

ugiat

Aplicaciones de la IA para la generación automática de resúmenes audiovisuales

Generación de resúmenes audiovisuales

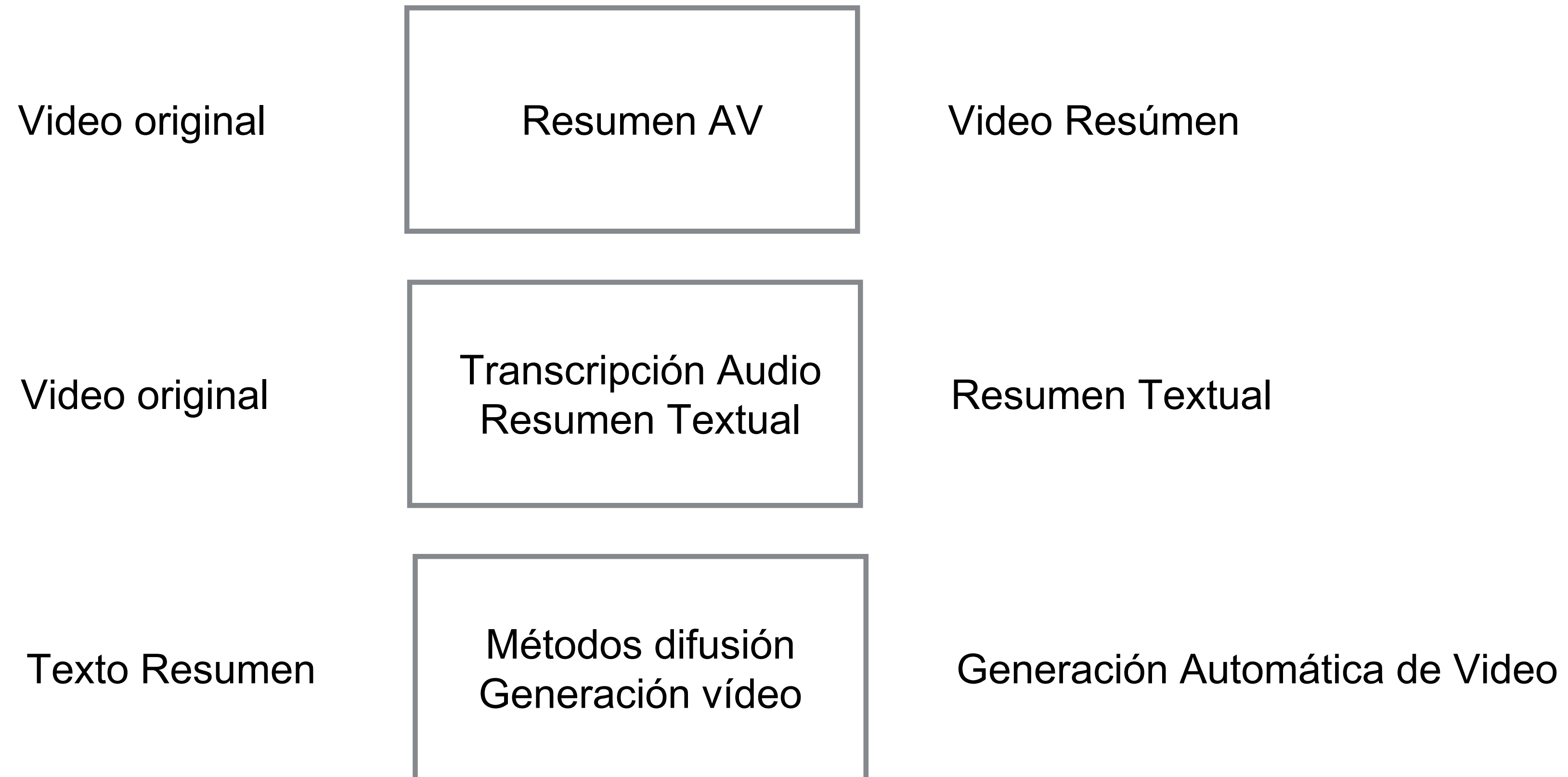
OBJETIVO: A partir de un vídeo original generar un resumen (Highlights) automático que sirva para obtener una síntesis y visión general del contenido.

APLICACIONES: El conjunto de aplicaciones es muy variado:

1. Resúmenes de acontecimientos deportivos.
2. Generación automática de cortos de películas.
3. Generación de resúmenes de noticias, entrevistas, etc.
4. Generación de resúmenes de capítulos anteriores de series, miniaturas de vídeo, etc.
5. Generación de minutados de TV

Generación de resúmenes audiovisuales

Diferentes contextos



Generación de resúmenes audiovisuales

TECNOLOGIAS

1. Métodos clásicos
2. Metodologías basadas en el análisis de imagen.
3. Metodologías basadas en el análisis multimodal
4. Metodologías basadas en el análisis textual

APLICACIONES: El conjunto de aplicaciones es muy variado:

1. Resúmenes de acontecimientos deportivos.
2. Generación automática de cortos de películas.
3. Generación de resúmenes de noticias, entrevistas, etc.
4. Generación de resúmenes de capítulos anteriores de series, miniaturas de vídeo, etc.
5. Generación de minutos de TV

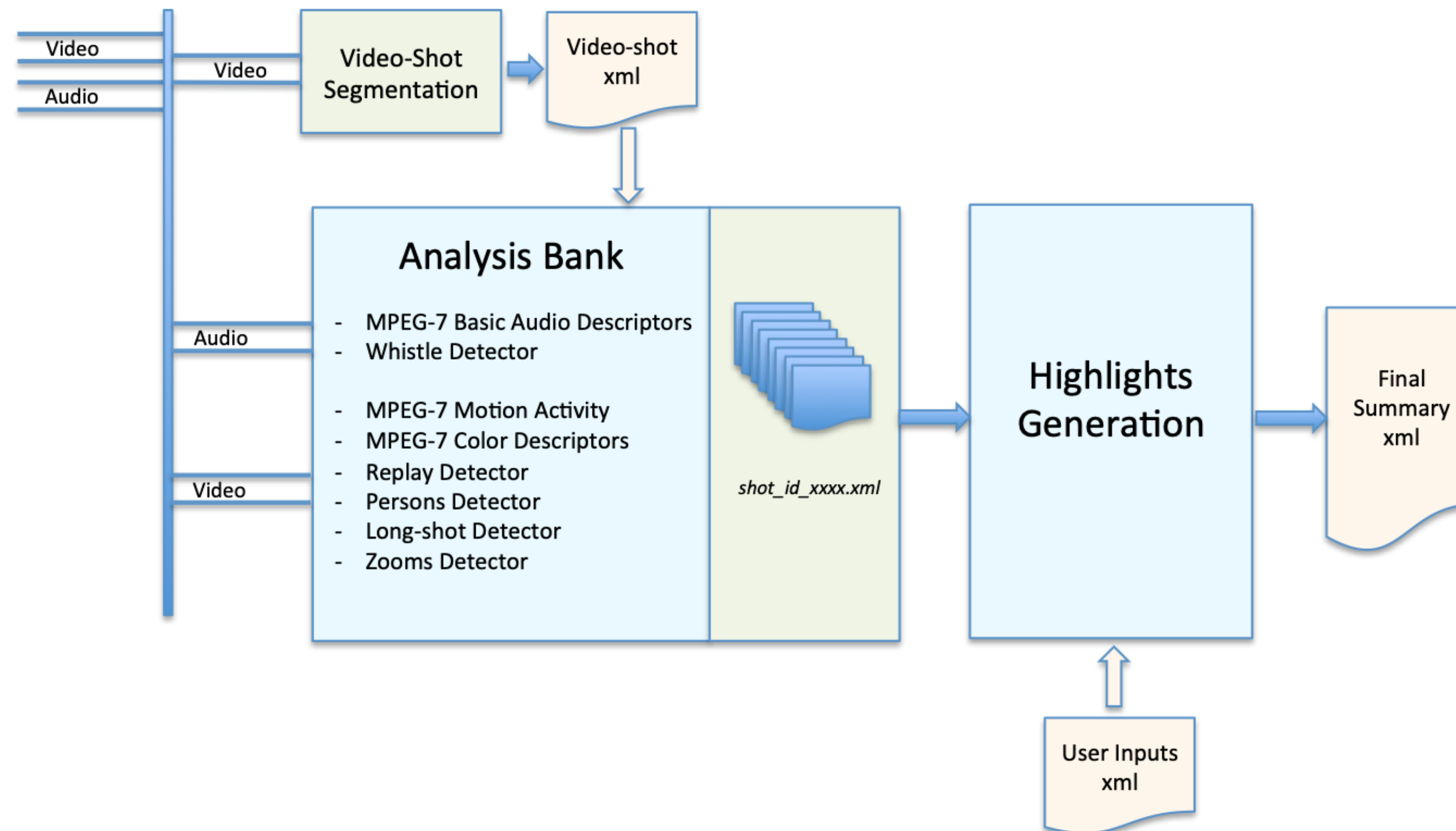
Tecnologías Clásicas

The video is segmented in shots and the Highlights will be a sequence of best shots.

Analysis per shot

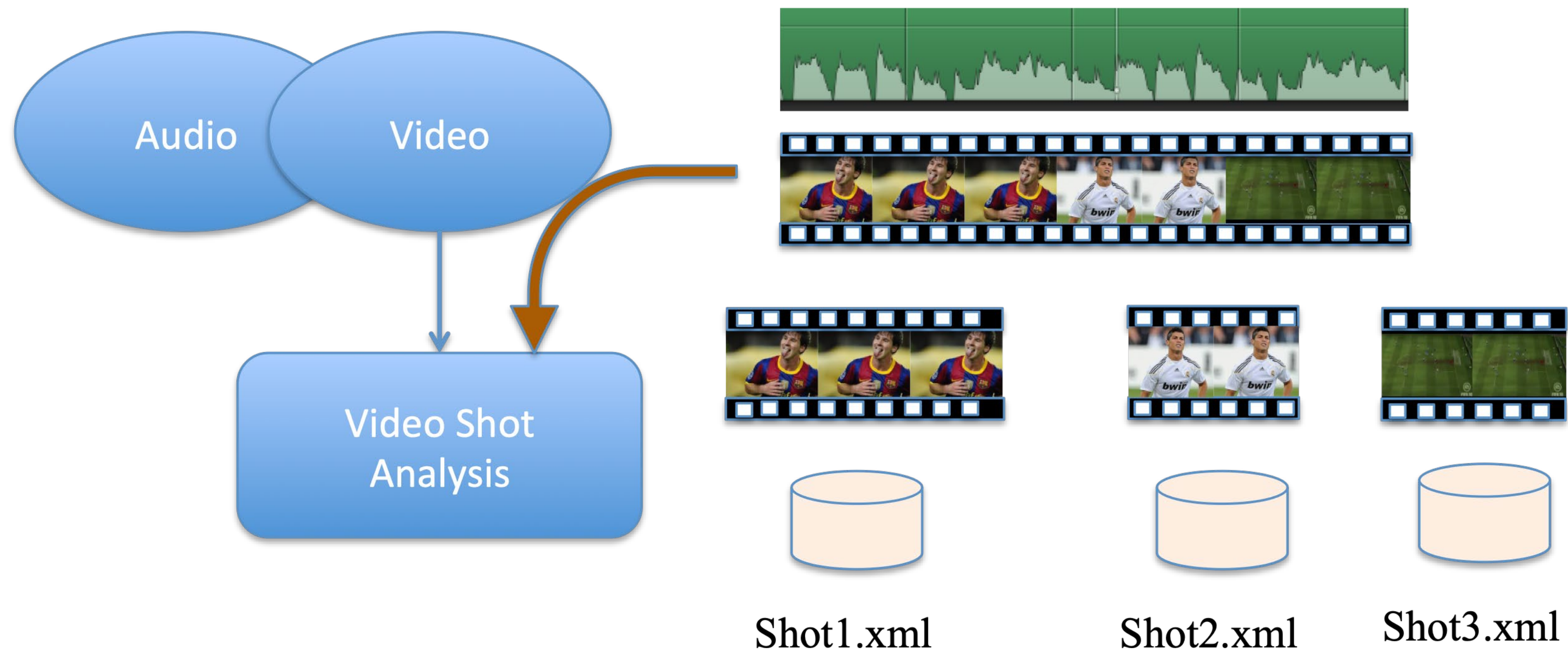
- Low level audio MPEG-7 descriptors
- Motion descriptors (Motion Activity & Camera Motion)
- Detection of Zoom-In & Zoom-Out
- List of Keyframes representing the video
- Color Layout, Color Structure & Dominant Color
- Faces detected in the key-frames
- Long shots detected in the key frames
- Replay Detection
- Referee's whistle time codes

