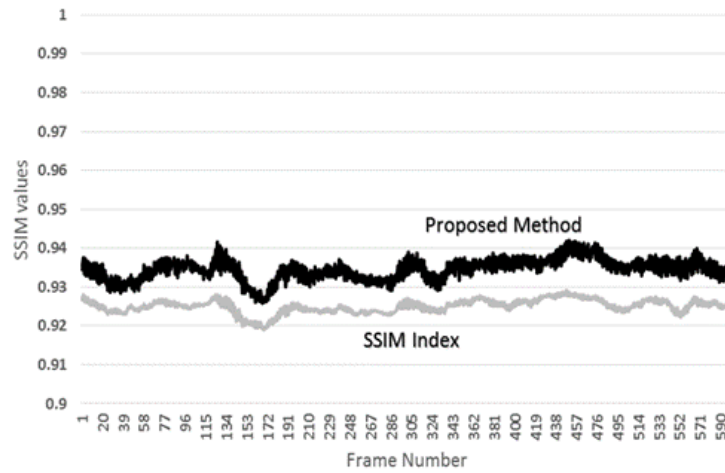


Figure 3.6 - Qualitative comparison of SRR and SSIM index applied on Kristen&Sara video sequence with QP=32

For the quantitative measurement of the performance of the proposed method, the Mean Absolute Percentage Deviation (MAPD) for each frame i between SRR and SSIM index is calculated. MAPD is a widely used metric for measurement of the accuracy of a prediction method, specifically in trend estimation, like the proposed one.

$$MAPD = \frac{1}{n} \sum_{i=1}^n \frac{|SSIM_i - Predicted_SSIM_i|}{SSIM_i} \quad (14)$$

Where $SSIM_i$ is the SSIM index per frame i and $Predicted_SSIM_t$ is the SRR value for the frame t , according to equation (7).

Table 3.1 presents the MAPD for the experimental set of the 40 video sequences at the three QP values (i.e, 12, 22 and 32) and also the mean value of MAPD for each QP.

Table 3.1 - MAPD for SRR and SSIM index for a white-video reference pattern

Video Name	Resolution	Frames	QP:12	QP:22	QP:32	QP:42
apocalypto1	352x288	990	0.003053	0.005581	0.033561	0.087794
apocalypto2	352x288	990	0.004793	0.005936	0.020171	0.055745
apocalypto3	352x288	990	0.005903	0.007334	0.021447	0.074754
apocalypto4	352x288	501	0.005136	0.009094	0.017331	0.092961
mission1	352x288	990	0.005578	0.007233	0.016888	0.065433
mission2	352x288	990	0.006228	0.006695	0.019181	0.053884
mission3	352x288	293	0.004423	0.005531	0.016902	0.077755
superman1	352x288	990	0.004448	0.006062	0.019283	0.075533

superman2	352x288	990	0.00459	0.0069	0.021657	0.043845
superman3	352x288	268	0.001328	0.003436	0.035995	0.053591
insidemana1	352x288	990	0.003798	0.004191	0.019316	0.066451
insidemana2	352x288	990	0.006988	0.008945	0.012651	0.031442
insidemana3	352x288	990	0.00526	0.006288	0.012146	0.04527
insidemana4	352x288	376	0.00125	0.001084	0.027495	0.050189
davinci1	352x288	990	0.0029	0.003661	0.014045	0.054189
davinci2	352x288	990	0.005846	0.007804	0.019533	0.043835
davinci3	352x288	990	0.005416	0.005909	0.012271	0.036922
davinci4	352x288	627	0.007838	0.009838	0.018827	0.044248
basic1	352x288	990	0.005603	0.006751	0.015259	0.06152
basic2	352x288	990	0.00652	0.009051	0.015006	0.048394
basic3	352x288	990	0.006734	0.00765	0.014208	0.040948
basic4	352x288	351	0.003842	0.004624	0.010905	0.063068
16blocks1	352x288	990	0.004517	0.004992	0.017711	0.044576
16block2	352x288	990	0.006962	0.006389	0.012574	0.035462
16block3	352x288	990	0.005295	0.005044	0.01225	0.05988
16block4	352x288	451	0.003401	0.00297	0.007796	0.038091
batman1	352x288	2659	0.010258	0.014053	0.042584	0.088853
batman2	352x288	913	0.00619	0.011052	0.070104	0.076392
bigbuckbunny	640x360	14315	0.01091	0.020527	0.039061	0.08247
elephantsdream	640x360	15691	0.008501	0.013748	0.040883	0.07121
basketballpass	416x240	501	0.011807	0.007371	0.022277	0.052565
Bqsquare	416x240	601	0.006849	0.028219	0.101372	0.040488
Bubbles	416x240	501	0.005385	0.015653	0.069205	0.045971
basketballdrill	832x480	501	0.016819	0.00973	0.0052	0.07707

Bqmall	832x480	601	0.008256	0.005319	0.009498	0.044177
Racehorses	832x480	300	0.009953	0.00794	0.022274	0.03739
Partyscene	832x480	501	0.005572	0.01613	0.075519	0.028829
Stockholm	1280x720	604	0.000514	0.030779	0.047909	0.037135
Kristen&Sara	1280x720	600	0.008666	0.006363	0.010176	0.036755
Foupeople	1280x720	600	0.009949	0.001049	0.001995	0.039007

Table 3.1 shows that the accuracy of the proposed method ranges from 0.62% (QP=12) to 2.56% (QP=32), which represents the worst-case performance, showing that the proposed SRR method maintains satisfactory performance across all the potential range of QP values, although better accuracy is achieved at lower QP values. The comparison between the correlation of QP and ground truth (i.e. MOS) versus the correlation of SRR scores and ground truth provides the result of 28.65, showing the advantage and the better performance of the proposed metric.

The experimental variation of MAPD vs QP is depicted in Figure 3.7(a), where the trend line is the dashed line. For comparison reasons, the theoretical MAPD is also depicted in Figure 3.7(b), which follows an exponential form as expected [Koumaras-2007_a].

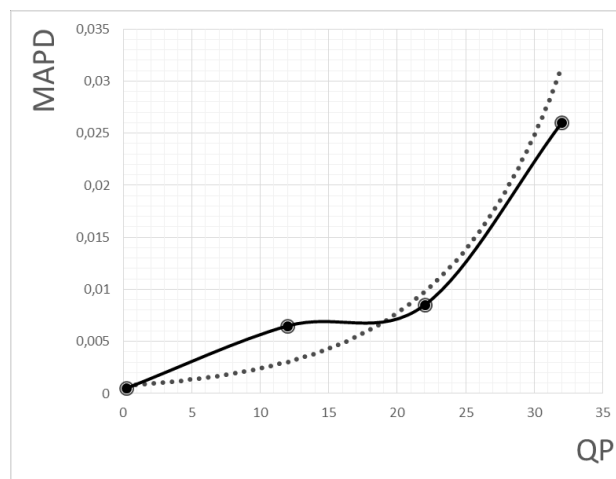


Figure 3.7 - Variation of MAPD vs QP: (a - solid line) Experimental, (b - dotted) Theoretical

It is calculated from equation (12), where $SSIM_i$ equals $SSIM_{ot}(QP)$, as calculated from equation (8). Also, Predicted_SSIMt equals SRR, i.e.:

$$SRR = \frac{SSIM_{or}(QP)}{SSIM_{tr}(QP)} \quad (13)$$

where SSIMtr(QP) corresponds to curve (c) of Figure 3.4, which refers to an exemplary practical case.

The slope of the two lines of Figures 3.7(a) and 3.7(b) are the same, which shows that the theoretically calculated SRR is very close to the experimental results.

3.5 PERFORMANCE COMPARISON OF SRR TO SUBJECTIVE DMOS AND OTHER ASSESSMENT METHODS

According to the video quality experts group (VQEG) research [42], in order to obtain a linear relationship between an objective assessment method score and its corresponding subjective score, each metric score x is mapped to $q(x)$. The non-linear best fitting logistic function $q(x)$ is given by the equation (14):

$$q(x) = \beta_1 \left(\frac{1}{2} - \frac{1}{1 + \exp(\beta_2(x - \beta_3))} \right) + \beta_4 x + \beta_5 \quad (14)$$

The parameters $\{\beta_1, \beta_2, \beta_3, \beta_4, \beta_5\}$ are calculated through minimizing the sum of squared differences among the subjective and the mapped scores. In order to compare the performance of a newly proposed SRR method with the existing ones, performance evaluation metrics are used, such as the Pearson's Linear Correlation Coefficient (LCC), which is the linear correlation coefficient between the predicted MOS and subjective MOS. LCC is a measure of prediction accuracy of an objective assessment metric, i.e., the capability of the metric to predict the subjective scores with low error. The LCC can be calculated via equation (15):

$$LCC = \frac{\sum_{i=1}^{M_d} (q_i - \bar{q})(s_i - \bar{s})}{\left(\sum_{i=1}^{M_d} (q_i - \bar{q})^2 \right)^{\frac{1}{2}} \left(\sum_{i=1}^{M_d} (s_i - \bar{s})^2 \right)^{\frac{1}{2}}} \quad (15)$$

Where s_i and q_i are the subjective score and the mapped score for the i th frame of a video of size M_d respectively, and \bar{s} and \bar{q} are the means of the mapped scores and subjective scores respectively. A good objective assessment metric is expected to have high LCC (close to 1) in contrast to MAPD, which should have low values (i.e. close to 0), as shown in previous subsection.

Moreover, the Spearman Rank Order Correlation Coefficient (SROCC), which measures the monotonicity of the proposed method against subjective human scores, was also applied. SROCC is a nonparametric measure of statistical dependence between two variables, which assesses how well the relationship between two variables can be described using a monotonic function. If there are no repeated data values, a perfect Spearman correlation (equal to 1) occurs when each of the variables is a perfect monotone function of the other.

Therefore, in order to evaluate the performance of the SSR method and derive the LCC, following the aforementioned methodology, the LIVE Video Quality dataset [Seshadrinathan-2010] was used. The LIVE Video Quality Database uses ten uncompressed high-quality videos with a wide variety of content as reference videos. A set of 150 distorted videos were created from these reference videos (15 distorted videos per reference) using H.264-based compression. Then each degraded video in the LIVE Video Quality Database was assessed by 38 human subjects in a single stimulus study with hidden reference removal, where the subjects scored the video quality on a continuous quality scale. The mean and variance of the Difference Mean Opinion Scores (DMOS) obtained from the subjective evaluations.

Scatter plot of proposed objective SRR scores vs. DMOS for all H.264 videos in the LIVE VQDB is shown in Figure 3.8 along with the best fitting logistic function.

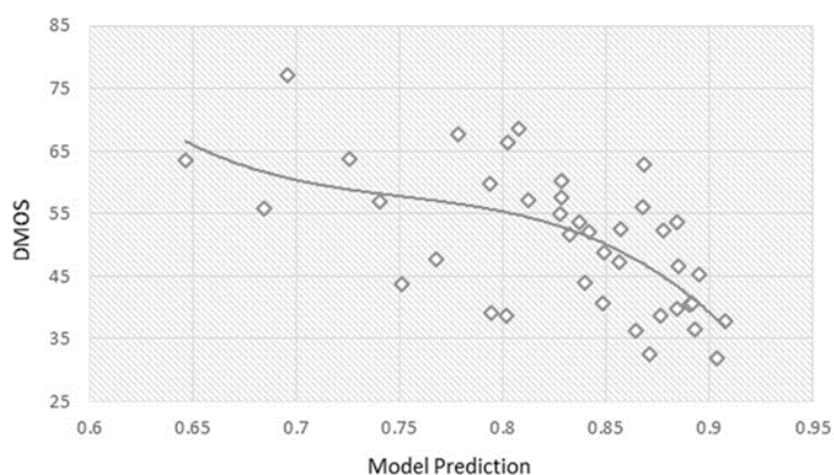


Figure 3.8 - Scatter plot of objective SRR scores vs. DMOS v with the best fitting logistic function

The SROCC and the LCC are computed between the objective scores and the subjective scores. Table 4.2 shows the performance of the proposed SRR model against other VQA methods, both FR and RR, in terms of the SROCC and LCC.

Table 3.2 - Comparison of the performance of VQA algorithms – LCC and SROCC

VQA Method	Type	LCC	SROCC
RR-LHS	RR	0.4557	0.4082
J.246	RR	0.4488	0.4157
PSNR	FR	0.5493	0.4585
Yang's RR VQA	RR	0.5654	0.5366
VSNR	FR	0.6216	0.6460
Proposed SRR Method	RR	0.6260	0.5862
VQM	FR	0.6459	0.6520
SSIM	FR	0.6656	0.6514
RR metric	RR	0.7567	0.7486

According to Table 3.2, the accuracy of the proposed SRR method is measured better than the RR VQA methods RR-LHS [Gunawan-2008], J.246 [ITU-T2008], Yang's RR VQA [Yang-2011] both in terms of LCC and monotonicity (SROCC), except from RR metric [Ma-2012], which provides better results. Similarly, the accuracy of the proposed SRR method is better than the PSNR and VSNR [Chandler-2007] full reference VQA methods in terms of LCC, while it is slightly lower than the performance of VQM and SSIM index, as expected, due to the reduce reference nature of the proposed methodology. In terms of monotonicity (SROCC), the proposed method performs better than the PSNR, but lower than the rest VQA methods, without however significant deviating from their performance range.

In addition to the above mathematical analysis a more qualitative evaluation has also been elaborated, based on a visual presentation of the correlation between SSIM and the proposed SRR method. Figures 3.9, 3.10 and 3.11 show the graphical presentation of SSIM and SRR vs the frame numbering, for a set of 3 video signals, Basketball Pass, BQSquare and BQMall, respectively. The video signals were encoded at 3 different QP values 12, 22 and 32 so as to cover a wide variety of visual degradations. In all figures, it is evident that the lower the QP (higher SSIM), the better accuracy of the SRR method. Another observation is that SRR follows the variations of SSIM for both low (e.g. figure 3.9 QP22 and figure 3.11 QP12) and high values of QP (e.g. figure 3.9 QP32 and figure 3.11 QP32). Finally, it is shown that SRR follows very reliably the abrupt variations of SSIM (e.g. figure 3.9 QP32 and figure 3.11 QP32), as well as the wide variations (figure 3.10 QP32 and figure 3.11 QP12), independently of the video quality.

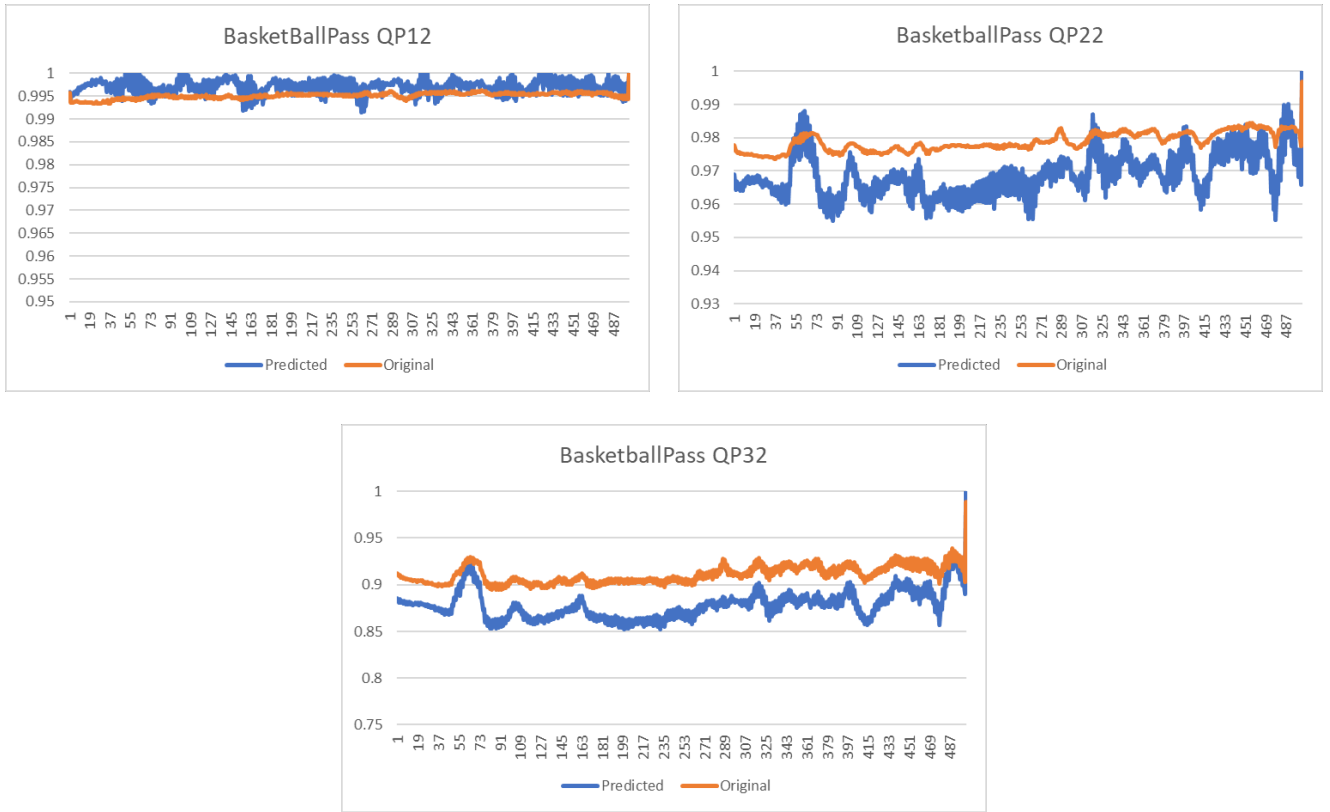


Figure 3.9 - Basketball Pass SSIM vs SRR for QPs 12-32

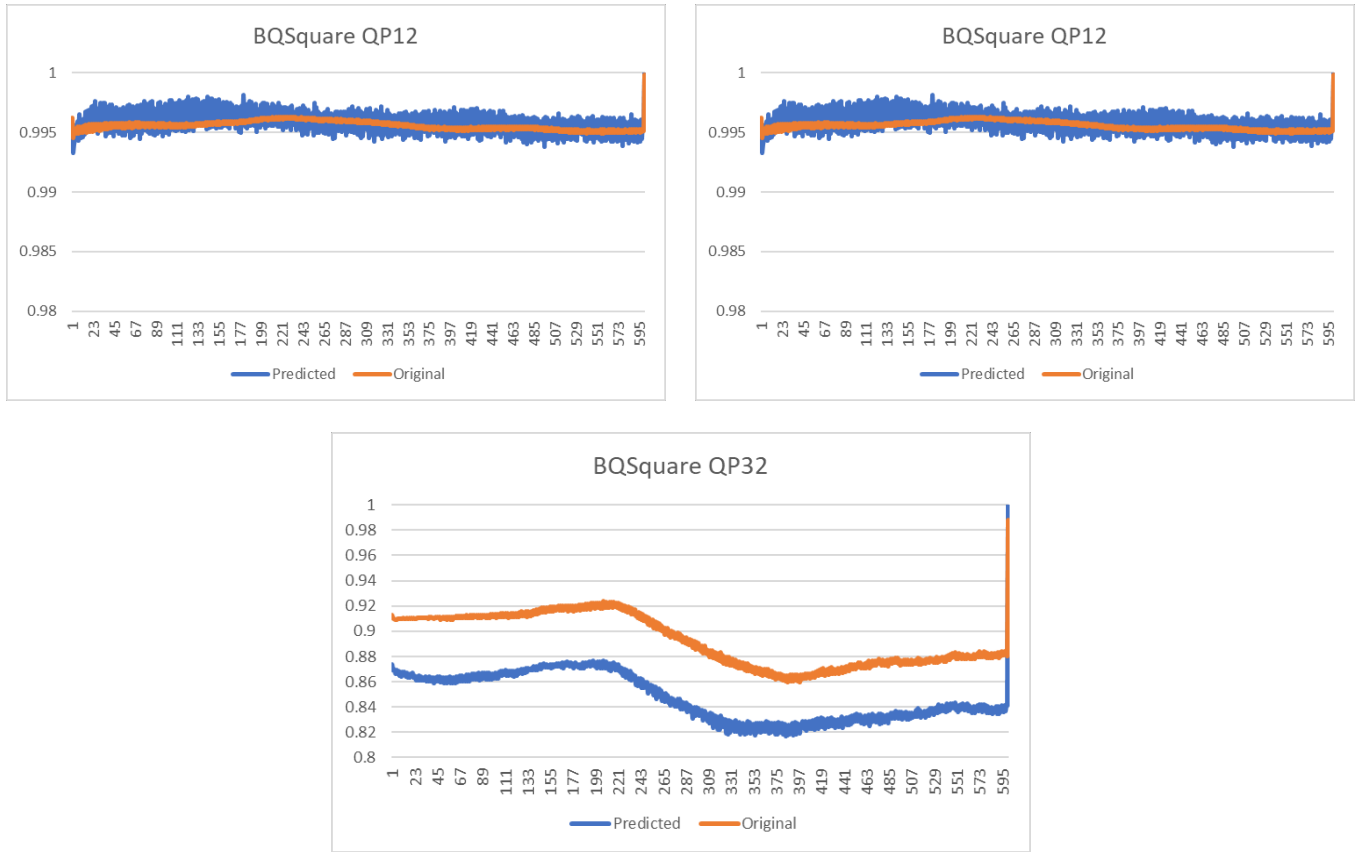


Figure 3.10 - BQSquare SSIM vs SRR for QPs 12-32

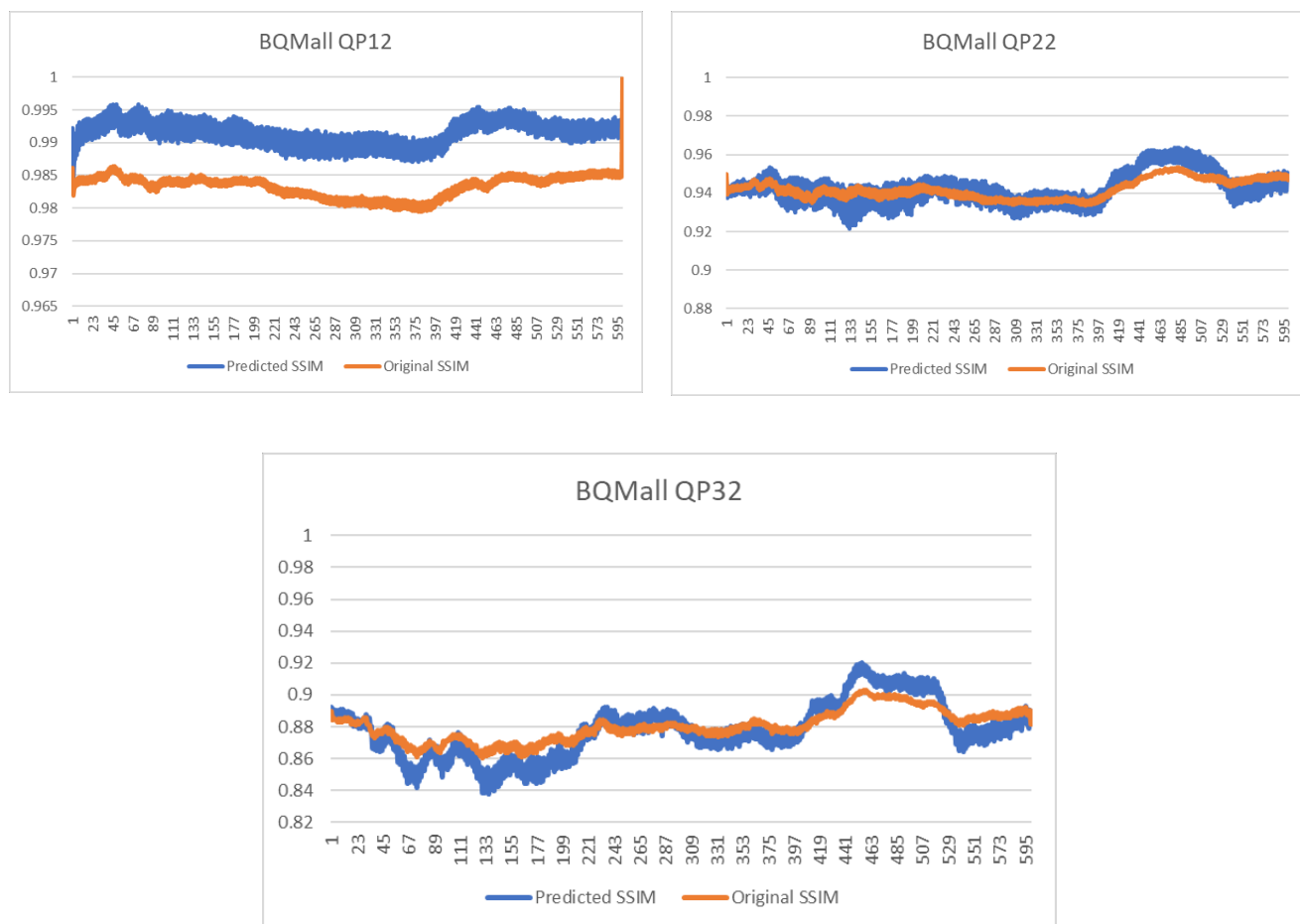


Figure 3.11 - BQMall SSIM vs SRR QPs 12-32

3.6 PERFORMANCE EVALUATION WITH OTHER RELATIVE COLORS

In all the previous performance measurements, a white video reference pattern has been used. However, it is interesting to investigate the performance of the proposed method, when other colors, beyond white, are used as reference video patterns. The selected color set shown in Table 3.3 portrays a distinct color diversity, which is essential to demonstrate and evaluate the behaviour of the proposed method on different relative reference video patterns.

In this section, the proposed method is applied on the experimental set of the 40 test signals at QP=12,22,32 and 42, each time utilizing a different color for the relative reference video pattern. For each color and QP value, the MAPD is calculated.

Table 3.3 - MAPD for the 40 Test Signals with Different Relative Colors

	QP:12	QP:22	QP:32	QP:42
BLACK	0.011049	0.030302	0.099677	0.300944
WHITE	0.007049	0.010328	0.032666	0.074436
RED	0.008156	0.013966	0.042319	0.114707
GREEN	0.009114	0.016522	0.047510	0.131344
BLUE	0.006726	0.014751	0.044131	0.117141
YELLOW	0.008185	0.013422	0.040684	0.124765
AQUA	0.007628	0.012956	0.040137	0.111483
PINK	0.006839	0.010283	0.032564	0.080557
GRAY	0.007039	0.010124	0.032484	0.072414

The experimental results of the process are provided on Table 3.3, where it is observed that the proposed method can be also applied with satisfactory accuracy utilizing also other colors, i.e. Pink, Gray, Blue, beyond the primarily selected white color. The alternative color patterns offer a satisfactory substitute, and in some cases it is noticed that the achieved accuracy outperforms the primarily used white color. On the other hand, other colors, i.e. Black, Green, perform notably worse than the rest, and are not a suitable candidate for the relative reference video pattern in the proposed method.

