

# MPEG-4

**Introducción a la  
codificación natural**

**Jose Antonio Cantón Claro**

**[jack.grim@gmail.com](mailto:jack.grim@gmail.com)**

# INDICE

- **Diferencias con MPEG-1 y MPEG-2**
- **Principales Usos**
- **Natural Video Coding**
  - **MPEG-4 Natural Video Standard**
  - **Principales Características y Funcionalidades**
  - **Bitrate Ranges**
  - **Estructura y Sintaxis**
  - **Niveles Jerárquicos en la escena visual**
  - **Shape Coding**
  - **Motion Estimation and Compensation**
  - **Texture Coding**
  - **Error Resilience**
  - **Scalability**

# Diferencias con MPEG1 y MPEG2

- **MPEG-1** y **MPEG-2** son estándares que se centran en la compresión y descompresión de flujos de video y audio.
- **MPEG-1** fue diseñado para proveer un estándar de compresión, **MPEG-2** para proveer mayor calidad en la transmisión de aplicaciones, centrándose en la TV Digital
- **MPEG-4** permite representar las escenas como un conjunto de **objetos audiovisuales independientes**
- Cada objeto independiente se le trata de forma distinta y se le aplican distintas técnicas de compresión para **obtener los mejores resultados**
- Aunque **MPEG-1** y **MPEG-2** son apropiados para los entornos para los que fueron diseñados **no son flexibles** para encargarse eficientemente de los requisitos de **aplicaciones multimedia**
- **MPEG-4** provee una plataforma común para un extenso rango de aplicaciones multimedia

# Principales Usos

- **Flujos de Medios Audiovisuales**
- **Distribución en CD**
- **Transmisión Bidireccional por videófono**
- **Televisión Digital**
- **Aplicaciones gráficas interactivas**
- **Móvil Multimedia**
- **Juegos**
- **Video Streaming**