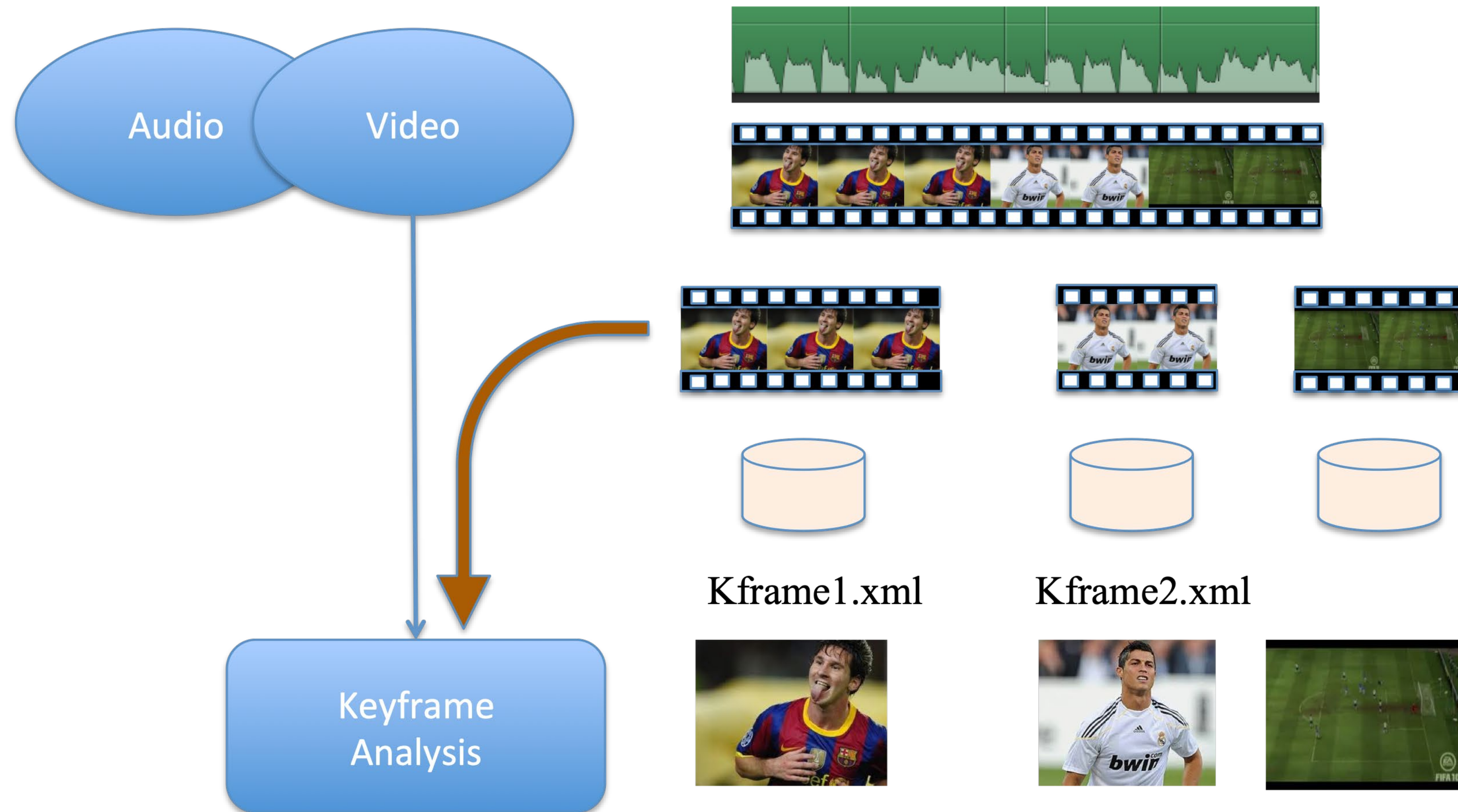
Tecnologías Clásicas

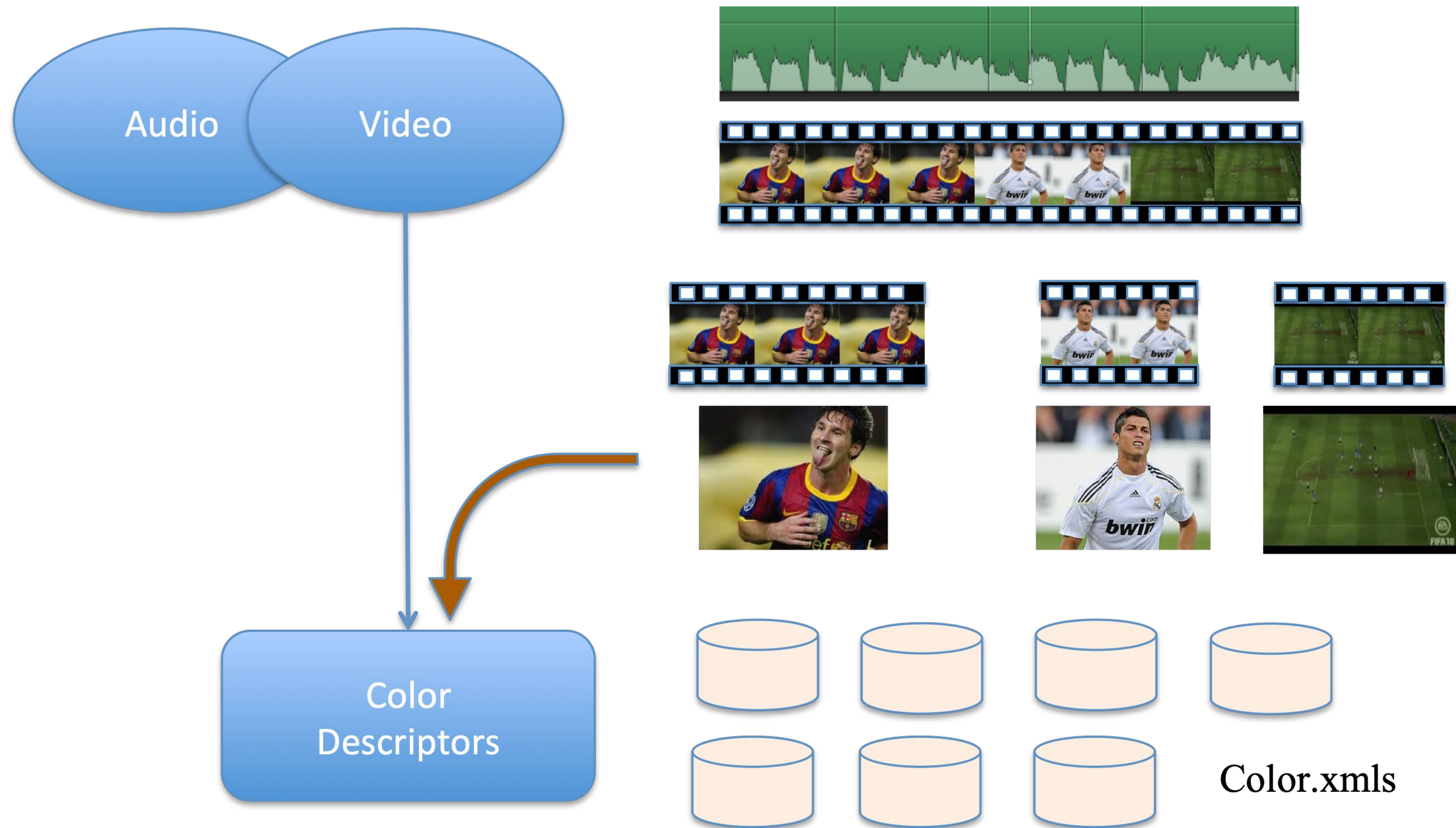


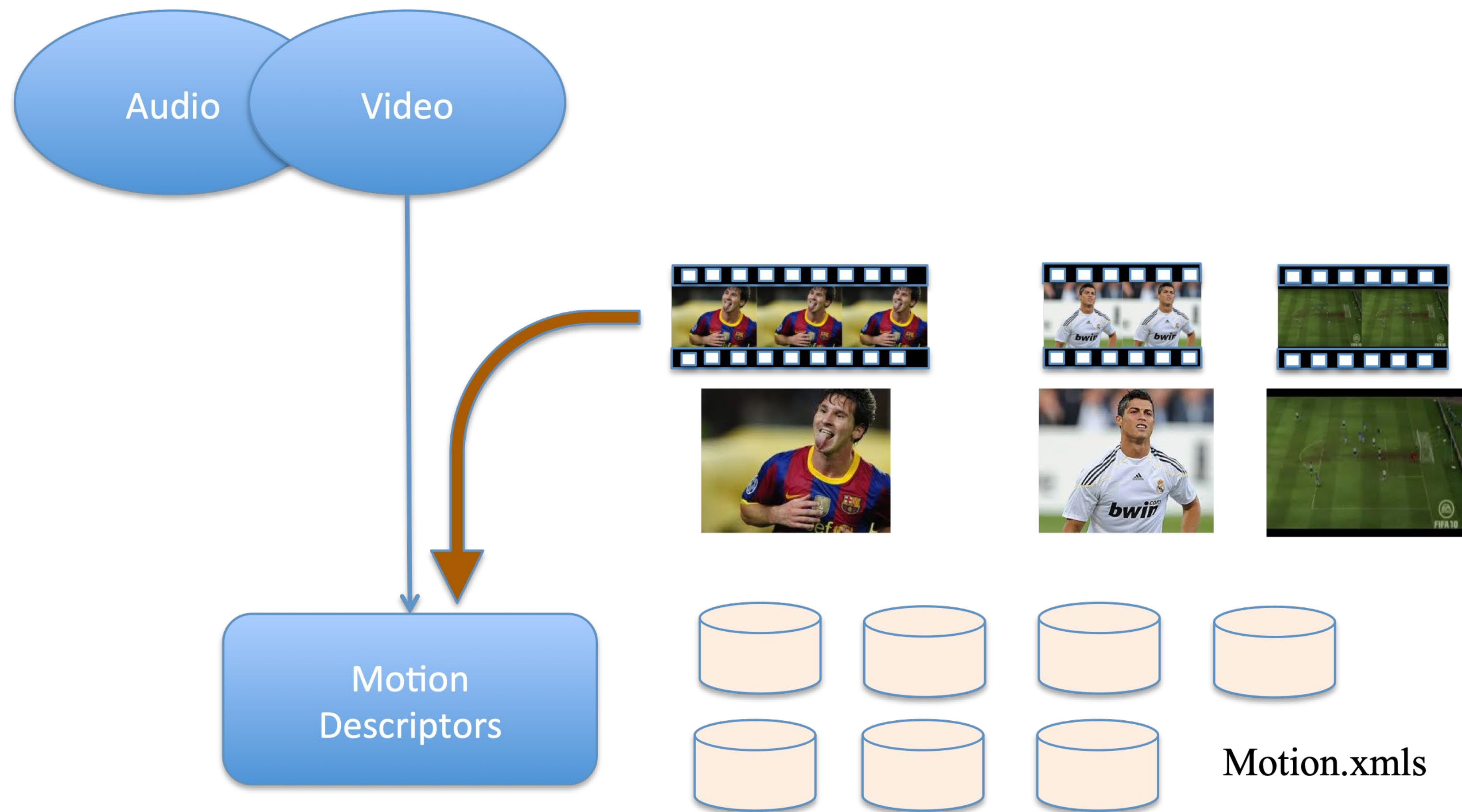
[1] A. Raventos, R. Quijada, L. Torres, F. Tarres, D. Girivet, and E. Carasusan, "The Importance of Audio Descriptors in Automatic Soccer Highlights Generation," in Proceedings of SSD'2014, 2014, pp. 1–6.

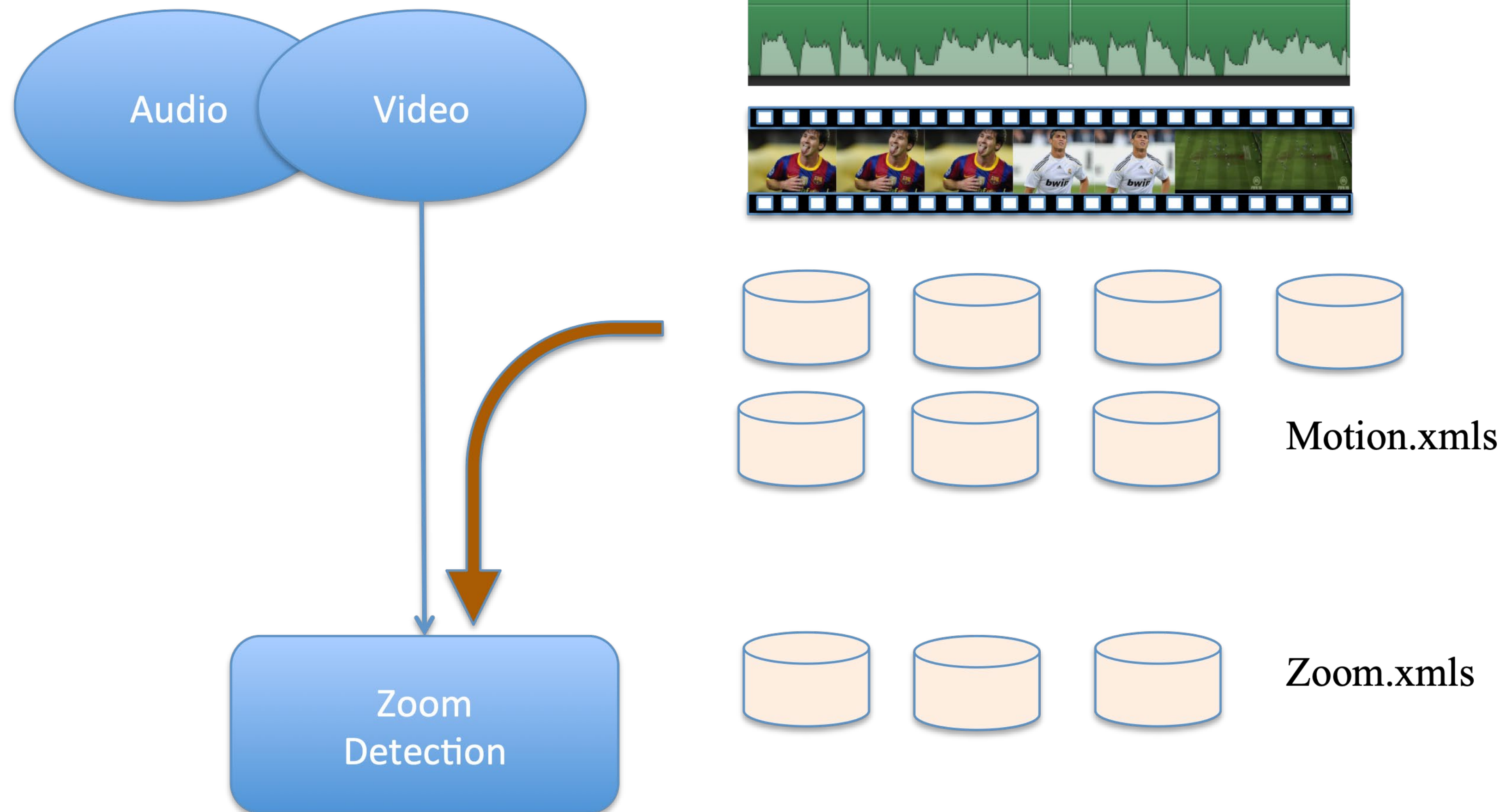
[2] A. Raventós, R. Quijada, L. Torres, and F. Tarrés, "Automatic summarization of soccer highlights using audio-visual descriptors," Springerplus, vol. 4, no. 1, 2015.

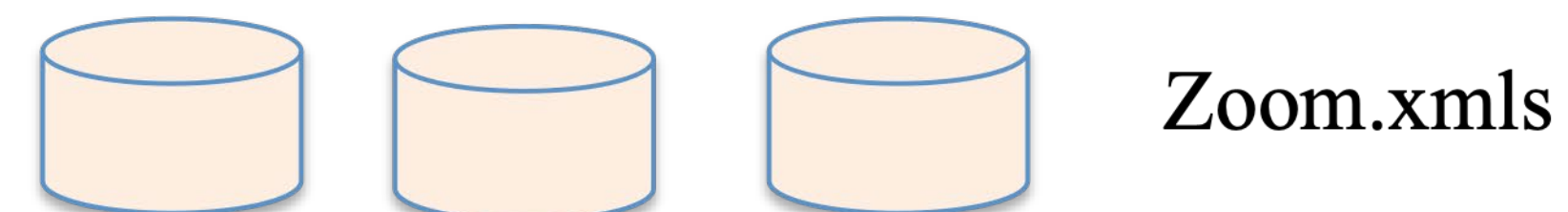
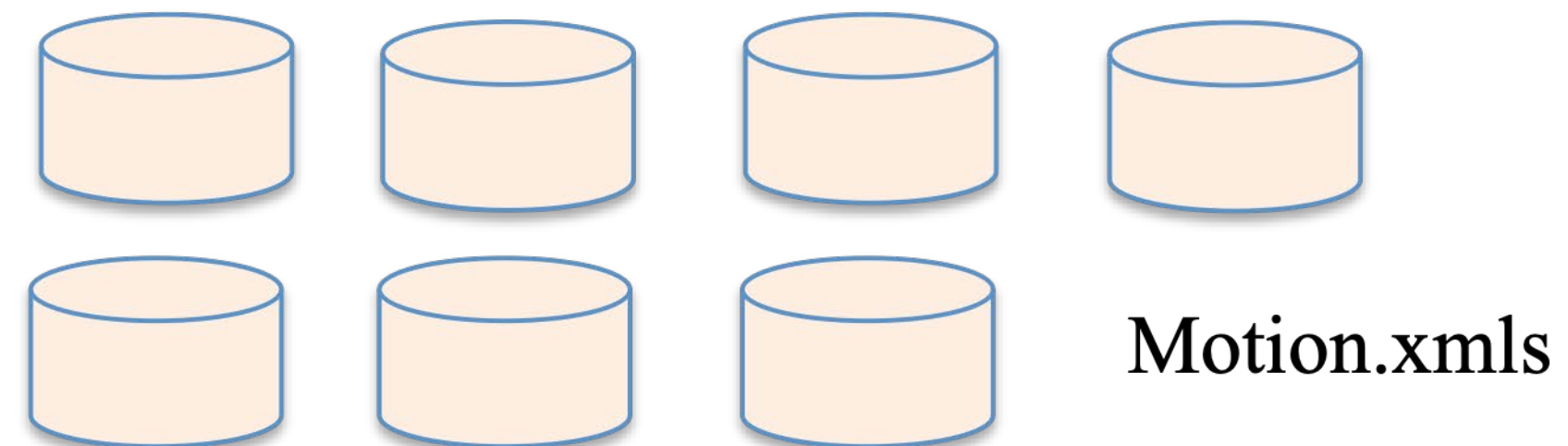
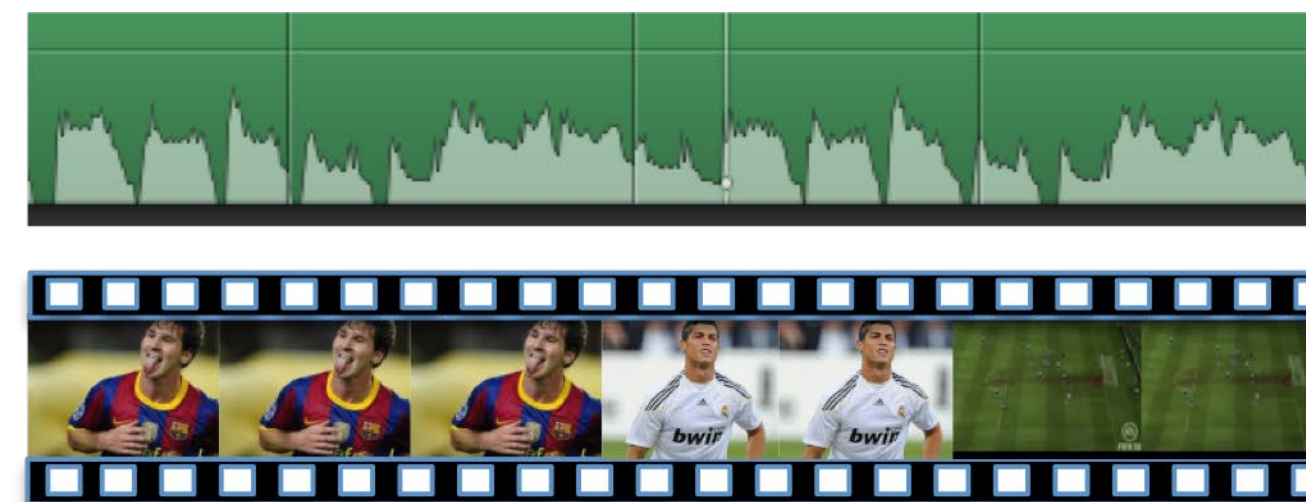
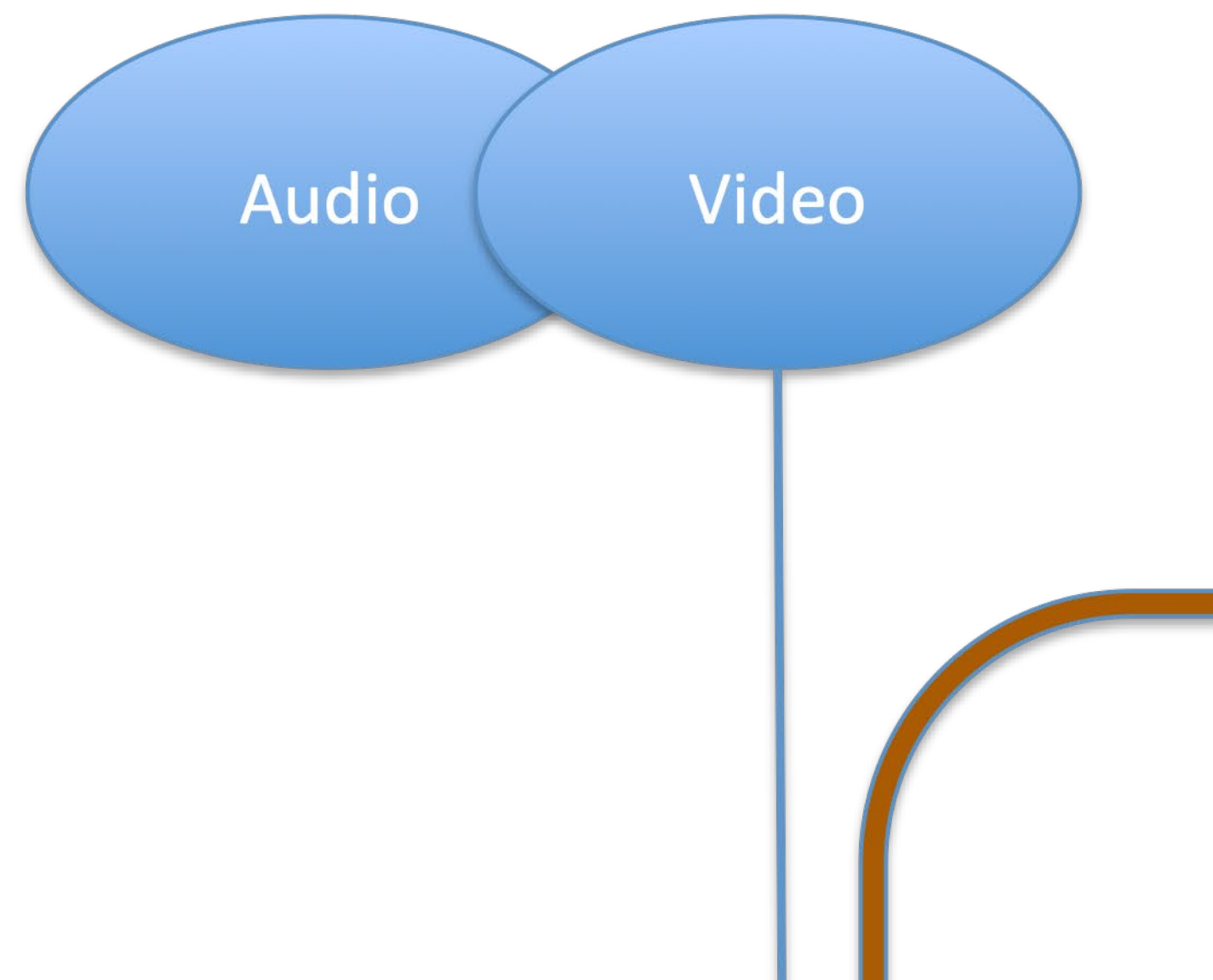


