

Plan de Pruebas Subjetivas Proyecto QoE4k

ALEJANDRO GODAY, FEDERICO PÁRAMOS, LAURA GOMEZ

15 de enero de 2019

Índice

1. Introducción	3
2. Observadores	4
2.1. Test de agudeza visual	4
2.2. Test de Daltonismo	4
3. Condiciones de las pruebas	5
3.1. Acondicionamiento de la sala	5
3.2. Especificaciones del Televisor	5
3.3. Distribución espacial de los observadores	6
4. Secuencias de video originales (SRCs)	7
4.1. Requerimientos	7
4.2. Complejidad espacial-temporal	7
4.3. Vector de Movimiento (MV) y Suma de Diferencias Absolutas (SAD)	8
5. Secuencias de video procesadas (PVS)	10
6. Método de Evaluación	13
7. Análisis de los resultados	14
7.1. Cálculo de MOS	14
7.2. Validación de calificaciones individuales	14
7.3. Cálculo del intervalo de confianza	14
7.4. Test t de Student	15
8. Referencias	17
A. Anexo 1 - Secuencias de video seleccionadas	18
B. Anexo 2 - Secuencias de video descartadas	20
C. Anexo 3 - Instrucciones para los observadores	22

1. Introducción

En este documento se detalla el procedimiento a seguir para realizar las pruebas subjetivas de calidad en Ultra Alta Definición, así como los materiales en ellas involucrados. Esto abarca:

- Requisitos para los observadores.
- Acondicionamiento de la sala.
- Especificación del televisor.
- Distribución espacial de los observadores.
- Descripción de los videos a utilizar.
- Descripción del método de evaluación.
- Análisis de resultados

La idea general del experimento es probar con cuántos observadores se puede realizar las pruebas, sin que sus votaciones individuales presenten diferencias estadísticamente significativas. Hacer los experimentos con varias personas en lugar de con uno o dos como es usual, abarataría costos relacionados al proceso de las pruebas subjetivas, así como también reduciría el tiempo que consumen las mismas.

2. Observadores

Quienes deseen participar de la evaluación no pueden ser expertos en calidad o procesamiento de video. A todos los candidatos a observadores se les tomarán los siguientes datos:

- Cédula de identidad.
- Posición (número de asiento)
- Edad.
- Sexo (Hombre/Mujer).
- Nivel educativo.
- Cantidad de horas que mira contenidos audiovisuales en su televisor/tablet/celular/PC por día.

Al momento de realizar la prueba, los observadores recibirán instrucciones sobre la misma. Los detalles e instrucciones se dan de forma oral y se detallan en el anexo C.

Además, los candidatos deberán pasar los tests de agudeza visual y daltonismo.

2.1. Test de agudeza visual

Este test se realiza con la gráfica de Snellen. Una persona tomándolo se cubre un ojo a tres metros de distancia de la gráfica, y lee en voz alta las letras de cada fila, comenzando por la fila de más arriba. La fila más pequeña que la persona puede leer, indica con exactitud la agudeza visual en ese ojo específico. El mínimo para pasar la prueba de agudeza visual y poder participar de la evaluación es de 20/30. En caso de que el candidato use lentes habitualmente, los usará también en la prueba.

2.2. Test de Daltonismo

Los observadores son luego examinados con el test de color Ishihara, para detectar daltonismo. Éste involucra 24 (o 38) placas pseudo-isocromáticas. Cada una de ellas muestra o bien un número, o algunas líneas. Las placas se sostienen a 75cm del sujeto, y son inclinadas de forma tal que el plano del papel esté a un ángulo recto respecto a la línea de visión. Los números que se ven en las placas 1-17 son declarados, y cada respuesta debería darse sin más de tres segundos de retraso. Si el sujeto es incapaz de ver los números, se usan las placas 18-24 y las líneas onduladas entre dos cruces son rastreadas con el cepillo. Cada rastreo debería ser completado en menos de 10 segundos. No es necesario en todos los casos usar la serie completa de placas. Las placas 16 y 17 pueden ser omitidas si el test es diseñado meramente para separar la gente que tiene visión de colores defectuosa de aquellos con apreciación de color normal. En un examen a gran escala, el test puede ser simplificado al uso de solamente seis placas. Para las pruebas subjetivas se utiliza este test abreviado de seis placas. Puede ser necesario variar el orden si se sospecha que el sujeto puede intentar engañar de forma deliberada.

3. Condiciones de las pruebas

3.1. Acondicionamiento de la sala

Las pruebas se realizarán en el Laboratorio de Televisión Digital del LATU. El acondicionamiento del espacio a utilizar, seguirá las pautas vistas en la recomendación ITU-T BT.500-13 [1]. En la tabla 1, se contrastan dichas pautas con las medidas de nuestra sala.

CUADRO 1
Condiciones de Sala

Parámetros	ITU-R BT.500	Medidas de la sala
Relación entre la luminancia de pantalla inactiva y el valor de cresta de la luminancia	$\leq 0,02$	
Relación entre la luminancia de la pantalla, cuando sólo se muestra el nivel del negro en una sala completamente oscura, y la correspondiente al blanco más intenso	$\simeq 0,01$	
Brillo y contraste de la imagen	Establecido vía PLUGE	
Ángulo máximo de observación con respecto a la normal (este valor se aplica a las pantallas de tubo de rayos catódicos (TRC), para otro tipo de pantallas se están estudiando los valores adecuados):	30 grados	
Relación entre la luminancia de fondo detrás del receptor de imágenes y el valor de cresta de luminancia de la imagen	$\simeq 0,15$	
Cromaticidad del fondo	D_{65}	
Otra iluminación de la sala	Débil	

3.2. Especificaciones del Televisor

Para las pruebas se utiliza un televisor LED marca SONY de 65", cuyo modelo es XBR-65X755D. El mismo soporta UHD (3840 x 2160) que es condición necesaria para las pruebas.