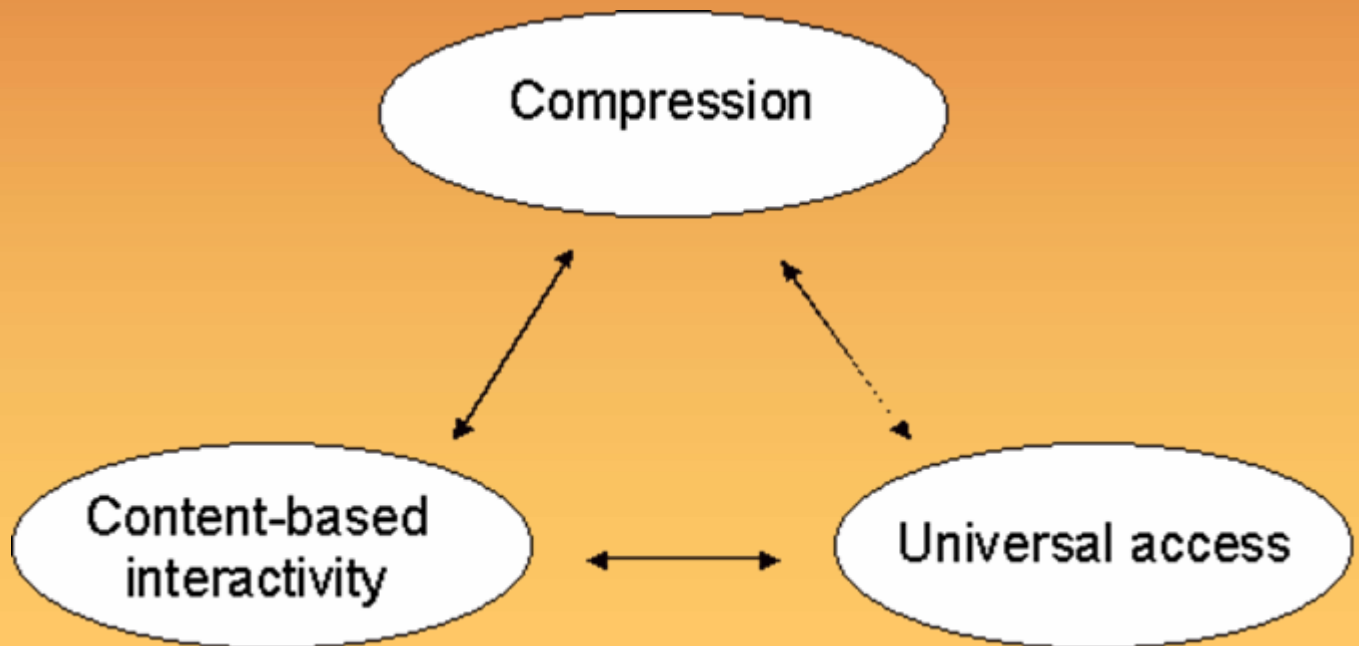
# NATURAL VIDEO CODING

## ● MPEG-4 Natural Video Standard

- El estandar visual **MPEG-4** se desarrolla para proveer a usuarios de un nuevo nivel de interacción con contenidos visuales
- Provee tecnología para **ver, acceder y manipular** objetos en vez de píxeles
- Consiste en una colección de herramientas que apoyan áreas de aplicación como: televisión digital, video streaming, multimedia y juegos en móviles
- **MPEG-4** nos provee de herramientas para **shape coding, motion estimation and compensation, texture coding, error resilience, sprite coding y scalability**

# NATURAL VIDEO CODING

- Principales Características y Funcionalidades



# NATURAL VIDEO CODING

## ● Características y Funcionalidades

- **Compression efficiency:** mejora en el rendimiento de la codificación y **añade codificación de múltiples flujos de datos concurrentes**
- **Content-based interactivity:** codificando y representando objetos de video en vez de frames permite la **creación de aplicaciones basadas en contenido**, una de las novedades más importantes del MPEG-4. El nuevo nivel de interactividad se basa en:
  - Representación eficiente de objetos
  - Manipulación de Objetos
  - Edición de secuencias de bits
  - Escalabilidad basada en objetos

# NATURAL VIDEO CODING

- **Características y Funcionalidades**

- **Universal Access:** la robustez en entornos propensos a errores permite al contenido codificado MPEG-4 ser accesible por una extensa gama de medios de comunicación.

- **Bitrate Ranges:**

- El estándar visual MPEG-4 ha sido explícitamente optimizado para los siguientes bitrate:

- **Por debajo de 64 kbit/sec**

- **64 - 384 kbit/sec**

- **384 - 4 Mbit/sec**

# NATURAL VIDEO CODING

## ● Estructura y Sintaxis

- **Concepto Central:** los **objetos audio-visuales** fundan la representación basada en objetos. Representación apropiada para aplicaciones interactivas y dan acceso directo a los contenidos de las escenas
- **Objeto Video:** consiste en una o más capas para dar soporte a codificación escalable. La **sintaxis de la escalabilidad permite la reconstrucción de video por capas**, empezando por una capa base independiente y añadiendo capas de mejora.
- **Escena Visual:** consiste en uno o más objetos video. Cada uno se caracteriza por información temporal y espacial de la forma, movimiento y textura

# NATURAL VIDEO CODING

- **Estructura y Sintaxis**

- **Visual bitstream:** provee una descripción jerárquica de una escena visual. Cada nivel de la jerarquía puede ser accedido en la secuencia de bits mediante valores especiales de código llamados **start codes**

# NATURAL VIDEO CODING

- **Niveles Jerárquicos en la escena visual:**

- **Visual Object Sequence (VS):** escena completa que puede contener cualquier cantidad de objetos 2-D o 3-D y sus capas de mejora

- **Visual Object (VO):** cada objeto 2-D en la escena

- **Visual Object Layer (VOL):** cada objeto video puede ser codificado de forma escalable (multi-capa) o no-escalable (capa única) dependiendo de la aplicación.

Estos Layer dan soporte a la codificación escalable pudiendose codificar un Video Object usando escalabilidad espacial o temporal.

Dependiente de parámetros como ancho de banda, poder computacional y preferencias de usuario