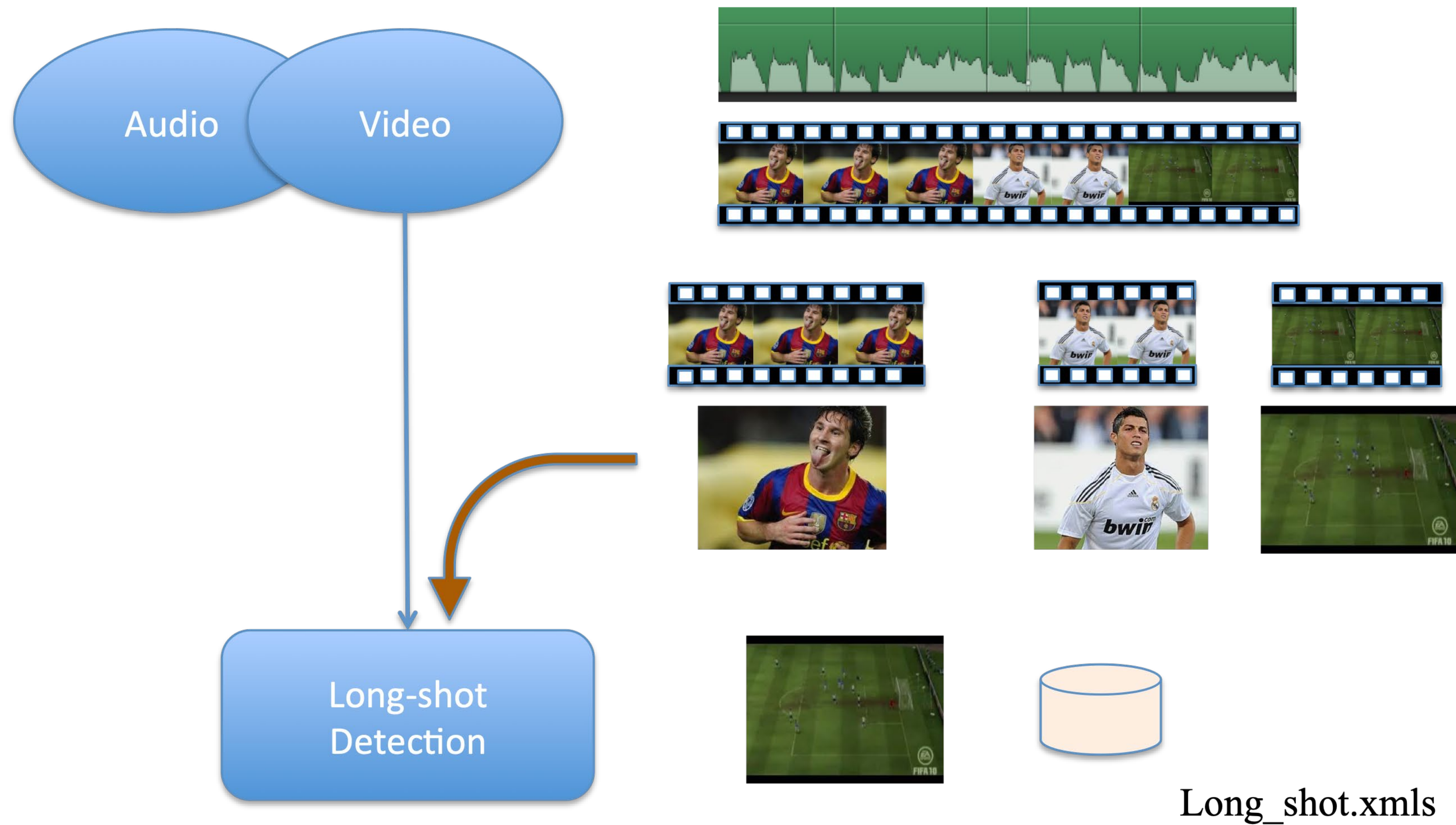


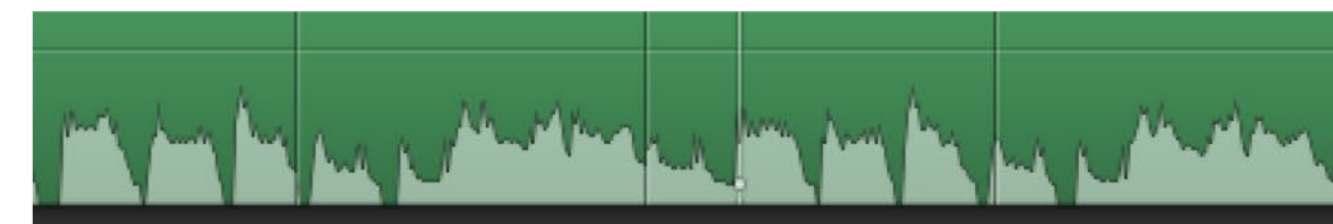
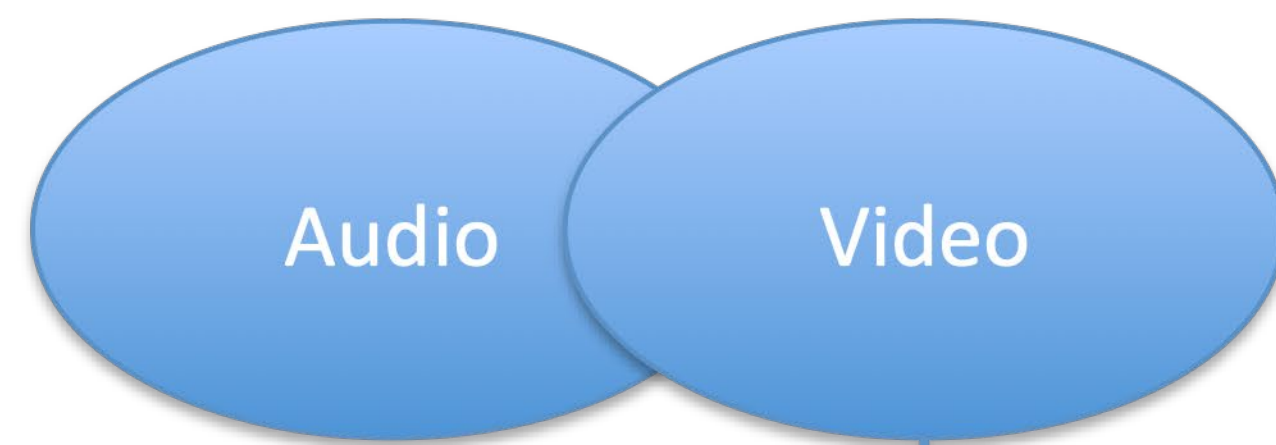






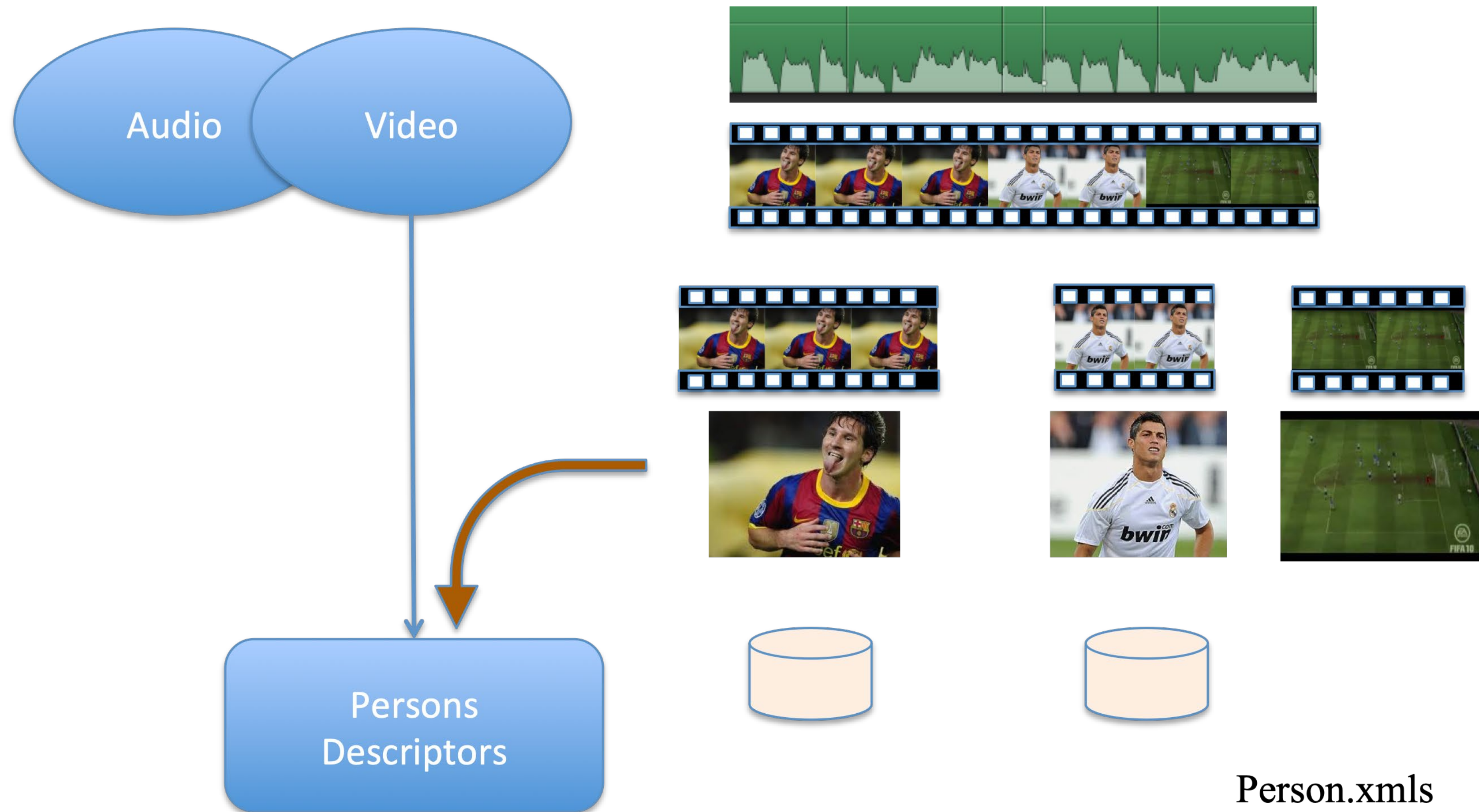


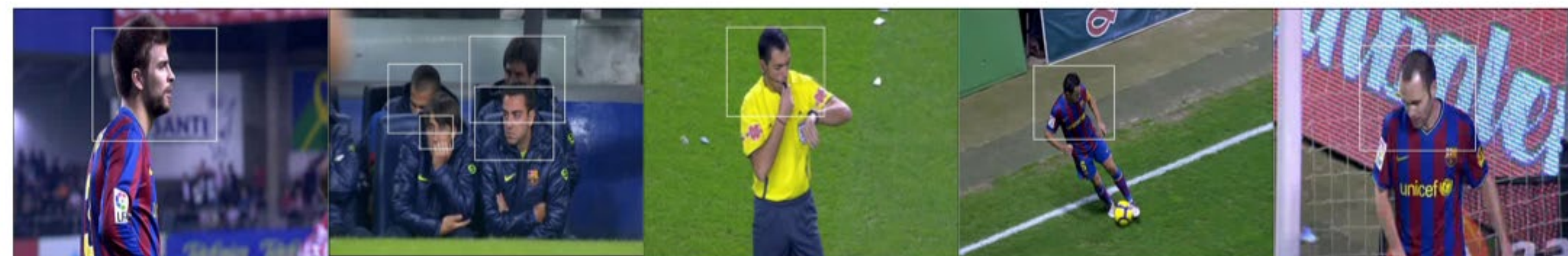
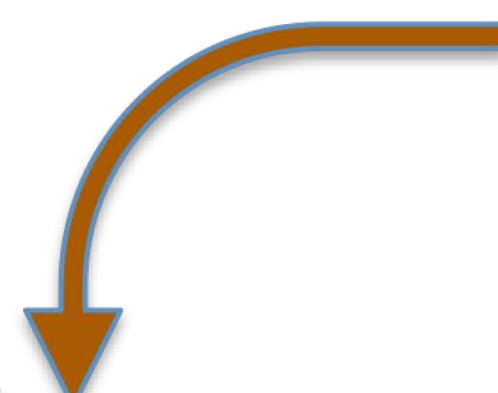
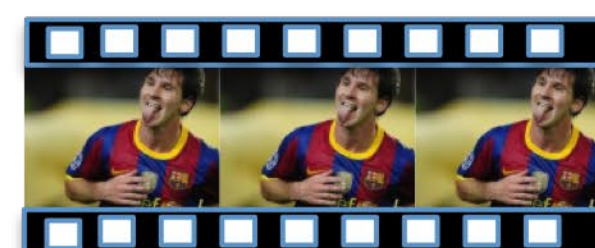
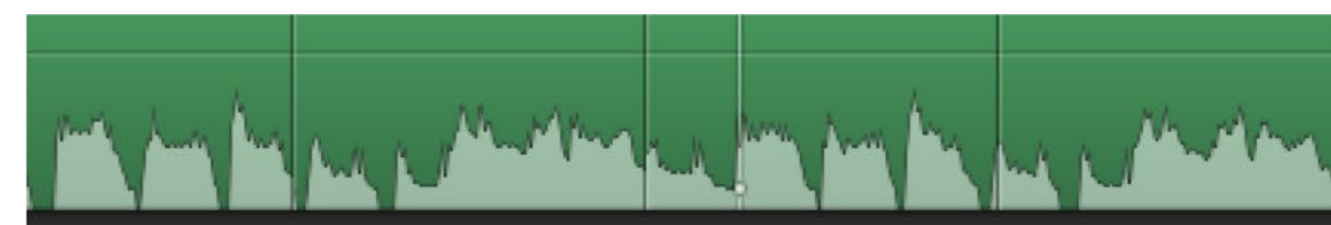
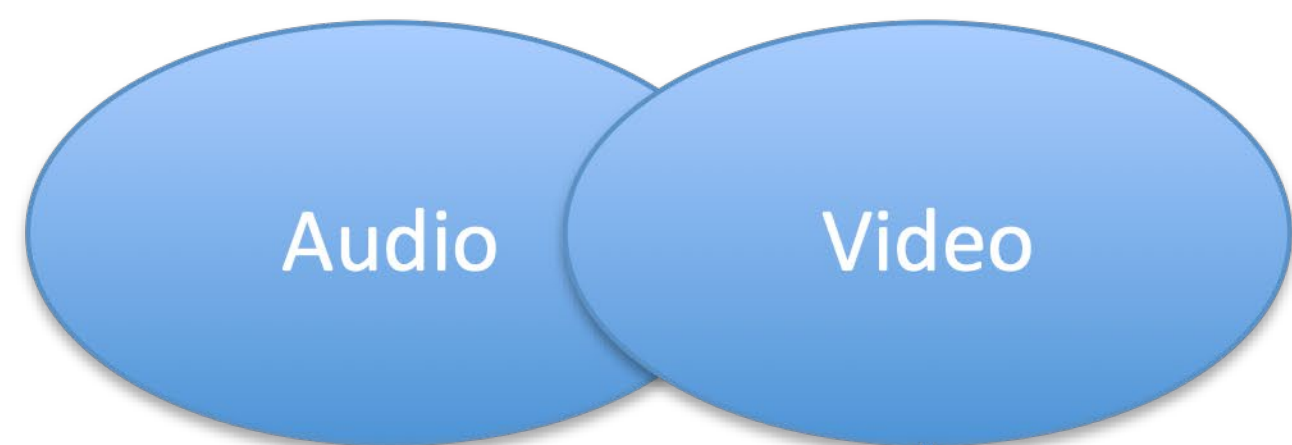