3.3. Distribución espacial de los observadores

La disposición de los observadores en la sala es la que se indica en la figura 1.

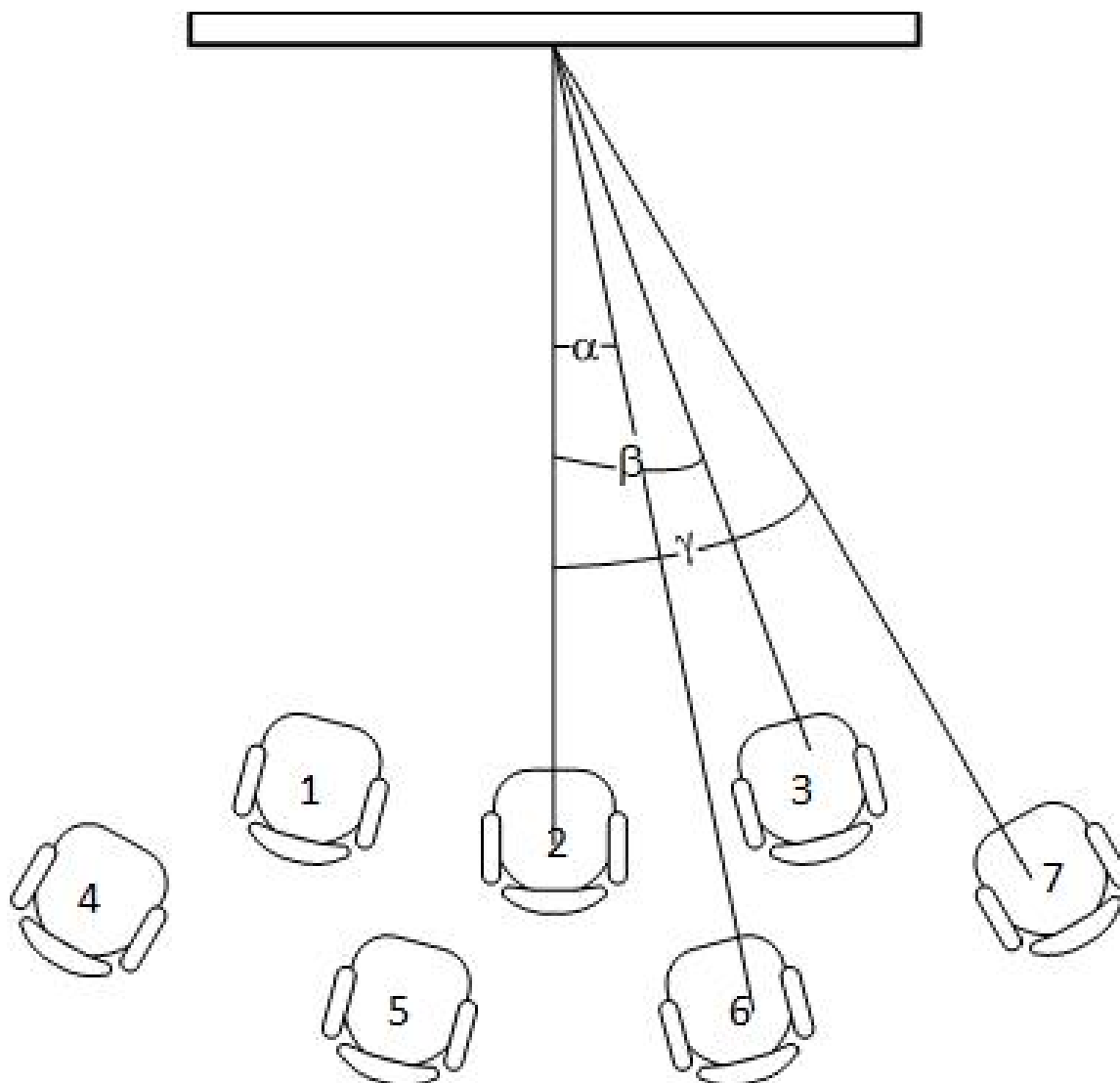


FIGURA 1
Distribución espacial de los observadores

La distancia d_2 será elegida en función de la altura de la pantalla como la distancia preferida de visión de acuerdo a la recomendación [1]. Por otro lado, los ángulos serán tales que el mayor de ellos (γ) respete el límite visto en la recomendación [1], o sea que habrá de ser menor a 30 grados. Los valores de las dimensiones a las cuales se hace alusión en la figura 1 se muestran en el cuadro 2.