# NATURAL VIDEO CODING

- **Niveles Jerárquicos en la escena visual:**

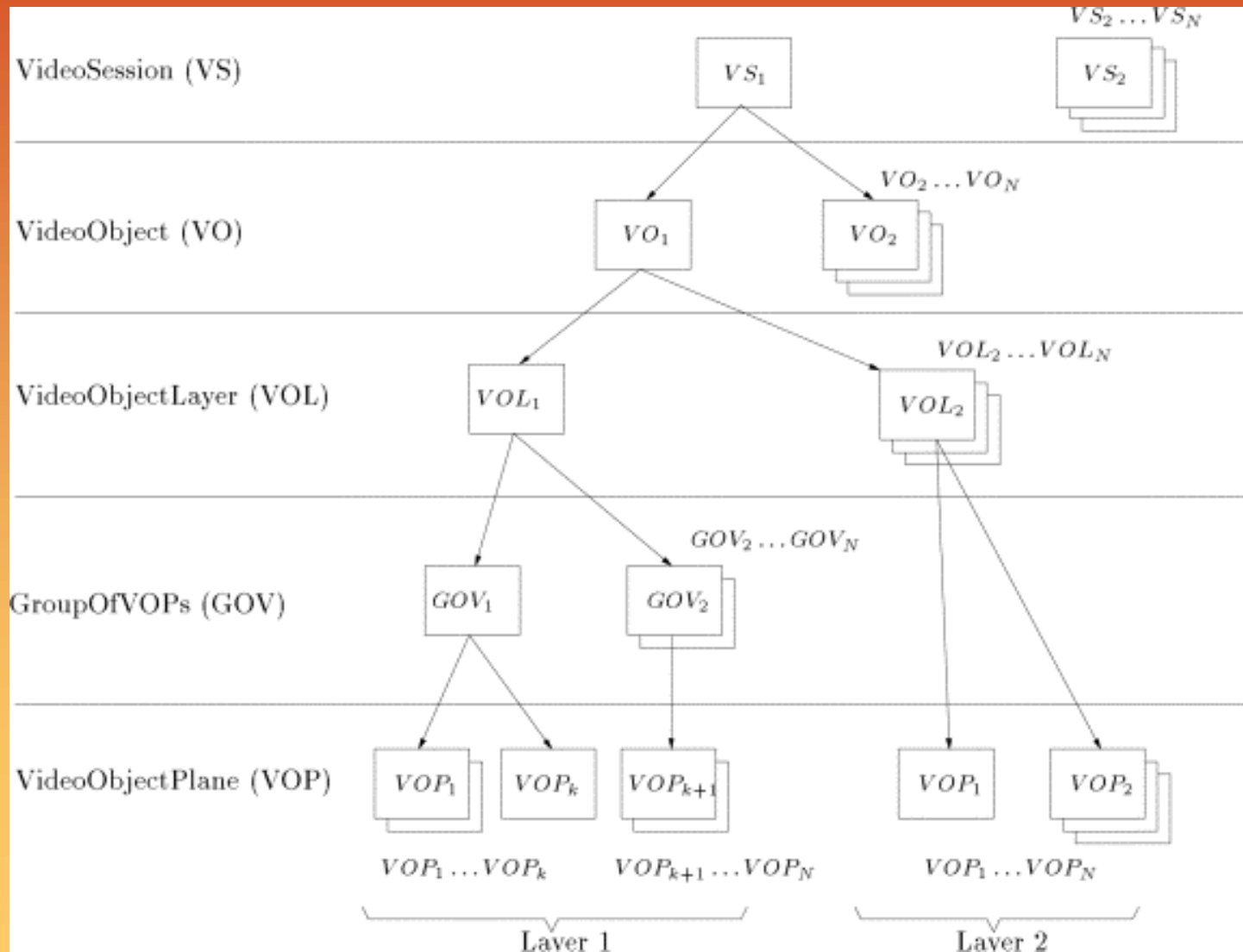
**Cada Video Object es muestreado en el tiempo**