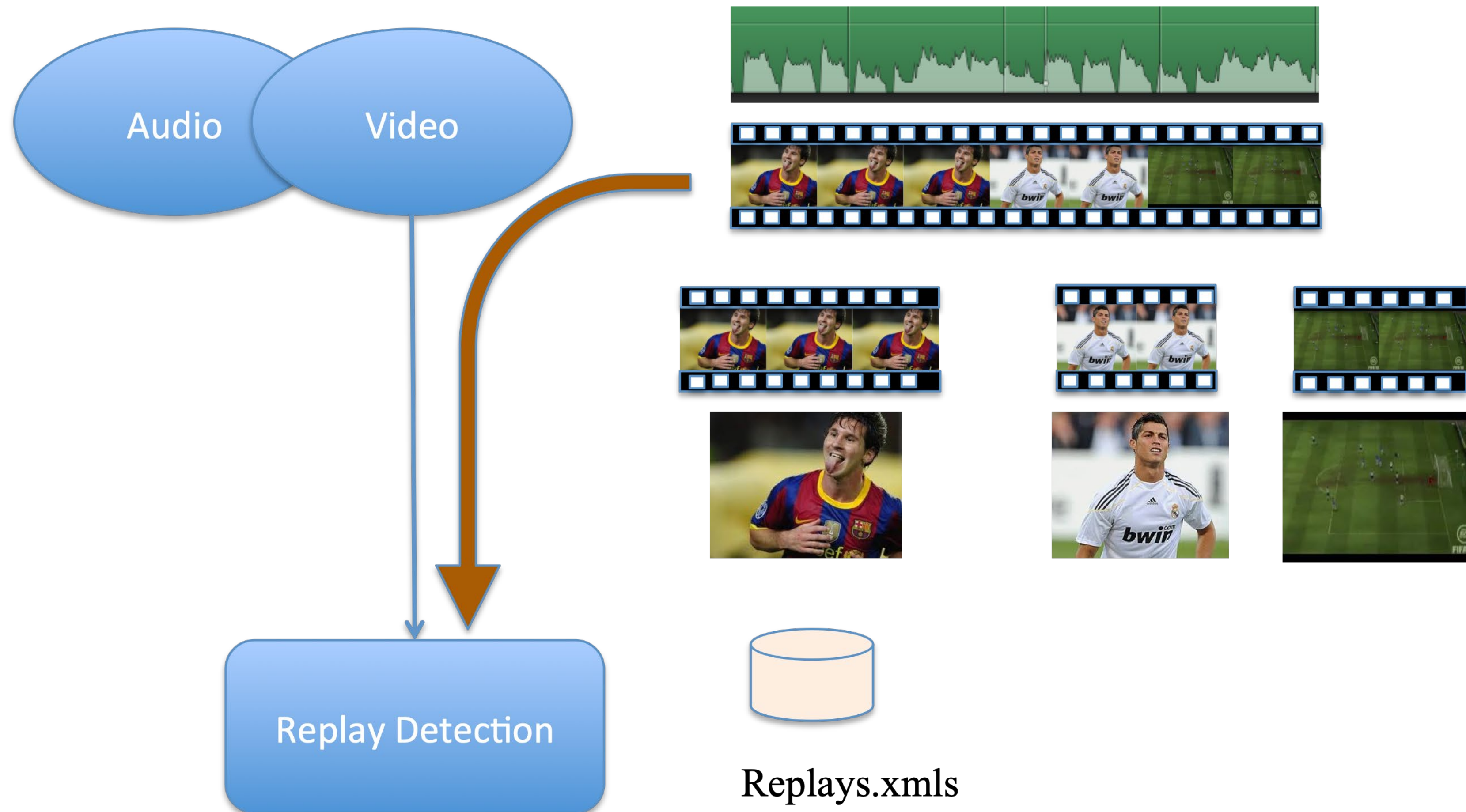


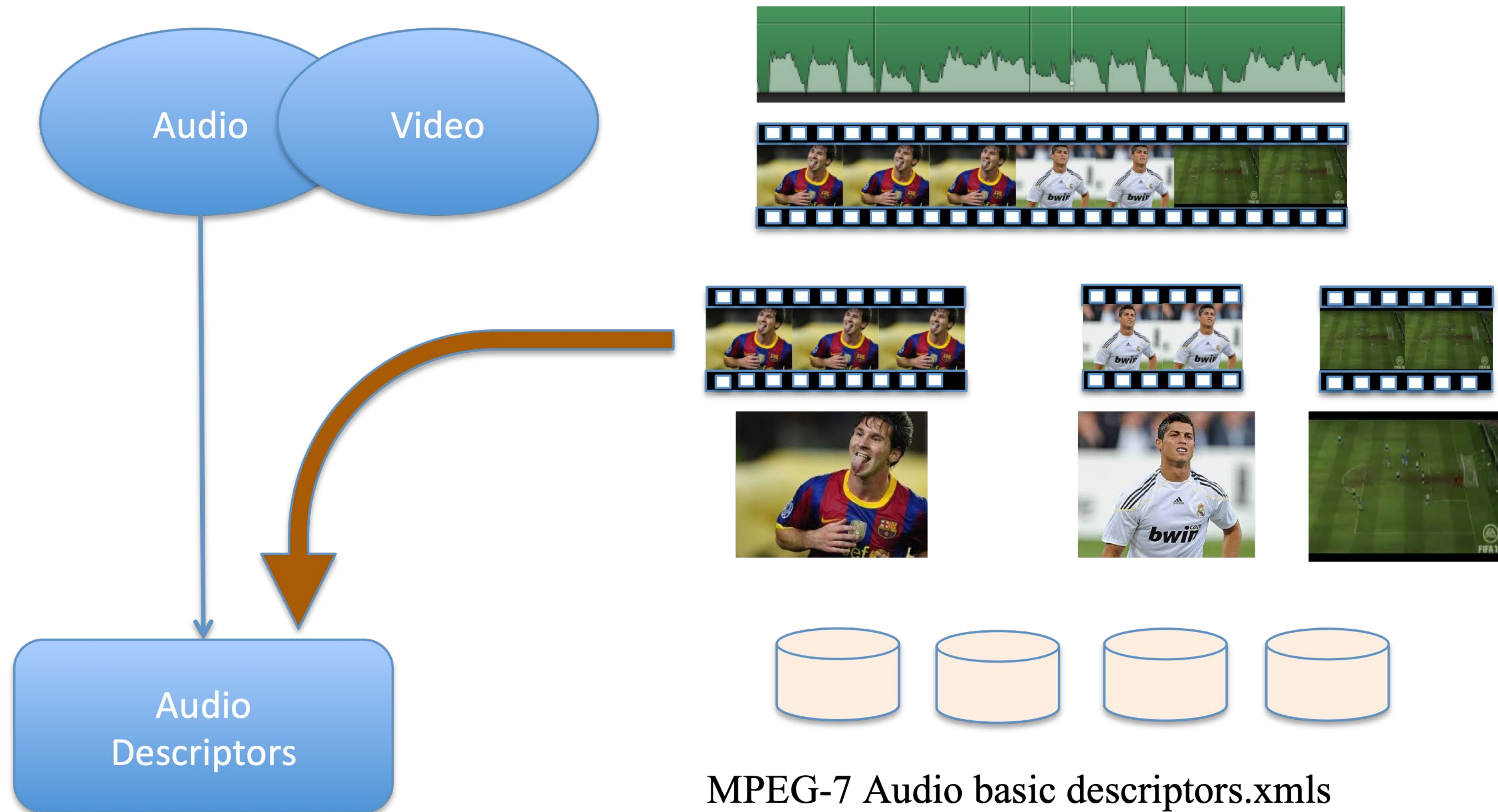
Long Shots

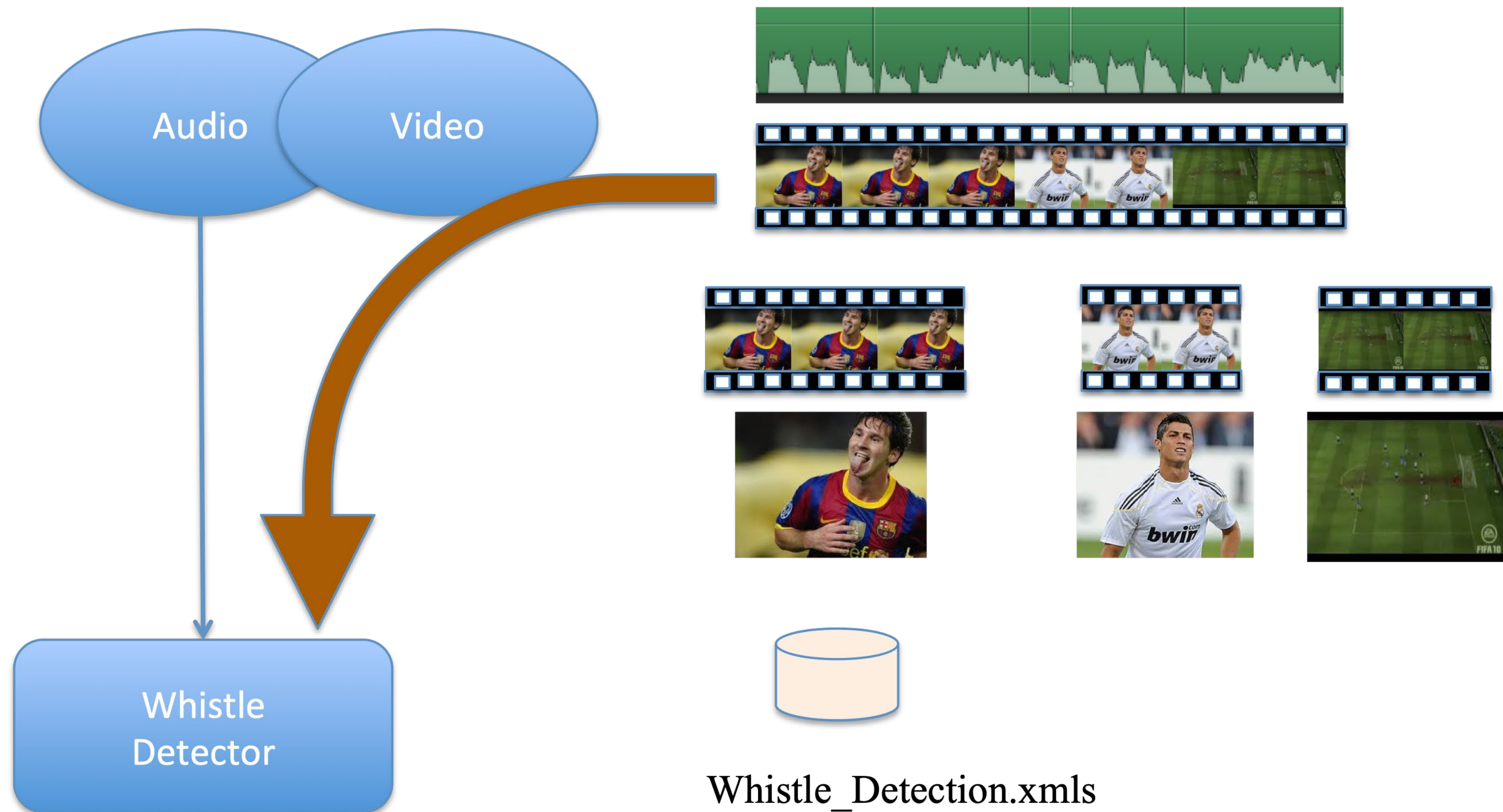
Mid Shots



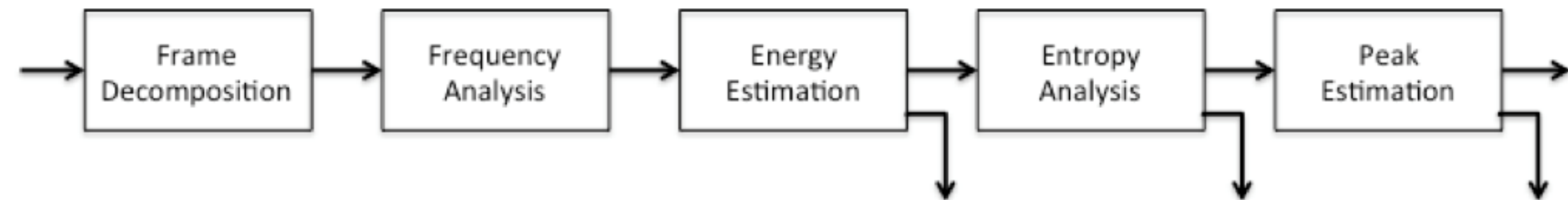




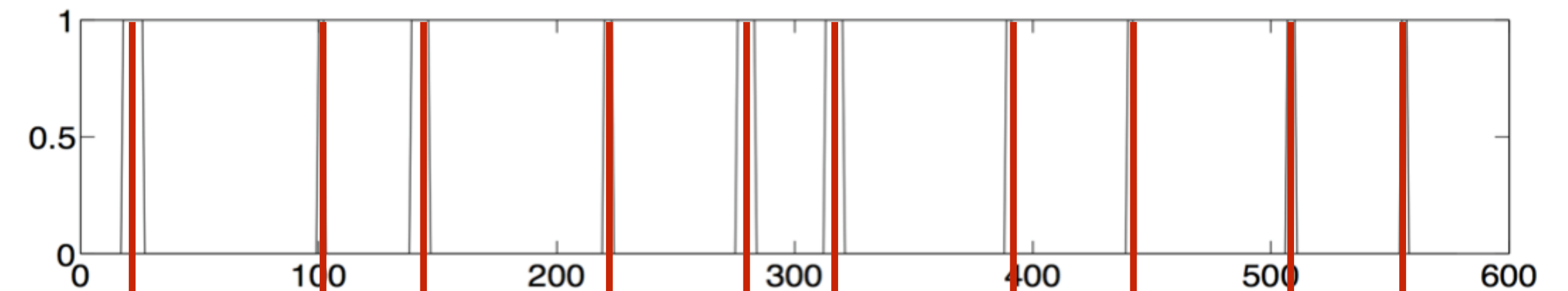




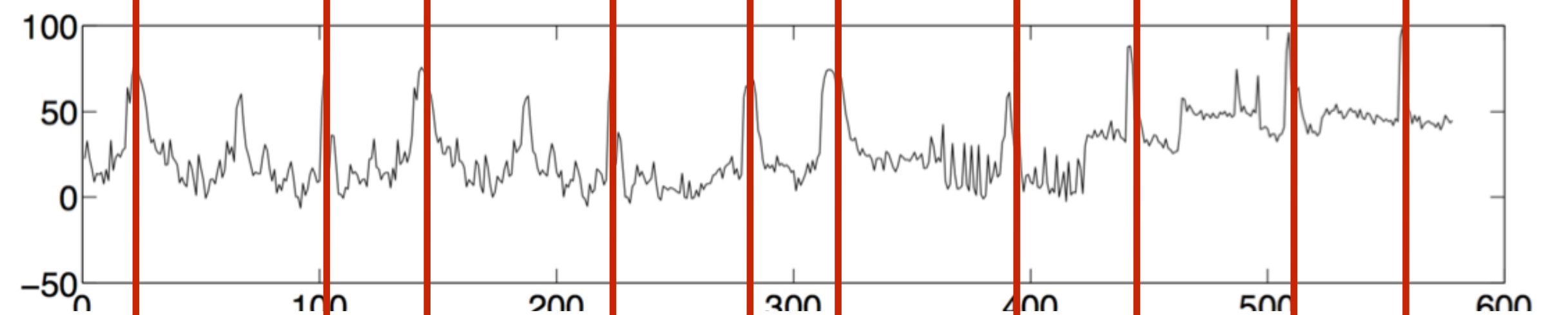
Whistle Detector



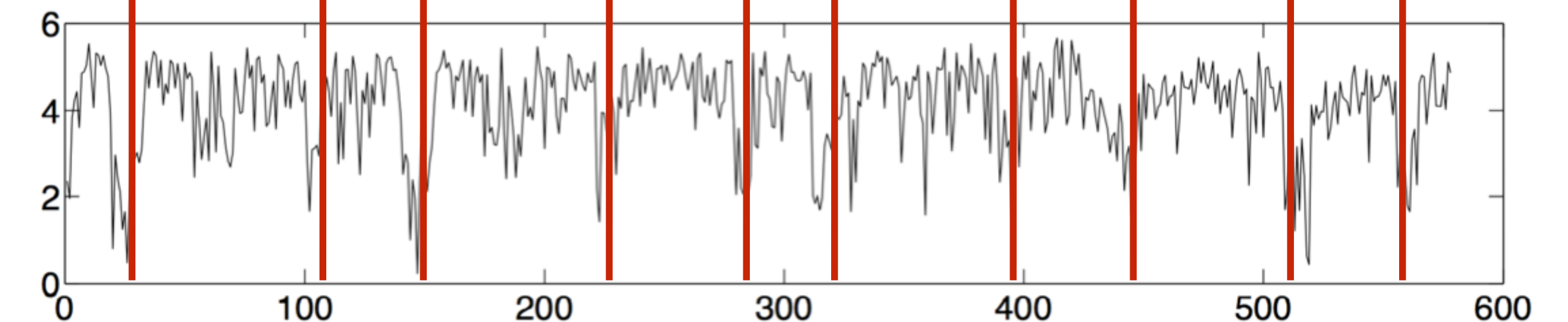
Whistle Groundtruth

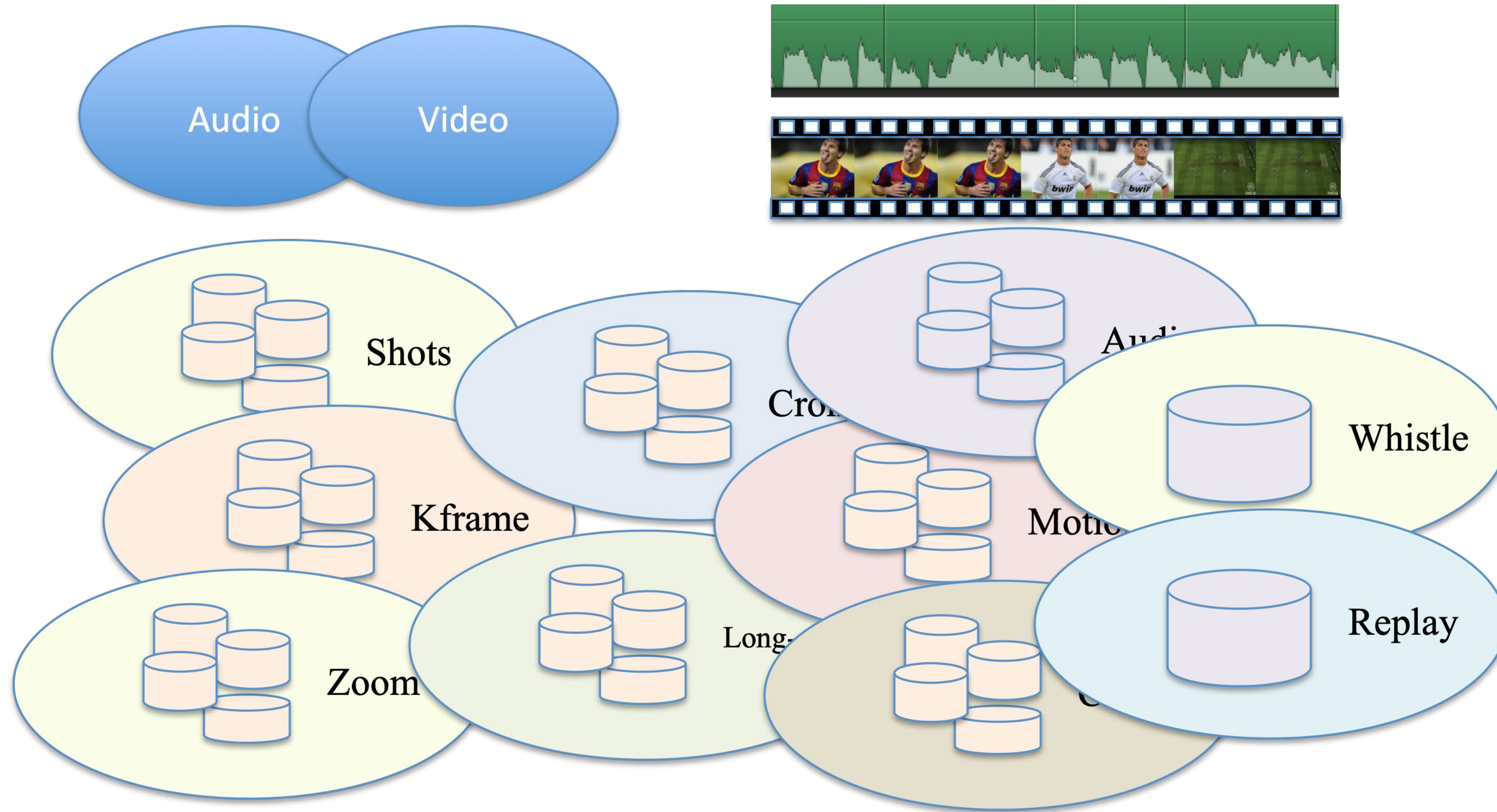


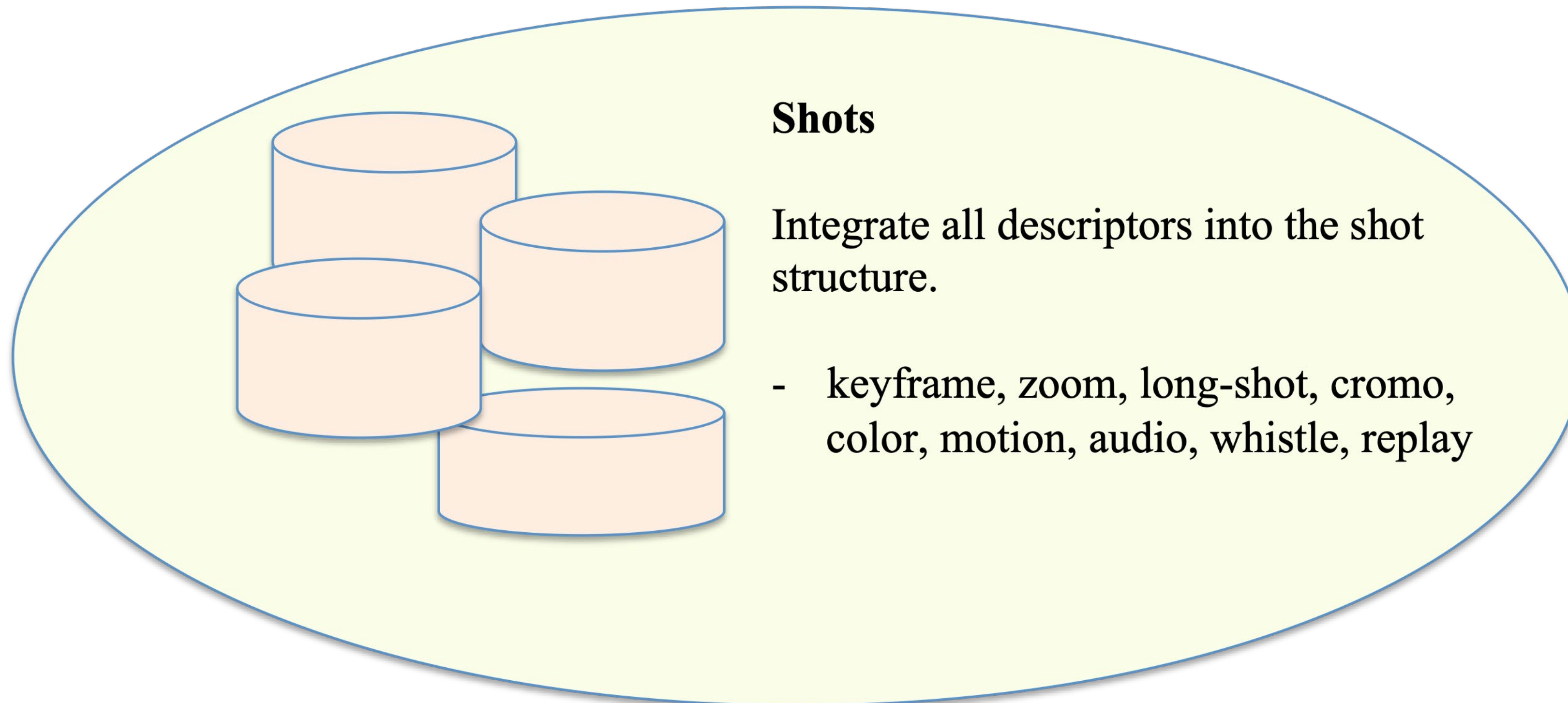
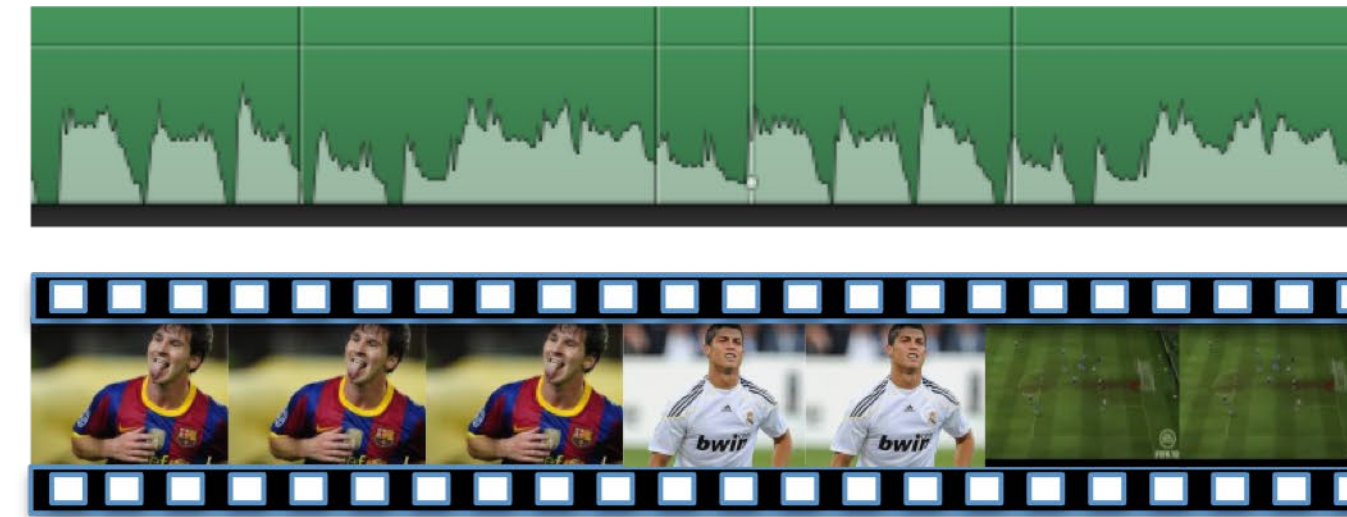
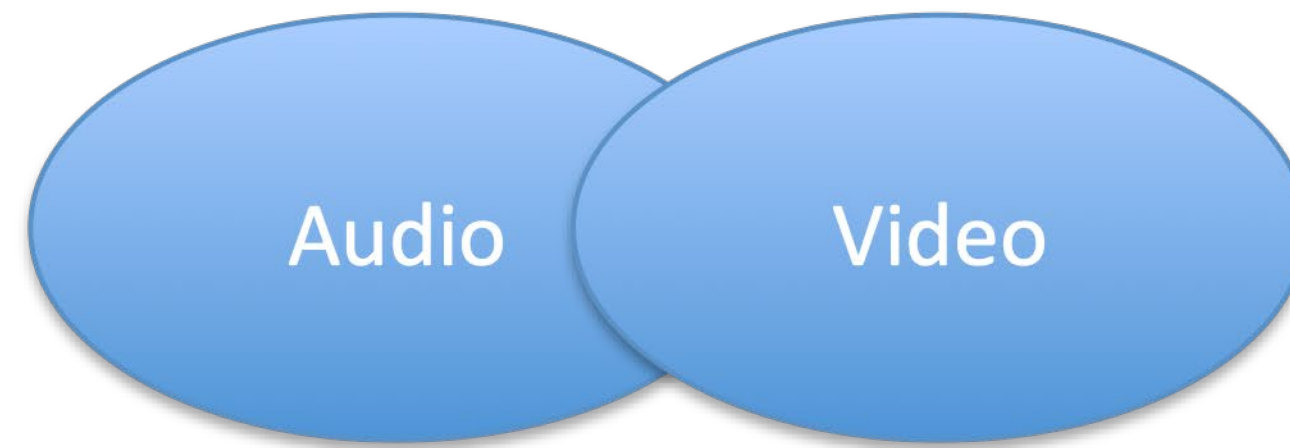
Energy in Band

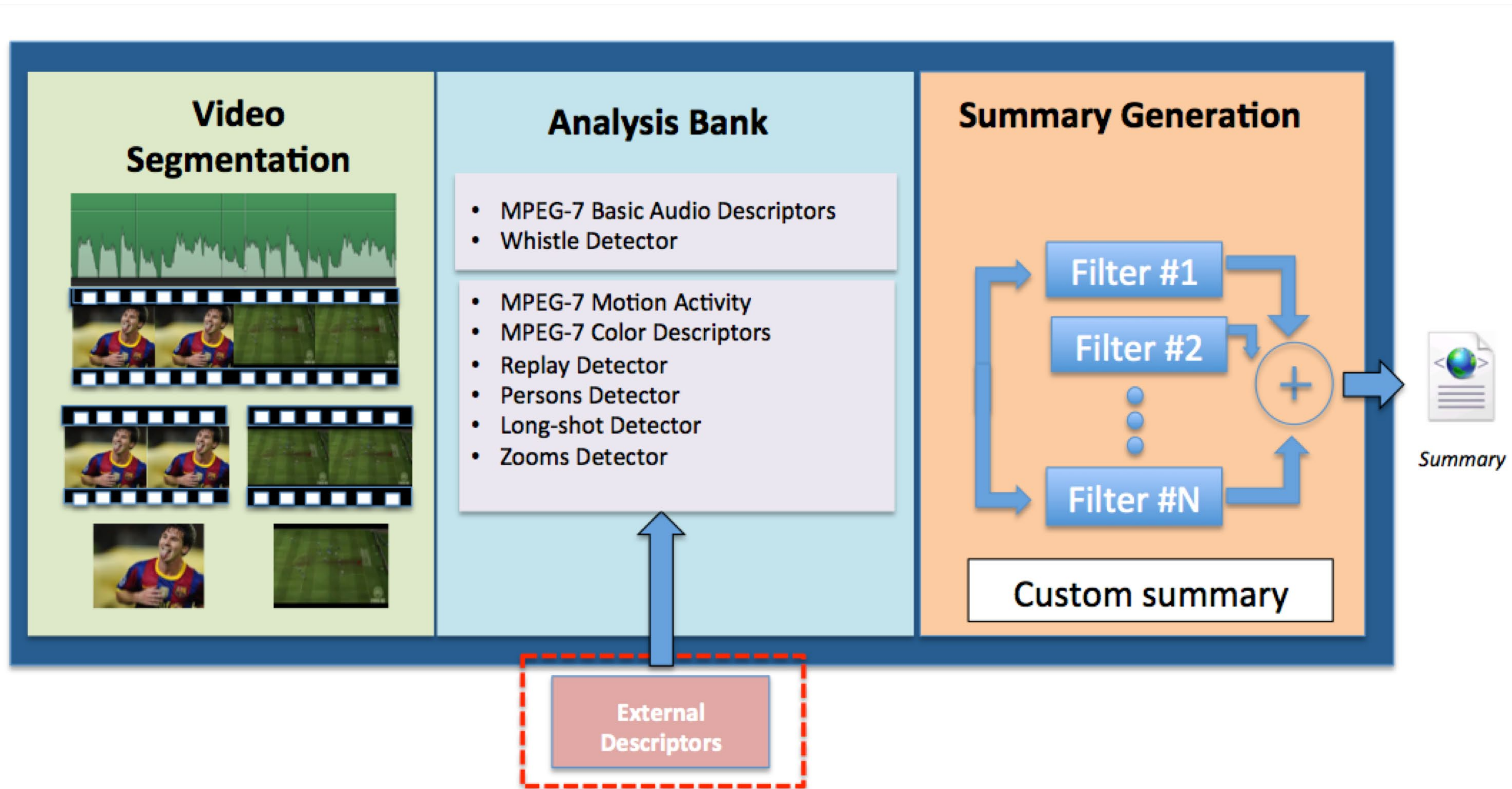


Spectral Entropy











EL MUNDO

Los momentos más destacados. No todos los partidos de fútbol reciben la misma atención mediática. Para hacer frente a aquéllos que no cubre un regimiento de periodistas, la Universitat Politècnica de Catalunya (UPC) está diseñando una tecnología que llamada a automatizar los resúmenes de los partidos incluyendo las jugadas más destacadas para que la información se obtenga de la forma más rápida posible. El sistema, desarrollado en el marco de un proyecto de investigación CENIT de la mano de Televisió de Catalunya (TVC), nace precisamente de la necesidad de la cadena autonómica de tener una herramienta que seleccionara las jugadas de mayor interés para resumir en un programa de 20 minutos partidos de tercera división o regional.

Aprovechamiento del 'big data'. El investigador principal del proyecto, Francesc Tarrés, explica que su tecnología «divide el partido en trozos y se asigna a cada uno una nota en función de su relevancia». Así, los que obtienen la puntuación más alta son los que entran en el resumen. ¿Y cómo saber lo que ha sucedido? Tarrés y su equipo han utilizado la información de bajo nivel que se extrae aplicando algoritmos matemáticos tanto a la imagen como al sonido. «Usamos el sonido ambiental, reconocemos el silbido del árbitro, los colores, detectamos caras y los movimientos de cámara, entre otros», explica Tarrés, quien ha conseguido que el sistema seleccione el 70% de goles. «Cuando marca el rival en tu campo la celebración no es la misma», argumenta, aunque ya trabajan para afinar más la tecnología.