CUADRO 2
Distribución espacial

Dimensión	Descripción	Valor
d_2	Distancia del centro de la pantalla a la posición 2	1,41 m
d_3	Distancia del centro de la pantalla a la posición 3	1,48 m
d_6	Distancia del centro de la pantalla a la posición 6	2,28 m
d_7	Distancia del centro de la pantalla a la posición 7	2,32 m
α	Ángulo entre las posiciones 2 y 6 al centro de la pantalla	14 °
β	Ángulo entre las posiciones 2 y 3 al centro de la pantalla	40 °
γ	Ángulo entre las posiciones 2 y 7 al centro de la pantalla	28 °

4. Secuencias de video originales (SRCs)

4.1. Requerimientos

A efectos de realizar las pruebas subjetivas, es necesario seleccionar secuencias de video “originales”, llamadas SRC (Source Reference). Dichas secuencias deben cumplir los siguientes requerimientos:

- No deben tener degradaciones visibles a criterio del grupo QoE4k
- El origen de la secuencia debe ser conocido y registrado
- Deben tener una duración entre 10 y 12 segundos
- En lo posible, deben cubrir diferentes rangos de actividad espacial y temporal

Se evaluaron veinticuatro posibles secuencias originales, de las cuales se seleccionaron diecisiete. Se pueden ver detalles de las veinticuatro secuencias en los anexos A y B. De las diecisiete secuencias seleccionadas, solo catorce se utilizan para las pruebas subjetivas, para que la prueba total no exceda los 45 minutos. Las siete secuencias restantes fueron descartadas por excesivo ruido. Todas las secuencias evaluadas son UHD (3840×2160).

4.2. Complejidad espacial-temporal

Para cuantificar la complejidad espacial y temporal se han utilizado las siguientes métricas:

- Índice de información espacial (SI)
- Índice de información temporal (TI)

Se procuró seleccionar las secuencias de forma tal que se abarcasen distintos grados de complejidad espacial y temporal.

La figura 2 es una gráfica de SI contra TI donde cada punto corresponde a una secuencia del conjunto de candidatas. Los valores de SI y TI se obtuvieron con el programa SITI-master [2], y se detallan en el Cuadro 3

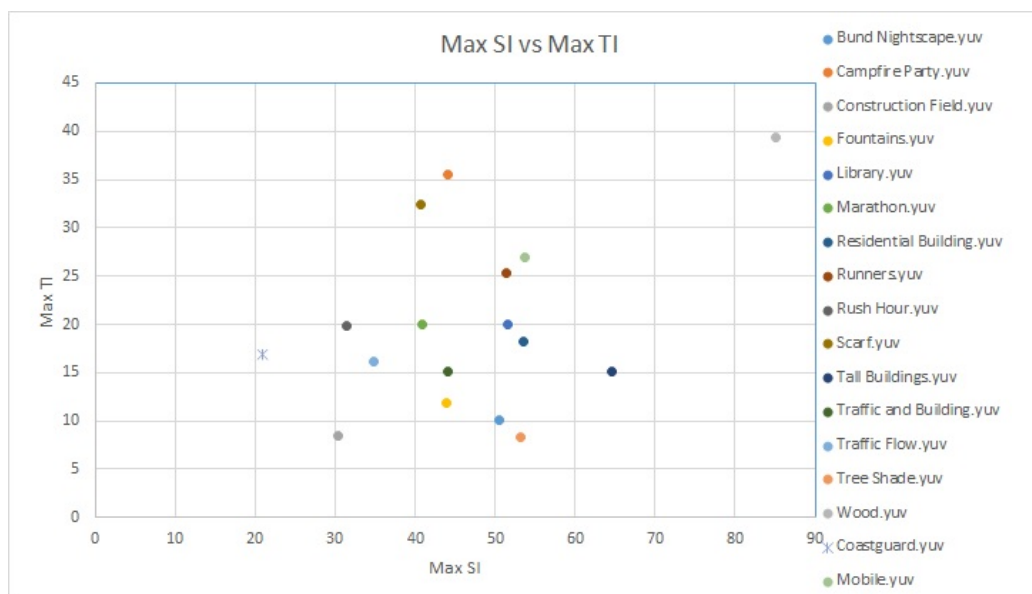


FIGURA 2

Gráfico de TI vs SI. Se muestran las complejidades espaciales y temporales para todas las secuencias candidatas.

CUADRO 3
Valores de SI y TI

Nombre del video	Origen	Max SI	Max TI
Bund Nightscape.yuv	SJTU	50,44	10,06
Campfire Party.yuv	SJTU	44,07	35,55
Construction Field.yuv	SJTU	30,39	8,45
Fountains.yuv	SJTU	43,93	11,82
Library.yuv	SJTU	51,64	20,03
Marathon.yuv	SJTU	40,85	19,05
Residential Building.yuv	SJTU	53,62	18,30
Runners.yuv	SJTU	51,40	25,35
Rush Hour.yuv	SJTU	31,43	19,91
Scarf.yuv	SJTU	40,72	32,44
Tall Buildings.yuv	SJTU	64,59	15,19
Traffic and Building.yuv	SJTU	44,04	15,13
Traffic Flow.yuv	SJTU	34,89	16,22
Tree Shade.yuv	SJTU	53,15	8,38
Wood.yuv	SJTU	85,16	39,38
Coastguard.yuv	Elemental	20,92	16,92
Mobile.yuv	Elemental	53,70	26,91

4.3. Vector de Movimiento (MV) y Suma de Diferencias Absolutas (SAD)

A efectos de cuantificar el movimiento presente en un video, es también posible calcular la magnitud de los vectores de movimiento (**MV** por Motion Vector) y la suma de diferencias absolutas (**SAD** por Sum of Absolute Differences) [4].