3.7 CONCLUSIONS

The proposed SRR method is a reduced reference video quality assessment one, which is based upon SSIM index as a tool, in order to extract features from both the original and the target video sequences, using a reference video pattern. The method is suitable for monitoring the video quality in real-time and across the service provision chain.

The performance of the proposed method was evaluated using a large experimental set of 40 reference and non-reference video sequences, with spatial resolution ranging from CIF up to HD, utilizing a static white video as a relative reference pattern. Results show that the accuracy of the proposed method ranges from 0.62% (QP=12) to 2.56% (QP=32), which represents the worst case performance. Thus, the proposed method maintains satisfactory performance across all the potential range of QP values, although better accuracy is achieved at lower QP values. Moreover, comparison to subjectively evaluated scores of LIVE Video Quality dataset shows that the accuracy of the proposed method is better than the average performance of reduced reference VQA methods and within the performance range of the full reference VQA methods.

Plots of SSIM and SRR vs the frame numbering are used for a more qualitative performance evaluation of the method. Plots demonstrate that : (a) the accuracy is

higher at better qualities, (b) SRR follows the variations of SSIM for both low and high values of QP and (c) SRR follows very reliably the abrupt variations of SSIM, as well as the wide variations, independently of the video quality.

Additionally, the performance of the proposed method has been evaluated, using non-white video reference signals. In general, the alternative color patterns offer a satisfactory performance. However, it was observed that the performance with colors sometimes is better than the one achieved with white and sometimes it is worse, depending on the content of the video under evaluation. The relation between the content and the color choice of the relative video pattern is a subject, outside the scope of this thesis.

Finally, an important advantage of the proposed method is analysed, which is the low bit rate reference information signal that needs to be sent to the end-user. The bit rate of the reference signal ranges between 400-600 bps and it is significantly lower than the 15kbps to 256kbps, required for other RR methods.

4 APPLICABILITY OF THE PROPOSED SRR AS A SERVICE ON TOP OF NFV/SDN-ENABLED NETWORKS

In the previous section the proposed video quality assessment method (SRR) was described, evaluated and compared to the original SSIM method, and its advantages were shown. This section examines how the SRR method can be applied and implemented as an in-service VQA method, combining the agility that Network Function Virtualisation (NFV) and Software Defined Networking (SDN) technologies offer.

As explained in section 3, the proposed SRR method is implemented in three phases: In phase I an initial SSIM_{or} value is evaluated at the service provider site, by comparing the original video sequence (V_{So}) with a reference video pattern (V_{Sr}). This phase can be implemented as a VNF at the service provider site. In phase II, an SSIM_{tr} value is evaluated at the consumer site, by comparing the received (test) video signal (V_{St}) with a reference white-video pattern (V_{Sr}). In phase III, the combination of SSIM_{or} and SSIM_{tr} provides the SSR metric SSIM_{ot}. Phase II and III can be implemented within one VNF at the consumer's site. Therefore, it is evident that the in-service implementation of the proposed method, requires the evaluation phases to be deployed as functions at appropriate nodes across the video traffic flow. Moreover, the traffic must be appropriately steered through the deployed nodes in order the necessary calculations to take place. Therefore, we propose an SDN/NFV-enabled networking architecture, which is capable of implementing the proposed method in-service and totally seamlessly to the end-user.

More specifically, this section presents a appropriate SDN/NFV-enabled networking architecture, which favors the implementation of the proposed SRR method and gives the ability perform Factory Acceptance Tests (FATs). For this purpose, an SDN/NFV infrastructure was deployed, supporting traffic steering of video flows through Open vSwitches (OVS: software-based Openflow-capable virtual switches) and the instantiation of the appropriate VNFs in an openstack environment, to support the required functions for the SRR method. In this way, the proposed SRR method can be enhanced to become an In-Service QoE method (IS-QoE), which can be implemented according to the concept of VNF as a Service (VNFaaS). The virtualisation of the proposed SRR enhances its applicability in commercial video distribution networks, because as a VNF it can be instantiated in many points within the video distribution chain. This feature is very important, because it enables the network operators to find the location within their distribution chain, where a degradation of the video quality takes place and thus apply specific remedy actions to fix the problem. In such a case the reaction time of the proposed SRR method to a video degradation is an important parameter for sustaining the end user's satisfaction to the desired level. The performance evaluation results show that the response time of SRR in video quality degradation is only 1 sec, which means that the viewer will suffer the video quality degradation for a very short period, until remedy actions are applied.

4.1 ANALYSIS OF NFV, SDN AND SRR AS A SERVICE

Network Function Virtualization (NFV) is a corner stone of software networks, and represents a large shift in how networks are built, deployed and managed. NFV does this by introducing a virtualization layer and decoupling the software from the hardware. The software-based assets then become the innovation and differentiating value, while the hardware becomes commodity. This is an attractive upgrade for network operators of various sizes and scope, whether they are a large communication service provider (CSP) or an enterprise running their own network. Commercial off-the-shelf (COTS) hardware for software networks will be generic by design, replacing what is now a proprietary landscape with hard dependencies between physical network functions (PNFs) and legacy telecommunication hardware.

In particular, the main focus in NFV is decoupling the physical appliances from the network functions that operate on them; in other words, the transition from hardware to software -based network function [Han-2015; Guerzoni-2012]. This paradigm has gained significant attention from the Telecommunication Service Providers (TSPs), due to its potential to provide a significant reduction in Operating Expenses (OPEX) and Capital Expenses (CAPEX) [Wu-2015; CMRI-2011] as well as widening Telco's service portfolios with novel added-value software network offerings.

The high level NFV framework is shown in fig 4.1. Network Functions Virtualisation envisages the implementation of NFs as software-only entities that run over the NFV Infrastructure (NFVI). [ETSI-NFV001], [ETSI-NFV002] Three main working domains are identified in NFV:

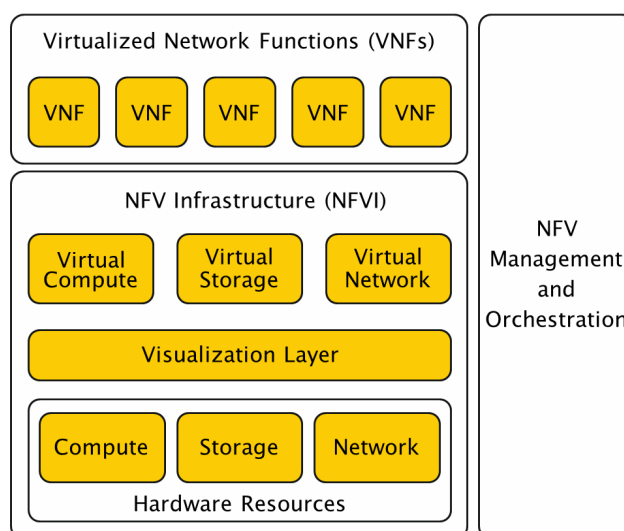


Figure 4.1 - NFV architecture according to ETSI

- Virtualised Network Function, as the software implementation of a network function which is capable of running over the NFVI.
- NFV Infrastructure (NFVI), including the diversity of physical resources and how these can be virtualised. NFVI supports the execution of the VNFs.
- NFV Management and Orchestration, which covers the orchestration and lifecycle management of physical and/or software resources that support the infrastructure virtualisation, and the lifecycle management of VNFs. NFV Management and Orchestration focuses on all virtualisation-specific management tasks necessary in the NFV framework.

NFVI is composed of NFV infrastructure points-of-presence (NFVI-PoPs), which host the VNFs and include resources for computation, storage, and networking. NFVI creates a virtualization layer that sits right above the hardware and abstracts the HW resources, so they can be logically partitioned and provided to the VNFs to perform their functions. NFVI networks interconnect the computing and storage resources contained in an NFVI-PoP.

According to this architecture, software appliances in the form of Virtualized Network Functions (VNFs) run on standard high-volume (SHV) servers, which consolidate the operation and management of various network devices on a common infrastructure. VNFs provide the ability to be dynamically initiated, reconfigured and even reallocated at different locations within the network, without the requirement of installing new hardware. NFV aims at delivering a more open and extensible network environment, where the deployment of new network services becomes easier and faster.

However, in order to deliver on these expectations, various fundamental developments have to be realized, many of which are still in a state of implementation and continuous evolution. The deployment of NFV and its adoption into Telco grade systems still subsumes various open issues identified in [ETSI-NFV002]. NFV will significantly challenge current network management systems, and will require additional levels of complexity over currently deployed systems. One of the key features in this transition is the management layer, which must be capable of supporting the unique features of an NFV-enabled system. The management layer in NFV environments, is known as the Management and Orchestration system (MANO) [ETSI-NFV003]. The primary aim of MANO is to coordinate NFVI resources and map them efficiently to various VNFs. In turn, VNFs can then be interconnected into chains to realize more complex Network Services (NS). An NFV NS can be seen as the evolution of a traditional telco connectivity service, as it is augmented by chains of VNFs which are dynamically inserted into network traffic paths. Although MANO aspects are not within the aim of this thesis, this section demonstrates that the proposed SRR method can be implemented by appropriate VNFs, which, could be managed by a MANO and therefore become components of a more complex network service, that fulfills specific needs of a network operator. In this way, VQA aspects, which are important for the end users' satisfaction, will be easier embedded within the evolving landscape of network virtualisation.

SDN is an emerging networking paradigm that changes the limitations of current network infrastructures [Kim-2013]. Firstly, it separates the network's control plane from the underlying network elements that forward the traffic (the data plane). Secondly, with the separation of the control and data planes, the control logic is implemented in a logically centralized controller. This change of control [Sgambelluri-2013] enables the underlying network infrastructure to be abstracted for various applications and network services and as a consequence of this the network can be treated as a logical or virtual entity. Some of the benefits of the SDN are: the centralized control of multi-vendor environments, the reduced complexity through automation, the increased network reliability and security, and the more granular network control. Thus, currently there is not any video quality assessment methods, which utilizes the advantages of the SDN in order to provide in a flexible way the quality assessment. SDN creates the opportunity (together with NFV) for expanding the current range of QoE assessment methods in order to involve new ones that will exploit the advances of the SDN and NFV. A possible approach, which is presented in this section, is based on the mapping of the different phases of the proposed method to network functions, which are virtualized in order to be deployed within an SDN/NFV-enabled network. Then, upon appropriate programming, the video traffic can be diverted through these virtual network functions (VNFs), providing finally the quality assessment measurements in seamless and agile way. So, the proposed approach in this thesis involves the deployment of the new generation of QoE models (such as the proposed SRR), as Virtual Network Functions (VNFs) over the SDN network, exploiting the concept of NFV. Moreover, the deployed video quality assessment VNFs can be further exploited as a Services by other service provides or network operators, maximizing in this way the utilization of the proposed method by different actors, which are involved in the media and entertainment industry.

The proposed approach has been inspired by the relative standardization activities that take place in the NFV-field, where ETSI NFV ISG has announced various NFV use cases that illustrate the application of NFV in combination with SDN [ETSI-NFV001]. Among the various use cases, the concept of the Virtual Network Function as-a-Service (VNFaaS) is defined, which prescribes the provision to the customer of an end-to-end connectivity service (virtual network) along with embedded VNFs. The connectivity service specifications would allow dictating certain limits for the provided QoS, but the QoS assurance will need monitoring mechanisms in place. There are numerous approaches to QoS assessment models, but the most dominant now is the Quality of Experience (QoE) concept because it provides a direct link to the user-satisfaction in relevance to a specific QoS-sensitive service (e.g. IPTV, VoD etc.) [Staelens-2010].

In this context, this section presents SRR according to the proposed SDN/NFV-enabled deployment, which creates from the one hand the opportunity for in-service video quality assessment method, and on the other hand the provision of the method as VNFaaS. This SDN/NFV-enabled QoE evaluation is able to be embedded into the network provisioning, and management processes and will give to the service provider and network operator the capability to efficiently manage the network resources by allocating to the virtual network instance only the ones that are necessary to maintain a

specific level of user satisfaction according to the QoE assessment made along the service delivery chain. Thus, such VNFaaS QoE methods are critical for vendors, network operators and service providers to sense the video degradation in time and apply specific measures, such as reduce the encoding bit rate or redirect the traffic through another network path and thus maintain the end users' QoE as high as possible.

4.2 DEPLOYMENT OF THE PROPOSED SRR METHOD AT AN NFV ENVIRONMENT

In order to validate and evaluate the proposed SRR method in the form of VNFaaS, over an SDN/NFV-enabled network architecture with scope to achieve in-service and seamless video quality assessment an experimental testbed was implemented and a real-time video streaming application was employed. The primary requirement for the testbed to be used for testing the proposed architecture is the support of Virtual Machines (VMs), where VNFs may run. Among the various virtualization environments that support VMs, we selected Openstack, because it is widely used for research purposes and also includes various components, such as Nova and Neutron, which allow the SDN-based networking experimentation that is required for the validation of the proposed architecture.

4.2.1 REQUIREMENTS OF THE PROPOSED SDN/NFV-ENABLED NETWORK

The proposed SDN/NFV-enabled network must satisfy specific requirements in order to be feasible the deployment of the proposed SRR in the form of VNFaaS. More specifically:

- The proposed networking architecture must be one SDN domain controlled by the same SDN controller in order to be feasible the programmability of the whole domain and therefore the video traffic steering to the virtual functions of the SRR method.
- The network domain must include NFV Infrastructures (NFVI) in the form of Points of Presence (PoPs) both close to the video server and the end-user in order to appropriately host the virtual functions of the SRR method, considering also the proximity requirements that the methods considers.
- The VNFs hosted at the different NFVI-PoPs of the proposed SDN/NFV network must be able to communicate with each other in order to send the intermediate data from the different phases of the SRR to the next virtual function.
- Once instantiated and configured, a VNF SHALL start its operation on a specific point in time, either on a pre-scheduled basis, or immediately upon user request.

4.2.2 PROPOSED SDN/NFV SRRaaS ARCHITECTURE

Based on the aforementioned requirements, which envisage the deployment of the proposed SRR method on a SDN/NFV-enabled network in order to achieve in-service

and seamless video quality assessment, the following reference architecture is proposed.

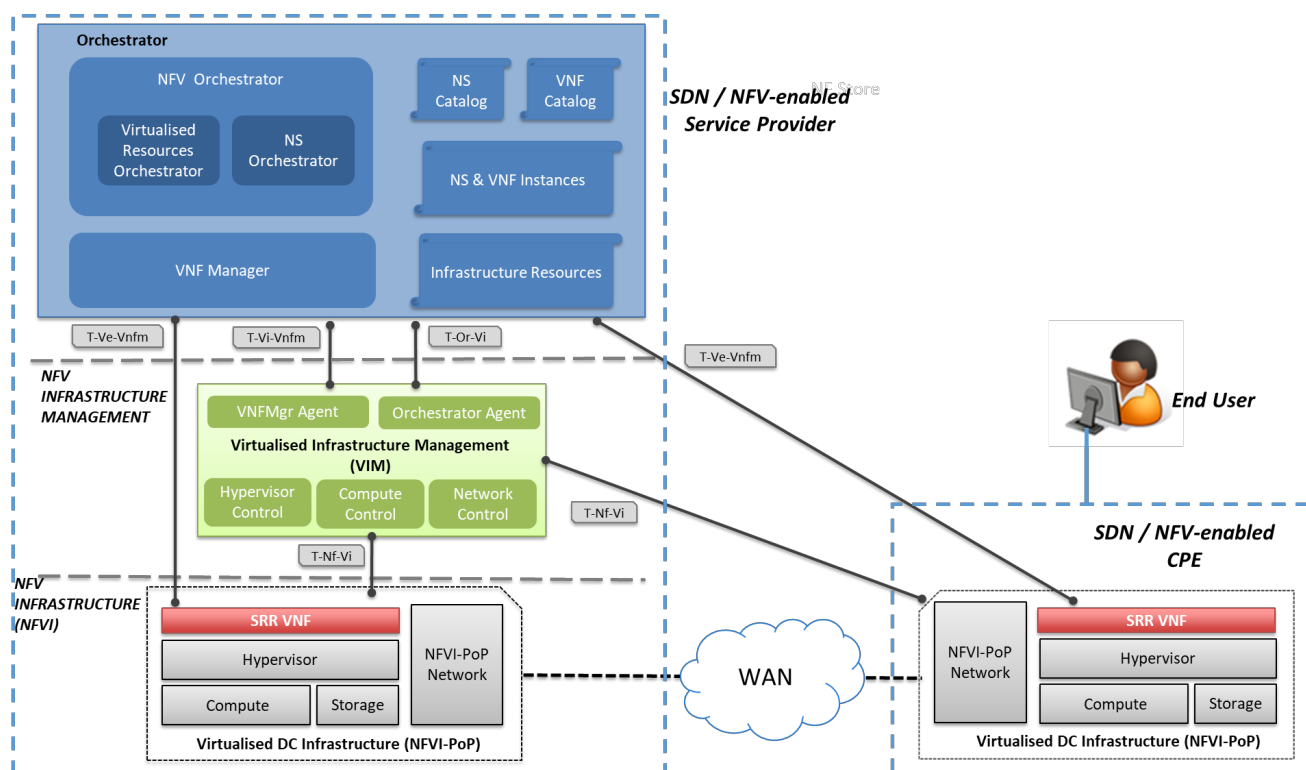


Figure 4.2 - Overview of the proposed reference SDN/NFV architecture

As it can be observed from the figure, the proposed architecture is split into three logical layers (from bottom to top):

- The Infrastructure layer, which primarily consists of the virtualisation-capable equipment, (i.e. the in-network Data Center platform or any computing infrastructure, namely the NFVI PoPs), on which the Virtual Network Functions (VNFs) of the VNF/SDN-enabled SRR are deployed (i.e. through the VNF instantiation) and the SDN/NFV-enabled CPE part, which continues to host phase 2 and 3 SRR functions. This is the core layer of the SRR implementation in a virtualized environment. The SRR VNF as depicted in Figure 4.2 is the integration of the proposed method of Section 3 in an NFV enabled system. Moreover, the infrastructure layer includes also SDN and non-SDN (i.e. WAN) based network elements, which provides the programmable network interfaces that will provide the connectivity establishment and will support the VNF chaining through traffic steering mechanisms. An NFVI compute cluster is inherently virtualization-capable and consists of three domains:
 - The Compute domain, which represents the lowest (physical) level, comprising the computing and storage equipment (standard high-volume servers with or without specialized hardware accelerations and storage

infrastructure). For standard-scale data centre implementations, servers based on the x86 architecture are a common choice. The adoption of features for hardware-assisted virtualization, such as DPDK (Data Plane Development Kit) support and SR-IOV (Single-Root I/O Virtualisation), seems quite promising for the enhancement of VNF performance and is thus recommended.

- The Hypervisor domain, which is responsible for the abstraction of the physical compute and storage resources (possibly aggregated across multiple physical elements) and their assignment/allocation to VNFs. The hypervisor domain mediates the resources of the computer domain to the virtual machines of the software appliances. Hypervisors as developed for public and enterprise cloud requirements place great value on the abstraction they provide from the actual hardware such that they can achieve very high levels of portability of virtual machines. In essence, the hypervisor can emulate every piece of the hardware platform even in some cases, completely emulating a CPU instruction set such that the VM believes it is running on a completely different CPU architecture from the actual CPU on which it is running. Such emulation, however, has a significant performance cost. The number of actual CPU cycles needed to emulate virtual CPU cycle can be large.
- The hypervisor commonly exposes a northbound interface for the interaction with the Management layer. Several choices are available for the hypervisor technology (Hyper-V, VMware, KVM, Xen etc.), heavily depending on the compatibility with the VIM and also with the physical infrastructure. Kernel-based Virtual Machine (KVM) would be a safe recommendation, given its openness, wide compatibility and full-featured integration with Openstack.
- The infrastructure network domain, which within the NFVI includes all networking elements, such as SDN and non-SDN switches and routers that interconnect all the compute/storage infrastructure of the compute domain.
- The NFV Infrastructure Management layer, which includes distributed management entities for the aforementioned parts of the infrastructure: the internal NFVI PoP management tools (i.e. the Cloud Computing Platform and the SDN controller of the internal virtual networks of the DC). The SDN/NFV enabled segments (NFVI-PoPs), as mentioned before, are managed by a bounded management entity of SDN controller and Cloud Computing platform, which is called Virtualised Infrastructure Management (VIM) entity. VIM, conforming to ETSI ISG NFV terminology, which is the functional entity responsible for controlling and managing the infrastructure (compute, storage and network) resources. The management scope of the VIM is generally restricted within a single NFVI-PoP. Thus, in a full deployment architecture, multiple VIMs may

operate across operator data centres providing multi NFVI-PoPs that can operate independently or cooperatively as required under the control of an Orchestrator. While a VIM, in general, can potentially offer specialisation in handling certain NFVI resources, in the proposed architecture context, the VIM is seen to encompass all management and control functionalities needed for the proper administration of the infrastructure, as well as the virtualised services running on top of it.

In specific, the following key tasks are performed by the VIM:

- Maintenance of a resource, capability and topology repositories/inventories, thus establishing a comprehensive “map” of the underlying hardware;
- Joint management of the infrastructure (compute, storage, networking) resources;
- Association/mapping of the virtualised services to the infrastructure resources;
- Basic network control services, including topology management and path computation;
- Management (create, query, update, delete) of service function chains i.e. the interconnections among VNFs, by creating and maintaining Virtual Links, virtual networks, sub-nets, and ports;
- Management of VM software images (add, delete, update, query, copy) that host VNFs;
- Management of virtual networks, tunnels and QoS, where applicable;
- Collection and communication of measurements and faults/events information relative to physical and virtual resources.
- In order to realise these tasks, the VIM needs to comprise the following components:
 - A Resource repository database – for maintaining a comprehensive landscape of the underlying infrastructure, the exposed capabilities and the available resources;
 - A Topology and service function chain management module – this component undertakes most network management tasks, including virtual network management, interconnection of virtualised components and tunnel establishment;
 - A Compute/hypervisor management module – this component undertakes the management of VMs/VNFs;
 - An integrated monitoring and event/alarm management framework – for efficient and effective collection of metrics and production of events/alarms;
 - A set of southbound interfaces for managing the infrastructure (compute nodes, hypervisors, network elements). These commonly come in form of “plug-ins” in order to accommodate multiple infrastructure technologies (such as several hypervisors, network management protocols etc.);

- A set of northbound APIs (commonly REST-based) for communication with the upper layer (NFV Management).

From the implementation point of view, a VIM is commonly realized by coupling a network controller and a cloud controller platform.

- The NFV Orchestrator is the management entity, which is responsible for the management of the VNF lifecycle, which includes the NFV instantiation, the dimensioning and the termination. The NFV manager receives appropriate commands from the upper orchestration/federation layer, which includes the NS descriptors, which will initiate the VNF instantiation with the appropriate network configuration internally in the NFVI PoP.

As seen, the Orchestrator platform comprises both catalogs/repositories as well as execution components:

- The VNF/PNF Catalogue represents the repository of all of available on-boarded VNF and PNF Packages, supporting the creation and management of the VNF Package (VNF Descriptor (VNFD) and PNF Descriptor (PNFD). The information contained in the VNFD/PNFD is defined by ETSI. Again, it is clarified that the VNF/PNF Catalogue contains a list of the available VNFs/PNFs which can be included in an NS, not the deployed VNFs themselves. In a similar way, a similar catalogue can be used –if necessary- in order to contain specific Network Services (i.e. combination of VNFs/PNFs) that are described by the respective NS Descriptors.
- The VNF Instances Repository contains information of all service instances, which have been actually deployed. The repository is frequently updated, to reflect the status and the lifecycle of the deployed virtualised services.
- The Infrastructure Resources Repository holds information about available/reserved/allocated NFVI resources as abstracted by the VIM across operator's Infrastructure Domains, thus supporting information useful for resources reservation, allocation and monitoring purposes.
- The VNF Manager (VNFM) is responsible for the lifecycle management of VNF instances. Each VNF instance is assumed to have an associated VNF Manager. A VNF manager may be assigned the management of a single VNF instance, or the management of multiple VNF instances of the same type or of different types. Operations carried out by the VNF Manager are VNF instantiation and feasibility checking; integrity management; VNF instance modification/scaling/healing/termination.
- The Infrastructure Resources is the component which mainly interacts with the VIM for resource discovery, allocation and management, allowing the NFV Manager platform to manage and control distributed resources across multiple NFVI-PoPs.

- Finally, the NFV Orchestration Logic is the core decision-making component, actually the “kernel” of the NFV manager. The NFV Orchestration Logic instantiates VNFs, which are part of network services orchestrated by FNRM, (using the VNF templates in the corresponding catalogues) and manages the whole VNF lifecycle. For this purpose, it communicates with the VNFM and the Infrastructure Resources for the control of VNF instances and the (re-) allocation of virtualised resources. This task includes the control of the network assets, for virtual network establishment and QoS provision.

Interfacing Entities	Reference point name	Description and Comments
Virtual Network Function Manager– Virtual Network Function Interface	T-Ve-Vnfm	This interface allows the VNF Manager to configure the VNF, collect monitoring/performance data, and be notified about faults within the VNF.
Virtual Network Function Management – VIM Interface	T-Vi-Vnfm	This interface allows the VNF Manager to request operations related to the NS lifecycle; and/or for the VIM to report the characteristics, availability, and status of VNF related infrastructure resources.
Orchestrator –VIM Interface	T-Or-Vi	This interface allows the Orchestrator to request reservation/allocation of resources and NS related lifecycles operations; and for the VIM to report the characteristics, availability, and status of infrastructure resources.
VIM – Network Interface	T-Nf-Vi/N	This interface dispatches management decisions from the VIM to the NFVI-PoP Network domain and communicates back Network status.