Football robot promises to get rid of the boring bits

By Steve Holden
Newsbeat reporter

5 December 2014 | Technology

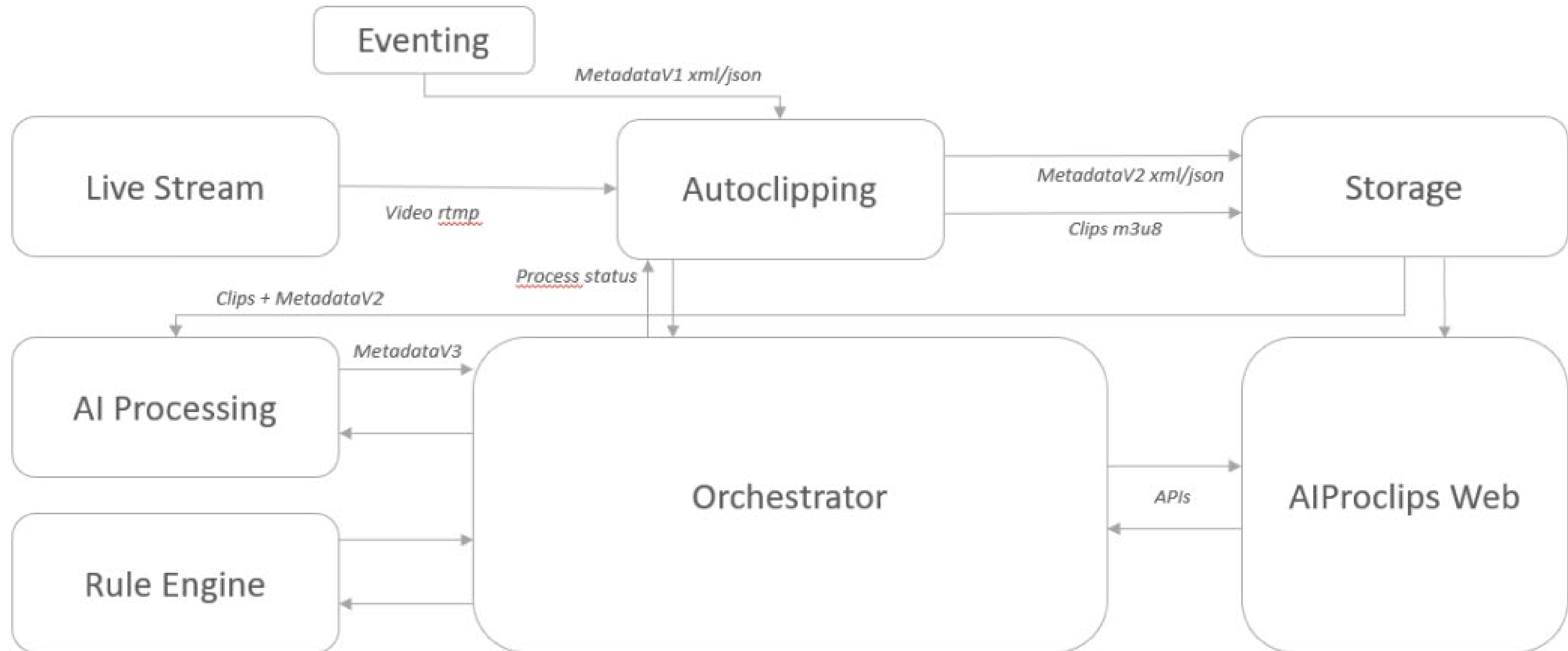


For years Match of the Day has satisfied impatient football fans by condensing a 90 minute match into a handy chunk of highlights.

But that requires a very patient man or woman to select the most important periods of play.

Now scientists in Spain are trying to make that work easier by getting a computer to do it.

The technology is being designed to automatically edit a whole game down to the key moments.

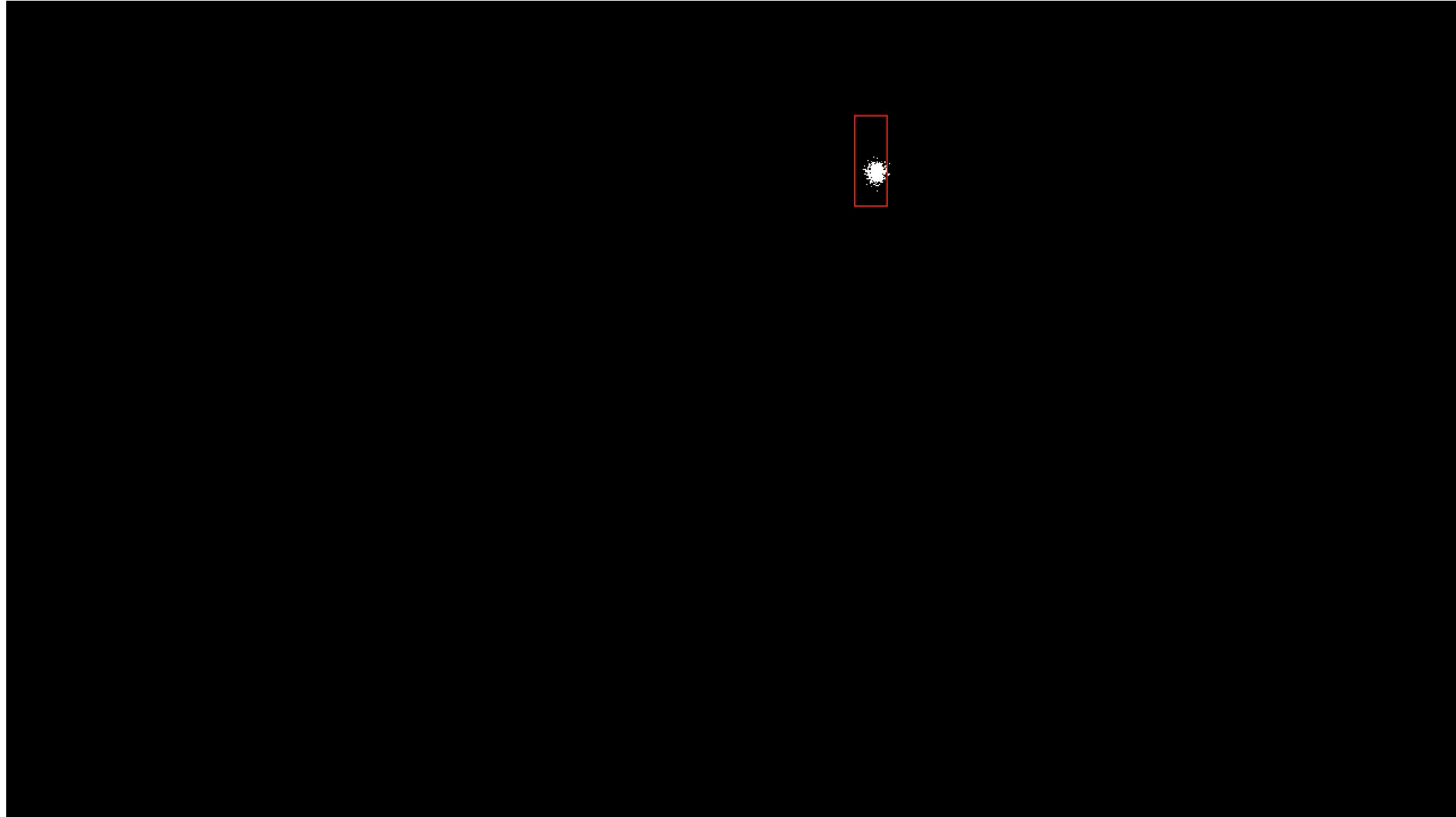


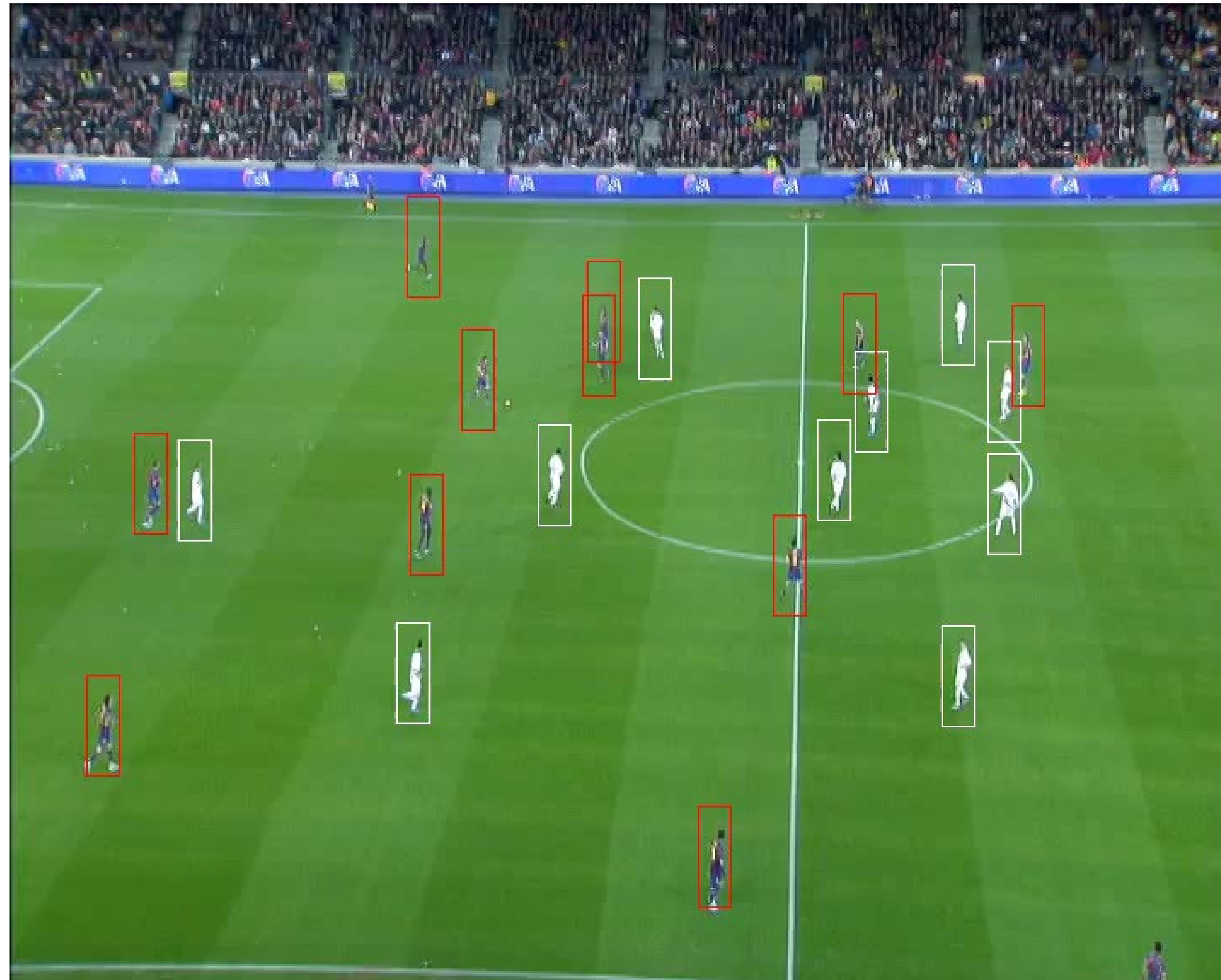
Ejemplos de reglas

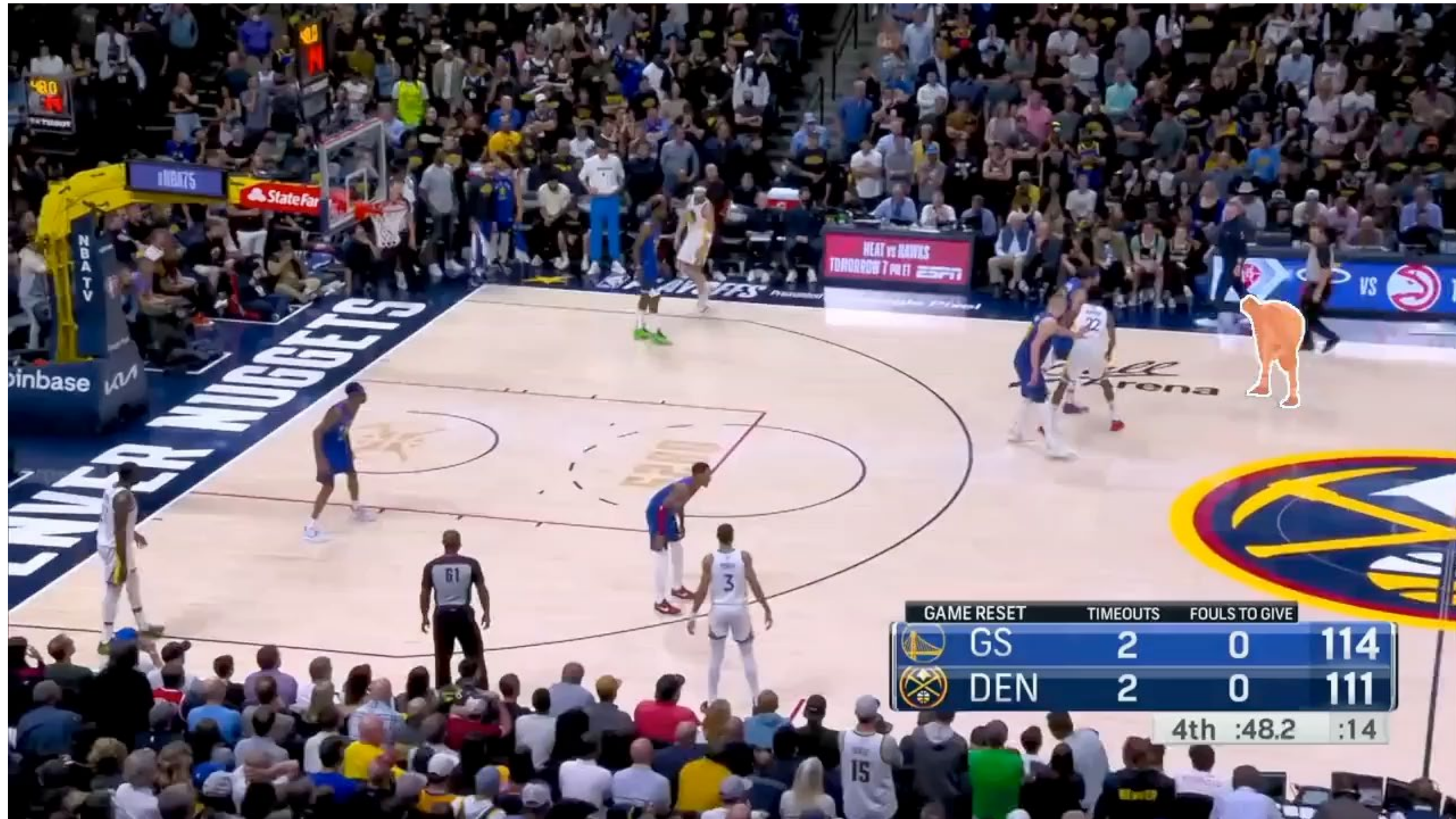
El usuario puede acceder a un web para seleccionar la tipología de resúmenes.

- Preferencia jugador
- Preferencia equipo
- Preferencia nacionalidad jugador
- Preferencia de tipo de acción del juego.

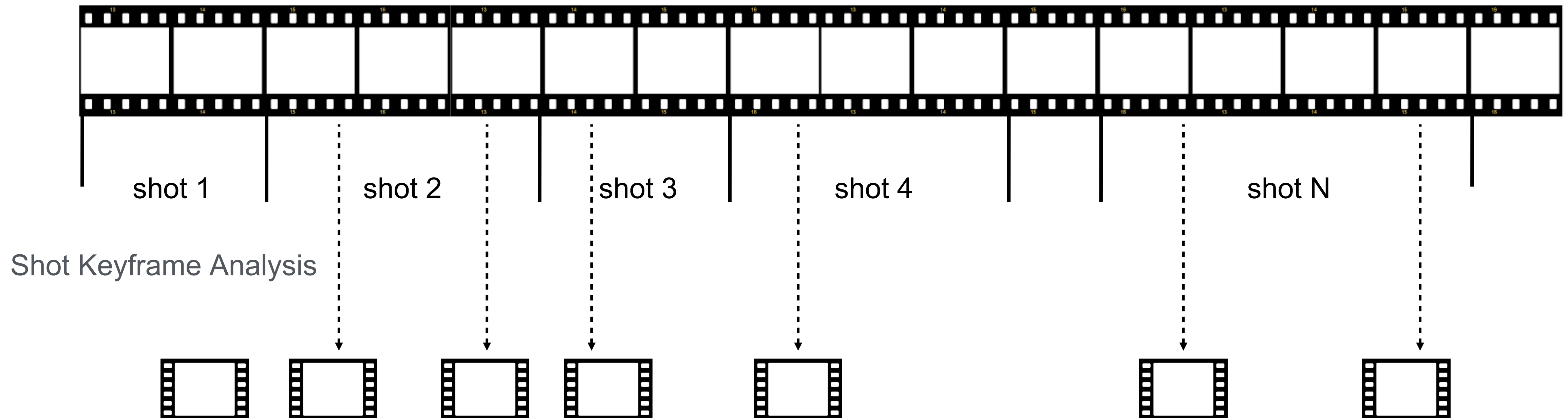
Descripción del requisito	Explicación del funcionamiento
1. Aparecerán todos los goles de todos los jugadores.	En fase de composición de playlists siempre que haya goles (es decir que no hayan sido filtrados o ponderados negativamente) el sistema asegurará que los goles se incluyen en el resumen.
2. Jugadas "VAR" con score máximo.	En fase de composición de playlists siempre que haya jugadas de este tipo (es decir que no hayan sido filtradas o ponderadas negativamente) el sistema asegurará que éstas se incluyen en el resumen. Se considerará score máximo un score por encima de 0,90 (valor configurable).
3. Jugadas "Saves" con score máximo.	
4. Acciones de Jugadores TOP con score máximo.	
5. Agrupaciones en este orden: 5.1. Regates (según preferencia) 5.2. Oportunidades (según preferencia) 5.3. Tarjetas (según preferencia) 5.4. Faltas (según preferencia) 5.5. Intercepciones (según preferencia) 5.6. Regates (score máximo) 5.7. Oportunidades (score máximo) excepto "Saves"	Para este requerimiento se agruparán los clips en 3 grupos según su score ponderado: Score Alto, Score Medio y Score Bajo. El score ponderado es el resultado de aplicar: - [Score de la jugada]*[Ponderación deseada para cada evento] - [Score de la jugada] cuando no hay ponderación. Es decir: $SF = SI*(1-p)+SI*P*p$