Para calcular esto se utilizan las siguientes herramientas: FFmpeg [5], Avisynth v.2.60 [6], Plugin MVTools (v.2.5.11.22) para Avisynth [7], VirtualDub v.1.10.4 [8], MVandSADReader.class (Programa JAVA provisto por equipo VQI).

Cálculo

1. Se usa FFmpeg [5] para pasar cada uno de los videos que tenemos con extensión YUV (<entrada>.yuv) a AVI (<entrada>.avi). A estos efectos se usa el comando:

```
ffmpeg -pix_fmt yuv420p -s 3840x2160 -i <entrada>.yuv -vcodec rawvideo -pix_fmt yuv420p -r 30 <entrada>.avi
```
2. Luego, para cada video se crea un script de Avisynth (archivo de texto con extensión .avs) con la forma siguiente:

```
clip = AviSource(<entrada>.avi)
clip2 = ConvertToYV12(clip)
clip3 = MSuper(clip2)
vectors = MAnalyse(clip3, isb = false, outfile = "<salida>", blksize = 8)
MShow(clip3, vectors)
```
3. Una vez que tenemos el script hecho, abrimos el programa VirtualDub, vamos a **File** → **Open Video File**, y ahí abrimos el script correspondiente al video cuyos MV y SAD queremos hallar. A continuación presionamos el botón de reproducir. Al finalizar la reproducción se genera un archivo binario (sin extensión), cuyo nombre es “<salida>” (como aparece en la cuarta línea del script correspondiente).

4. Para poder pasar la información de dicho archivo a formato .txt se utiliza el programa *MVandSADReader.class*.

Resultados

La figura 3 es una gráfica de MV promedio contra SAD promedio, donde cada punto corresponde a una secuencia del conjunto de candidatas. Los valores de MV promedio y SAD promedio, se obtuvieron con el a partir de los .txt generados, y se detallan en el cuadro 4.

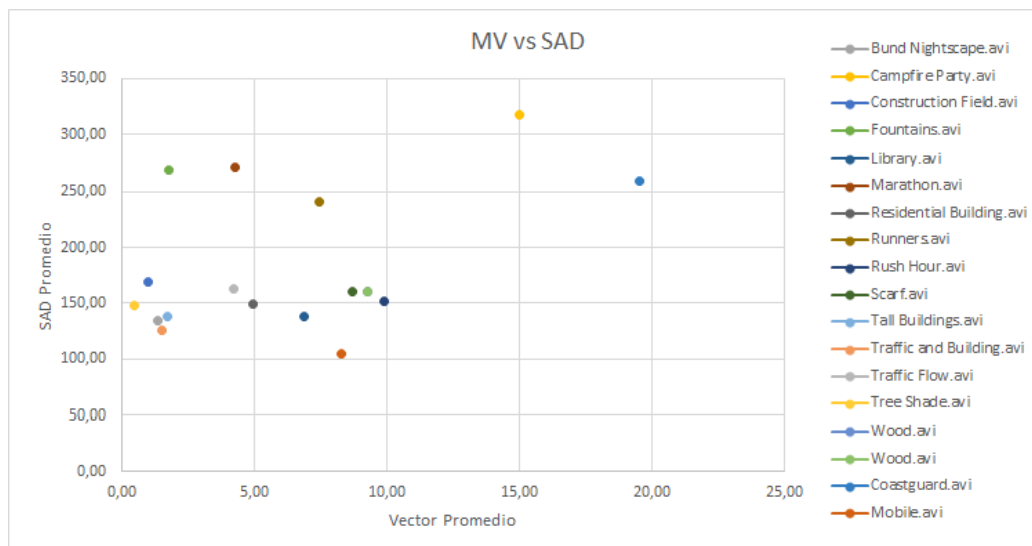


FIGURA 3

Gráfico de MV vs SAD. Se muestran el promedio de los vectores de movimiento y la suma de diferencias absolutas promedio para todas las secuencias candidatas.

CUADRO 4

Valores de MV y SAD

Nombre del video	Origen	Vector Promedio	SAD Promedio
Bund Nightscape.avi	SJTU	1,35	134,12
Campfire Party.avi	SJTU	15,00	317,10
Construction Field.avi	SJTU	0,99	168,80
Fountains.avi	SJTU	1,75	268,81
Library.avi	SJTU	6,88	138,10
Marathon.avi	SJTU	4,25	272,03
Residential Building.avi	SJTU	4,95	149,53
Runners.avi	SJTU	7,45	240,26
Rush Hour.avi	SJTU	9,89	151,44
Scarf.avi	SJTU	8,68	159,82
Tall Buildings.avi	SJTU	1,73	138,58
Traffic and Building.avi	SJTU	1,50	126,18
Traffic Flow.avi	SJTU	4,22	162,25
Tree Shade.avi	SJTU	0,48	147,70
Wood.avi	SJTU	9,26	160,04
Coastguard.avi	Elemental	19,51	259,80
Mobile.avi	Elemental	8,28	105,04

5. Secuencias de video procesadas (PVS)

Para la realización de pruebas subjetivas se generan secuencias de video degradadas, partiendo de los videos originales (SRC) y procesándolos según los distintos circuitos de referencia (HRC) establecidos. Las secuencias de video obtenidas se denominan PVS (Processed Video Sequence).