VIM – Compute Interface	T-Nf-Vi/C	This interface dispatches management decisions from the VIM to the Compute domain and communicates back the Compute domain status.
-------------------------	-----------	--

The implementation of the SRR method as a Service on top of the proposed architecture considers the deployment of phase 1 and phase 2 functions of the SRR at the ingress NFVI-PoP of the proposed architecture, while the deployment of phase 3 function of the SRR is considered at the egress NFVI-PoP. The deployment of SRRaaS is performed according to the ETSI NFV specifications [ETSI-NFV001], as to provide a compatible operating network service to an NFV enabled system. Then, the steps towards providing the SRR as a Service on top of the proposed architecture are:

- The video server initially hosts the test signals in their original MPEG-4 form, namely, Quantization Parameter (QP) is equal to one (QP=1). They are considered as being the reference signals for the video quality assessment.
- The end-user requests a specific video from the video server;
- The video server initiates the streaming process, which for the experimental validation was selected unicast UDP-based streaming.
- The video is streamed from the video server towards the end-user over the SDN-enabled network domain.
- At the ingress NFVI-PoP the video stream is steered through the VNF1, which performs the calculation of the SSIM_{ow} of the original signal. The SSIM_{ow} value is sent to VNF3.
- Then the traffic is further steered within the ingress NFVI-PoP to the VNF2, which transcodes the MPEG-4 signal in real time at lower video quality levels, by altering the QP value of the streaming video to either 12, 22, 32 or 42.
- The video stream is further forwarded outside of the ingress NFVI-PoP and over the SDN-enabled network till the egress NFVI-PoP.
- At the egress NFVI-PoP the video stream is steered via VNF3, which calculates in real time the SSIM_{tw} and the ratio of SSIM_{ow}/SSIM_{tw} (i.e. the proposed IS-QoE) is calculated (utilizing the SSIM_{ow} value that has been sent by VNF1).
- The video stream is further forwarded outside the egress NFVI-PoP and reaches finally the end-user.

4.3 PROPOSED ARCHITECTURE IMPLEMENTATION

The proposed architecture shown in Figure 4.2 is experimentally implemented, as depicted in Figure 4.3. The proposed implementation meets the provided requirements and considers at the ingress (service provider) and egress (end user) points of the network domain two SDN-compatible Open Virtual Switches (OVS), which are under the

management and control of the OpenDaylight SDN controller. At both ingress and egress points of the domain, an Openstack cloud platform is installed to support an NFVI-PoP, which is capable of instantiating VNFs and performing also the appropriate network traffic steering in order to support service chaining (i.e. the forwarding of the traffic seamlessly from the video server through the VNFs of the NFVI-PoPs and then to the end user. The NFVI-PoPs are based on the Openstack open source cloud computing platform. The release used was Liberty, which was the latest stable version during the time of the experimental tests.

The video files to be used for the tests are hosted in a video server, at the service provider's site. Files contain the video content in the original encoded format, at a high quality. For the needs of the tests to be performed, a real time video adaptation server is included, which decreases the transmission bit rate, to fit in the requirements of the connection link between the two NFVI-PoPs. The transcoder is based on the widely used FFMPEG [FFMPEG-2017] and is instantiated as VNF2 at the ingress NFVI PoP.

The proof-of-concept implementation of the proposed SRR method, considering the proposed SDN/NFV architecture is split into two functions: The first one is instantiated as VNF1 and is located at the ingress NFVI-PoP, calculating the SSIMor value at the service provider site, by comparing the original video sequence (VSo) with a reference video pattern (VSr) (i.e. phase 1 of the method). The second function is instantiated as VNF3 and calculates the SSIMtr value at the end user site, by comparing the received (target) video sequence (VSt) with a reference video pattern (VSr) (i.e. phase 2 of the method).

The value of SSIMor can be sent to VNF3 in various ways, as for example embedded in the video frame, as metadata of the video signal or being saved to a local relational database, as is the case for this thesis. The database exposes a REST webservice interface, so that SSIMor can be retrieved through an HTTP REST-based request initiated by the VNF3, at the end user's NFVI-PoP.

The value of SSIMor for each frame is retrieved at the VNF3, where SSIMor/SSIMtr (i.e. phase 3) is calculated in order to provide the estimated SSIM value at the end user's site.

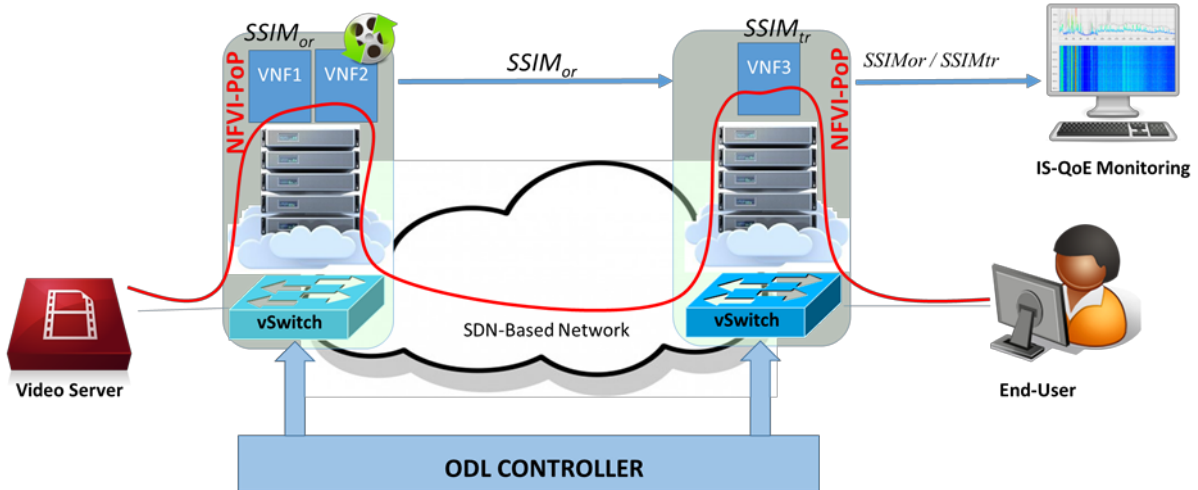


Figure 4.3 - An overview of the experimental topology

The evaluation tests consider that a unicast video service is delivered from the video server, through the two NFVI-PoPs to the end user terminal. Video traffic is steered from the video server to VNF1, then VNF2, within the first NFVi-PoP and through the connection link to VNF3 and finally to end user's terminal. The steering of the video stream through the two NFVI-PoPs and the corresponding VNFs is totally seamless to the end-user.

Considering the SDN-based traffic steering mechanism within the NFVI-PoPs, Figure 4.4 zooms in the Openstack cloud computing platform of the ingress NFVI-PoP, where VNF1 and VNF2 are hosted. More specifically, the figure presents the complex L2/L3 traffic steering process, which is achieved using SDN/Openflow mechanisms to alter per hop in the traffic flow the values of the destination IP address, destination MAC address and source IP address. For better representation of these SDN-based network programming, the label of Fig. 4.5 has been used at each hop of Fig. 4.4.

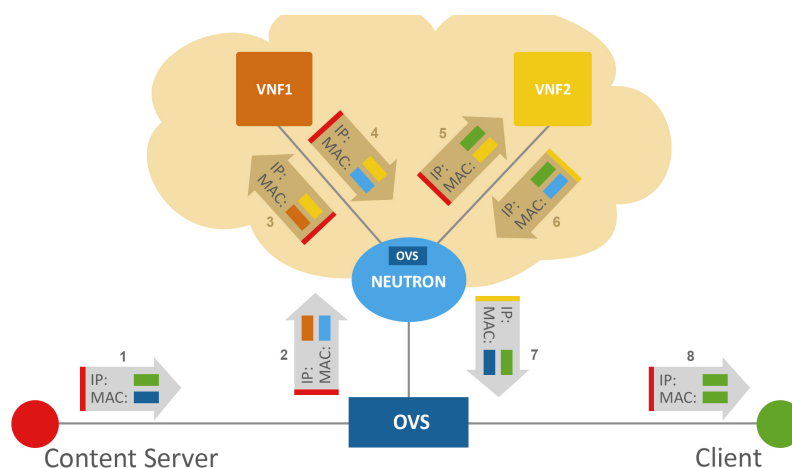


Figure 4.4 - Openflow-based traffic steering within Openstack platform

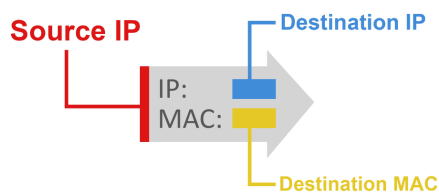


Figure 4.5 - Label used to show the SDN-based traffic steering commands per hop

As seen in Fig. 4.4, we use Open vSwitches (OVS: software-based Openflow-capable virtual switches), in order to divert the traffic through VNF1 and VNF2 by altering the destination MAC and IP address fields of the packets. The OpenStack network node (Neutron) plays an active role in the process.

4.4 PERFORMANCE EVALUATION OF THE PROPOSED IS-QOE METHOD

The aim of the tests is to assess the appropriateness of the proposed SDN/NFV-enabled architecture for deployment purposes of the SRR as a Service. For this reason, various video quality degradations are created utilizing background traffic that results saturation of the connection link capacity, thus causing packet losses of the delivered video service.

In this respect, the following processes are performed:

- The SDN controller configures the OVSs in the two NFVI-PoPs, to create the network path between the service provider and the end user's terminal through the above mentioned NFV1, NFV2, NFV3.
- A video service is initiated and video is streamed from the video server towards the end-user over the SDN-enabled network domain. For the experimental validation unicast UDP-based streaming was used.
- At the ingress NFVI-PoP the video stream is steered through the VNF1, which performs the calculation of the SSIMor and its value is stored in a local relational database.
- Then the traffic is further steered within the ingress NFVI-PoP to VNF2, which transcodes the original video signal in real time. Transcoding utilizes an MPEG-4 video codec using Simple Profile with spatial resolution 640x480, frame rate 24 fps and at ~1024 kbps.
- The video stream is further forwarded outside of the ingress NFVI-PoP and over the SDN-enabled network till the egress NFVI-PoP.
- At the egress NFVI-PoP the video stream is steered via VNF3, which calculates the SSIMtr for each frame and also retrieves the value of SSIMor through an HTTP REST-based request from the data base. Finally, the ratio of SSIMor/SSIMtr is calculated for each frame.
- The video stream is further forwarded outside the egress NFVI-PoP and reaches finally the end-user's terminal.

The test scenario includes the following steps:

- After video transmission is initiated, the video quality of the delivered video is calculated in real time at the end user's site, using the SRR method
- At a specific time point, background network traffic is added in the connection link, thus degrading the video quality
- The video quality, as evaluated by SRR, degrades and the time elapsed is measured
- At a specific time point, background network traffic is removed and the video quality is restored in its previous value
- The video quality, as evaluated by SRR, increases and the time elapsed is measured

According to the story line, background traffic is added in the link between the two NFVI-PoPs, in order to saturate it and thus force quality degradation (due to packet loss) of the delivered video service. For the experimental needs of the scenario, the maximum available bandwidth of each NFVI-PoP interface was reduced to 10 Mbit, in order to be easily saturated with background traffic.

Towards flooding the connection link with background traffic, synthetic UDP traffic is generated by a Linux virtual machine utilizing the iperf command. Iperf generates approximately traffic of 10Mbit, which is enough in order to flood the link and therefore create significant degradation to the delivered video service.

Upon the introduction of the background traffic, significant degradation is observed in the perceived quality of the delivered service, which is made practically unviewable due to multiple error propagations. In terms of video quality assessment, the respective average QoE level (as measured by the SRR metric) drops from the approx. 0.83 value down to 0.21.



Figure 4.6 - Video quality degradation due to packet loss

The metrics to be measured is the both the accuracy and the responsiveness of the SRRaaS. Upon the execution of an experimental set of ten repetitions, the responsiveness of SRR to the traffic congestion, is about 1 sec, as shown in Figure 4.6, which shows that the metric can be also used appropriately in real-time conditions for

monitoring purposes. Moreover, the values measured capture satisfactorily the quality degradation cause by the introduction of the background traffic.

By comparing the measured quality in-service with the quality that the SRR provides for the captured video, it is deduced that the SRR performance is not affected by its virtualization and distributed deployment. However, the SRR components in the distributed implementation operate in different locations, which leads to the task of synchronizing the results from each VNF in order to deduct the final index.

The synchronization process is basically carried out in 2 main steps. Firstly, in VNF1-2 the SSIMor is calculated and stored in a Redis NoSQL database for further use, in order to combine it with SSIMtr and deduct the SRR index. The key point in this implementation is the usage of the Redis database for reasons of speed and performance. In [Kabakus-2017] the performance of Redis for read operations is measured and the experimental results show that Redis can fetch a value from a 100,000 record database in approximately 8ms. In our implementation videos with 25-30 fps are used to evaluate our proposed SRR metric, so in order to achieve real-time video quality evaluation a total time of 240ms are needed to fetch all the SSIMor values from a Redis database, which is significantly below the 1sec threshold.

Additionally, a provision has been made to ensure the stability and efficiency of the system. As the communication with the database is done over a best-effort IP network, there may occur incidents that will disrupt the normal service of the evaluation process. This is why a short timeout for the response has been set in the VNF which combines SSIMor and SSIMtr, and if this timeout expires a "not available" output is generated for the particular frame(s), until the system/network recovers.

In the opposite situation, where the background traffic is removed, Figure 4.7 depicts the improvement in the QoE level, as expressed by the SRR values. By executing again an experimental set of ten repetitions, the responsiveness of SRR to the absence of traffic congestion, is about 3 sec, as shown in Figure 4.7. Again in terms of comparison between the offline SRR metrics and the SRRaaS metric, by repeating the experiment offline with the captured video signal, it is not observed any deviation, proving that the metric can be efficiently used both in offline and in-service conditions.

In terms of video quality assessment, the respective average QoE level (as measured by the SRR metric) increases from 0.22 up to its previous level of 0.83.



Figure 4.7 - Video Quality is reinstated

4.5 CONCLUSIONS

This section examined the provision of the SRR metric as a Service (SRRaaS) and proposed a suitable architecture for the deployment of the SRR in a distributed way on top of SDN/NFV-enabled network domain. In order to achieve this, the proposed architecture must consider the proximity requirements of the three phases of the method, which means that two NFVI-PoP should be considered: one close the video server and the other one close to the user (ideally integrated with the CPE). Following this proposed SDN/NFV-enabled infrastructure, a proof-of-concept testbed was implemented, in order to perform Factory Acceptance Tests (FATs), utilizing Openstack and Opendaylight for enabling the SDN/NFV technologies, as well as Open Virtual Switches. By performing some experiments by introducing background traffic which resulted in video quality degradation, we have shown that the proposed SRR method can successfully be deployed in a distributed way as a group of VNFs, without this split to affect the performance or the responsiveness of the method.

By offering the proposed SRR method as an In-Service QoE metric (IS-QoE), which can be implemented as a VNF (VNFaaS) enhances its applicability in commercial video distribution networks. SRRaaS can be instantiated in any point within the video distribution chain, which is a very important feature for network operators that allows them to find the location within their network, where a video quality degradation takes place and apply specific remedy actions to fix the problem. Such reaction can be to transcode the video in lower bit rate or send the video traffic through another network path.

5 VALIDATING THE SRRaaS METHOD OVER A SMALL CELL MOBILE NETWORK

5G networks promise significantly reduced latency and increased capacity for delivering high bandwidth data streams between high densities of people and things at low energy and with high reliability. The focus of 5G research so far has been largely on the required advances in network technologies: spectrum, radio access, SDN, NFV and cloud infrastructure, flexible management and control architectures and development and operations systems. Less attention has been put on the applications and services that will make use of and exploit advanced 5G network capabilities. Media applications are amongst the most demanding services requiring huge quantities of network capacity for high bandwidth audio-visual streams as well as extremely low latency for immersive, responsive and tactile user experiences. While many of the technologies that will underpin and enable 5G are still under development or have yet to be decided upon, there was early industry consensus that small cells will be a critical building block of 5G networks.

This section validates the SRRaaS method, which was tested in the previous chapter in a FAT environment, a real 5G network based on a small cell architecture in order to expand the validation process in Site Acceptance Tests (SATs). More specifically, we show that the SRRaaS method is suitable for deployment as a VNF within an IT infrastructure located close to the small cell. It enables the in-service monitoring of the delivered video quality, which is a very useful tool for the mobile network operators, to monitor their customers' satisfaction. An advantage of the proposed method, when applied in such a network, is that the complex and power consuming process of video quality assessment is performed at the edge of the network, and not at the UE itself, thus significantly reducing its power consumption.

For the SAT experimentation, a Small Cell commercial product was used for the implementation and performance evaluation of the SRRaaS. The experimental results in the field showed that the proposed method is able to monitor the video quality efficiently as expected from the FAT experimentation in the lab.

5.1 VIDEO QUALITY ASSESMENT IN 5G SMALL CELL NETWORKS

The proliferation of mobile wireless broadband technologies during the last decade, has triggered the ascension of 5G Mobile Networks, which is designed to ensure scalability, efficiency and versatility [Andrews-2014]. Although the 5G related standards have not yet been fully completed, there are a few assumptions already being agreed about 5G, which are virtualization, small cells and expansion into high frequency bands [SCF-2015-055].

The virtualization of network functions has been initially applied to large IT data center and has turned data-centers into service-oriented architectures that are able to rapidly respond to the dynamic business environment. Network function virtualization (NFV) techniques have been proven to offer great benefits to the world of IT, in terms of

sharing compute, storage and network resources, as well as service agility and ultimately bringing higher revenues and competitiveness.

NFV is impacting all service provider segments and is therefore affecting the realization of future mobile networks. Mobile network operators having realized the benefits, encourage the development and deployment of NFV techniques in their networks. Today, virtualization is a well-examined topic within core networks and macro-cell access networks (Cloud Radio Access Network C-RAN). Nevertheless, the impact of virtualization for a small cell network have received very limited attention and its benefits have not been considered thoroughly. Much of the work on virtualization has been undertaken by ETSI, including the definition of virtualization use cases. In particular, use case #6 'Virtualization of the mobile base-station' is of interest to a small cell network.

The Small Cell concept has become pivotal in today's mobile networks; Small Cells provide improved cellular coverage, capacity and applications for homes and enterprises as well as dense metropolitan and rural public spaces [Andrews-2013], [4GAMERICAS-2012]. Their role is crucial for providing services in populated areas like stadiums, shopping malls, concert venues, and generally, places with (tactic or sporadic) high end-user density [Osseiran-2014], [Tehrani-2014]. In such cases, normally each telecom operator deploys its own infrastructure, acting complementary to the macro cell network. Normally, Small Cell provisioning requires a number of time and money consuming procedures as e.g., provisioning of installation site, power supply and so on. Operators must also face the costs of establishing dedicated, high-capacity backhaul connections, not to mention radio resource management and interference mitigation techniques, all translating to extra costs and efforts. However, this static approach based on the ownership of the physical Small Cell infrastructure not only increases operators' CAPEX and significantly hampers business agility, but also is it unable to cope with dynamic scenarios. For example, one should consider the case where sporadic flash crowd events arise not only at predefined venues (e.g., shopping malls, urban areas, stadiums, etc.) but also at arbitrary areas with minor infrastructure in place, resulting in traffic overflow and signal outage. In order to respond to this dynamicity, network operators may wish to deploy for some time a Small Cell network to serve e.g., a sporadic flash crowd event, without really owning the underlying infrastructure. The latter could even be provided by a third party, i.e., the owner/operator of the venue. Such sharing scenarios have been already identified in 3GPP [3GPP-TS23251] and are expected to play vital role in 5G networks. From above, it can be deduced that the small cell architecture will play an important role in the deployment of the upcoming 5G networks.

The applicability of NFV techniques to small cell base stations has specific individualities, which are examined in the Small Cell Forum (SCF). A basic requirement is that a subset of small cell functionalities that support at least the RF functions is run on a physical network function (PNF) [SCF-2015-055]. The hardware on which the PNFs will run is a small cell radio unit consisting of at least a single cell radio transceiver (with multiple antennas as required). The remaining functionality is run as one or more VNFs on

virtualized compute platforms. The split of the functions, i.e. which will remain as PNFs and which will become VNFs is a challenge that is thoroughly examined in SCF, but it is out of the scope of this thesis.

There are specific small cell functions, which refer to layers L1, L2 and L3 and functions related to application services (such as content caching, firewalls, QoS monitoring etc) and small cell management functions, (such as Radio Resource Management -RRM). The proposed SRR is an application service, suitable for small cell networks, that enables the in-service monitoring of the end users' satisfaction.

Another individuality of small cell networks is their unreliable backhaul. It is anticipated that the non-ideal backhaul networks, which are commonly found in today's small cell deployments, are highly applicable to 5G scenarios too [SCF-2014-088]. Considering that the connectivity of small cells with the core network of the LTE system (i.e. the Evolved Packet Core (EPC)) is based on best-effort/Internet links, its quality cannot be guaranteed. As a consequence, neither the end-user satisfaction, nor the quality of the provided services can be guaranteed. On the other hand, the quality of the Radio Frequency (RF) link, between the User Equipment (UE) and the small cell, is generally considered satisfactory, due to the short distance between UE and the small cell and the low number of mobile subscribers connected to the specific small cell.

Among the various services that are planned to be provisioned over the future 5G networks, video services are expected to be the most sensitive to network impairments, and their quality assurance is a significant factor for the wider penetration of 5G networks. In the small cell network (SCN) deployments, the mobile subscriber may experience degraded quality of the delivered service mainly due to the unreliable and unmanaged backhaul link. Thus, it has created the need by the mobile operators to consider Quality of Experience (QoE) aspects within the 5G architecture, allowing the monitoring, assessment and adaptation of the delivered service, aiming at the improvement of the delivered quality and the experience of the mobile subscribers.

Towards envisaging an appropriate in-service QoE assessment solution within the 5G architecture, the proposed SRRaaS method, which has been validated in FAT environment in the previous chapter, will be specially tailored for the needs of a small cell network in this chapter, performing SAT validation. The advantage of the proposed method is that the complex and power consuming process of video quality assessment is performed at the edge of the network, and not at the UE itself, thus significantly reducing the impact into UE's battery life. Nevertheless, it is equivalent to the corresponding video quality assessment at the UE, because, as earlier explained, the main reason for video quality degradation in 5G small cell architectures is due to the potentially congested backhaul link.

5.2 DEPLOYMENT OF THE PROPOSED SRR METHOD AT AN SMALL CELL ENVIRONMENT

As explained in section 4, the proposed video quality assessment metric, is a VNF, which can be decomposed in three functions, i.e. it comprises of three VNFCs, one of which is instantiated in the NFVI-PoP of the EPC, while the other two are instantiated in the NFVI-PoP of the Central Small Cell.

The target of a video quality assessment metric is to measure the video quality at the consumer's terminal, which in this case is the UE. Actually, SRR metric can run in a UE. However, the SRR running in the UE would consume much of the battery power because, as all video processing methods, it requires a lot of CPU power. Furthermore, it is not necessary to run SRR in the UE, because the main reason for quality degradation in SCNs is the backhaul link (which is usually over a non-reliable networks such as the Internet) and not the RF link between the SC and the UE. Thus, the measurement of the video quality in the SC is the same as in the UE. This makes the proposed method a useful tool for video service providers and mobile network operators that wish to provide video services over small cell networks, in order to assess the quality of the provided services and probably take appropriate measures.

The insertion of an NFVI-PoP between the EPC and the Central Small Cell has to deal with the GPRS Transport Protocol (GTP) [3GPP-TS29060]. The traffic exchanged between the EPC and the Small Cell is encapsulated using GTP, for the reasons of multiplexing and scalability. So, when an NFVI-PoP is inserted in this link, it has to perform both the GTP de-capsulation and re-encapsulation processes of the IP packets travelling from EPC to small cell and reverse. The de-capsulation is required to retrieve the pure IP packets, which are travelling from the EPC or the small cell to the NFVI-PoP, while the re-encapsulation is required for the pure IP packets from the NFVI-PoP to the EPC or the small cell. The implementation of the GTP en/de-capsulation processes, is also examined and a solution for deployment is provided in section 5.4.

5.2.1 REQUIREMENTS OF THE PROPOSED NFV-ENABLED SMALL CELL ARCHITECTURE.

The proposed NFV-enabled Small Cell system must satisfy specific requirements in order to be feasible the deployment at the edge of the proposed SRRaaS. More specifically:

- The proposed networking architecture must be SDN compatible in order for the traffic forwarding of the SRRaaS to work. The architecture's network must accommodate both the VNFs and the PNFs of the underlying system.
- The proposed networking architecture must be able to accommodate the hosting of the proposed SRRaaS in its corresponding VIM in a data center near the small cell. This requires the adoption of a local-breakout mechanism in order to route properly LTE network packets towards the data center.

5.2.2 PROPOSED NFV-ENABLED SMALL CELL ARCHITECTURE

As referred above in section 5.1, mobile network operators have realized the benefits of virtualisation and promote the development and deployment of NFV techniques in their networks. For the operators deploying small cell networks there are additional issues to be confronted. Applying NFV techniques to a small cell base station still necessitates a physical network function (PNF), which is responsible at least for supporting the RF functions of the base station. This requires a different architecture and ETSI-NFV and 3GPP have created an architectural framework that support combined PNF and VNF systems, a simplified version of which is shown in Figure 5.1. According to Figure 5.1, the small cell functionality is decomposed into physical (PNF) and virtual (VNF) network functions.

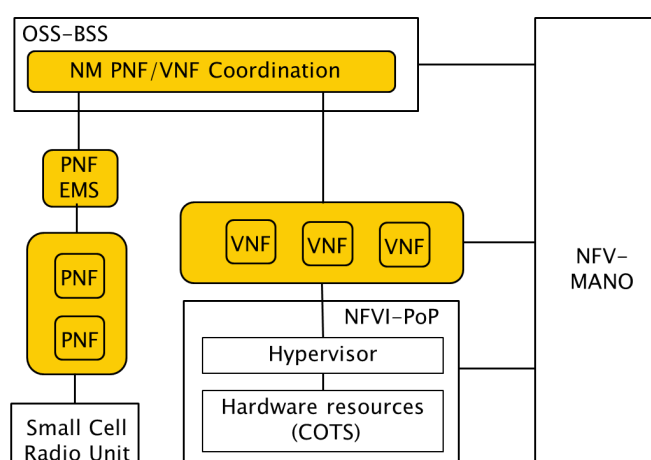


Figure 5.1 - Architectural framework for combined PNF and VNF systems

The application of the above architectural framework to small cells has been analyzed by SCF through several small cell virtualization use cases [SCF-2015-055], which examine the impact and benefits of virtualizing different layers and functions of a small cell. To facilitate the analysis, a small cell is split into two components; a Central Small Cell where functions are virtualized (VNF), and Remote Small Cell with non-virtualized functions (PNF), (see Figure 5.2). According to this approach functions are split in two types:

- Functions which are within L1, L2 and L3 layers, referred as PHY (Physical), MAC (Media Access Control), RLC (Radio Link Control), and PDCP (Packet Data Convergence Protocol).
- Functions that are not part of L3, L2 or L1 and include service functions, such as content caching, firewalls, QoS monitoring etc. and small cell management functions, such as RRM and Self Organizing Network (SON) features.