**Cada muestra de un Video Object es un Video Object Plane (nivel de objeto de video)**

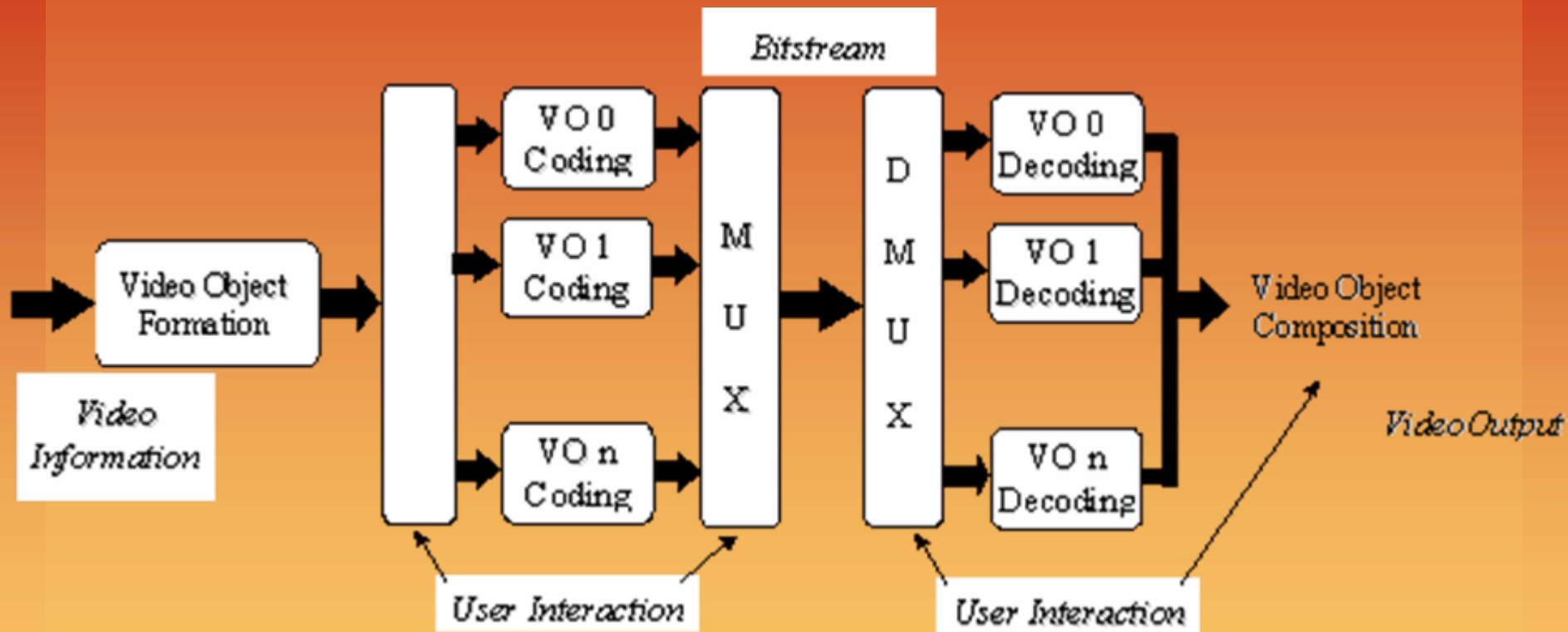
- **Group of Video Object Planes (GOV):** agrupa los Video Object Planes y puede proveer puntos en la secuencia de bits donde éstos son codificados independientemente. Opcionales

- **Visual Object Plane (VOP):** es una muestra en el tiempo de un Video Object. Pueden ser codificados independientemente o depender unos de otros mediante el uso de compensación de movimiento

# Niveles Jerárquicos en la escena visual



# Diagrama de Bloques de Video MPEG-4



Cada Video Object se codifica por separado

Por razones de eficiencia y compatibilidad hacia atrás, son codificados via sus correspondientes Video Object Planes

# NATURAL VIDEO CODING

- **Niveles Jerárquicos en la escena visual:**

Un **Video Object Plane** (nivel de objeto de video) se puede usar de varias maneras:

**En la más común** el VOP contiene la codificación de datos de video de una muestra de un Video Object

Los parámetros de movimiento, la información de forma y los datos de textura son codificados mediante **macroblocks** (también usados para codificar **sprites**)

Los **sprites** son Video Object normalmente mayores que el video visualizado y continuo en el tiempo (se usan para representar áreas más o menos estáticas como backgrounds)

# NATURAL VIDEO CODING

- **Niveles Jerárquicos en la escena visual:**

Un **macroblock** contiene una sección para la componente **luminosidad** y las componentes de **chrominancia**