A estos efectos, se utiliza el programa *Ffmpeg* [5], al cual se le dio como entrada un video con formato raw (YUV), y el cual dio como salida un archivo mp4 codificado en HEVC según el circuito de referencia utilizado. Dicho programa se ejecuta por línea de comandos y los comandos utilizados tienen la siguiente forma:

```
ffmpeg.exe -f rawvideo -video_size 3840x2160 -pixel_format yuv420p -i <entrada>.yuv
-c:v hevc -b:v <bitRate>k -maxrate <bitRate>k -minrate <bitRate>k -r 30 -x265-params
-an -y <salida>.mp4
```

Donde para cada entrada, se utilizaron los valores de Bit Rate de los cuadros 5, 6 y 7, así obteniendo seis versiones con distintos grados de degradación.

CUADRO 5
Valores de Bitrates

Salida	Nivel de Tasa	Bit Rate (kbps)
Bund_Nightscape1.mp4	Muy bajo	2341.39
Bund_Nightscape2.mp4	Bajo	5069.66
Bund_Nightscape3.mp4	Medio	7964.73
Bund_Nightscape4.mp4	Medio alto	11994.09
Bund_Nightscape5.mp4	Alto	18011.08
Bund_Nightscape6.mp4	Muy alto	24072.74
Campfire_Party1.mp4	Muy bajo	1908
Campfire_Party2.mp4	Bajo	3566
Campfire_Party3.mp4	Medio	5916
Campfire_Party4.mp4	Medio alto	7817
Campfire_Party5.mp4	Alto	12300.38
Campfire_Party6.mp4	Muy alto	16231.04
Coastguard0.mp4	Muy bajo	1491.21
Coastguard1.mp4	Bajo	1996.5
Coastguard2.mp4	Medio	3498.1
Coastguard3.mp4	Medio alto	5956.84
Coastguard4.mp4	Alto	9886.47
Coastguard5.mp4	Muy alto	11882.27
Construction_Field2.mp4	Muy bajo	927.59
Construction_Field3.mp4	Bajo	1432.58
Construction_Field4.mp4	Medio	2420.89
Construction_Field5.mp4	Medio alto	3917.54
Construction_Field6.mp4	Alto	9869.63
Construction_Field7.mp4	Muy alto	14877.93
Fountains1.mp4	Muy bajo	2404.09
Fountains2.mp4	Bajo	5240.66
Fountains3.mp4	Medio	8167.25
Fountains4.mp4	Medio alto	12190.31
Fountains5.mp4	Alto	18221.92
Fountains6.mp4	Muy alto	24236.71

CUADRO 6
Valores de Bitrates cont.

Salida	Nivel de Tasa	Bit Rate (kbps)
Marathon1.mp4	Muy bajo	2054.05
Marathon2.mp4	Bajo	4056.19
Marathon3.mp4	Medio	7514.98
Marathon4.mp4	Medio alto	11057.18
Marathon5.mp4	Alto	15046.79
Marathon6.mp4	Muy alto	18043.88
Mobile1.mp4	Muy bajo	821.53
Mobile2.mp4	Bajo	1111.46
Mobile3.mp4	Medio	1788.37
Mobile4.mp4	Medio alto	2323.05
Mobile5.mp4	Alto	5377.13
Mobile6.mp4	Muy alto	8447.09
Runners1.mp4	Muy bajo	2124.51
Runners2.mp4	Bajo	5354.24
Runners3.mp4	Medio	6753.5
Runners4.mp4	Medio alto	9930.42
Runners5.mp4	Alto	12320.76
Runners6.mp4	Muy alto	18263.62
Rush_Hour1.mp4	Muy bajo	1573.02
Rush_Hour2.mp4	Bajo	2737.13
Rush_Hour3.mp4	Medio	4798.39
Rush_Hour4.mp4	Medio alto	5988.94
Rush_Hour5.mp4	Alto	8126.02
Rush_Hour6.mp4	Muy alto	12287.15
Scarf0.mp4	Muy bajo	377.06
Scarf1.mp4	Bajo	1537.41
Scarf2.mp4	Medio	2886.03
Scarf3.mp4	Medio alto	5111.91
Scarf4.mp4	Alto	7025.3
Scarf5.mp4	Muy alto	11723.19
TrafficAndBuilding0	Muy bajo	1087.67
TrafficAndBuilding1	Bajo	2256.33
TrafficAndBuilding2	Medio	4966.11
TrafficAndBuilding3	Medio alto	7845.67
TrafficAndBuilding4	Alto	11828.24
TrafficAndBuilding5	Muy alto	17838.98
Traffic_Flow1.mp4	Muy bajo	899.09
Traffic_Flow2.mp4	Bajo	1959.23
Traffic_Flow3.mp4	Medio	4009.65
Traffic_Flow4.mp4	Medio alto	5000.24
Traffic_Flow5.mp4	Alto	7975.35
Traffic_Flow6.mp4	Muy alto	11850.27
Tree_Shade0.mp4	Muy bajo	1019.83
Tree_Shade1.mp4	Bajo	2218.25
Tree_Shade2.mp4	Medio	4871.96
Tree_Shade3.mp4	Medio alto	7627.09
Tree_Shade4.mp4	Alto	11475.88
Tree_Shade5.mp4	Muy bajo	17325.7

CUADRO 7

Valores de Bitrates cont.