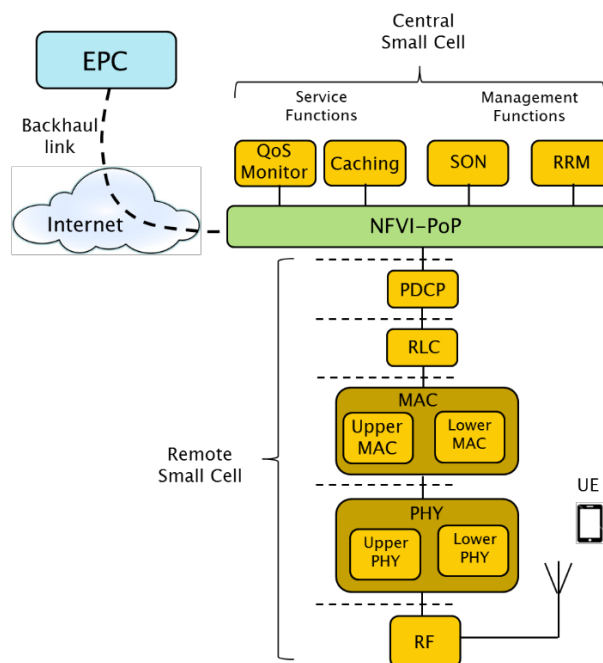


Figure 5.2 - Virtualization of a Small Cell and split of functions between PNF and VNF

The virtualization of small cell layers and functions are investigated with a bottom-up approach, where gradually more functions are moved from the remote small cell to the central small cell. The split points are shown as dashed lines in Figure 5.2. For example, PDCP may be either a PNF running in Remote Small Cell or a VNF running in the NFVI-PoP of Central Small Cell. A key differentiator for the split points is the front haul link (i.e. the link between the Central and the Access Small Cells) in terms of latency and bandwidths requirements.

In a more general architecture, there may be many NFVI-PoPs distributed among the EPC and the Central Small Cells. Also, a VNF may be comprised of more than one components (Virtual Network Function Components-VNFCs). Depending on the VNF and the network topology, the VNFCs may be instantiated within the same NFVI-PoP or in different ones.

Furthermore, among the ETSI NFV ISG [ETSI-NFV002] use cases, the concept of the Virtual Network Function as-a-Service (VNFaaS) is defined, which prescribes the provision to the customer of an end-to-end connectivity service (virtual network) along with embedded VNFs. Of course, the automatic deployment of a large number of VNFs comprised of many VNFCs requires an orchestrator to manage the location of the deployment of the VNFCs and their network interconnections.

5.3 PROPOSED NFV-ENABLED SMALL CELL ARCHITECTURE IMPLEMENTATION

This section describes the experimental testbed that was implemented for the validation of the proposed SRR method using real-time video streaming. As the

standards for 5G are currently evolving, there are not yet commercial implementations of 5G networks. So, the experimental testbed that was implemented for the needs of the thesis was an enhanced LTE network, supporting small cells and virtualisation capabilities. The testbed as Figure 5.3 shows, implemented a fully operational mobile network domain and it consisted of an EPC, a small cell, a UE, two NFVI-PoPs and a video server. The UE was implemented by a laptop equipped with a commercial LTE dongle, which included the appropriate USIM (UMTS Subscriber Identity Module) card. The experimental testbed was a full stack end to end network, where the UE is able to retrieve video content from the video server. As the SRR is implemented in the NFVI-PoPs, the usage of LTE equipment, instead of the not yet available, 5G equipment does not affect the validation or the performance evaluation of the SRR.

In order to implement SRR, the testbed was able to calculate the SSIM_{or} of the original video located at the video server, the SSIM_{tr} of the video at the small cell and by dividing these two values to calculate SRR metric.

The LTE network was built upon OpenAirInterface (OAI) [OPENAIR-2017] wireless technology platform. The OAI EPC software run on a 64 bit x86 based computer, while the small cell was implemented using a B210 Ettus card, installed on a similar computer, running the appropriate OAI eNode B software, which for our case was acting as the Remote Small Cell. The UE was based on a laptop equipped with 4G LTE USB Adapter, which included a USIM card with the appropriate keys stored in it, so that the UE could be authenticated by the EPC. The two computers (EPC and small cell) were interconnected over an S1 interface [3GPP-TS36413], which provided all the control signalling and data transport between EPC and eNode B.

As Figure 5.3 shows, the first NFVI-PoP was located between the video server and the EPC and the other was the Central Small Cell NFVI-PoP, located between the EPC and the Remote Small Cell. The virtualization platform of both the NFVI-PoPs were supported by the Openstack open source cloud computing platform. The release used was Liberty, which was the latest stable version during the time of the experimental tests.

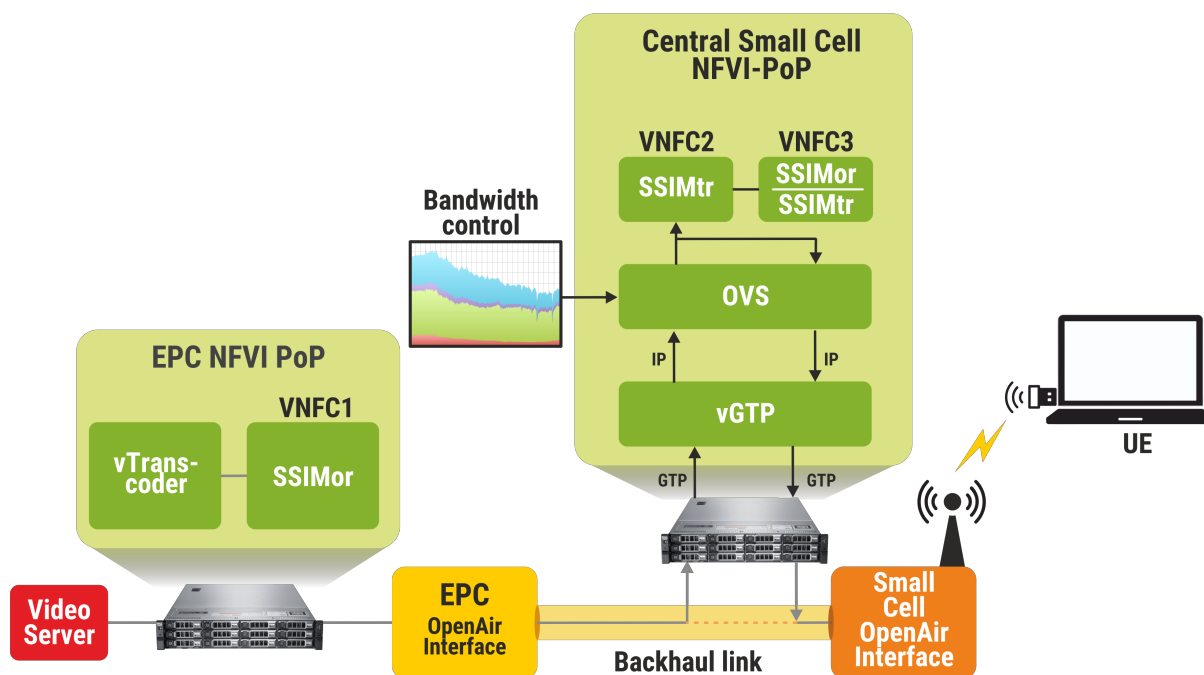


Figure 5.3 - An overview of the experimental testbed

Both NFVI-PoPs were capable of instantiating VNFs and performing also the appropriate network traffic steering, in order to support service chaining (i.e. the forwarding of the traffic seamlessly from one VNFC to the next) and finally to the UE. The video server in the testbed hosted the video files to be tested at their original encoded format. In order to stream the video files through the network and also adapt their bit rate to the desired level, a vTranscoder was instantiated at the EPC NFVI-PoP. The vTranscoder implementation was based on the widely used FFmpeg [FFmpeg-2017] and was instantiated as a VNF in the EPC NFVI-PoP.

The implementation of the proposed SRR method was split into three VNFCs, distributed between the two NFVI-PoPs: The first VNFC1 (figure 5.3), instantiated the calculation of the SSIMor (as explained in section 3) at the EPC NFVI-PoP. It also stored the calculated values in a local relational database (not shown in Figure 5.3 for simplicity reasons). The second VNFC2 instantiated the calculation of the SSIMtr (also see section 3) and was located at the Central Small Cell NFVI PoP. The third VNFC3 calculated the SRR as the ratio $SSIMow/SSIMtw$ for each frame. This VNFC was also responsible to HTTP REST-based request to the Redis (NoSQL) database, where the values of SSIMor were stored, as described in detail in section 4.4. It was located at the Central Small Cell NFVI-PoP, as well. The proposed method can be also realized as a VQA assessment method that can be virtualized as a network service. The proposed VNF already provides QoS and QoE feedback and can be complimentary to any system that provides multimedia content and wants to have additional QoE relevant input from the backhaul network. The SRRaaS can be deployed in a lightweight virtual machine, requiring at least 512 MB of memory and 1 virtual CPU. However, if the workload is increased significantly, it would require a larger amount of resources, to maintain a satisfactory quality of operation.

In order to perform the tests, a mechanism to vary the bandwidth of the backhaul link (and thus its quality) was required. For this reason, an Open vSwitch (OVS) was deployed in the testbed, as shown in Figure 5.3. It was an OVS licensed under the open source Apache 2.0 license. One of the functions that is able to perform is the control of the bandwidth of the IP traffic among its ports. An OVS was installed in the Central Small Cell NFVI PoP, controlling the traffic between the vGTP and the VNFCs. The deployed OVS was able to receive commands through OpenFlow protocol [McKeown-2008]. So, through OpenFlow commands to the OVS, it was possible to emulate conditions, where the backhaul link was congested and there was limited bandwidth for the video service, causing its quality degradation.

In the typical LTE architecture, the data traffic exchanged between the EPC and the Small Cell, is encapsulated using the GPRS Transport Protocol (GTP), as previously explained. On the other hand, the SRRaaS as well as the OVS require pure IP packets, so it is necessary for the traffic to be de-capsulated from its GTP headers and then forward the inner IP packets, which contain the actual video service data, to the upper modules. For the needs of the tests, a vGTP decapsulation and re-encapsulation software has been implemented, running on top of the widely used packet processing library PF_RING [PF_RING-2017].

The software was running as a VNF, in the Central Small Cell NFVI-PoP forwarding the traffic both directions. It also passed through all the control and data traffic between EPC and small cell, thus preserving the connectivity of the two nodes. The video traffic was filtered from the rest of the control traffic, the GTP header was removed from the filtered packets, which were then forwarded to the OVS. There, a bandwidth regulation rule was applied and the output IP traffic was forwarded to the VNFC2 for calculating the SSIMtr value. The bandwidth regulated IP traffic, was sent back to the vGTP through OVS, where it was re-encapsulated with the valid GTP header and further on forwarded to its original path, to arrive at the UE. As the decapsulation and re-encapsulation operations can introduce a penalty to the performance of the system, it was performed in a parallel manner to the rest of the process. The control and the rest of the traffic were forwarded using the zero-copy PF_RING library, and only the video service packets were copied to memory, as they needed to be further processed by a GTP agnostic mechanism. The GTP header storage for the re-encapsulation is considered insignificant as the GTP header, merely allocates 8 bytes to the memory.

The functionality provided by the vGTP was vital to the proposed framework, as it handled GTP traffic and delivered it in a valid IP format to the SRRaaS VNFCs to process it. The vGTP enables the integration of the Small Cell architecture and environment into a multimedia-over-IP environment seamlessly, by handling the Small Cell GTP traffic.

5.4 PERFORMANCE EVALUATION OF SRR IN SMALL CELL NETWORKS

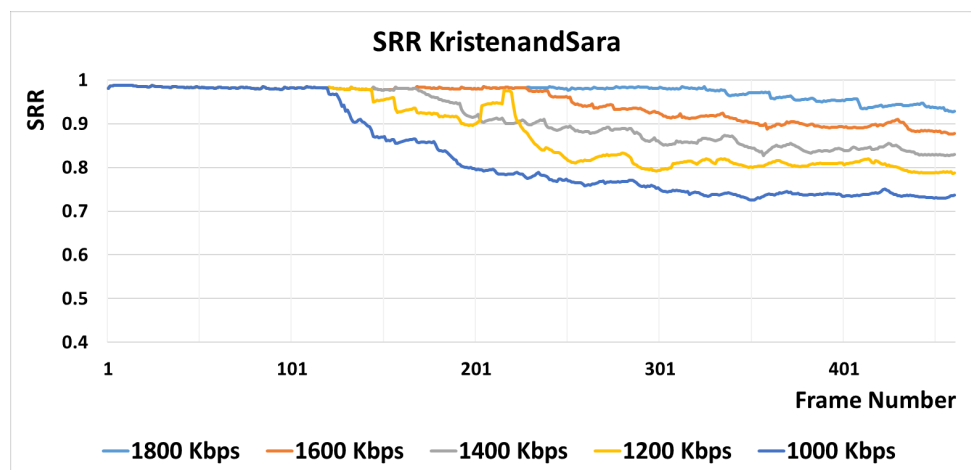
The purpose of this section is to test the capability of the proposed SRR method to detect the degradation of the video quality, when the backhaul link becomes congested and its bandwidth is restricted.

For the experimental needs of this section, two reference video signals were selected: the KristenandSara and the BasketballDrill sequences, which can be found from database [XIPH-MEDIA]. They were selected as representative ones for the two extreme cases in video content categorization: the former with low spatial and temporal activity (KristenandSara, a talk show) and the latter with high activity (BasketballDrill, a basketball play).

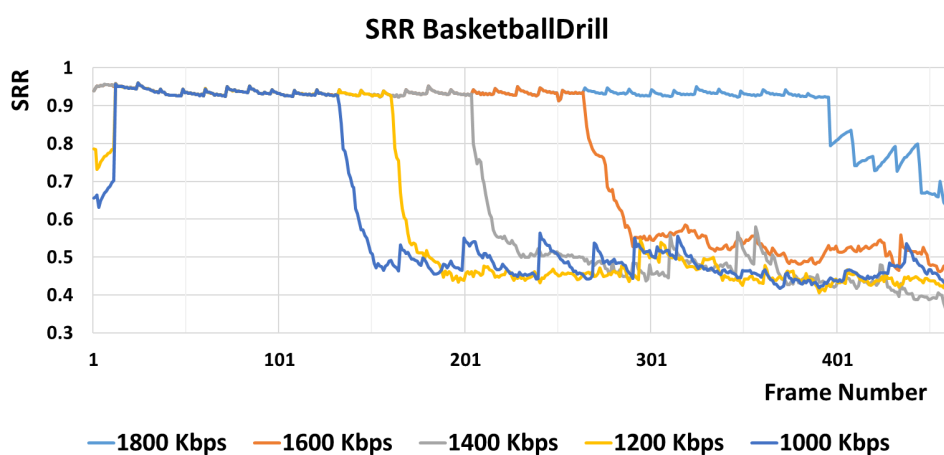
The two reference videos were stored in the video server of the testbed. The vTranscoder transcoded each video sequence to MPEG-4 format with bit rate of about 2.000 Kbps. The transcoded videos were considered as the reference signals for the video quality assessment.

For each video, the vTranscoder initiated a unicast UDP streaming over the experimentation testbed with destination IP, the IP address of the UE. The video stream passed through all the VNFCs described in section 5.3, where the appropriate calculations were performed. Finally, the value of SRR was calculated for each frame. The unicast streaming was repeated 5 times with the same bit rate (2.000 Kbps) for each video and each time a different backhaul bandwidth was set to the OVS, through OpenFlow commands. For the experiments, the bandwidth ranged from 1800 Kbps to 1000 Kbps at a step of 200 Kbps.

The results are shown in Figure 5.4 (a) and (b) for KristenandSara and BasketballDrill, respectively. This Figure shows a plot of SRR vs the sequence number of the frames, for various values of the bandwidth for the backhaul link.



(a)



(b)

Figure 5.4 - SRR variation vs frame sequence number without source adaptation for videos (a) KristenandSara and (b) BasketballDrill for various values of the backhaul bandwidth

From Figure 5.4 it is evident that after a short delay, which is explained in the following, the video quality is reduced, when the backhaul link is degraded. In the case of low spatial and temporal activity, Figure 5.4 (a) shows that the quality is associated to the bandwidth reduction. For example, the SRR drops to 0,9 for bandwidth 1.800 Kbps, while it drops to 0,7 for bandwidth 1.000 Kbps, i.e. the lower the bandwidth the lower the quality of the video. However, in the case of high activity video, Figure 5.4 (b) shows that the quality drops significantly, without being linked to the bandwidth reduction.

This is due to the different statistical significance of each frame between the two test signals in the decoding process. More specifically, for low dynamic video signals, each frame differs very little from the next one, which results to very small residual information. Thus, the loss of such a frame does not affect significantly the decoding process and thus the quality is degraded gradually (in an analogous manner to the

decrease of the backhauling bandwidth). However, in the case of high dynamic video content, each video frame contains significantly higher residual information, which means that it is much more important for the proper decoding process. Thus, a loss of such a frame, which contains significant residual information, affects seriously the decoding process, causing error propagations and decoding artifacts that cause the quality to drop very fast.

From both Figures 5.4 (a) and (b) it is evident that initially, and although the bandwidth reduction command has been applied to the OVS, the video quality remains high and it drops after a delay of a few hundreds of frames. This is because the OVS follows the Leaky Bucket algorithm [Turner-1986] to limit the bandwidth. This means that there is a buffer which receives the IP packets to be sent and forwards them to their destination at a constant bit rate, determined by the bandwidth limit. At the beginning, while the buffer is not full, all IP packets received at the buffer are forwarded to their destination. So, there is a short time interval (until the buffer is full) where the bandwidth limitation is not applied (no packets are dropped) and SRR remains unaffected. After some time (or equivalently some frames) the buffer becomes full and bandwidth decreases, causing the gradation of the video quality. It is evident that the lower the bandwidth, the shorter is the time required to fill the buffer, as verified in Figure 5.4.

5.5 CONCLUSIONS

The work presented in this section focused on the applicability, implementation and performance evaluation of SRR to next generation (5G) mobile networks, following the small cell deployment architecture, enhanced with virtualisation capabilities. An enriched LTE-small cell experimental platform has been implemented, to test and evaluate the performance of the proposed method. The testbed has been enhanced by two NFVI-PoPs, which host the three VNFCs, out of which the SRR is comprised. The problem of GTP en/decapsulation process of the data packets in S1 link is analysed and a solution has been implemented. The performance of SRR is evaluated when the bandwidth of the backhaul link is reduced. The experimental results show that the proposed method is able to detect successfully the video quality reduction when the backhaul link is degraded. Another conclusion is that the video quality reduction is associated to the bandwidth variation for low activity videos, while it is not linked to the bandwidth variation, for high activity videos. The proposed SRR method can be offered as a Service to the mobile networks operators and it provides them with a tool to monitor their customers' satisfaction.

6 CONCLUSIONS AND FUTURE LINES

6.1 MAIN CONTRIBUTIONS

Over the recent years, with the advent of multimedia applications and their increasing popularity, a need for novel design and implementation ideas was created. In the networking field the paradigms of NFV and SDN have opened a new path of reshaping existing infrastructure and further extend its capabilities. This was the motivation for the birth of 5G in the telecommunications area, which proposed a holistic approach to improve and accelerate our current network. However, user experience still remains an important factor in next generation networks, and future technologies in general. Thus, the assessment of QoS and QoE for multimedia services plays an important role in the design and implementation of network systems.

Video Quality Assessment methods enable the network operators to measure the end users' perceived quality and thus adapt the available network resources to satisfy the users' expectations. Although subjective VQA methods are very accurate in expressing the end users' satisfaction, they involve an audience of people to watch the video under examination and collect their opinions, which are processed at a later stage, to evaluate the quality. Such a no-real-time process is not useful for networks operators, which want to know in real time the quality of a video distributed to their customers.

So, the scientific community has devoted particular attention in developing objective metrics that automatically evaluate the quality of the video, trying to model the subjective analysis of humans. Such a real time process is very useful for network operators. Among the two main categories of objective methods the FR one is more accurate, but still not very appropriate for implementation within the operator's network, because they need both the original and the distorted video signal at the end user's terminal, in order to perform the comparison. The RR methods are less accurate but they are appropriate for implementation in commercial video distribution networks, because they required only a few features extracted from the original video at the end user's terminal to assess the quality. Another requirement for the VQA methods is to be easily applicable to modern emerging and future networks.

Initially the thesis provided a survey on various categories of video quality assessment methods and analysed their advantages and disadvantages. From the analysis it is deduced that in order to increase accuracy in VQA, more information from the original video signal is required at the end user's terminal. In this case, the required bandwidth and the complexity make such methods less attractive for application in commercial video distribution networks. On the other hand the less information from the original signal the less accuracy, but the applicability in commercial networks is increased. Therefore it is highlighted the compromise between accuracy and appropriateness for commercial video distribution networks.

Challenges	Objectives	Contributions
<p><u>Challenge 1</u></p> <p>Define an efficient yet accurate objective VQA method: This thesis will address this challenge by proposing a VQA method that is based on the evaluation of a FR metric (Structural Similarity Index Metric-SSIM), but it is also suitable to be implemented in video distribution networks, since it does not require the original video signal as the user's site, but only a single feature extracted from it.</p>	<p><u>Objective 1</u></p> <p>To define, implement and evaluate an accurate and efficient Reduced Reference video quality evaluation metric, based on the calculation of the Structural Similarity Index Metric (SSIM), which is a par excellence FR metric.</p>	<p>The thesis addressed Challenge 1 by proposing a VQA method that is based on FR metric SSIM, and is also suitable to be implemented in video distribution networks, since it does not require the original video signal as the user's site, but only a single feature extracted from it. The proposed method is presented and evaluated thoroughly in Section 3.</p>
<p><u>Challenge 2</u></p> <p>Ensure the applicability of the proposed method in NFV/SDN environments. Most VQA methods do not take under consideration their applicability in virtualised environments, based on NFV/SDN technologies.</p>	<p><u>Objective 2</u></p> <p>To implement the proposed SSIM Reduced Reference (SRR) metric as a VNF, suitable to be instantiated in an NFV infrastructure (NFVI), and assess its performance.</p>	<p>The thesis addressed Challenge C2 (efficient deployment) by implementing the proposed method as a VNF, which can be instantiated in a virtualized Network Function Virtualization Infrastructure (NFVI). Furthermore, being implemented as a VNF the proposed method was evaluated in an actual virtualized testbed, by monitoring the video quality in real-time and across the service provision chain. The integration and evaluation process of this contribution are presented in detail in Section 4.</p>

<p><u>Challenge 3</u></p> <p>Analyse how multimedia applications should be coupled and interwork with the 5G network components and how to deliver a fully operational 5G-ready NFV enabled system with QoS/QoE capabilities. Little effort has been allocated on the multimedia applications and services that will make use of and exploit advanced 5G network capabilities.</p>	<p><u>Objective 3</u></p> <p>To implement and evaluate the performance of the proposed SRR method in a 5G network, based on small cells topology, enhanced with virtualisation capabilities, in order to deliver a fully operational 5G NFV enabled system with QoE capabilities .</p>	<p>The thesis addressed Challenge C3, by integrating the proposed SRRaaS in a 5G network architecture based on small cells, supported by a virtualised environment. The main reason for this selection is that any VQA method is a complex and power consuming process, so it is better to be performed at the edge of the network, rather than at the UE itself, because it will significantly improve the overall network provision of the system. The integration of SRRaaS in 5G is carried out in Section 5, along with the evaluation results in 5G-ready environment.</p>
--	--	--

The thesis addresses all the above challenges by proposing a new RR VQA method, which not only requires minimal original video signal data to operate, but also offers a performance comparable to FR VQA metrics. The method combines the advantages of both FR and RR methods. It has an accuracy compared to FR (as it is based on the calculation of the SSIM) and the applicability of RR methods in video distribution networks as it requires only a single feature from the original video signal to be available at the end user's terminal : the SSIM metric of the original video with a white video reference pattern. At the end user's terminal the received video is compared frame by frame with a similar white reference pattern and the ratio of the two values is a very accurate estimation of SSIM metric.

The performance of the proposed SRR method was then evaluated using a large experimental set of 40 reference and non-reference video sequences, with a variety of spatial resolution ranging from CIF up to HD. Though these experimental measurements it is shown that proposed SRR metric has a value very close to the SSIM index, with a Mean Absolute Percentage Deviation lower than 2.56 %. Experiments also show that the accuracy of the proposed method increases with the quality of the video signal. In this respect, accuracy ranges from 0.62% (for QP=12) to 2.56% (QP=32), which represents the worst case performance. Thus, the proposed method maintains satisfactory performance across all the potential range of QP values, although better accuracy is achieved at higher video qualities (i.e. lower QP values).

Furthermore, the performance of the proposed method was compared to a set of other VQA algorithms, namely RR-LHS, J.246, PSNR, Yang's RR VQA, VSNR, VQM, SSIM and RR metric. In order to achieve this, the LIVE Video Quality Database was employed. The performance evaluation metrics used to compare the performance of the proposed SRR method with the above ones, were LLC and SROCC. Results show that the accuracy of the proposed SRR method is better than RR-LHS, J.246, Yang's RR VQA, both in terms of

LCC and SROCC, except from RR metric, which provides better results. Additionally, the accuracy of the proposed SRR method is better than the FR methods PSNR and VSNR in terms of LCC, while it is slightly lower than the performance of VQM and SSIM index, as expected, due to the reduce reference nature of the proposed methodology. In terms of monotonicity (SROCC), the proposed method performs better than the PSNR, but lower than the rest VQA methods, without however significant deviating from their performance range. From the above analysis it is evident that the proposed SRR method ranks very well among other VQA methods.

In order to enhance the comparison, a more qualitative evaluation has also been elaborated, based on a visual presentation of the correlation between SSIM and the proposed SRR method. Plots of SSIM and SRR vs the frame numbering demonstrate that :

- the accuracy is higher at better qualities
- SRR follows the variations of SSIM for both low and high video quality levels
- SRR follows very reliably the abrupt variations of SSIM, as well as the wide variations, independently of the video quality.