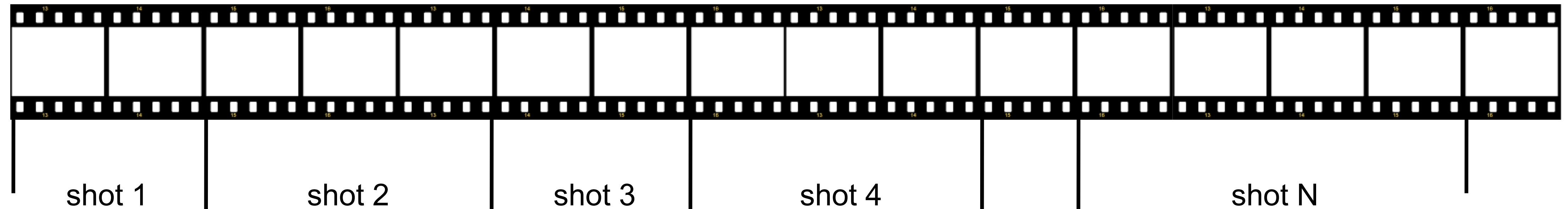




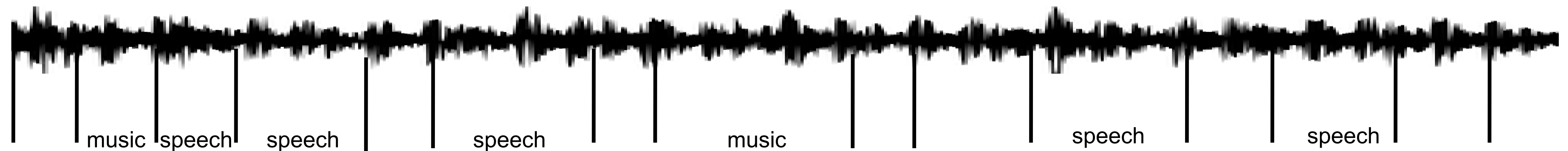
Shot Analysis



Best shot keyframes are automatically selected



Speech detector, music detector, speaker diarization



Speech-to-Text

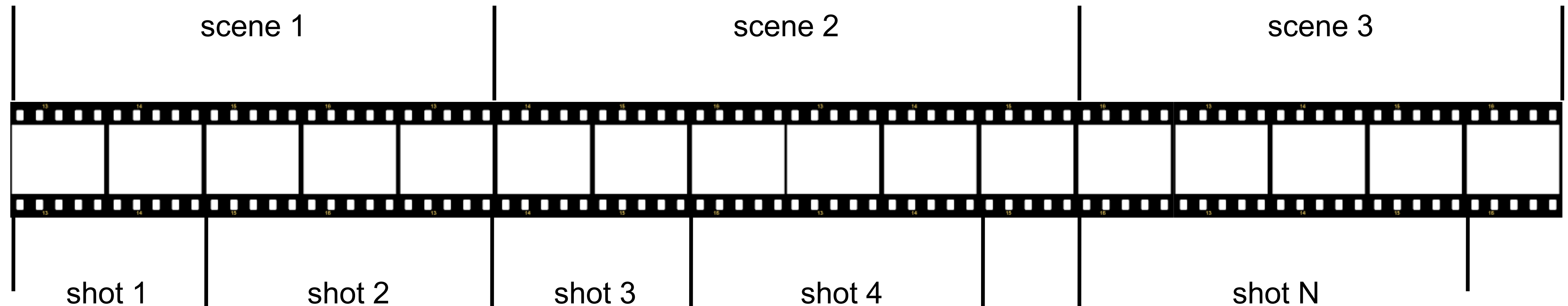
We need to go deeper!

You're gonna need a bigger boat

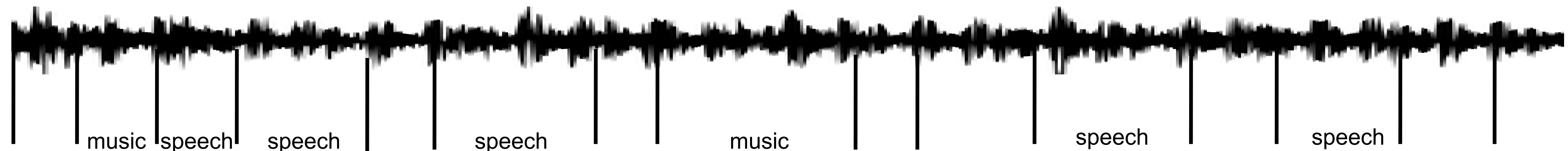
Luke, I am your father

We will always have Paris

Scene Segmentation



Speech detector, music detector, speaker diarization



Speech-to-Text

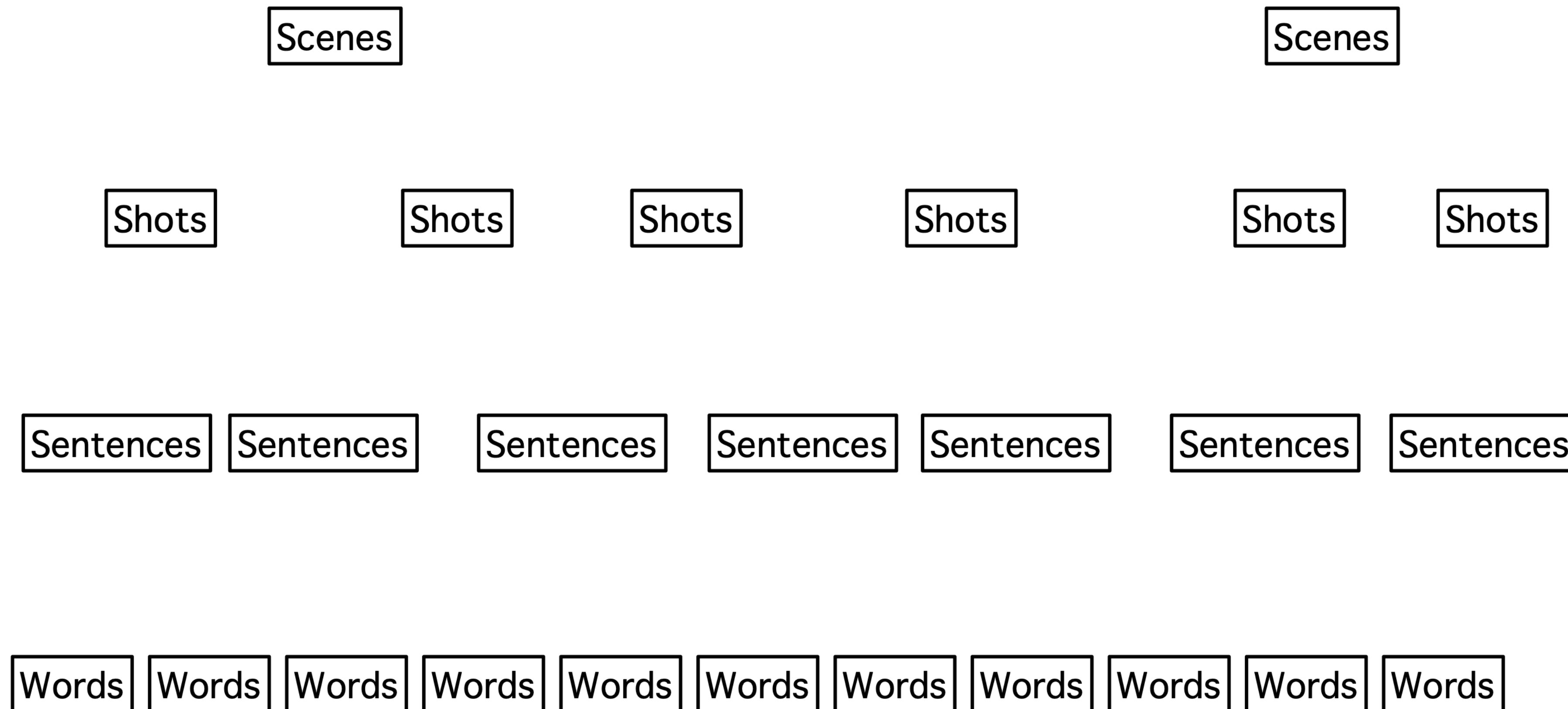
We need to go deeper!

You're gonna need a bigger boat

Luke, I am your father

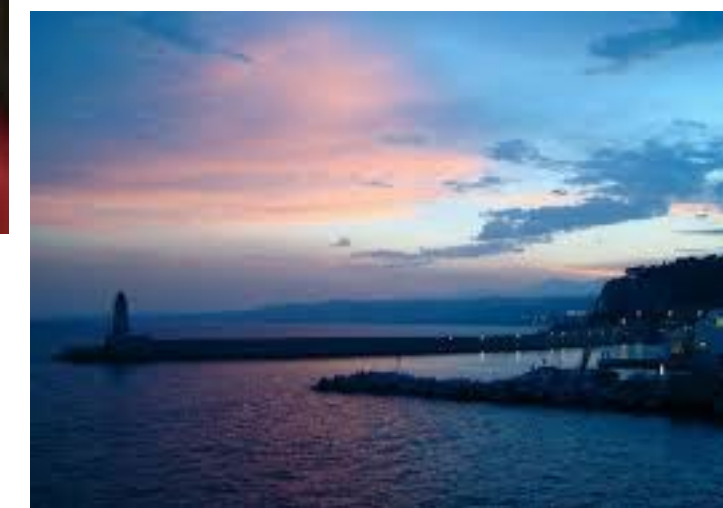
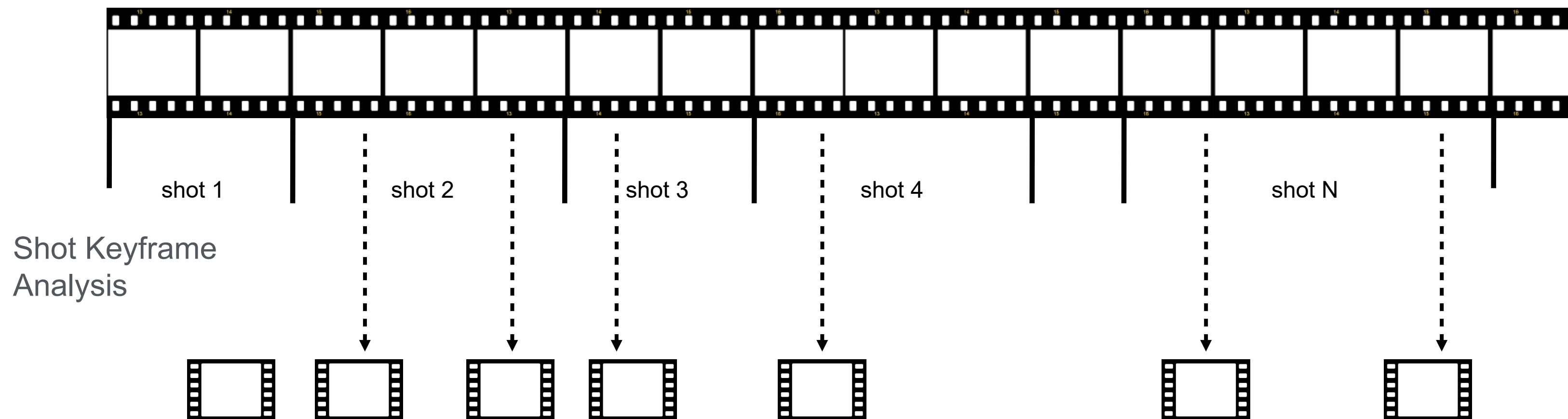
We will allways have Paris

Hierarchical Segmentation



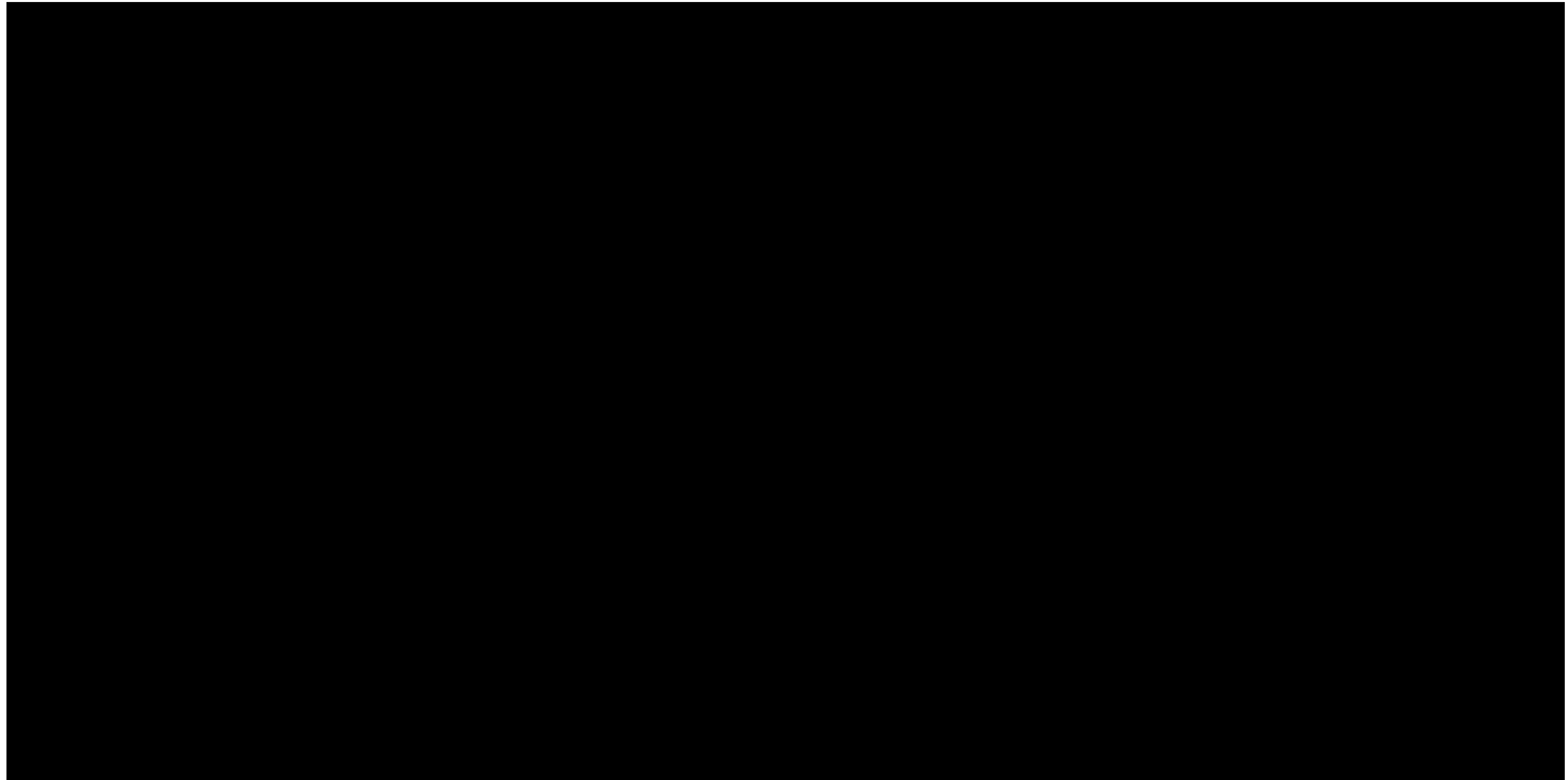
High Level Metadata Extraction by means of Deep Learning

Shot Analysis



For every shot we extract:

- Places (beach, hospital, bar, etc.)
- Objects (person, dog, cat, bottle, etc.)
- Faces (Actors, Politics, Sportswomen)
- Aesthetics



Tecnologías Basadas en Análisis Multimodal

Ejemplos: Minutador

The screenshot displays the GECA CONTENT interface. At the top, the breadcrumb path is 'LA1 → Telediario_1_Fin_de_semana → TVE1_20220116_145905'. The main video player shows a news anchor, Nico de Vicente, with a green bounding box around his face. The video progress is at 20:44 / 1:00:55. To the right, a table lists video segments with columns for Hora, Sección, Eje Temático, and Tema. Below the video player, there are several analysis panels: PERSONAS (listing Nico de Vicente), OCR (no data), KEYWORDS (listing Patrones, Entidades, and Sustantivos), SPEECH-TO-TEXT (transcribing a segment about socialists in Castilla and León), RESUMEN MINUTO (providing a one-minute summary), and TIPOS DE ESCENA (identifying the scene as a Plató). At the bottom, there are playback controls, a current video time of 15:19:49, and buttons for INICIO, FINAL, FINALIZADO, and Exportar.