Salida	Nivel de Tasa	Bit Rate (kbps)
Wood1.mp4	Muy bajo	1779.16
Wood2.mp4	Bajo	3689.95
Wood3.mp4	Medio	5180.78
Wood4.mp4	Medio alto	8090.87
Wood5.mp4	Alto	11952.7
Wood6.mp4	Muy alto	14928.66

6. Método de Evaluación

El método a utilizar en las pruebas es el denominado “ACR”, también conocido como “Índice por categoría absoluta”. En esta sección se presenta una descripción de dicho método, extraída de la recomendación ITU-T P.910 [3].

Descripción del método ACR

El método de los índices por categorías absolutas es un juicio de categorías en el que las secuencias de prueba se presentan una por vez y se califican independientemente en una escala de categorías. (Este método se denomina también método de evaluación con un solo estímulo.) El método especifica que después de cada presentación se invite a los sujetos a evaluar la calidad de la secuencia mostrada. En la Figura 1 se ilustra el diagrama de tiempos de la presentación del estímulo. Si se utiliza un tiempo de votación constante entonces el tiempo de votación debe ser igual o inferior a 10 s.

En la Figura 4 (extraída de [3]) se ilustra el diagrama de tiempos de la presentación del estímulo.

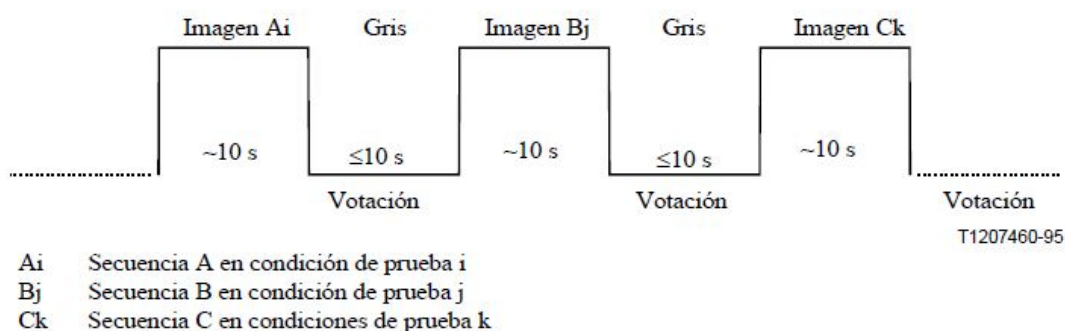


FIGURA 4
Presentación del estímulo en el método ACR

El tiempo de votación deberá ser igual o inferior a 10 s. El tiempo de presentación puede reducirse o aumentarse en función del contenido del material de prueba. Para evaluar la calidad global se debe utilizar la siguiente escala de cinco niveles, como se muestra en la tabla 8:

CUADRO 8
Puntajes y Niveles

Puntuación	Nivel
5	Excelente
4	Buena
3	Aceptable
2	Mediocre
1	Mala

7. Análisis de los resultados

7.1. Cálculo de MOS

El primer paso al analizar las evaluaciones subjetivas de calidad consiste en hallar el valor de MOS final correspondiente a cada contenido de video. Para cada contenido j se define MOS_j como:

$$\text{MOS}_j = \bar{u}_j = \frac{1}{N} \sum_1^N u_{ij} \quad (1)$$

donde u_{ij} representa la evaluación del observador i para el contenido j y N es el número de observadores válidos.

En nuestro caso, se calculan valores de MOS promediando las opiniones de distintos observadores ubicados en la misma posición. En principio, no se calcula el MOS como promedio de opiniones de observadores en distintas posiciones.

7.2. Validación de calificaciones individuales

Se establece un criterio para validar las calificaciones individuales. Para ello se calcula la correlación entre las calificaciones individuales y las calificaciones promedio de todos los evaluadores por posición. Esta correlación se define según la ecuación (2) :

$$r_1(i) = \frac{\sum_{j=1}^K \bar{u}_j u_{ij} - \frac{\left(\sum_{j=1}^K \bar{u}_j\right) \left(\sum_{j=1}^K u_{ij}\right)}{K}}{\sqrt{\left(\sum_{j=1}^K \bar{u}_j^2 - \frac{\left(\sum_{j=1}^K \bar{u}_j\right)^2}{K}\right) \left(\sum_{j=1}^K u_{ij}^2 - \frac{\left(\sum_{j=1}^K u_{ij}\right)^2}{K}\right)}} \quad (2)$$

Donde:

- j es el índice de PVS
- \bar{u}_j es el MOS calculado con todos los evaluadores por PVS j
- u_{ij} es la calificación individual del evaluador i para ese PVS j
- K es el número de PVSs

Para cada posición se calcula $r_1(i)$ para cada observador i , y se excluye a los que tienen $r_1(i)$ menor a 0,75. En caso de excluirse observadores, se vuelve a calcular el MOS de (1). Este proceso se realiza una única vez.

7.3. Cálculo del intervalo de confianza

Cuando se presenten los resultados de una prueba, todas las notas medias (MOS) deberán tener un intervalo de confianza asociado, que se obtiene a partir de la desviación típica y el tamaño de cada muestra. Se utiliza en este caso un intervalo de confianza del 95 %, que viene dado por las ecuaciones (3) , (4) y (5):

$$\text{IC} = [\bar{u}_j - \delta_j, \bar{u}_j + \delta_j] \quad (3)$$

donde:

$$\delta_j = 1,96 \cdot \frac{S_j}{\sqrt{N}} \quad (4)$$

y

$$S_j = \sqrt{\sum_{i=1}^N \frac{(\bar{u}_j - u_{ij})^2}{N-1}} \quad (5)$$

donde:

- j es el índice de la PVS
- i es el índice del observador
- N es la cantidad de observadores
- \bar{u}_j es el **MOS** _{j}
- u_{ij} es la calificación individual del observador i para la PVS j .