A qualitative interpretation of the proposed method is also elaborated. From previous works it has been shown that the variation of SSIMot vs QP is an exponential declining function. As explained in section 3, the variation of SSIMtr vs QP is an exponential ascending function with the same slope as the previous. Considering that for QP very low, the target video resembles to the original and for QP very high it resembles to the white reference pattern, it is deduced that the product of SSIMot and SSIMtr is a constant equal to SSIMor. Therefore, SRR is equal to SSIM. Since the variations of SSIMot and SSIMtr are not ideally exponential, this deviation affects the performance of the proposed method.

Another issue that this thesis addresses is the investigation of the performance of the proposed SRR method when non-white video reference signals are used. Experimental results with black, red, green, blue, yellow aqua, pink and gray colors, show that the alternative color patterns offer a satisfactory performance. It was also observed that the performance with colors sometimes is better than the white one and sometimes worse, depending on the content of the video under evaluation.

Another advantage of the proposed method is the required bit rate to send the reference signal to the end-user. Since it is only a single number less than 1, the required bit rate ranges between 400-600 bps and it is significantly lower than the 15kbps to 256kbps, required for other RR methods.

No matter how efficient a VQA method can be, it is important to examine its applicability in modern telecommunication networks. Addressing this challenge the thesis investigated the implementation of the SRR method in an NFV/SDN environment and evaluated its performance. The SRR method was decomposed in three functions, two of which were grouped. So the method was implemented with two VNFs, capable to be instantiated in a virtualised infrastructure. An experimental testbed was employed,

based on SDN/NFV technologies. It included two NFVI-PoPs, where the above two VNFs were instantiated. The testbed was implemented on an Openstack platform and video traffic steering was achieved through Openflow capable OVSs under the management and control of the OpenDaylight SDN controller. The complex L2/L3 traffic steering process and associated issues were analysed in the thesis and a solution was implemented using the SDN mechanisms to alter per hop the values of the destination IP address, destination MAC address and source IP address.

The objective was not only to demonstrate the applicability of the proposed SRR method as VNFs over virtualized infrastructures, but also to evaluate its performance. Among the main goals of this thesis is to offer the proposed SRR method as an in-service QoE metric which can be instantiated as a VNFaaS in any point within the video distribution chain. This is a very important feature for network operators, which enables the identification of the location within their network, where a video quality degradation takes place and help them to efficiently apply specific remedy actions to restore the video quality. In this respect, the response time of the proposed method to an abrupt variation of the video quality is an important performance metric.

A test scenario was designed and executed over the testbed, which included :

- The initial configuration of the OVSs, in order to address all networking issues and properly steer the video traffic from the video server, through the VNFs, from one NFVI-PoP to the other and finally to the end user's terminal.
- The instantiation of the appropriate VNFs at the service provider's and at the end user's NFVI-PoPs , which enable the implementation of the SRR method over the testbed
- The measurement of the video quality through SRR, in real time
- The insertion of background traffic in the connection link between the two NFVI-PoPs, in order to emulate the video quality degradation, which is caused by network flooding with IP traffic. Iperf was used to generate synthetic background UDP traffic
- The measurement of the response time of the SRR metric to the degradation of the video quality
- The removal of the background traffic
- The measurement of the response time of SRR to the restoration of the video quality

Upon the execution of experimental sets of ten repetitions, results show that the response time of SRR to an abrupt video degradation is about 1 sec. In terms of video quality, the respective average QoE level (as measured by the SRR metric) drops from the approx. 0.83 value down to 0.21. On the other direction, the response time to an abrupt video quality restoration is about 3 sec. Video quality as measured by SRR metric, increases from 0.22 up to its previous level of 0.83.

Between the two response times (video degradation and reinstatement) the first one is more important, because it is more useful to know the video quality degradation as soon as possible and timely apply actions that restore it, at a certain level. Such actions can be the transcoding of the video in lower bit rate or sending the video traffic through another

network path. The timely notification and reaction minimizes the viewer's dissatisfaction.

5G is the next big breakthrough of telecomm operators and research efforts are mainly focused on advances in radio and network technologies. However, less effort has been allocated on the multimedia applications and services that will make use of and exploit advanced 5G network capabilities. Addressing the challenge of how multimedia applications should be coupled and interwork with the 5G network components, the thesis investigated the implementation and performance evaluation of the SRR method in 5G networks, supporting small cells and virtualisation capabilities. In such an architecture, SRR is not implemented in the UE itself, but in the small cell. The SRR metric as extracted from the small cell is equivalent to the measurement at the UE, because the main reason for quality degradation in small cells is the backhaul and not the radio link. Additionally, the battery consumption of the UE is reduced.

As there are not yet available 5G commercial equipment an experimental testbed was deployed based on an enhanced LTE network, supporting small cells and virtualisation capabilities. The testbed was a full stack end to end network, consisted of an EPC, a small cell, a UE, two NFVI-PoPs and a video server. The VNFCs that comprise the SRR method were instantiated in the two NFVI-PoPs, which are based on openstack. The issue of GTP tunnelling en/decapsulation process that is applied on the data packets in S1 link is analysed and a solution was implemented.

The objective was to evaluate the performance of SRR method for an imperfect backhaul link, which is the main reason for video quality degradation in small cell architectures. The performance was evaluated for two reference video signals, the former with low and the latter with high spatial and temporal activity. Also, a vTranscoder transcoded each video to MPEG-4 format with a rate of about 2.000 Kbps. Using an OVS, the bandwidth of the backhaul link ranged from 1800 Kbps to 1000 Kbps at a step of 200 Kbps. Experimental results showed that for the low spatial and temporal activity video, the quality degradation is associated to the bandwidth reduction, i.e. the lower the bandwidth the lower the quality of the video. However, for high activity video the quality dropped significantly, without being linked to the bandwidth reduction. This difference is explained by the fact that frames in low dynamics videos differ very little and thus, frame losses do not affect significantly the decoding process and the quality is degraded gradually, as frame loss increases. However, in high dynamics videos the frames differ significantly and thus a frame loss affects seriously the decoding process, causing the quality to drop very fast.

In commercial video distribution networks, where the customer's QoE has to be retained as high as possible, a significant video quality reduction would not be allowed. When the backhaul link degrades specific actions need to be taken. In order to adapt to such a degradation event, further experiments were performed using a video source rate adaptation mechanism. Experimental results showed that when the source video signal is transcoded to a bit rate appropriate for the capacity of the degraded backhaul link, the video quality degradation was much lower. More specifically, when the low activity video was transcoded, SRR remained higher than 0.98, even for bit rate as low as 1000 Kbps.

Similarly, the SRR of the high activity video was higher than 0.94, even at 1000 Kbps bit rate.

The proposed SRR method can be offered as a Service to the mobile networks operators and it provides them with a tool to monitor their customers' satisfaction. An additional advantage of the proposed method is that the complex and power consuming process of video quality assessment is performed at the edge of the network, and not at the UE itself, thus significantly reducing its power consumption.

6.2 FUTURE LINES

This section provides an overall analysis of the future research activities and next steps, which can be based on the presented thesis.

- Future challenges related to NFV and 5G with regards to the presented method.
- MANO orchestration services with regards to VQA in NFV and 5G.
- Initial experimental results of a Forward Error Correction (FEC) mechanism with source rate adaptation based on the proposed VQA method and measured on 5G Small Cell Network.

The thesis proposed a new VQA method with a very good accuracy, which can be implemented as a VNF and be offered as a service over NFV infrastructures and 5G networks. The thesis presented how the SRR method can be instantiated in a virtualized infrastructure but does not refer to the interaction of this VNF with the upper layers of management and control of the virtualized environment.

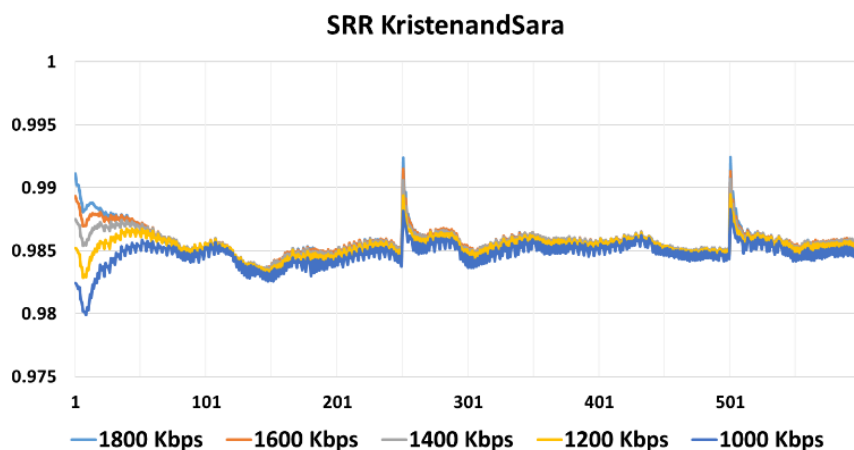
One of the primary challenges associated with NFV is the automated management of the service lifecycle. This can be achieved by the Management and Orchestration (MANO) system, which operates with OpenStack and OpenDaylight controllers and has the in-built functionality to automate the key phases of the NFV service lifecycle, like service mapping, service deployment and monitoring. So, a future work, beyond the scope of this thesis, is the development of the VNFs that comprise SRR to be compatible with MANO architecture. So, the VNFs under consideration must be accompanied by the proper descriptors that will provide the MANO with all the required resources that are needed for the instantiation on these VNFs in the appropriate NFVI-PoPs.

A basic challenge to be tackled by a MANO system is automated deployment. Network functions are no longer bound to a physical machine; instead they reside in a shared resource environment. As a consequence, various considerations arise when it comes to the deployment (VNF "placement") decision. In the specific case of SRR, the placement of the appropriate VNFs should match the requirements of the network operator to monitor the video quality at specific locations within his network, in conjunction with the location of the existing NFVI-PoPs and the distribution of the end users.

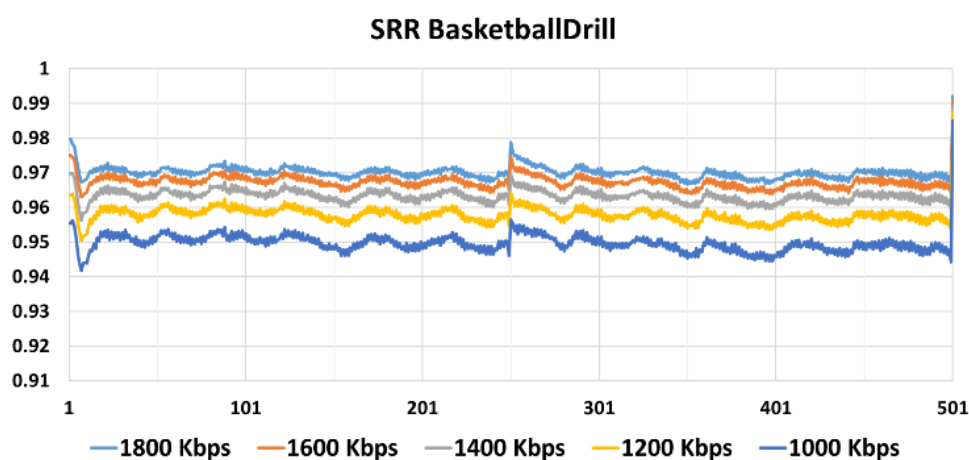
The existing placement algorithms are already complex. As the hosting computing infrastructures become more heterogeneous with various add-on features enabling hardware acceleration (i.e. PCIe, GPU) which can support various system enhancements and capabilities, this increases the complexity of the VNF placement decision significantly. The inclusion of additional placement requirements for monitoring VQA will increase the complexity of the MANO orchestration service, even more. However, it results in more optimal resource allocation, increased VNF performance and better VQA measurements to fulfill the needs of the network operators. So, a line for future work is towards enhancing MANO placement algorithms to include VQA aspects.

Another important feature of MANO is monitoring. Monitoring not only covers NFVI resources, but also the status of VNF services. Comprehensive monitoring also facilitates service mapping, billing/charging, and SLA conformance/violation. For all the aforementioned reasons, the development of an integrated monitoring framework for NFV, collecting metrics from physical infrastructure resources as well as specific virtualized services metrics, such as VQA, is considered crucial for any NFV infrastructure. An effective NFV monitoring framework should expose a holistic awareness of the status and performance of the deployed services as well as the underlying infrastructure to all management entities in order to allow the latter to take proper and timely decisions. In this context, a future line is to include the SRR metric in a more general NFV monitoring framework, which will provide VQA metrics to the MANO, which in turn will enforce remedy actions. Especially in 5G networks, metrics collected from the radio/cloud/software elements, along with the SRR metrics will enable a more efficient management of 5G network resources and higher QoE for end users.

Another future direction of the proposed SRRaaS method in a commercial deployment in the provision in a closed loop mechanism. More specifically, in this future approach the video quality would not be allowed to drop significantly, in order to retain the QoE of the customers. A mechanism of source rate adaptation, loss protection (e.g. FEC) or error concealment techniques would be applied, to keep video quality degradation as small as possible. In order to adapt to such an event, further experiments were performed, in the presence of a source rate adaptation mechanism through a transcoder. So, for various backhaul bandwidths ranging from 1800 to 1000 Kbps, the initial video source rate of 2000 Kbps was transcoded down to the corresponding bit rate to match the available bandwidth and SRR was evaluated. The results are shown in Figure 6.1 (a) and (b) for KristenandSara and BasketballDrill, respectively. Figure 6.1 shows a plot of SRR vs the sequence number of the frames, for various values of the bandwidth of the backhaul link, when the video rate matches the available bandwidth. From figures 6.1 (a) and (b) it is evident that the video quality is slightly degraded and in the case of KristenandSara remains higher than 0.98, while for BasketballDrill is higher than 0.94.



(a)



(b)

Figure 6.1 - SRR variation vs frame sequence number with source adaptation for videos (a) KristenandSara and (b) BasketballDrill for various values of the backhaul bandwidth

6.3 ACKNOWLEDGEMENTS

The research activities leading to the results presented in this document have been performed in the Networking, Quality and Security (NQaS) research group of the University of the Basque Country (UPV/EHU), in the scope of the European project SESAME.

Beyond the institutional acknowledgements, the credits of the document must include several colleagues who assisted and participated in the research study of this document:

- To Dr. Fidel Liberal, as PhD supervisor and principal researcher of the cited projects, who actively and passionately participated in the whole process of conception, development and identification of the main results, and also provided valuable editorial reviews to the document.

- To Dr. Harilaos Koumaras, who collaborated and provided useful guidance in various research papers and supported the entirety of the work.
- To George Xilouris, networking zen master who contributed in various papers, or provided meaningful comments and suggestions throughout the study.
- To Dr. George Gardikis, who contributed in various steps of the process with helpful contributions and positive feedback.

The PhD applicant is also thankful and grateful to his family and friends for all their support during the fruitful and interesting PhD years.

7 PUBLICATIONS OF THE AUTHOR

Table 7.1 - Resulting publications and relation to the PhD work

Scientific publication	Relation to the PhD work
<p>Journal Paper (Section 3)</p> <p>Kourtis, M.-A.; Koumaras, H. & Liberal, F.</p> <p>Reduced-reference video quality assessment using a static video pattern</p> <p>Journal of Electronic Imaging, SPIE-Intl Soc Optical Eng, 2016, 25, 043011</p>	<p>This paper presents the novel RR VQA method SRR, which is the core topic of this thesis. The proposed algorithm is evaluated extensively over its performance against related VQA metrics, and its correlation to the SSIM VQA metric.</p>
<p>Conference Paper (Section 4)</p> <p>Koumaras, H.; Kourtis, M.-A.; et al.</p> <p>In-service Video Quality assessment based on SDN/NFV techniques</p> <p>2016 23rd International Conference on Telecommunications (ICT), IEEE, 2016</p>	<p>This paper presents the implementation of the proposed RR VQA method over an NFV/SDN environment. The presented method is evaluated and measured for its performance over a heterogeneous environment based on a multimedia scenario.</p>
<p>Journal Paper (Section 4)</p> <p>Kourtis, M.-A.; McGrath, M. J.; Gardikis, G.; Xilouris, G.; Riccobene, V.; Papadimitriou, P.; Trouva, E.; Liberati, F.; Trubian, M.; Batalle, J.; Koumaras, H.; Dietrich, D.; Ramos, A.;</p>	<p>This paper presents the current trend of NFV in networking and proposes an architectural framework for VNFs. The manuscript not only demonstrates a functional NFV architecture, but also evaluates the performance of a prototype VNF under various workloads.</p>

<p>Riera, J. F.; Bonnet, J.; Pietrabissa, A.; Ceselli, A. & Petrini, A. T-NOVA: An Open-Source MANO Stack for NFV Infrastructures IEEE Transactions on Network and Service Management, Institute of Electrical and Electronics Engineers (IEEE), 2017, 14, 586-602</p>	<p>The contributions of the paper in regard to VNF performance serve as a basis for the latter integration of SRR into an NFV environment.</p>
<p>Journal Paper (Section 5) Kourtis, M.-A.; Koumaras, H.; Xilouris, G. & Liberal, F. An NFV-Based Video Quality Assessment Method over 5G Small Cell Networks IEEE MultiMedia, Institute of Electrical and Electronics Engineers (IEEE), 2017, 24, 68-78</p>	<p>This paper presents the integration of the proposed SRRaaS into a 5G enabled environment, along with its experimental performance evaluation. The proposed VNF is evaluated in a 5G environment. The manuscript demonstrates the modularity and of the proposed method in various heterogeneous environments, and its efficiency.</p>
<p>Journal Paper (Section 5) Rhodashenas, P. S.; Blanco, B.; Kourtis, M.-A.; Taboada, I.; Xilouris, G.; Giannoulakis, I.; Jimeno, E.; Trajkovska, I.; Fajardo, J. O.; Kafetzakis, E.; Lloreda, J. G.; Liberal, F.; Whitehead, A.; Wilson, M. & Koumaras, H. Service Mapping and Orchestration Over Multi-Tenant Cloud-Enabled RAN IEEE Transactions on Network and Service Management, Institute of Electrical and Electronics Engineers (IEEE), 2017, 14, 904-919</p>	<p>This paper proposes a Cloud-enabled architecture for Small Cells in the scope of 5G. Various topics in regard to 5G are covered in this manuscript, and more specifically the paradigm of MEC and the convergence of NFV to 5G. The work serves as a basis for the latter integration of the proposed SRRaaS into 5G. Additionally, the subject of GTP packet handling in 5G is discussed in detail, which is vital in the merging of NFV and Access technologies.</p>
<p>Journal paper (Section 4) Trajkovska, I.; Kourtis, M.-A.; et al. SDN-based service function chaining mechanism and service prototype implementation in NFV scenario</p>	<p>This paper presents an SDN-based service function chaining mechanism and demonstrates its functionality in an NFV enabled environment. The manuscript serves as an introduction to an NFV enabled VQA scenario.</p>

Computer Standards & Interfaces, Elsevier BV, 2017, 54, 247-265	
Patent (Section 3) Reduced reference method for video quality assessment	Issued on Sep 1, 2014. Patent issuer Industrial Property organization Greece with number ΔE1008606

8 ANNEX

This Annex provides a basic overview of basic concepts of Video encoding methods that are needed to fully understand the aspects addressed in this thesis.

8.1 VIDEO PROCESSING BASICS

Video coding is defined as the process of compressing and decompressing a raw digital video sequence, which results in lower data volumes, besides enabling the transmission of video signals over bandwidth limited means, where uncompressed video signals would not be possible to be transmitted. The use of coding and compression techniques leads to better exploitation and more efficient management of the available bandwidth.

Video compression algorithms exploit the fact that a video signal consists of sequence series with high similarity in the spatial, temporal and frequency domain. Thus, by removing this redundancy in these three different domain types, it is possible to achieve high compression of the deduced data, sacrificing a certain amount of visual information, which however it is not highly noticeable by the mechanisms of the Human Visual System, which is not sensitive at this type of visual degradation.

Thus, the research area of video compression has been a very active field during the last years by proposing various algorithms and techniques for video coding. As we mentioned earlier, video coding techniques compress the data volume of initial raw video signal with the cost of degrading the perceived quality of the video service. By enhancing the encoding algorithms and techniques, the latest proposed coding methods try to perform in a more efficient way both the data compression and the maintenance of the deduced perceived quality of the encoded signal at high levels. In this framework, many of these coding techniques and algorithms have been standardized, encouraging by this way the interoperability between various products designed and developed by different manufactures.

This chapter deals with the evolution of video coding process from the very early standards to the latest one, discussing all the fundamental concepts and ideas behind the video coding evolution. Following the evolution of the video compression efficiency, this chapter presents also the methods for video quality assessment, both subjectively and objectively.

Following this introductory section, the rest of this chapter is organized as follows: Section 2 describes the basic principles of video coding. Section 3 provides an analytical description of the video coding standards evolution, both in ITU-T and MPEG. Section 4 provides information on the video quality assessments methods that are available. Finally, Section 5 provides some future trends, while Section 6 concludes this chapter.

8.2 PRINCIPLES OF VIDEO CODING

This section presents the basic principles of video coding procedure, which the reader must understand in order to be able to continue further reading in this chapter.

All forms of video-coding that have compression as a primary goal, try to minimize redundancy in the media. A video consists of a number of frames, meaning separate pictures, which given the fact they are projected one after the other at a particular rate, they give the human eye the feeling of continuous movement. This leads us to the fact that we can have 2 kinds of redundancy, spatial and temporal. Spatial redundancy refers to intraframe coding techniques, which means that we use neighboring similar pixels of the same frame to encode it. Temporal redundancy has to do with interframe coding, meaning the usage of past and future frames to encode our current frame.

Therefore video compression techniques are divided into 2 categories based on the redundancy type. The temporal phase exploits the similarities between successive frames with scope to reduce the temporal redundancy in a video sequence. The spatial stage exploits spatial similarities located on the same frame, reducing by this way the spatial redundancy. Then the output parameters of the temporal and spatial stages are further quantized and compressed by an entropy encoder, which removes the statistical redundancy in the data, producing an even more compressed video stream. Thus, all the video coding standards are based on the same basic coding scheme, which briefly consists of the following phases: The temporal, the spatial, the transform, the quantization and the entropy coding phase.

Finally, it must be also noted that in more simplistic systems every frame is coded separately, so the intraframe redundancy method is the best, because the loss of one frame does not affect the coding of the other frames. Due to the simplicity of these systems, this frame specific methodology is not analysed in the chapter, because it is a part of the more complicated coding systems that use both spatial and temporal techniques in order to achieve greater effectiveness and efficiency in video compression ratio.

8.2.1.1 Compression at the Temporal Plane

As input to the temporal stage of the encoding process, the uncompressed video sequence is used, which contains a lot of redundancy between its successive frames. The scope of this stage is to remove this redundancy by constructing a prediction of the each frame based on previous or future frames, enhanced by compensating for fine differences between the selected reference frames. Depending on the prediction level, by which each frame is constructed, each frame is classified to three discrete types, namely:

Intra-frame (I), Predictive (P) and Bidirectional predictive (B), widely referred as I, P and B. The I frames are also called Intra frames, while B and P are known as Inter frames.

- I frames do not contain any prediction from any other coded frame.

- P frames are coded based on prediction from previously encoded I or P frames.
- B frames are coded based on prediction from previously or future encoded I or P frames.

The pattern of successive types of frames like IBBPBBPBBP... forms a Group of Pictures (GOP), whose length is mainly described by the distance of two successive I frames (see Figure 8.1).

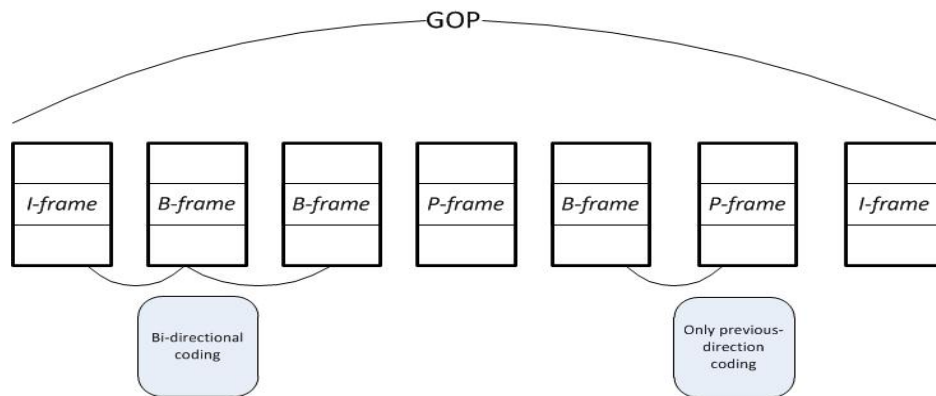


Figure 8.1 - GOP Structure

Therefore, in order to perform this temporal compression, two discrete processes are performed at this stage: The motion estimation and motion compensation. Both these processes are usually applied on specific rectangular regions of a frame, called blocks if their size is 8x8 pixels or MBs if they are 16x16 rectangular pixel regions. At the latest standards (i.e. H.264) as Figure 8.2 depicts, variable block sizes are used for motion compensation depending on the content, achieving better coding efficiency.

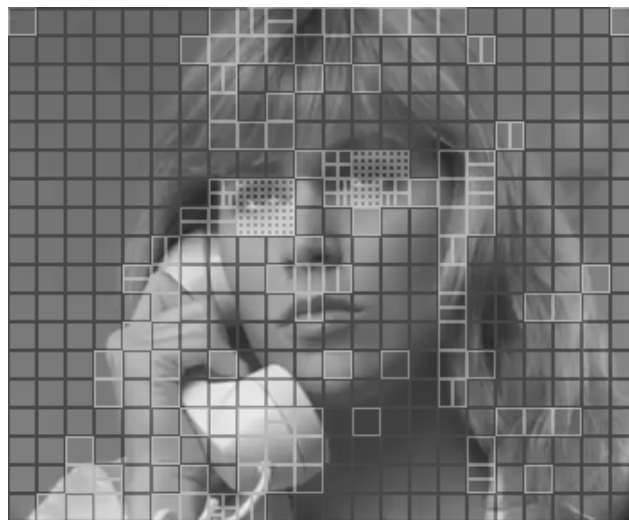


Figure 8.2 - Example of variable block size coding

During motion estimation, the encoding algorithm searches for an area in the reference frame (past or future frame) in order to find a corresponding matching region. The process of specifying the best match between a current frame and a reference one,

which will be used as a predictor of the current frame, is called motion estimation. This is performed by comparing specific rectangular areas (i.e. blocks/MBs) in the reference and current frame, until the best match to be detected. Due to this, their spatial differences are calculated, using the Sum of Absolute Differences (SAD) or Sum of Absolute Errors (SAE), which is defined as:

$$SAD(d_x, d_y) = \sum_{i=0}^{15} \sum_{j=0}^{15} |f(i, j) - g(i - d_x, j - d_y)|$$

where $f(i, j)$ and $g(i, j)$ denote the luminance pixels of the current rectangular area (in this case a MB) and the reference one respectively. The reference area is relatively defined by the current one using the motion vectors (d_x, d_y) , denoting the position of the best matching region (see Figure 8.3).