HORA	SECCIÓN	EJE TEMÁTICO	TEMA
15:17		Centro cultural San Pablo (suspensión maniqués vestidos)	
15:18		Pandemia coronavirus (Repaso situación global pandemia)	
15:19	Nacional	Conex. Dto. León (Asistencia líder PP Pablo Casado)	
15:20		Conex. Dto. Palacio Moncloa (Precampaña electoral socialistas)	
15:21		Conex. Dto. Ampudia, Palencia (Presentación España asdoiuiuasdoiusdf)	
15:22		Elecciones Castilla y León (Cifras candidaturas diferentes)	
		Ganaderos Galicia de carne Suprema	

PERSONAS
Nico de Vicente

OCR
No se han encontrado captions...

KEYWORDS
Patrones
No hay datos...
Entidades
Castilla, León, Ponferrada, León Luis Tudanca, Presidencia, Félix Bolaños
Sustantivos
socialistas, cita, candidato, ministro

SPEECH-TO-TEXT
Los socialistas de Castilla y León se han dado cita en Ponferrada en León Luis Tudanca Su candidato a las autonómicas ha estado acompañado del ministro de la Presidencia, Félix Bolaños

RESUMEN MINUTO
Conex. Dto. Palacio Moncloa (Precampaña electoral socialistas)
Conex. Dto. Palacio Moncloa (Acto PSOE y C's)

TIPOS DE ESCENA
Plató

HORA ACTUAL VÍDEO: 15:19:49
Botones: INICIO, FINAL, FINALIZADO, Exportar

Tecnologías Basadas en Análisis Multimodal

Ejemplos: Minutador

Objetivos:

1. Extraer para cada minuto de tiempo real la categoría y un resumen del contenido.
2. Utilizar estos minutos para correlar las categorías/resúmenes con las audiencias.

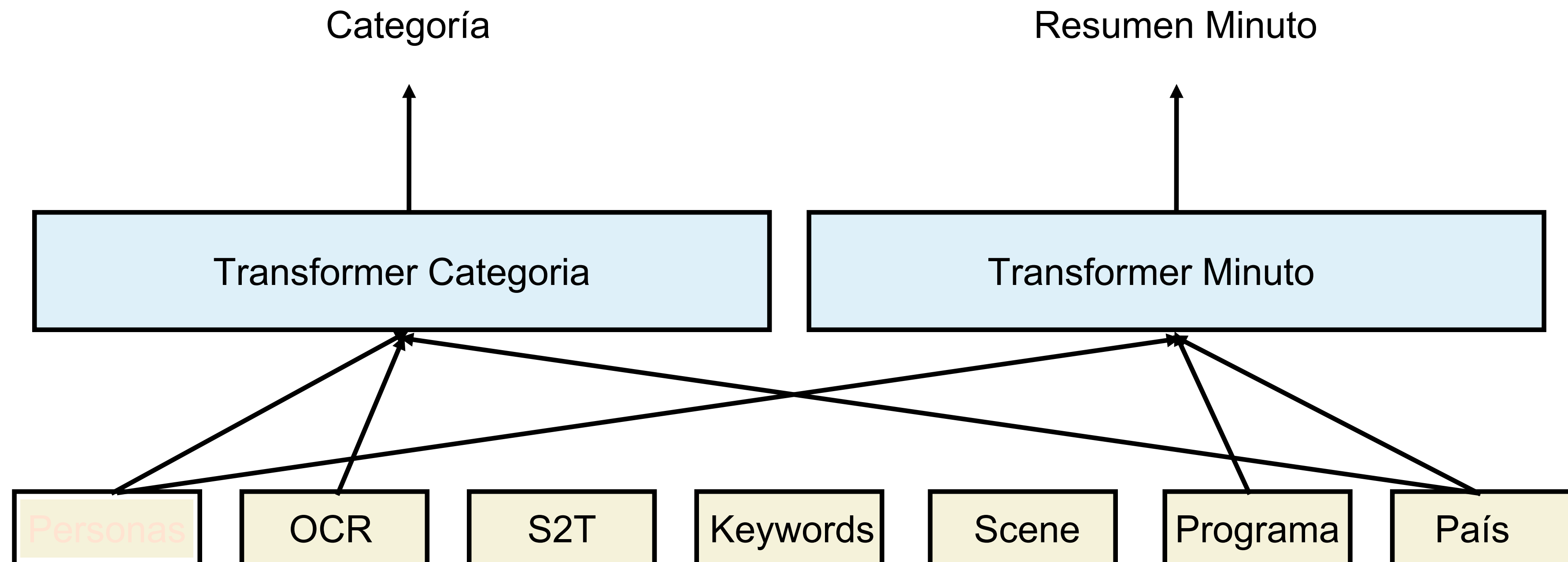
Procesado automático incluye:

1. Detección/Reconocimiento de personas.
2. Reconocimiento de caracteres en pantalla.
3. Transcripción del contenido.
4. Extracción de palabras clave a partir de la transcripción (ontologías).
5. Extracción de palabras clave a partir de la transcripción (transformers)
6. Extracción de la tipología de escenario (plató/interior/periodista/publicidad)

Tecnologías Basadas en Análisis Multimodal

Ejemplos: Minutador

Extracción multimodal de categorías y resúmenes:



Tecnologías Basadas en Análisis Multimodal

Ejemplos: Minutador

Tecnologías Basadas en Análisis Textual

El resumen se descompone en diferentes etapas:

1. Transcripción textual del contenido del vídeo original.
2. Resumen (sumario) de la transcripción textual del vídeo.
 - a. Selección de las mejores frases
 - b. Selección/Descomposición en capítulos

Ejemplos:

Conferencias

Presentaciones académicas

Tecnologías Basadas en Análisis Textual

Ejemplo: Discurso Steve Jobs Standford



Tecnologías Basadas en Análisis Textual

Transcripción

Traducción

The screenshot displays the DoblAI Beta web interface. At the top left is the logo, and at the top right are user controls for 'Francesc' and 'Log out'. Below the logo is a 'Select language' dropdown menu set to 'Spanish'. To the right of this are 'Save all' and 'Generate new video' buttons. The main content area is split into two columns. The left column features a video player showing a man in a black academic gown speaking at a podium. The video player includes a progress bar, volume control, and a 'Sync video with text' checkbox which is checked. The right column contains a list of transcribed text segments. Each segment includes a play button, a time range (e.g., 00:01:09.500 - 00:01:10.599), a speaker identifier (e.g., S33 (es-US | male)), and a red trash icon. The text segments are as follows:

- 00:01:09.500 - 00:01:10.599 Speaker: S33 (es-US | male) Entonces, ¿por qué abandoné?
- 00:01:12.300 - 00:01:13.800 Speaker: S33 (es-US | male) Comenzó antes de que yo naciera.
- 00:01:15.360 - 00:01:20.940 Speaker: S33 (es-US | male) Mi madre biológica era una joven estudiante de posgrado soltera y decidió darme en adopción.
- 00:01:22.459 - 00:01:30.540 Speaker: S33 (es-US | male) Ella estaba muy convencida de que yo debería ser adoptado por graduados universitarios, así que todo estaba listo para que yo fuera adoptado al nacer por un abogado y su esposa.
- 00:01:32.099 - 00:01:36.340 Speaker: S33 (es-US | male) Excepto que cuando salí, decidieron en el último minuto que realmente querían una niña.
- 00:01:37.900 - 00:01:42.160 Speaker: S33 (es-US | male) Así que mis padres, que estaban en lista de espera, recibieron una llamada en medio de la noche preguntando:

FR

Can you select 10 sentences to produce a video summary. Give me the sentences and the timecodes



Sure! Here are 10 sentences along with their corresponding timecodes to create a video summary:



1. Sentence: "Me siento honrado de estar con ustedes hoy para su graduación de una de las mejores universidades del mundo."
Timecode: 00:00:27,960 - 00:00:30,499
2. Sentence: "Hoy quiero contarles tres historias de mi vida."
Timecode: 00:00:48,520 - 00:00:52,045
3. Sentence: "La primera historia es sobre conectar los puntos."
Timecode: 00:00:56,000 - 00:00:58,760
4. Sentence: "La segunda historia es acerca de amor y perdida."
Timecode: 00:07:04,920 - 00:07:09,029
5. Sentence: "La tercera historia es acerca de la muerte."
Timecode: 00:09:08,880 - 00:09:14,737
6. Sentence: "No puedes conectar los puntos mirando hacia adelante, solo puedes conectarlos mirando hacia atrás."
Timecode: 00:05:58,831 - 00:06:03,180
7. Sentence: "La única manera de hacer un gran trabajo es amar lo que haces."
Timecode: 00:08:36,182 - 00:08:39,796
8. Sentence: "Recordar que vas a morir es la mejor forma que conozco de evitar la trampa de pensar que tienes algo que perder."
Timecode: 00:09:40,620 - 00:09:46,021
9. Sentence: "A veces la vida te va a golpear en la cabeza con un ladrillo. No pierdas la fe."
Timecode: 00:06:00,681 - 00:06:06,380
10. Sentence: "Espero que todos puedan seguir su corazón y encontrar lo que aman hacer en la vida."
Timecode: 00:12:10,620 - 00:12:17,301

Please note that the timecodes are approximate and may vary slightly depending on the video editing software and the actual video content.

FR

quiero dividir el video de steve jobs en standford en 3 capítulos. A partir de los subtítulo que he pasado me puedes decir los códigos de tiempo en los que empieza cada capítulo y un título de capítulo?



Basándome en los subtítulos proporcionados, puedo ayudarte a dividir el discurso de Steve Jobs en la Universidad de Stanford en tres capítulos junto con sus respectivos códigos de tiempo y títulos de capítulo:



Capítulo 1: "Conectar los puntos"

- Código de tiempo de inicio: 00:00:48,520
- Código de tiempo de finalización: 00:07:04,920
- Título del capítulo: "Conectar los puntos"

Capítulo 2: "Amor y pérdida"

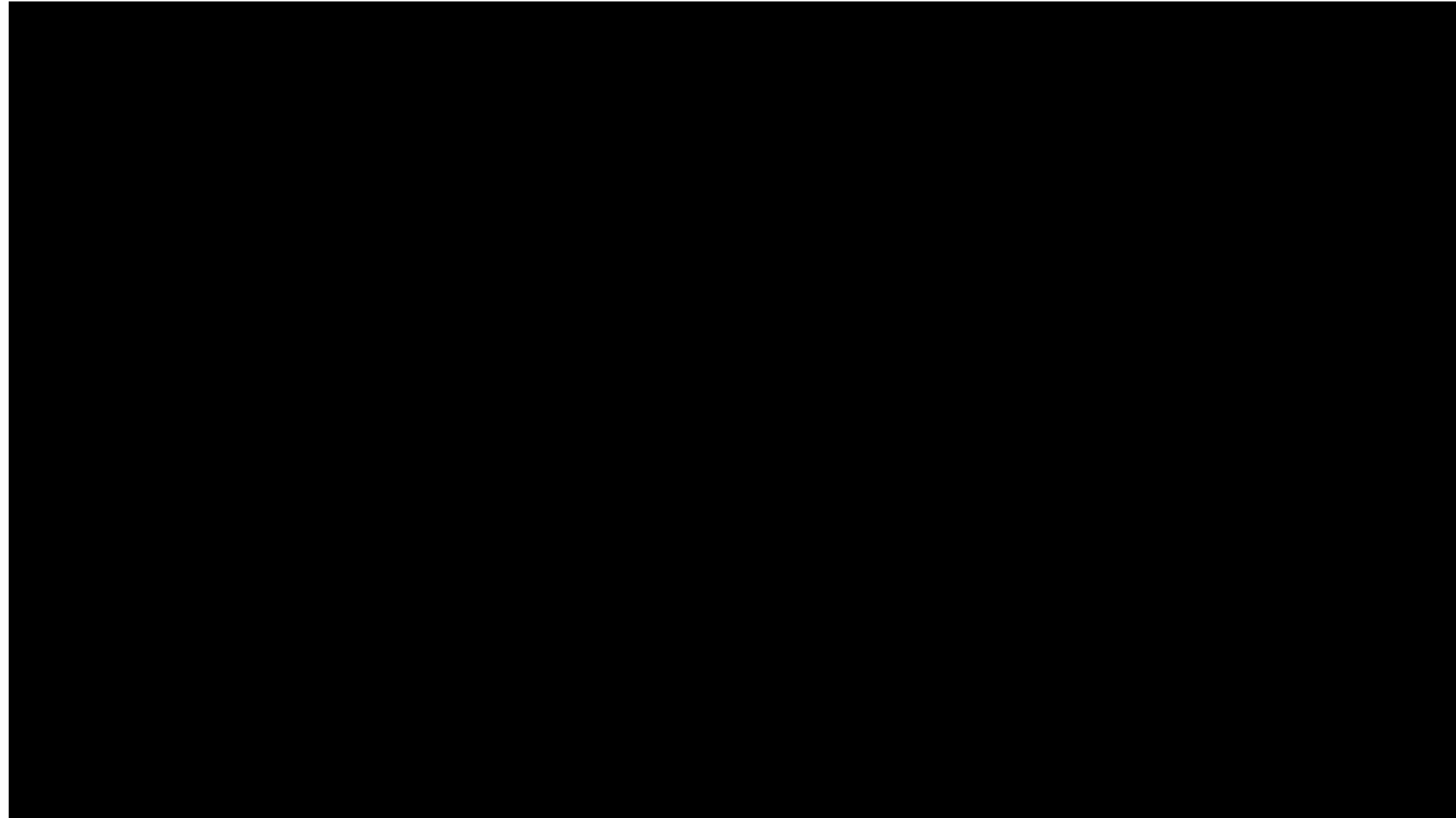
- Código de tiempo de inicio: 00:07:04,920
- Código de tiempo de finalización: 00:09:08,880
- Título del capítulo: "Amor y pérdida"

Capítulo 3: "Enfrentando la muerte"

- Código de tiempo de inicio: 00:09:08,880
- Código de tiempo de finalización: 00:12:17,301
- Título del capítulo: "Enfrentando la muerte"

Ten en cuenta que los códigos de tiempo pueden variar ligeramente dependiendo de la versión y edición del video. Asegúrate de verificar los códigos de tiempo con precisión al realizar la edición del video.

Tecnologías Basadas en Análisis Textual



FR

Quiero hacer un vídeo con 5 capítulos utilizando el texto que te he proporcionado. Puedes darme los códigos de tiempo en los que empieza cada capítulo y los títulos de los capítulos



¡Claro! A continuación, te proporcionaré los códigos de tiempo aproximados y los títulos para cada capítulo del video basados en el texto que me proporcionaste:



Capítulo 1: Aceptar la realidad (0:00 - 2:15)

Capítulo 2: La importancia de la formación del carácter (2:16 - 5:42)

Capítulo 3: Creer en las posibilidades de victoria (5:43 - 8:10)

Capítulo 4: La lección de no justificarse constantemente (8:11 - 11:28)

Capítulo 5: Reflexiones finales del entrenador (11:29 - 14:36)

Estos tiempos son solo aproximados y pueden variar dependiendo de la velocidad de lectura y el ritmo del video. Puedes ajustarlos según tus necesidades y preferencias.



En este extracto, el entrenador de tenis y tío de Rafael Nadal habla sobre su experiencia y enfoque como entrenador. Destaca que ha sido el entrenador de Nadal durante muchos años y que reunía dos condiciones importantes: ser su tío y ser el entrenador más barato del circuito. Aunque reconoce que es más difícil echar a un familiar que a una persona ajena, también menciona que su relación como tío y entrenador ha sido beneficiosa. Comenta que se ha enfocado en formar el carácter de Nadal y en fomentar la autocrítica en lugar de justificaciones. Destaca la importancia de asumir la responsabilidad y evitar el engaño. También menciona anécdotas relacionadas con las raquetas y las condiciones de juego en diferentes torneos.



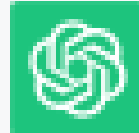
En resumen, el entrenador destaca su papel como tío y entrenador de Rafael Nadal, menciona su enfoque en la formación del carácter y la importancia de la autocrítica.

También comparte anécdotas relacionadas con las condiciones de juego y las raquetas en diferentes torneos.

Tecnologías Basadas en Análisis Textual



Que fragmentos del texto original elegirías para hacer un resumen de 2 minutos del video







Para hacer un resumen de 2 minutos del video, podrías seleccionar los siguientes fragmentos del texto original:



1. "Aceptar la realidad. Una cosa bastante complicada hoy en día."
2. "Yo he rehuido siempre de una sobrevaloración de los chicos a los que entrenaba, sobre todo de Rafael."
3. "He sido un entrenador que me preocupé casi más de la formación del carácter de Rafael, formarle un carácter fuerte que en sí de formarle bien técnicamente."
4. "Yo siempre he creído en las opciones de victoria, aunque a veces la verdad es que no ha salido muy bien."
5. "Yo he procurado que nunca me dieras justificaciones, ni a derrotas ni a nada de lo que pasaba. Es demasiado fácil justificarse constantemente."

Tecnologías Basadas en Análisis Textual

 Selecciona las 5 frases que creas más representativas del texto

 Aquí están las cinco frases que considero más representativas del texto:   

1. "Aceptar la realidad. Una cosa bastante complicada hoy en día."
2. "Yo he rehuido siempre de una sobrevaloración de los chicos a los que entrenaba, sobre todo de Rafael."
3. "Yo siempre he creído en las opciones de victoria, aunque a veces la verdad es que no ha salido muy bien."
4. "He sido un entrenador que me preocupé casi más de la formación del carácter de Rafael, formarle un carácter fuerte que en sí de formarle bien técnicamente."
5. "Yo he procurado que nunca me dieras justificaciones, ni a derrotas ni a nada de lo que pasaba. Es demasiado fácil justificarse constantemente."