Con una probabilidad del 95 %, el valor absoluto de la diferencia entre la nota media experimental y la nota media “verdadera” (para un número de observadores muy elevado) es menor que el intervalo de confianza del 95 %, siempre que la distribución de las notas individuales cumpla ciertos requisitos.

Calculando los intervalos de confianza para cada PVS en cada asiento, podremos tener una idea de cuan similar es la percepción de calidad del usuario que se encuentra en una posición, a la de un usuario que se encuentra en otra.

7.4. Test t de Student

Un análisis más riguroso que el presentado en la sección anterior, es realizar un test t de Student de dos muestras y varianzas desiguales, usando una distribución de dos colas para determinar si en efecto las calidades subjetivas dadas por los valores promedio de las muestras de un par de asientos no son iguales. La hipótesis nula H_0 en este caso sería que los observadores en diferentes asientos perciben la misma calidad para una PVS, y la hipótesis alternativa H_a es que los observadores en distintos asientos no perciben la misma calidad para una PVS. Para comparar las medias de dos poblaciones, puede usarse la estadística t, la cual se expresa como:

$$t = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{s_1^2/n_1 + s_2^2/n_2}} \quad (6)$$

donde \bar{X}_i , s_i^2 , n_i denotan la media de la muestra, la varianza de la muestra, el tamaño de la i -ésima muestra y $i \in 1,2$.

Se calculan \bar{X}_i y s_i^2 como:

$$\bar{X}_i = \frac{1}{n_i} \sum_{k=1}^{k=n_i} X_k \quad (7)$$

$$s_i^2 = \frac{1}{n_i - 1} \sum_{k=1}^{k=n_i} (X_k - \bar{X}_i)^2 \quad (8)$$

donde X_k es la k -ésima opinión en un asiento para una PVS, e $i \in 1,2$.

Al calcular la estadística t de esta forma y aproximándola con una distribución t de Student cuyo grado de libertad DF se define así

$$DF = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{\frac{(s_1^2/n_1)^2}{n_1-1} + \frac{(s_2^2/n_2)^2}{n_2-1}} \quad (9)$$

se puede calcular un valor de probabilidad p a partir de la estadística t que indica el grado al cual las medias de las dos poblaciones se consideran diferentes. Cuanto más pequeño es el valor p , más significativa es la diferencia entre las distribuciones de las poblaciones.

Un valor p menor a 0,05 indica una probabilidad muy baja de cometer un error tipo I (esto es, rechazar la hipótesis nula cuando es cierta). En tal caso, la hipótesis nula puede ser rechazada con seguridad, y puede concluirse que hay significado estadístico en que las dos ubicaciones perciben calidades diferentes. Un valor p mayor o igual a 0,05 significa que la hipótesis nula no puede ser rechazada con confianza. Sin embargo, todavía existe la posibilidad de cometer un error tipo II (esto es, no poder rechazar la hipótesis nula cuando de hecho la hipótesis alternativa es cierta).








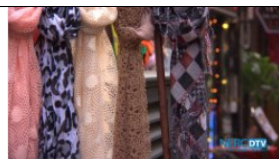

El test será aplicado a todas las degradaciones de cada uno de los videos, tomando pares de ubicaciones. Cada ubicación será comparada con la ubicación central, ubicación 2 en la figura 1. Con este criterio se obtendrán 504 resultados del test.

8. Referencias

- [1] ITU-R BT.500-13, “Methodology for the subjective assessment of the quality of television pictures, ” International Telecommunication Union, 2012.
- [2] <https://github.com/Telecommunication-Telemedia-Assessment/SITI>
- [3] ITU-T, “Subjective video quality assessment methods for multimedia applications”, Recommendation ITU-R P 910, Sep 1999.
- [4] E. G. Richardson, Iain (2003). H.264 and MPEG-4 Video Compression: Video Coding for Next-generation Multimedia. Chichester: John Wiley and Sons Ltd.
- [5] <https://www.ffmpeg.org/>
- [6] Sitio web de Avisynth: <http://avisynth.nl>
- [7] Sitio web de MVTools <https://avisynth.org.ru/mvtools/mvtools2.html>
- [8] Sitio Web de VirtualDub: <http://www.virtualdub.org/>
- [9] *Shanghai Jiao Tong University*. [Online]. Available: <http://medialab.sjtu.edu.cn/web4k/index.html>
- [10] *Ultra Video Group*. [Online]. Available: <http://ultravideo.cs.tut.fi/>

A. Anexo 1 - Secuencias de video seleccionadas

Las secuencias de video utilizadas fueron descargadas del sitio de *Shanghai Jiao Tong University* [9]

Título	Descripción	Primer Cuadro
Bund Nightscape	Plano aéreo de una ciudad en cámara rápida, durante la noche.	
Campfire Party	Muestra llamas ante el equipo de NERC-DTV en una fiesta con una fogata.	
Construction Field	Muestra una excavadora en un sitio de construcción.	
Fountains	Muestra chorros verticales de una fuente de agua frente a un edificio alto.	
Marathon	Muestra la escena de las primeras etapas de la 2012 Shanghai International Marathon Race.	
Runners	Muestra muchos corredores en medio de la 2012 Shanghai International Marathon Race.	
Rush Hour	Muestra muchos estudiantes yendo a la cantina o al dormitorio después de clases.	
Scarf	Muestra unas bufandas moviéndose al viento.	
Traffic and Building	Muestra tráfico con una ciudad de fondo.	