When the best match has been performed, then the motion compensation follows. During this process the selected optimal matching region in the reference frame (i.e. the region that sets the SAD minimum) is subtracted from the corresponding region in the current frame with scope to produce a luminance and chrominance residual block/MB that is transmitted and encoded along with the reference motion vectors. The deduced frame by the motion compensation process is called residual frame, which contains the result of the subtraction of the reference regions from the corresponding ones of the current frame. In the residual frame the static areas correspond to difference equal to zero, while darker areas denote negative differences and lighter areas positive differences respectively. A typical example of a residual frame is represented in Figure 8.4.

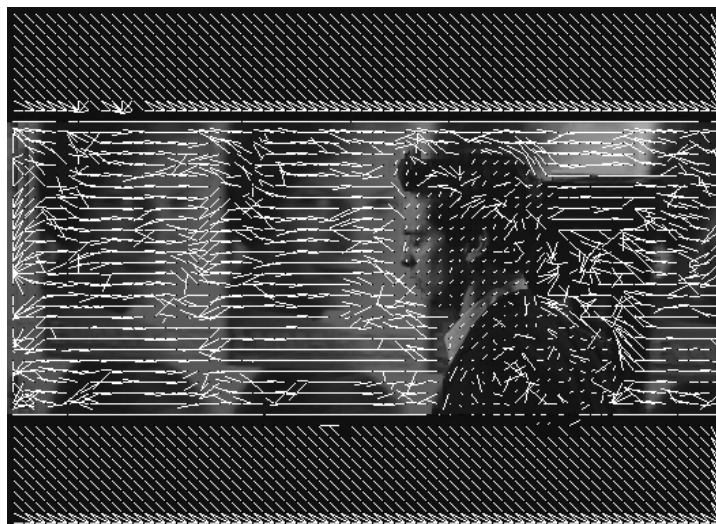


Figure 8.3 - A frame where motion vectors appear denoting the position of the best matching region



Figure 8.4 - A residual frame (on the left) denoting the differences between two successive frames for the fireman reference sequence (shown on the right).

Thus motion compensation enhances the efficiency of the motion estimation by adding at the predicted frame the fine differences that may contain the motion estimated predicted regions in comparison to the actual frame. Thus, during motion estimation the best matches between reference and current frames are detected and this match is further improved by motion compensation, which calculates the residuals of the motion estimated frame and the actual frame. So, adding this motion compensated residual information on the motion estimated frame, an accurate and efficient prediction of the current frame can be performed, using regions of past or future frames.

8.2.1.2 Compression at the Spatial Plane

Similarly, to the temporal stage, where predictive coding is performed between successive frames, a prediction of an image region may be also performed based on samples located within the same image or frame, which is usually referred as Intra coding. At spatial stage, the encoder performs a prediction for a pixel based pattern on combination of previously-coded pixels located on the same frame. Especially for frames that contain homogeneous areas, the spatial prediction can be quite efficient. The process in the terminology of video coding is call intracoding prediction. In the case of a good prediction then the residual energy is small and the corresponding compression ratio high.

More specifically, the spatial stage predicts the pixels of the current block using reconstructed blocks of neighboring blocks interpolated along different orientations, which results in a closely characteristics related image correlation. The computational complexity of this process significantly increases because of the number of different block partitioning modes and prediction directions. In order to reduce the computational resources needed, various complexity reduction strategies, along with novel hardware accelerators are used.

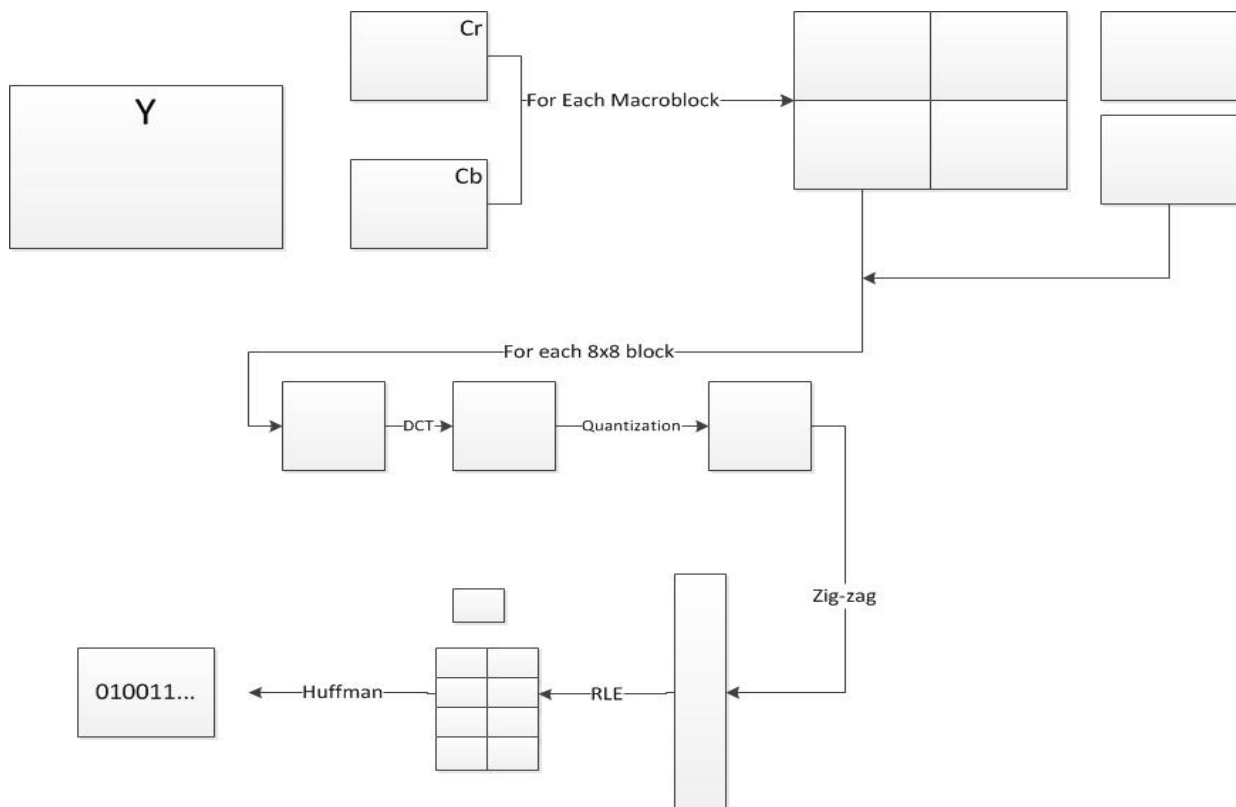


Figure 8.5 - Steps of the Intra-Coding Prediction

Based on the input signal, each frame is divided into Macroblocks (MBs), which are 16x16 pixel areas on Y plane of original frame. Each MB unit consists of 4 Y blocks, 1 Cr block and 1 Cb block. So the steps for Spatial Plane Compression (as it is called Intra Coding Prediction) are depicted on Figure 8.5 and are the following:

- Sub-divide picture into 16x16 pixel blocks: Macroblocks
- Apply DPCM, i.e. intra-prediction of the (16) 4x4 pixel blocks inside one Macroblock
- Residual transform (e.g. DCT algorithm), quantization and redundancy reduction

In natural sequences, the optimal intraprediction mode usually represents the texture direction in a block. Therefore, the directional prediction can reduce the texture redundancy significantly. H.264/AVC introduces directional intraprediction in the spatial domain. Intra_4x4 and Intra_8x8 both have up to nine prediction modes, namely mode 0 - mode 8. Except the DC mode (mode 2), the other eight modes correspond to different prediction directions as illustrated in Figure 8.6.

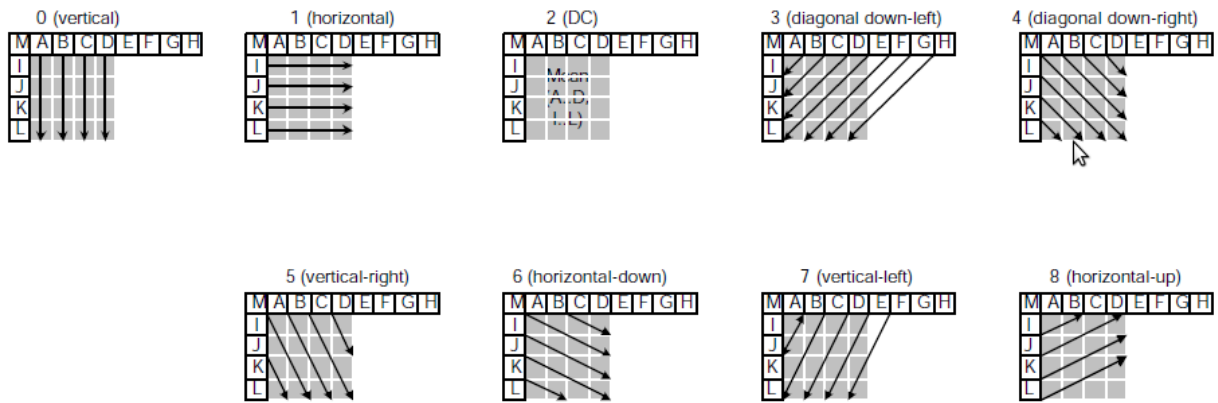


Figure 8.6 - H.264/AVC Intracoding Prediction Modes

An example of intracoding prediction is depicted on Figure 8.7 [Richardson, 2003], where a specific macroblock identified with white border has been selected to be intracoded predicted.



Figure 8.7 - H.264/AVC prediction example

Based on the possible predictions, intracoding algorithm will select the best mode by choosing the one with the smallest SAE. In addition, if multiple modes have the same, smallest SAE, then multiple solutions exist.

8.2.1.3 Transform Coding Phase

At this stage the spatially/temporally encoded frames or the motion-compensated residual data are converted into another domain, usually called as the transformed domain, where the optically correlated data become decorrelated. The use of transformation facilitates the exploitation in the compression technique of the various psycho-visual redundancies by transforming the picture to a domain where different

frequency ranges with dissimilar sensitivities at the Human Visual System (HVS) can be accessed independently. [Winkler, 2005].

The most commonly used transformation is the Discrete Cosine Transformation (DCT). The DCT operates on an X block of $N \times N$ image samples or residual values after prediction and creates Y , which is an $N \times N$ block of coefficients. The action of the DCT can be described in terms of a transform matrix A . The forward DCT is given by:

$$Y = AXA^T$$

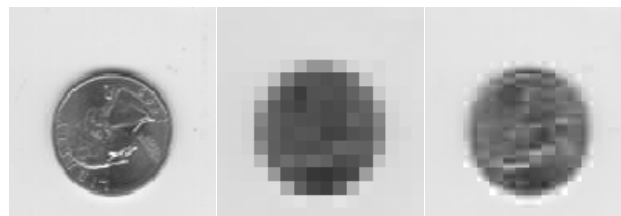
Where X is a matrix of samples, Y is a matrix of coefficients and A is an $N \times N$ transform matrix. The elements of A are:

$$A_{ij} = C_i \cos \frac{(2j+1)i\pi}{2N} \quad \text{Where } C_i = \begin{cases} \sqrt{1/N}, & i=0 \\ \sqrt{2/N}, & i>0 \end{cases}$$

Therefore the DCT can be written as:

$$Y_{xy} = C_x C_y \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} X_{ij} \cos \frac{(2j+1)y\pi}{2N} \cos \frac{(2i+1)x\pi}{2N}$$

The advantage of the DCT transformation is that it is possible to reconstruct quite satisfactorily the original image, applying the reverse DCT on a subset of the DCT coefficients, without taking under consideration the rest coefficients with insignificant magnitudes (see Figure 8.8).



a

b

c

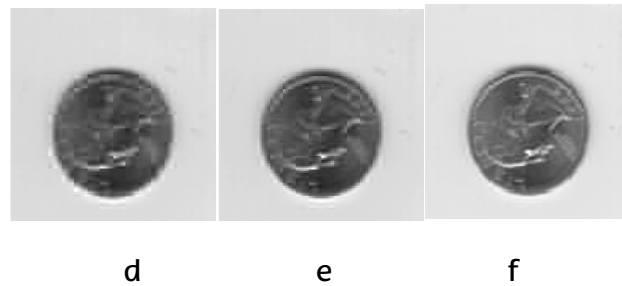


Figure 8.8 - Example of DCT efficiency. Figure a is the source image, while b is reconstructed using only 1

DCT coefficient, b exploits 4 coefficients, c uses 8 coefficients, d uses 12, e uses 18 and f 32 out of the 64 total DCT Coefficients for each block (i.e. 8x8).

Thus, with cost of some quality degradation, the original image can be satisfactorily reconstructed with a reduced number of coefficient values. This DCT property is exploited by the following stage where quantization of the DCT coefficients is performed.

Quantization Phase

Quantization is the process of approximating the continuous range of DCT coefficients by a relatively-small set of discrete integer values. The best-known form of quantization is the scalar quantizer, which maps one sample of the input to one quantized output value. A scalar quantization operator $Q()$ can be mathematically represented as

$$Q(x) = g(\lfloor f(x) \rfloor)$$

Where

- x is a real number
- $\lfloor x \rfloor$ is the floor function
- $f(x)$ and $g(i)$ are arbitrary real-valued functions.

The integer value $i = \lfloor f(x) \rfloor$ is the representation that is usually stored or transmitted, but the final interpretation may be further modified using also $g(i)$. Thus, typically during a scalar quantization process, it is performed the rounding of a fractional number to its nearest integer:

$$Y = QP \text{ round} \left(\frac{X}{QP} \right)$$

Where QP is the quantization parameter (i.e. quantization step size), X the initial integer value and Y the deduced quantized number. Table 8.1 depicts some representative examples of the scalar quantization process for various Quantization Parameters.

Table 8.1 - Quantization Examples

X	y		
	QP=1	QP=2	QP=3
0	0	0	0
1	1	0	0
2	2	2	3
3	3	2	3
4	4	4	3
5	5	4	6
6	6	6	6
7	7	6	6
8	8	8	9
9	9	8	9

Applying quantization on the aforementioned DCT coefficients is the main reason for the quality degradation and the appearance of artifacts, like the blockiness effect, at the digitally encoded videos. The blockiness effect refers to a block pattern of size 8x8 pixels in the compressed sequence, which is the result of the independent quantization of individual blocks of block-based DCT. Due to the quantized DCT coefficients, within a block (8x8 pixels), the luminance differences and discontinuities between any pair of adjacent pixels are reduced. On the contrary, for all the pairs of adjacent pixels, which are located across and on both edge sides of the border of adjacent DCT blocks, the luminance discontinuities are increased, by the coding process. This happens because

the quantization process is lossy (i.e. not totally reversible) since it is not possible to determine the accurate fractional number from the deduced rounded integer. So, it is somewhat equivalent with the case of not exploiting the entire DCT coefficient set for the reconstruction of the original image, as in Figure 8.4, because some low DCT values may have been quantized to zero.

It must be noted that the quantization stage is the only lossy stage at the described coding chain and is mainly responsible for any visual artifact and quality degradation, which may appear on the deduced coded video signal.

8.2.1.4 Entropy coding phase

At this final stage it is performed a transformation of the video sequence symbols into a compressed stream. The term video sequence symbol stands for all the aforementioned encoding parameters, such as quantization coefficients, motion vectors etc. Basically, two widely known variable length coding techniques are exploited at this stage: The Huffman Coding [Huffman, 1952] and the Arithmetic Coding [Witten et al., 1987]. Both methods are briefly analyzed hereafter:

Huffman coding technique creates a binary tree of nodes. Each node contains the symbol itself, and the weight (appearance frequency) of the symbol, and optionally a link to a parent node. The internal nodes contain symbol weight, links to their child nodes (two), and a link to the parent node to make the reverse tree reading easier. Usually the bit '0' represents the left child and bit '1' the right child. The basic algorithm to create the Huffman binary tree is:

1. Create a node for every symbol, and place all of them in group.
2. Remove from the group the nodes with the smallest weight, make them children of a new node that will have the weight sum of the two nodes, and place the new node to the group.
3. Repeat recursively step (2) until the group of nodes contains one node containing the weight sum of all nodes.

The decoding afterwards begins by the root of the binary tree and for each symbol we want to encode we memorize the '0','1' bits we encounter on the route to the symbol, the string of '0' and '1' we have memorized is the encoded symbol.

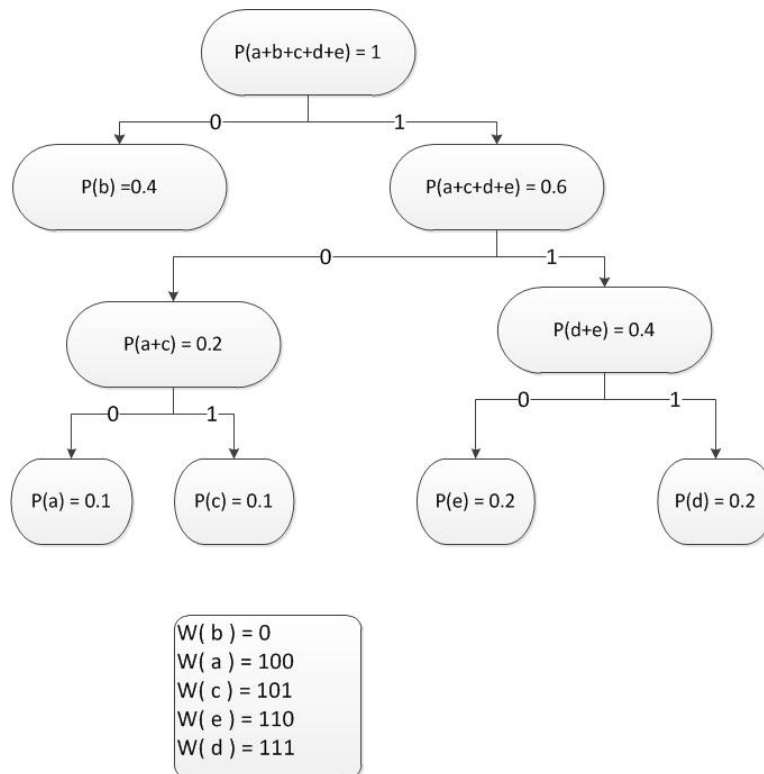


Figure 8.9 - Huffman Coding Algorithm

Arithmetic coding differs from Huffman coding, because it doesn't separate the input into component symbols and replace each with a code, it encodes the entire message into a single number, a fraction f where $(0.0 \leq f < 1.0)$. Apart from the weight of its symbol arithmetic coding demands that the terminating symbol does not appear in any other position of the input string. The coding process is the following:

1. All n symbols of input s are sorted, usually in alphabetical order.
2. Every symbol x_i with weight $p(x_i)$ is assigned an interval $[a_i, b_i)$ where $b_i - a_i = p(x_i)$.
3. First symbol is assigned the $[0.0, p(x_1))$ interval.
4. Every next symbol is assigned the $[p(x_1), p(x_1) + p(x_2))$ interval.
5. The last symbol has the interval $[1.0 - p(x_n), 1.0)$

The algorithm for the arithmetic coding technique is:

```

LowValue = 0.0;

HighValue = 1.0;

Do {

    inputString s;

    range = HighValue – LowValue;

    HighValue = LowValue + range * highRangeArray[s];

    LowValue = LowValue + range * lowRangeArray[s];

}

While(symbol != '$')

Print any number in [LowValue, HighValue);

```

The variable length coding methods assign to each video sequence symbol a variable length code, based on the probability of its appearance. Symbols appearing frequently are represented with short variable length codes while less common symbols are represented with long variable length codes. Over a large number of encoded symbols, this replacement of video sequence symbols by variable length codes lead to efficient compression of the data [Held, 1991].

8.3 VIDEO CODING STANDARDS EVOLUTION

Till today all the commonly known video compression methods/standards have been developed and approved either by ITU-T or ISO. The majority of the standards have followed a discrete development phase by either the ITU-T or the MPEG, while some joint standards have been proposed by both bodies. Figure 8.10 provides a timeline of the video coding evolution by each standardization body.

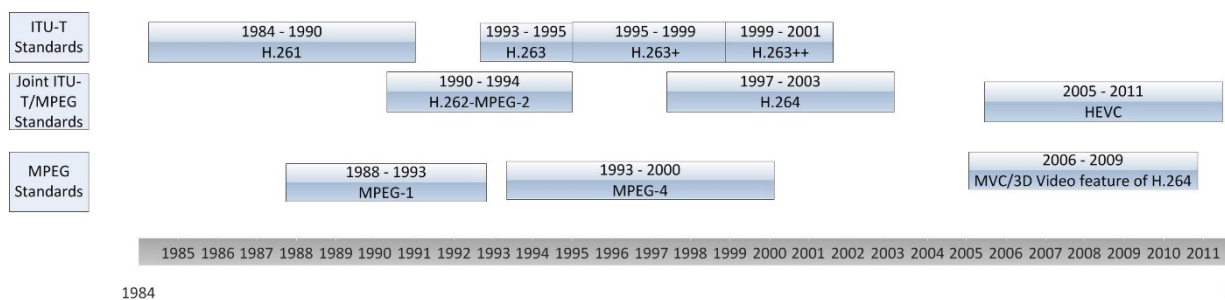


Figure 8.10 - Timeline of video coding standards evolution

The next subsections provide a brief presentation of the standards developed by ITU-T and MPEG respectively, focusing on the new features that each standard developed and introduced in the coding community.

ITU-T Standards

8.3.1.1 H.261

H.261 is a video standard by ITU-T, ratified in November 1988. It is the first member of the H.26x family of video coding standards in the domain of the ITU-T(VCEG), and was the first video codec that was used in commercial terms. H.261 was originally designed for transmission over ISDN lines on which data rates are multiples of 64 kbit/s. The coding algorithm was designed to operate at video bit rates between 40 kbit/s and 2 Mbit/s. The standard supports two video frame sizes: CIF (352×288 luma with 176×144 chroma) and QCIF (176×144 with 88×72 chroma) using a 4:2:0 sampling scheme.

The basic processing unit of the design is called a MB and H.261 was the first standard in which the MB scheme appeared. Each MB consists of a 16×16 array of luma samples and two corresponding 8×8 arrays of chroma samples, using 4:2:0 sampling and a YcbCr sampling.

The inter-frame prediction reduces temporal redundancy, with motion vectors used to help the codec compensate for motion. Whilst only integer-valued motion vectors are supported in H.261, a blurring filter can be applied to the prediction signal partially mitigating the lack of fractional-sample motion vector precision. Transform coding using an 8×8 discrete cosine transform (DCT) reduces the spatial redundancy. Scalar quantization is then applied to round the transform coefficients to the appropriate precision determined by a step size control parameter, and the quantized transform coefficients are zig-zag scanned and entropy coded (using a "run-level" variable-length code) to remove statistical redundancy. Figure 8.11 presents the steps of H.261 coding and decoding.

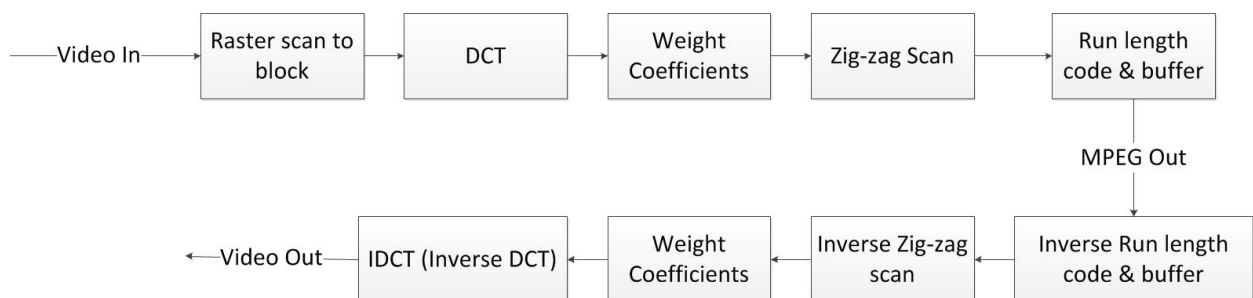


Figure 8.11 - H.261 Coding and Decoding steps

The H.261 standard actually only specifies how to decode the video. Encoder designers were given the liberty to design their own encoding algorithms, as long as their output is properly compatible to be decoded by any decoder implemented according to the standard. Encoders are also left free to perform any pre-processing they want to their input video, and decoders are allowed to perform any post-processing they want to their decoded video prior to display.

One major post-processing feature that was part of the H.261 is the deblocking filter. This technique is applied to blocks in the decoded video to give a more smooth visual texture by reducing the appearance of edgier blocks of the frame, which are caused by the block-based motion compensation and spatial transform of the design. Deblocking filtering has since become an important part of the successor of H.261 the H.264. In which a further post-processing can be done and enhance further the deduced video quality.

Future design optimizations included in later “family” standards have significantly reduced the compression level compared to the H.261, making it out-of-date, although, it is still used in some conferencing systems, internet videos and as a backward-compatibility mode. Despite that, H.261 is indisputably a milestone in the ever-evolving video-coding field.

8.3.1.2 H.262

The H.262 or MPEG-2 Part 2 (also known as MPEG-2 Video) is a digital encoding and compression standard developed and maintained jointly by ITU-T Video Coding Experts Group (VCEG) and Moving Picture Experts Group (MPEG). H.262 is very similar to MPEG-2 standard, which will be discussed later, but also provides support for interlaced video. In lower bitrates (less than 1Mbits) H.262 is not very efficient, but outperforms MPEG-1 at bitrates higher than 3 Mbits. All H.262/MPEG-2 conformed standards have full compatibility for MPEG-1 video playback.

H.262 supports a wide range of applications from high quality HD editing to mobile. It's unrealistic or too expensive for several applications to support the entire standard. In order to support these applications too, H.262 defines a set of profiles and levels. A profile defines a set of features i.e compression algorithm, chroma – luma format, etc. A level defines the set of quantitative capabilities i.e maximum frame size, maximum bitrate. The specifications set by an application in terms of profile and level, mean that a given player supports up to main profile and main level, can also playback any MPEG encoded up to this profile and level or less. The set of profiles is SP (Simple profile), MP (Main profile), SNR (SNR scalable profile), HP (High profile), 422 (4:2:2 profile), and MVP (Multi-view profile). Accordingly the set of levels is LL (Low Level), ML (Main Level), H-14 (High 1440), HL (High Level). Also any given profile – level combination is allowed by the standard.

8.3.1.3 H.263

H.263 video standard is the successor of H.261 and H.262, it maintains its predecessor's basic structure, but also has additional compression capabilities and error recovery techniques. This makes H.263 suitable for unreliable networks and lower data transfer rates, such as PSTN (public switched telephone network), where we experience high signal to noise ratio.

H.263 can handle very low data transfer rates, supporting SQCIF (sub-QCIF) video frame size. From the aspect of video compression H.263 extends the capabilities of H.261 as it also uses B-frames, which can use both previous and forward frames (I or P frames) for data reference. Additionally the motion vectors can point to areas that are outside of the reference frame. This can prove very useful particularly in low resolutions supported by H.263.

In the area of error detection and recovery, where H.263 really differs from its predecessor, has the error tracking, the reference picture selection and the independent segment decoding techniques. In the first one when the decoder spots an error on a group of blocks (GOB) sends an error message to the encoder with the ID of the faulty GOB. The encoder uses this information to find the MBs that referenced on the faulty GOB and transmits those MBs with intraframe coding instead of interframe, to stop the error transmission.

The codec in the reference picture selection is able to encode an MB not only from the first past frame, but from a group of past frames. This means that when the decoder finds a faulty GOB and messages accordingly the encoder, the encoder simply excludes that "past" frame from the references and uses "older" frames that have reached the receiver without errors.

The last method for error recovery in H.263 is the independent segment decoding. In this technique the encoder handles every GOB as an independent part of the frame. Meaning inter-frame coded MBs can only refer to MBs that belong to the same GOB. As an example, we encode a frame in QCIF resolution, the frame is separated in 3 different GOBs (parts), and each one is coded as a unique GOB using no reference to the other GOBs, indicating that an error on the first GOB will not affect the other GOBs of the frame on encoding.

8.3.1.4 H.263+

H.263+ can be referenced as H.263 version 2, as it includes all of H.263 features and decoding and encoding algorithms, but also adds twelve new optional modes for motion image processing quality improvement. The additional modes are: Annex I, which employs a new VLC table for quantized coefficients encoding for intra MBs. Annex J imports a deblocking filter inside the coding loop, resulting in better prediction and reduction in blocking artifacts. Annex K divides the picture into segments containing numbers of MBs. Annex L provides the decoder supporting features and functionalities

within the video bitstream. Annex M uses additional forward and backward predictors that improve P and B frame changes that may occur between pictures. Annex N maintains good picture reproduction by reducing error propagation between corrupted pictures. Annex O specifies techniques for temporal, snr, and spatial scalability capabilities. Annex P provides the algorithm to warp the reference frame prior to its prediction. Annex Q is used for detailed background in highly active motion scenes. Annex R prevents the propagation of errors, which results in efficient error resilience and recovery. Annex S, when enabled, uses the intra VLC table described in Annex I for inter block coding. Annex T has three features. First, rate control methods for increased flexibility in the MB layer quantizer change, second is the chrominance quality enhancement by specifying a finer quantizer step size, and third, improved picture quality by extending the range of representable quantized DCT coefficients. All the aforementioned modes are independent from one another, so according to every case's requirements we could select the best combination.

8.3.1.5 H.263++

H.263++ can be described best as the H.263 version 3 or the 2000 version. It supports three additional annexes. These and an additional annex (that specified profiles), were originally published separately from the standard's main body. The annexes are: Annex U that provided an improved reference picture selection mode. Annex V introduced the Data-partitioned slice mode, which reduces the error-prone intra frame prediction. Annex W supports additional supplemental enhancement information specification, and last, Annex X (originally specified in 2001) provided profile and level specification.

8.3.1.6 H.264

The intent of the H.264/AVC(Advanced Video Coding) standard was to create a standard capable of providing good video quality at lower bit rates than previous standards (i.e H.263), with an economic design both in terms of complexity and implementation. Another goal of the standard was to be able to be easily applied to a wide variety of applications on a wide variety of networks and systems, including low and high bit rates, low and high resolution video, broadcast, DVD storage, RTP-IP packet networks, and ITU-T multimedia telephony systems.

Due to its wide application on networks and systems H.264 had to be able to provide equal and stable solutions for many of these cases. So it formed a solution based on different profiles, which technically are different configuration files for the encoder. Each profile sets different parameters for the encoder, aiming to provide the suitable solution. H.264 defines a set of three profiles, each supporting a particular set of coding functions and each specifying what is required of an encoder or decoder that complies with the Profile.

The *Baseline Profile* supports intra and inter-coding (using only I-slices and P-slices) and entropy coding with context-adaptive variable-length codes (CAVLC). The *Main Profile* includes support for interlaced video, inter-coding using B-slices, inter coding using weighted prediction and entropy coding using context-based arithmetic coding (CABAC). The *Extended Profile* does not support interlaced video or CABAC, but adds more efficient switching between coded bitstreams (SP- and SI-slices) and is more resilient to errors(Data Partitioning). Potential applications of the Baseline Profile include videotelephony, videoconferencing and wireless communications; potential applications of the Main Profile include television broadcasting and video storage; and the Extended Profile is particularly useful for streaming media applications. A brief analysis of each of the aforementioned profiles follows hereafter.

8.3.1.6.1 The Baseline Profile of H.264

The Baseline Profile supports coded sequences containing only I- and P-slices. I-slices contain intra-coded MBs in which each 16×16 or 4×4 luma region and each 8×8 chroma region is calculated and predicted from previously-coded parts of the same slice. P-slices may contain intra-coded, inter-coded or skipped MBs. Inter-coded MBs in a P slice are predicted from a number of previously coded pictures, using motion compensation with quarter-sample (luma) motion vector accuracy.

After prediction, the residual data for each MB is DCT transformed using a 4×4 integer transform and quantised. Quantised transform coefficients are reordered and the syntax elements are entropy coded. In the Baseline Profile, a context-adaptive variable length coding scheme (CAVLC) is used for entropy coding transform coefficients, whereas syntax elements are coded using fixed-length or Exponential-Golomb Variable Length Codes.

Quantised coefficients are scaled, inverse transformed, reconstructed (added to the prediction formed during encoding) and filtered with a de-blocking filter before (optionally) being stored for possible use in reference pictures for further intra- and inter-coded MBs.

A filter similar to H.261 one is applied to each decoded MB to reduce block-shaped artifacts. The deblocking filter is applied after the inverse transform in the encoder (right before reassembling and storing the MB for future prediction use) and in the decoder (before reassembling and displaying the MB). The filter creates smoother block edges, improving the appearance of the decoded frames. This is proven to be very beneficial as the filtered image will be used for motion-compensated prediction of future frames and this improves the compression performance because a more detailed and smoother filtered reference frame is more faithful reproduction of the original frame, than an unfiltered one. Additionally, it is possible for the encoder to alter the filter strength or to disable the filter.

8.3.1.6.2 The Main Profile of H.264

The Main Profile is a more complete version of the Baseline Profile, except that multiple slice groups, ASO (Arbitrary Slice Order, i.e. no particular frame decoding order) and redundant slices (all supported by the Baseline Profile) are not included. The additional tools provided by Main Profile are B slices (bi-directional prediction slices for greater coding efficiency), weighted prediction (providing increased flexibility in creating a motion-compensated prediction block), support for interlaced video (coding of separate fields as well as frames) and CABAC (an alternative entropy coding method based on Arithmetic Coding). Suitable applications for the Main Profile include (but are not limited to) broadcast media applications such as digital television and stored digital video.

In B slices every MB partition in an intercoded MB can be predicted from one or two reference images, before or after the current picture in temporal order. In relation to the reference pictures stored in the encoder and decoder, this provides the alternative to choose the MB prediction references in a B MB type, for example using: (i) one future and one past, (ii) two past, (iii) two future references to predict a B-frame type image.

Unlike the baseline profile the main profile supports the interlaced video, also used in analog video transmission. The interlaced video is a technique in which we double the perceived video frame rate, efficiently without the cost of extra consumed bandwidth. The interlaced video contains two fields of a video frame shot taken from two different times and it improves the viewer's video perception by reducing flickering and taking advantage of the persistence of vision effect (the phenomenon where the eye by which an afterimage is thought to persist for approx. 1/25 of a sec of the retina). We can achieve this effect by doubling the temporal resolution. In the main profile specifically a specific slice is coded as 'MB pair'. The encoder afterwards will encode each MB either as two frame MBs or two field MBs and can select the optimum pair. Because of that further modification may be required to a several encoding and decoding steps, for example we may need to modify the P and B prediction depending on whether the M Bs are coded in frame or field mode.

8.3.1.6.3 The Extended Profile of H.264

The Extended Profile (also known as the X Profile in earlier versions of the draft H.264 standard) can be very helpful in video-streaming applications [Schwarz-2007]. It includes all of the features of the Baseline Profile (i.e. it is a supreme superset of the Baseline Profile, unlike Main Profile), but also uses B-slices, Weighted Prediction and unique features that allow it to be extra-efficient over network stream services such internet video-streaming. SP and SI slices become a helpful tool for switching between different coded streams and 'VCR-like' (the 'SETUP', 'PLAY', 'PAUSE', 'TEARDOWN') functionality and Data Partitioned slices provide improved performance error susceptible data transmission environments.

SP and SI slices are specially-coded slices that allow video decoders to switch efficiently and rapidly between video streams and provide an efficient random access for video

decoders. One major requirement for a reliable video stream application is the flexible and quick switching between video streams, to maintain continuous playback. For example, a video stream application in order to provide an uninterrupted video playback has to handle data throughput drop cases, or network traffic issues. A solution this area of problems is to have the same video material in two different coded versions, so when an error occurs, our application should be capable of switching between the high-quality stream to the low-quality one efficiently, this is where SP and SI slices aid with the stream switching.

The decoded data that forms a slice is separated into three Data Partitions supposedly A, B and C, each one contains a part of the coded slice. The slice header and the header data is contained in slice A, B contains coded residual data for intra and SI frame MBs and C the inter-coded MB residual data (bi-directional). Therefore each partition can be put into a different NAL unit and be transported separately. If partition A is lost, it is most likely very difficult or even impossible to reconstruct the slice, which leads to the fact that partition A is very sensitive to transmission errors. However, with a careful choice of parameter we can make partitions B and C independently decodable, thus decoding only A and C or only A and B, gaining flexibility and efficiency in an error-prone environment.

8.3.1.6.4 H.265

High efficiency Video Coding is a standard for video compression, successor of the H.264/MPEG-4 AVC, currently under-development by the MPEG and ITU-T Video Coding Experts Group(VCEG) collaboration team named Joint Collaborative Team on Video Coding(JCT-VC).

The goal of HEVC is to substantially improve coding efficiency compared to AVC High Profile and H.264 extended profile, i.e. improve significantly bitrate to image quality ratio, probably increasing the computational complexity. Depending on the application requirements, HEVC is designed to maintain the balance computational complexity, compression rate against proneness to errors and processing delay time.

HEVC is designed to support next-generation HDTV displays and content capture systems that feature progressive scan frame rates and display resolutions from QVGA (320x240) up to 1080p and Ultra HDTV (7680x4320), as well as techniques that guarantee improved picture quality in terms of noise level, color gamut and dynamic range.

MPEG standards

In contrast to H.26x standards, which solely covers video-coding, MPEG standards support both video and audio coding and also the multiplexing of several media streams in one.

8.3.1.7 MPEG-1

MPEG-1 is a video standard designed by Motion Pictures Experts Group (MPEG), which is formed by ISO. It is the first of the MPEG standards family that covers the fields of coding high-quality audio and video, and multiplexing for data storing or network transmission. The first goal of MPEG-1 was the total bandwidth of transmitted audio and video not to exceed 1.2 – 1.5 Mbps (i.e the bandwidth of T-1 transfer line).

MPEG-1 is a more complex variation of H.261, whereas H.261 aimed towards supporting bi-directional symmetric applications, MPEG-1 was designed for multimedia distribution applications. This is because MPEG-1 is usually asymmetrically implemented, as in multimedia distribution applications the time-consuming compression process is performed only once and the video playback constantly.

MPEG-1 Video codec's compression methods significantly reduce the video stream's required data rate. It reduces or discards information in certain frequencies and areas of the picture that they are not recognized by the human eye. It also exploits temporal (time related) and spatial (area related) redundancy a commonly used method to achieve more efficient compression that would be impossible otherwise.

The first frame category in MPEG-1 is the I-frame that are intra-coded, with no reference to other previous or forward frames. This category fully-supports random access, as the video playback can instantly begin from an I-frame. The second frame category is the P-frame, which coded by using a reference on their previous frame I or P frame. For every P-frame MB the codec searches the corresponding area on the reference frame. The standard does not dictate the search method, but the search range via the motion vector. If no relative area is found, then the frame is intra-coded, just like an I-frame. In an alternate case when the differences between the current and the reference MBs are few the blocks are omitted, otherwise they are DCT transformed, quantised and RLE and entropy coded.

The B-frames are the third frame category, which can use both their next, as well as their previous I or P frame as a reference frame. The reference frame area can also be calculated from the average of a previous and next frame area. The coding of B-frames is almost identical to the P-frame one. P and B frames can use the same quantization parameter array, whereas I-frames use a different array. The fourth frame category is the D-frames that are intra-coded just like I-frames but only using DC and DCT coefficients.

In MPEG-1 standard the entity of frame grouping is reflected upon the picture group (group of pictures – GOP) structure. Every GOP sequence begins with an I-frame and contains all the remaining P and B frames, until the next I-frame. GOP is the main synchronization unit of MPEG-1, since random access is feasible only every I-frame of a GOP. Additionally every GOP is divided into smaller frame groups, which start with an I-frame and end with a P-frame.

MPEG-1 being a full audio and video coding system determines, not only determines the stream of every media type, but also the media multiplexing into a single stream. From

the video aspect the media stream is hierarchically ordered and consists of 6 levels. The structure of the stream is: 1.Sequence Header, 2.GOP header, 3.Frame Header, 4.MB Header, 5.Block Header, 6.Block coefficients. The independent media streams are multiplexed into a single stream, which is called pack. The pack contains a timestamp header which assists the codec to remain synchronized with the encoder.

8.3.1.8 MPEG-2

The MPEG-2 (or H.262) design began, even before the standardization of MPEG-1 was complete. The initial goal was the support of video quality relative to CCIR-601 to support digital television systems, while simultaneously designing MPEG-3 for future use on HDTV. During the development of MPEG-2, MPEG-3 was incorporated, as it was generalized to include the support of HDTV systems.

Due to the wide variety of application support, the MPEG-2 standard states a series of profiles and levels which are related to one another. The profiles cover several complexity categories. Every level states the video parameters, meaning that a level covers a quality category. The levels are the "low level", which supports MPEG-1 television signals with data rate up to 4 Mbps, whilst "main level" supports quality relative to the CCIR-601 video with data rate up to 15-20 Mbps. The "high profile-1440" supports HDTV quality with aspect ratio 4:3 and data rate 60-80 Mbps, and the "high-profile" HDTV quality with aspect ratio 16:9 and data rate 80-100 Mbps.

Profile-wise the "simple profile" is the first one, a simpler version of MPEG-1 aimed for low-cost devices. Next comes the "main profile" which supports all of MPEG-1 capabilities and its common use is on DVDs. The "main profile" characteristics are extended to several "scalable profiles" which support projection devices with different capabilities. The scalability can be achieved in two ways, the signal-to-noise scalability, which allows layered picture coding so that they can be coded in different quality levels, and the spatial scalability which allows the video decoding in different horizontal and vertical resolutions.

In the P and B frame coding, accordingly, three options are provided, the "field mode", in which the MBs are coded by either their next or previous field reference. In the "frame mode" the MBs are coded by either their next or previous same type field reference, meaning the even fields are refer to their previous even field and the odd ones respectively. However in the "mixed mode" the MBs are coded by either their next or previous same type field reference, but taking into consideration which option offers them better compression.

MPEG-2 standard, like MPEG-1, determines the data stream, not only for every individual type of media, but for multiplexed media streams. The difference between the 2 standards is that MPEG-2's media stream can be more complex when used from a scalable profile. Every type of media, including audio and video, is organized into a unique Packetized Elementary Stream (PES). In the next level the PES are multiplexed

into a single transport stream (TS), and transmitted. Additionally the program stream does not include extra synchronization information, since all the PES derive from the same clock. MPEG-2 specifies that the raw frames are compressed into three types of frames: intra-coded frames (I-frames), predictive-coded frames (P-frames) and bidirectionally-predictive-coded frames (B-frames).

8.3.1.9 MPEG-4 Part 2

MPEG-4 Part 2 is a video compression standard developed by MPEG, of the MPEG-4 ISO/IEC standards family. Similarly to previous standards MPEG-1 and MPEG-2, it is DCT (Discrete Cosine Transform) based. Popular applications that implement this standard are Divx, Xvid and Nero Digital. It is also H.263 compatible, i.e. a H.263 bit stream can be successfully decoded by an MPEG-4 decoder. MPEG-4 uses 2 video object layer types, the video object layer which fully supports the MPEG-4 functionality and a reduced functionality video object layer with short headers, compatible with baseline H.263.

MPEG-4 focuses on lower bitrates, in which the main goal is the compression efficiency. Since the standard is designed to support different type terminals with scalable data streams, and error-prone networks with fault detection and recovery techniques, an important service field for MPEG-4 is the mobile telephony.

The quintessential innovation of MPEG-4 differs from its predecessors due to its user-interactive platform support. This enables the user to dynamically interact with the content. In order to achieve this, video and audio signals are handled as single but independent data streams. The content in MPEG-4 is structured as an object-oriented hierarchy, which can be modified by the recipient. The objects can be created uniquely or synthetically, and in their turn consist of simpler objects. Every object is coded according to its type, thus the object hierarchy contributes to low transmission bitrates.

Due to its design MPEG-4 does not only support coding and media multiplexing schemes, but also techniques to build complex scenes from individual objects. In order to achieve efficient media coding MPEG-4 offers various playback options along with coding techniques, unlike MPEG-1 and 2 which coded the video and audio streams integrally.

Every scene in MPEG-4, meaning a sequence of frames with identical content, consists of an audio/visual object (AVO) group. The AVOs are organized hierarchically (i.e. an AVO can consist of simpler AVOs). The scene model used in MPEG-4 is based on the VRML (virtual reality modeling language) with a few modifications like BIFS (binary format for scenes), which is an alternative scene coding method that enables interaction between the user and the AVO, permitting the user to delete or modify any existing AVOs in the scene.

In order to encode every AVO, they separate each of their components so that everyone is compressed differently. Every AVO can be projected as a VOP (Video Object Plane), which is composed by an integer-valued sum of MBs in every dimension. VOPs are

decoded individually based upon shape, motion and texture. The multiplexer synchronizes every single VOP and AVO accordingly, in order to create a single media stream.

The de-multiplexer from its perspective retrieves each VOP, the audio data and the pattern of the scene from a single data stream, which can be originated from a network or a storage media object. After that every media is decoded separately based on its content, and the scene synthesis is performed according to the BIFS information given from the data stream. Optionally the recipient can alter the scene, using the BIFS model, by adding, removing, moving or modifying its objects. Concluding that MPEG-4's main difference from its predecessors is that every media consists of many objects instead of being considered as a single entity during the encoding and decoding.

The MPEG-4 specifications main purpose is the support of several types of bitrate, quality, resolution and service. Since a scene is composed by an object hierarchy, instead of encoding wholes frames, MPEG-4 encodes the object-hierarchy and sends the modifications that need to be performed on every object, based upon the frame motion. The hierarchy itself can change, as new objects enter the scene or current objects exit. If the hierarchy creator allows it, the user can interact with it. Possible interaction cases are the shift of objects, the change in the point of view or the behavioral object modification.

MPEG-4 provides techniques for structure, communication and implementation of the object classes. In order to decode an MPEG-4 media stream, the standard based upon the object's decoding defines the class structure phase. At the communication phase it transfers the classes that are missing and completing the already installed ones to decode all the objects. When all the class descriptions, including their data structures are transmitted, the standard proceeds to their initialization. In the end all the video and audio objects are de-multiplexed, synchronized and decoded by the recipient according to their relative classes.

The video-object coding supports spatial and temporal scalability, allowing the recipient to decode partially the video-reconstructing data. Scalability allows video decoding with reduced spatial resolution, reduced temporal resolution or both, but with lower quality in exchange. This feature is useful for video transmission applications over networks with limited bandwidth, or for services in which the recipient does not want or is not able to display the video at high resolution and quality. In cases when the screen resolution or the network has limited capabilities, scalability can radically improve the system's efficiency, although subjected to the aforementioned limitations.

MPEG-4 standard provides two technical support solutions for channels with high noise and low bitrates. The first one is the use of fixed-length video packets, which allow the recipient resynchronization at the beginning of the next packet and not at the start of the next part, unlike MPEG-1. Generally every packet contains some MBs, meaning the resynchronization of the decoder is performed at MB level. In order to develop a fault-tolerant system, the packets separate the one packet's motion vectors so that we can

exploit the rest of it. The packets can also contain a timestamp, so that the decoder won't dispatch the whole frame in case errors appear at its beginning.

The second fault-tolerant technique is the use of RVLC (reversible variable length codewords). These codewords have the characteristic that their limits can be recognized by reading the data from end to start. During the RVLC procedure for DCT coefficient coding, if the decoder detects errors while reading the frame from start to end, it will not dispatch the entire packet but restart to read it end to start this time. Therefore prior or subsequent to the error codewords are encoded normally, thus affecting the frame only partially, and allowing the efficient error recovery even from partially destroyed packets.

8.3.1.10 HEVC (H.265)

High efficiency Video Coding is a standard for video compression, successor of the H.264/MPEG-4 AVC, currently under-development by the MPEG and ITU-T Video Coding Experts Group (VCEG) collaboration team named Joint Collaborative Team on Video Coding (JCT-VC). H.265 is used as a nickname for the standard.

HEVC's goal is to substantially improve coding efficiency compared to AVC High Profile and H.264 extended profile, i.e. improve significantly bitrate to image quality ratio, probably increasing the computational complexity. Depending on the application requirements, HEVC is designed to maintain the balance computational complexity, compression rate against proneness to errors and processing delay time.

HEVC is designed to support next-generation HDTV displays and content capture systems that feature progressive scan frame rates and display resolutions from QVGA (320x240) up to 1080p and Ultra HDTV (7680x4320), as well as techniques that guarantee improved picture quality in terms of noise level, color gamut and dynamic range.

Deblocking filter as we mentioned earlier is the technique that alleviates the blocking artifacts caused by the DCT transform. HEVC employs two filters the deblocking filter, which is inherited from the H.264 standard, and the Adaptive Loop Filter (ALF) which is used to restore the encoded and degraded frame in entire-frame or MB level, after the deblocking process. The main difference between these 2 filters is that the deblocking filter improves the subjective quality, whereas ALF the objective quality, as a result these two techniques are mutual complements. The spatial redundancy and information loss caused by the deblocking filter is restored at some percentage by ALF.

The HEVC still the block-based hybrid video coding framework, although the MB size is extended to (64x64) compared to H.264. Three units of various sizes are used to describe the overall coding structure, coding unit (CU), prediction unit (PU) and transformation unit (TU). The CU is basically the MB in H.264, it can have various square-shaped sizes. A CU forms an hierarchical tree whose leaves are the PU s, like the H.264 standard two different terms are used to describe the prediction method, PU type and PU splitting. As

a result different PU splittings correspond to different PU types, which consist of skip, intra and inter. This feature enables us to use different prediction techniques in one frame, achieving better quality, while not sacrificing bandwidth for video stream, although costing as in computational resources because of the greater complexity.

In accordance to the CU and PU definitions the TU is defined to transform and quantize. The TU structure resembles the quadtree design of the CU. TU have different splittings for low complexity and high efficiency configurations.

In this section we report the basic new features of the new HEVC video encoding standard. Considering the standard is still under development, possible more features will be added in the near future or the reported ones may be altered in the next versions of the codec versions.

8.3.1.10.1 Block-Based Coding

The HEVC continues to implement the block-based hybrid video coding framework, with the exception of the increased macroblock size (up to 64x64) compared to AVC.

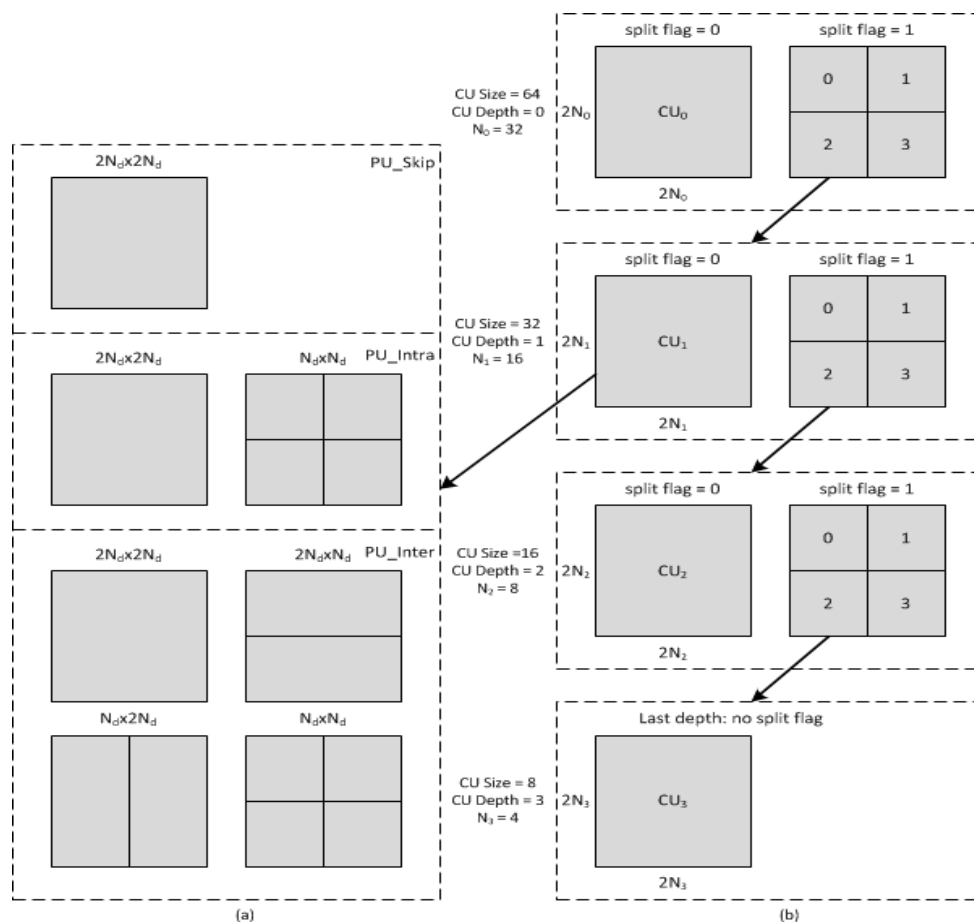


Figure 8.12 - Recursive quadtree representation of CU

Also, three novel block concepts are introduced, namely: the coding unit (CU), the prediction unit (PU) and the transform unit (TU). CU is the basic coding unit like the H.264/AVC's macroblock and can have various sizes but is restricted to be square shaped.