<p>Traffic Flow</p>	<p>Muestra caminos con automóviles moviéndose en distintas direcciones</p>	
<p>Tree Shade</p>	<p>Muestra un lugar con mucha vegetación y un árbol cuyas hojas se mueven al viento.</p>	
<p>Wood</p>	<p>Muestra un bosque en el campus de SJTU con rayos del sol penetrándolo.</p>	
<p>Mobile</p>	<p>Muestra un tren de juguete circulando y otros objetos en movimiento.</p>	
<p>Coastguard</p>	<p>Muestra un barco navegando bajo el sol.</p>	

B. Anexo 2 - Secuencias de video descartadas

Las siguientes secuencias de video descartadas fueron descargadas del sitio de *Ultra Video Group* [10].

Título	Descripción	Primer Cuadro
Beauty	Muestra la cara de una mujer que pestaña y cuyo pelo se mueve con el aire.	
Bosphorous	Muestra un bote navegando en un río, con un puente y una ciudad de fondo.	
HoneyBee	Muestra a una abeja moviéndose entre distintas flores.	
Jockey	Muestra a un jinete cabalgando en una carrera.	
ReadySetGo	Muestra el comienzo de una carrera de caballos.	
ShakeNDry	Muestra a un perro sacudiéndose el agua para secarse.	
Yacht Ride	Muestra una pareja en un yate navegando.	

Las siguientes secuencias de video descartadas fueron descargadas del sitio de *Shanghai Jiao Tong University* [9]

Library	Muestra el exterior de una biblioteca con gente pasando.	
Residential Building	Muestra el exterior de un edificio residencial.	
Tall Buildings	Muestra edificios altos en Lujiazui, Pudong New District Shanghai	

C. Anexo 3 - Instrucciones para los observadores

Al inicio de la prueba se les explica a los observadores la forma en que se realizará la misma, así como también los detalles para votar.

1. Se indica a que red de wifi deben conectar sus smartphones y las credenciales de acceso. En caso de que alguien no tenga un celular con navegador, se le proveerá uno.
2. Se exhorta a los participantes que utilizan celular propio a silenciar sus celulares durante la prueba y a prestar atención solamente a la prueba (no a mensajes entrantes, etc.).
3. Se indica a cada participante cuál es su número de asiento.
4. Se les explica como registrarse a la aplicación que recoge los datos y los votos. Para esto se provee la IP y se les comenta que deben completar los datos que se piden.
5. Se explica sucintamente como funciona el programa en cuanto a la reproducción de los videos y como transcurrirá la votación.
6. Se aclara que la votación será exclusivamente sobre la calidad de imagen del video y no sobre el contenido.

La exposición oral inicial se redacta a continuación, para que la prueba sea la misma, sin importar que cambie la persona que da las indicaciones.

Estimados participantes:

Sean bienvenidos a la prueba de evaluación de calidad de video, en el marco de nuestro proyecto de fin de carrera en la Facultad de Ingeniería de la Universidad de la República. Desde ya agradecemos su participación.

Si alguien necesita salir de la sala (por ejemplo, para ir al baño), que lo haga ahora. Una vez comenzada la prueba no se podrá abandonar su asiento hasta que la misma finalice.

Pausa para ir al baño.

Pasamos a comentarles cómo se desarrollará la prueba. Se mostrará una serie de 84 videos sin sonido de 12 segundos de duración cada uno. Después de cada video, el mismo deberá ser calificado en una escala de 5 posibles categorías:

- Excelente
- Bueno
- Aceptable
- Mediocre
- Malo

Solo se podrá calificar una vez cada video. Los mismos contenidos se verán varias veces, con diferentes tipos de degradaciones o problemas de calidad. Por favor, realice cada calificación en forma independiente de las anteriores.

Desde el navegador web de los celulares que les vamos a proveer a cada uno, completen los datos pedidos por la aplicación. El número de cédula tendrá que ingresarse sin puntos y sin el guion. El número de asiento que les pedirá la aplicación, es el que está pegado a cada una de sus sillas.

Pausa. Se le entregan los celulares a todos los usuarios.

Esta misma aplicación es la que usarán para emitir sus votos luego de cada video. En el televisor se les indicará cuándo comienza el siguiente video, y cuando el video finaliza les pedirá que lo califiquen. En ese momento aparecerán en sus celulares las 5 categorías para que califiquen. No es necesario apurarse a votar, ya que la aplicación espera a recibir todos los votos antes de pasar al siguiente video.

Pausa. También se supervisa y verifica que los usuarios completen satisfactoriamente esta etapa.

Les vamos a solicitar encarecidamente que silencien sus celulares. A aquellos que estén usando sus propios celulares para la prueba, por favor, rogamos se abstengan de usarlos con otros propósitos que no sea el de votar.

Haremos especial hincapié, en que se deberá evaluar únicamente la calidad de cada video, independientemente de su contenido, el cual puede llamar más o menos nuestra atención dependiendo de nuestros gustos.

¿Alguna pregunta?

Muchas gracias.