The general outline of the coding structure is formed by various sizes of CUs, PUs and TUs in a recursive manner, once the size of the largest coding unit (LCU) and the hierarchical depth of CU are defined. Given the size and the hierarchical depth of LCU, CU can be expressed as a recursive quadtree representation as it is depicted in Figure 8.12, where the leaf nodes of CUs can be further split into PUs or TUs.

The introduction of larger block structures is one of the most important elements for higher compression performance in high resolution videos, due to the flexible sub-partitioning mechanisms. Respectively, the model defines CUs which sub-partition a frame into equal or variable size rectangular regions. At the PU level, either intra-frame or inter-frame can be selected.

8.3.1.10.2 Intra-Prediction in HEVC

The current intra prediction technique in HEVC unifies two simplified directional intra prediction methods: the Arbitrary Direction Intra and the Angular Intra Prediction. The unified intra prediction technique enables a lower-complexity method in which parallel processing can be achieved, where samples of already decoded adjacent PUs are used, the signaling of the mode is produced from the modes of adjacent PUs (horizontal, vertical or depending on the block size up to 28 angular directions) and syntax indicators.

8.3.1.10.3 Inter-Prediction in HEVC

The inter prediction in HEVC uses the frames stored in a reference frame buffer (with a display order independent prediction, as in AVC), which allows multiple bi-direction frame reference. A reference picture index and a motion vector displacement are needed in order to select reference area. The merging of adjacent PUs is possible, by the motion vector, not necessarily of rectangular shape as their parent CUs. In order to achieve encoding efficiency, skip and direct modes similar to the AVC ones are defined, and motion vector derivation or a new scheme named motion vector competition is performed on adjacent PUs. Motion compensation is performed with a quarter-sample motion vector precision. At TU level (which commonly is not larger than the PU), an integer spatial transform (with range from 4x4 to 64x64) is used, similar in concept to the DCT transform. In addition a rotational transform can be used for block sizes larger than 8x8, and apply only to lower frequency components. In AVC scaling, quantization and scanning of transform are performed in a similar way.

At CU level, an adaptive loop filter (ALF) can be applied prior to copying the frame into the reference picture buffer. This is a FIR filter whose main purpose is to minimize distortion relative to the original picture, and its filter coefficients which are encoded at slice level. Additionally a deblocking filter is operated within the prediction loop (similar to the AVC deblocking filter design). After applying these 2 filters the display output is written to the picture buffer.

8.3.1.10.4 Entropy Coding in HEVC

The HEVC defines 2 context-adaptive entropy coding patterns, one for the higher-complexity mode and one for the lower-complexity mode. The lower-complexity mode is based on a variable length code (VLC) table selection for all the syntax elements, while using a particular code table which is picked in a context-based scheme depending on previous decoded values. This design is very similar to the CALVC pattern from AVC, but enables even simpler implementation according to its more systematic structure. A re-sorting of code table elements can be used as a supplementary compression improvement.

The higher-complexity design uses a binarization and context adaptation pattern similar to the AVC entropy coder, CABAC, but with the difference of using a set of variable-length-to-variable-length codes (indexing a variable number of bins into a variable number of encoded bits) instead of using an arithmetic coding engine. This is performed by applying a bank of parallel VLC coders – each of which is responsible for a certain range of odds of binary events (which area referred to as bins). The coding performance can be better parallelized and has higher throughput per processing cycle in software or hardware implementation than CABAC, although being very similar to it. It must be noted that the compression performance of this design can be significantly higher than the lower-complexity VLC.

9 TECHNICAL AND GENERAL PURPOSE REFERENCES

[3GPP-TS23060]	3GPP, "Technical Specification. 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects," TS 23.060.V5.2.0, 2000.
[3GPP-TS23251]	3GPP TS 23.251 v12.0.0, "Network sharing; Architecture and functional description (Release 12), December, 2013
[3GPP-TS29060]	3GPP, "GPRS Tunneling Protocol (GTP) across the Gn and Gp Interface," TS29.060. v3.3.0, 2000.
[3GPP-TS36413]	3GPP TS 36.413 V9.5.0 3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access Network (E-UTRAN); S1 Application Protocol (S1AP).
[4GAMERICAS-2012]	4G Americas, "Developing and Integrating a High Performance HET-NET", White Paper, October, 2012
[CLOUDRAN-2015]	Ericsson White Paper, "Cloud-RAN – the benefits of virtualization, centralization and coordination", (2015).
[CMRI-2011]	China Mobile Research Institute, "C-RAN: The Road Towards Green RAN. White Paper. Version 2.5." October 2011.
[ECOMP-2017]	http://about.att.com/content/dam/snrdocs/ecomp.pdf
[ETSI-MEC-2014]	ETSI, "Mobile-edge computing: Introductory technical white paper", 2014.
[ETSI-NFV001]	European Telecommunications Standards Institute, Network Functions Virtualisation (NFV); Use Cases, ETSI group specification NFV 001, 2013.
[ETSI-NFV002]	ETSI Industry Specification Group (ISG) NFV, "ETSI GS NFV 002 V1.2.1: Network Functions Virtualisation (NFV); Architectural Framework," http://www.etsi.org/deliver/etsi_gs/NFV/001_099/002/01_02.01_60/gs_NFV002v010201p.pdf , December 2014.
[ETSI-NFV003]	ETSI Industry Specification Group (ISG) NFV, "ETSI GS NFV 003 V1.2.1: Network Functions Virtualisation (NFV); Terminology for Main Concepts in NFV," http://www.etsi.org/deliver/etsi_gs/NFV/001_099/003/01_02.01_60/gs_NFV003v010201p.pdf , December 2014

[FFMPEG-2017]	https://www.ffmpeg.org
[HUGEPAGES-2017]	https://www.kernel.org/doc/Documentation/vm/hugetlbpage.txt
[ITU-R2012]	Recommendation ITU-R BT.500-13 (2012), Methodology for the subjective assessment of the quality of television pictures, [Online]. Available: https://www.itu.int/rec/R-REC-BT.500
[ITU-T1999]	Recommendation ITU-T P.910 (1999), Subjective video quality assessment methods for multimedia applications., [Online]. Available: http://www.itu.int/rec/T-REC-P.910-200804-I
[ITU-T2008]	Perceptual Visual Quality Measurement Techniques for Multimedia Services Over Digital Cable Television Networks in the Presence of a Reduced Bandwidth Reference, ITU-T Rec. J.246, Aug. 2008 [Online]. Available: http://www.itu.int/rec/T-REC-J.246/en
[ODL-2017]	https://www.opendaylight.org/
[ONAP-2017]	Open Network Automation Platform (ONAP), on-line: https://www.onap.org
[OPENAIR-2017]	http://www.openairinterface.org/
[OPENBATON-2017]	http://openbaton.github.io/
[OPENO-2017]	https://www.open-o.org/
[OPENSTACK-2017]	https://www.openstack.org/
[OPENSTACK-HEAT]	OpenStack. (2015). Heat Orchestration Template (HOT) Guide. Available: http://docs.openstack.org/developer/heat/template_guide/hot_guide.html
[OPNFV-2017]	OPNFV Open Platform for NFV Project, https://wiki.opnfv.org/ , accessed: 04-2017.
[OSM-2017]	https://osm.etsi.org/wikipub/index.php/Main_Page
[RFC2544-1999]	[RFC 2544] S. Bradner, J. McQuaid, "Benchmarking Methodology for Network Interconnect Devices", March 1999.
[SCF-2014-088]	Small Cell Forum, "Urban Small Cell Network Architectures," document SCF088, June 2014; http://scf.io/documents/088 .

[SCF-2015-055]	Small Cell Forum, "Small Cells and 5G Evolution: A Topic Brief," document SCF055, June 2015; http://scf.io/documents/055 .
[TNOVA-2017]	http://www.t-nova.eu/
[TNOVA-SOURCE]	http://github.com/T-NOVA
[VQEG-2000]	"Final Report From the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment", http://www.vqeg.org .
[XIPH-MEDIA]	https://media.xiph.org/video/derf/

Scientific Publication References

[Andrews-2013]	J. G. Andrews, "Seven Ways that HetNets are a Cellular Paradigm shift", IEEE Communications Magazine, March, 2013, pp. 136-144
[Andrews-2014]	J.G. Andrews et al., "What Will 5G Be?" IEEE J. Selected Areas in Comm. (JSAC), vol. 32, no. 6, 2014, pp. 1065–1082.
[Barlow-1961]	H. B. Barlow, "Possible principles underlying the transformation of sensory messages," in Sensory Communication, W. A. Rosenblith, Ed. Cambridge, MA: MIT Press, 1961, pp. 217–234.
[Buxton-1995]	W. Buxton, (1995) "Integrating the periphery and context: A new taxonomy of telematics", in Proceedings of Graphics Interface 1995, pp. 239-246, 1995.
[Carnec-2003]	M. Carnec, P. Le Callet, and D. Barba, "An image quality assessment method based on perception of structural information," in Proc. IEEE Int. Conf. Image Process., vol. 3. Sep. 2003, pp.185–188.
[Carnec-2005]	M. Carnec, P. Le Callet, and D. Barba, "Visual features for image quality assessment with reduced reference," in Proc. IEEE Int. Conf. Image Process., vol. 1. Sep. 2005, pp.421–424.
[Chandler-2007]	Chandler, D. M. and Hemami, S. S., "VSNR: A wavelet-based visual signal-to-noise ratio for natural images," IEEE Transactions on Image Processing 16(9), 2284–2298 (2007).
[Chono-2008]	K. Chono, Y.-C. Lin, D. Varodayan, Y. Miyamoto, and B. Girod, "Reduced-reference image quality assessment using distributed source coding," in Proc. IEEE Int. Conf. Multimedia Exposit., Apr. 2008, pp.609–612

[Cover-1991]	T. M. Cover and J. A. Thomas, Element of Information Theory. New York, Wiley, 1991.
[Farias-2002]	Farias MCQ, Mitra S, Carli M, Neri A (2002) A comparison between an objective quality measure and the mean annoyance values of watermarked videos. In: Proc. IEEE international conference on image processing, vol III. Rochester, NY, pp 469–472.
[Giotis-2015]	K. Giotis, Y. Kryptis, and V. Maglaris, "Policy-based orchestration of NFV services in Software-Defined Networks," in (NetSoft), 2015 1st IEEE Conference on, 2015, pp. 1-5.
[Guerzoni-2012]	R. Guerzoni, "Network Functions Virtualisation: An Introduction, Benefits, Enablers, Challenges and Call for Action. Introductory white paper," in SDN and OpenFlow World Congress, June 2012.
[Gunawan-2003]	Gunawan,I.P. Ghanbari,M., "Reduced-reference picture quality estimation by using local harmonic amplitude information", Proc. London Communications Symposium 2003, University College London, UK, 8-9 September 2003 pp 137-140.
[Gunawan-2008]	Gunawan,I., Ghanbari,M, "Reduced-Reference Video Quality Assessment using Discriminative Local Harmonic Strength with Motion Consideration", IEEE Trans. on Circuits and Systems for Video Technology IEEE, Jan. 2008, pp. 71-83.
[Han-2015]	B. Han, V. Gopalakrishnan, L. Ji, and S. Lee, "Network function virtualization: Challenges and opportunities for innovations," Communications Magazine, IEEE, vol. 53, no. 2, pp. 90–97, Feb 2015.
[Kim-2013]	H. Kim and N. Feamster, "Improving network management with software defined networking," Communications Magazine, IEEE, vol. 51, no. 2, pp. 114–119, 2013.
[Koumaras-2004]	H. Koumaras, E. Pallis, G. Xilouris, A. Kourtis, D. Martakos, J. Lauterjung, (2004) "Pre-Encoding PQoS Assessment Method for Optimized Resource Utilization" 2nd Inter. Conference on Performance Modelling and Evaluation of Heterogeneous Networks, Het-NeTs04, Ilkley, United Kingdom, 2004.
[Koumaras-2007_a]	H. Koumaras, "Method for predicting perceived quality of encoded video in relation to encoding bit rate and spatiotemporal content dynamics," PhD Dissertation, University of Athens, Computer Science and Telecommunications Department, Athens, Greece (2007).

[Koumaras-2007_b]	H. Koumaras et al., "Quantified PQoS assessment based on fast estimation of the spatial and temporal activity level," <i>Multimedia Tools Appl.</i> 34(3), 355–374 (2007).
[Kourtis-2016]	M.A Kourtis, H. Koumaras, and F. Liberal, "Reduced-Reference Video Quality Assessment Using a Static Video Pattern," <i>SPIE J. Electronic Imaging</i> , vol. 25, no. 4, 2016, article 043011.
[Kusuma-2003]	T. Kusuma and H.-J. Zepernick, "A reduced-reference perceptual quality metric for in-service image quality assessment," in <i>Proc. Joint 1st Workshop Mobile Future Symp. Trends Commun.</i> , Oct. 2003, pp. 71–74
[Le-Callet-2010]	P. Le Callet, C. V. Gaudin, and D. Barba, "Continuous quality assessment of MPEG2 video with reduced reference," in <i>Proc. Int. Workshop Video Process. Quality Metrics Consumer Electron.</i> , Jan. 2005, pp. 11–16.
[Li-2002]	X. Li, "Blind image quality assessment," in <i>Proc. IEEE Int. Conf. Image Process.</i> , vol. 1. Sep. 2002, pp. 449–452.
[Li-2009]	Q. Li and Z. Wang, "Reduced-reference image quality assessment using divisive normalization-based image representation," <i>IEEE J. Select. Top.Signal Process.</i> , vol. 3, no. 2, pp. 201–211, Apr. 2009.
[Lin-2015]	T. Lin et al., "Optimal Network Function Virtualization Realizing End-to-End Requests," <i>Proc. IEEE GLOBECOM</i> , San Diego, (2015).
[Lu-2002]	L. Lu, Z. Wang, A. C. Bovik and J. Kouloheris, "Full-reference video quality assessment considering structural distortion and no-reference quality evaluation of MPEG video," in <i>Proceedings. {IEEE} International Conference on Multimedia and Expo (2002)</i> .
[Lucas-2017]	Lucas, L. F. R.; da Silva, E. A. B.; de Faria, S. M. M.; Rodrigues, N. M. M. & Pagliari, C. L. <i>Efficient Predictive Algorithms for Image Compression</i> Springer International Publishing, 2017
[Ma-2011]	L. Ma, S. Li, F. Zhang, and K. N. Ngan, "Reduced-reference image quality assessment using reorganized DCT-based image representation," <i>IEEE Trans. Multimedia</i> , vol. 13, no. 4, pp. 824–829, Aug. 2011.
[Marziliano-2004]	P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, "Perceptual blur and ringing metrics: Application to JPEG2000," <i>Signal Process.: Image Commun.</i> , vol. 19, no. 2, pp. 163–172, 2004.
[McKeown-2008]	N. McKeown et al., "OpenFlow: Enabling Innovation in Campus Networks," <i>SIGCOMM Computer Comm. Rev.</i> , vol. 38, no. 2, 2008, pp. 69–74.

[Mullin-2001]	J. Mullin, L. Smallwood, A. Watson and G. Wilson, (2001) "New techniques for assessing audio and video quality in real-time interactive communications" Third International Workshop on Human Computer Interaction with Mobile Devices, Lille, France, 2001.
[Oechsner-2015]	S. Oechsner and A. Ripke, "Flexible support of VNF placement functions in OpenStack," in Network Softwarization (NetSoft), 2015 1st IEEE Conference on, 2015, pp. 1-6.
[Olson-1994]	J. Olson (1994), "In a framework about task-technology fit, what are the tasks features", Proceedings of CSCW '94: Workshop on video mediated communication: Testing, Evaluation & Design Implications, 1994.
[Osseiran-2014]	A. Osseiran et.al. "Scenarios for 5G Mobile and Wireless Communications: The Vision of the METIS Project", IEEE Communications Magazine, May, 2014, pp. 26-35.
[Redi-2010]	J. A. Redi, P. Gastaldo, I. Heynderickx, and R. Zunino, "Color distribution information for reduced-reference assessment of perceived image quality," IEEE Trans. Circuits Syst. Video Technol., vol. 20, no. 12, pp.1757–1769, Dec.2010.
[Rehman-2012]	Abdul Rehman, Zhou Wang, "Reduced-Reference Image Quality Assessment by Structural Similarity Estimation". IEEE Transactions on Image Processing 21(8): 3378-3389 (2012).
[Riccobene-2016]	Vincenzo Riccobene, Michael J. McGrath, Michail-Alexandros Kourtis, George Xilouris and Harilaos Koumaras, Automated Generation of VNF Deployment Rules Using Infrastructure Affinity Characterization, 2nd IEEE Conference on Network Softwarization (NetSoft 2016), Seoul, Korea, June 6-10, 2016.
[Riera-2016]	Riera, J. F.; et al. "TeNOR: Steps towards an orchestration platform for multi-PoP NFV deployment", 2016 IEEE NetSoft Conference and Workshops (NetSoft), IEEE, 2016
[Ross-1993]	J. Ross Quinlan, C4.5: Programs for Machine Learning, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1993.
[Savi-2016_a]	M. Savi et al., "Impact of Processing Costs on Service Chain Placement in Network Functions Virtualization," Proc. IEEE NFV-SDN, San Francisco, (2016).
[Savi-2016_b]	M. Savi, A. Hmaity, G. Verticale, S. Host and M. Tornatore, "To distribute or not to distribute? Impact of latency on Virtual Network Function

	distribution at the edge of FMC networks," in 2016 18th International Conference on Transparent Optical Networks (ICTON), 2016.
[Schwarz-2007]	H. Schwarz, D. Marpe, and Th. Wiegand (2007) "Overview of the Scalable Video Coding Extension of the H.264/AVC Standard", IEEE Transactions On Circuits And Systems For Video Technology, Vol. 17, No. 9, September 2007, pp.1103-1120.
[Seshadrinathan-2010]	Seshadrinathan, K, Soundararajan, R., Bovik, A.C., Cormack, L.K, "Study of Subjective and Objective Quality Assessment of Video", vol. 19, issue 6, pp. 1427-1441, 2010.
[Sgambelluri-2013]	A. Sgambelluri, A. Giorgetti, F. Cugini, F. Paolucci, and P. Castoldi, "OpenFlow-based segment protection in Ethernet networks," IEEE/OSA Journal of Optical Communications and Networking, vol. 5, no. 9, pp. 1066-1075, 2013.
[Sheikh-2005]	H. R. Sheikh, A. C. Bovik, and L. Cormack, "No-reference quality assessment using natural scene statistics: JPEG2000," IEEE Trans. Image Process., vol. 14, no. 11, pp. 1918-1927, Nov. 2005.
[Simoncelli-2001]	E. P. Simoncelli and B. Olshausen, "Natural image statistics and neural representation," Annu. Rev. Neurosci., vol. 24, pp. 1193-1216, May 2001.
[Soundararajan-2011]	R. Soundararajan and A. C. Bovik, "RRED indices: Reduced reference entropic differencing framework for image quality assessment," in Proc. IEEE Int. Conf. Acoust. Speech Signal Process., May 2011, pp. 1149-1152
[Staelens-2010]	N. Staelens , S. Moens , W. Van den Broeck , I. Marien , B. Vermeulen , P. Lambert , R. Van de Walle and P. Demeester "Assessing quality of experience of IPTV and video on demand services in real-life environments", IEEE Trans. Broadcast., vol. 56, no. 4, pp.458 -466 2010.
[Sugimoto-2000]	Sugimoto O, Kawada R, Wada M, Matsumoto S (2000) Objective measurement scheme for perceived picture quality degradation caused by MPEG encoding without any reference pictures. In: Proc. SPIE, vol. 4310, p 932.
[Tagliasacchi-2010]	M. Tagliasacchi, G. Valenzise, M. Naccari, and S. Tubaro, "A reduced-reference structural similarity approximation for videos corrupted by channel errors," Multimedia Tools and Applications, vol. 48, pp. 471-492, 2010.
[Tehrani-2014]	M. N. Tehrani, M. Uysal, H. Yanikomeroglu, "Device-to-Device Communication in 5G Cellular Networks: Challenges, Solutions, and

	Future Directions”, IEEE Communications Magazine, May, 2014, pp. 86-92
[Trajkovska-2017]	Irena Trajkovska, Michail-Alexandros Kourtis, et al. “SDN-based service function chaining mechanism and service prototype implementation in NFV scenario”, Computer Standards & Interfaces, Volume 54, Part 4, 2017, Pages 247-265
[Turner-1986]	J. Turner, “New Directions in Communications (or Which Way to the Information Age?),” IEEE Comm. Magazine, vol. 24, no.10, 1986, pp. 8–15.
[Wang-2000]	Z. Wang, A. C. Bovik, and B. L. Evans, “Blind measurement of blocking artifacts in images,” in Proc. IEEE Int. Conf. Image Process., vol. 3. Sep. 2000, pp. 981–984.
[Wang-2002]	Z. Wang, H. R. Sheikh, and A. C. Bovik, “No-reference perceptual quality assessment of JPEG compressed images,” in Proc. IEEE Int. Conf. Image Process., Sep. 2002, pp. 1-477–1-480.
[Wang-2003]	Z. Wang, H. R. Sheikh, and A. C. Bovik, “Objective video quality assessment,” in The Handbook of Video Databases: Design and Application, B. Furht and O. Marqure, Eds. Boca Raton, FL: CRC Press, Sep. 2003, pp. 1041–1078.
[Wang-2004_a]	Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli, Image quality assessment: From error visibility to structural similarity, Trans. Img. Proc. 13 (4) (2004) 600–612.
[Wang-2004_b]	Z. Wang, L. Lu, and A. C. Bovik, “Video quality assessment based on structural distortion measurement,” Signal Process.: Image Commun., vol. 19, no. 2, pp. 121–132, Feb. 2004.
[Wang-2005]	Z. Wang and E. P. Simoncelli, “Reduced-reference image quality assessment using a wavelet-domain natural image statistic model,” in Proc. SPIE Hum. Vision Electron. Imaging, Jan. 2005, pp. 149–159.
[Wang-2006]	Z. Wang, G. Wu, H. R. Sheikh, E. P. Simoncelli, E. Yang, and A. C. Bovik, “Quality-aware images,” IEEE Trans. Image Process., vol. 15, no. 6, pp. 1680–1689, Jun. 2006.
[Wang-2011]	Z. Wang and A. C. Bovik, “Reduced and no-reference image quality assessment: The natural scene statistic model approach,” IEEE Signal Process. Mag., vol. 28, no. 6, pp. 29–40, Nov. 2011.
[Webster-1993]	Webster AA, Jones CT, Pinson MH, Voran SD, Wolf S (1993) “An objective video quality assessment system based on human

	perception". In: SPIE human vision, visual processing, and digital display IV, vol 1913, pp 15–26
[Wolf-1999]	S. Wolf, Pinson MH (1999) "Spatial–temporal distortion metric for in-service quality monitoring of any digital video system". In: Proc. SPIE, vol 3845, pp 266–277.
[Wolf-2005]	S. Wolf and M. H. Pinson, "Low bandwidth reduced reference video quality monitoring system," in Proc. Int. Workshop Video Process. Quality Metrics Consumer Electron., Jan. 2005, pp. 76–79.
[Wu-1997]	H. R. Wu and M. Yuen, "A generalized block-edge impairment metric for video coding," IEEE Signal Process. Lett., vol. 4, no. 11, pp. 317–320, Nov. 1997.
[Wu-2015]	J. Wu, Z. Zhang, Y. Hong, and Y. Wen, "Cloud radio access network (C-RAN): a primer," Network, IEEE, vol. 29, no. 1, pp. 35–41, Jan 2015.
[Xilouris-2014]	G. Xilouris, E. Trouva, F. Lobillo, J. Soares, J. Carapinha, M. McGrath, G. Gardikis, P. Paglierani, E. Pallis, L. Zuccaro, Y. Rebahi, and A. Kourtis, "T-nova: A marketplace for virtualized network functions," in Networks and Communications (EuCNC), 2014 European Conference on, June 2014, pp. 1–5.
[Yang-2011]	S. Yang, "Reduced reference MPEG-2 picture quality measure based on ratio of DCT coefficients," Electronics Letters, 47(6), 382–383, (2011).
[Yoshida-2014]	M. Yoshida, W. Shen, T. Kawabata, K. Minato, and W. Imajuku, "MORSA: A multi-objective resource scheduling algorithm for NFV infrastructure," in Network Operations and Management Symposium (APNOMS), 2014 16th Asia-Pacific, 2014, pp. 1-6.
[Yu-2002]	Z. Yu, H. R. Wu, S. Winkler, and T. Chen, "Vision-model-based impairment metric to evaluate blocking artifact in digital video," Proc. IEEE, vol. 90, no. 1, pp. 154–169, Jan. 2002.
[Zeng-2010_a]	K. Zeng and Z. Wang, "Temporal motion smoothness measurement for reduced-reference video quality assessment," in Proc. IEEE Int. Conf. Acoust. Speech Signal Process., Mar. 2010, pp. 1010–1013.
[Zeng-2010_b]	K. Zeng and Z. Wang, "Quality-aware video based on robust embedding of intra and interframe reduced-reference features," in Proc. IEEE Int. Conf. Image Process., Sep. 2010, pp. 3229–3232.