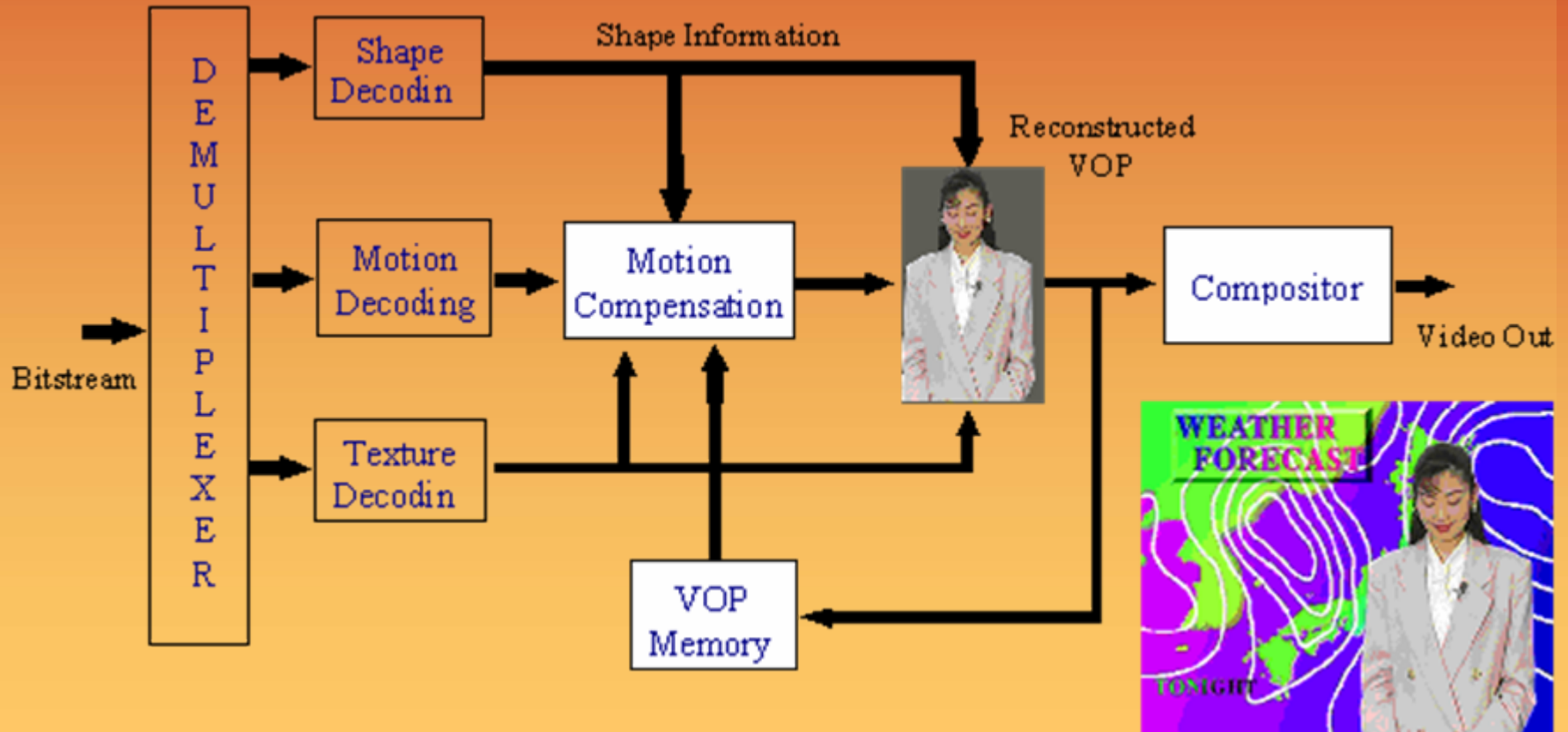
MPEG-4 da soporte para un solo formato de chrominancia (componente de la señal de video que contiene la información de color), el 4:2:0

En este formato cada macrobloque contiene **4 bloques de luminosidad** y **2 bloques de chrominancia**

Cada bloque contiene **8x8 pixels** y se codifica usando la **Transformada Discreta del Coseno (DCT)**

El macrobloque transporta información de forma, de movimiento y de textura

# Ejemplo: Decodificación VOP



# NATURAL VIDEO CODING

## ● Shape Coding

- **Modo Binario:** la forma de cada objeto es descrita por una máscara binaria.

Esta máscara normalmente se representa como una matriz del mismo tamaño que la caja delimitante del VOP, cuyos elementos pueden tomar 2 valores dependiendo si el pixel se encuentra dentro o fuera del Video Object

- **Modo Escala de Grises:** la forma es descrita de manera similar al canal alpha, permitiendo transparencia y reduciendo aliasing

La información de forma se representa por 8 bits, en lugar de un valor binario

# NATURAL VIDEO CODING

## ● Motion Estimation and Compensation

- La estimación y compensación de movimiento se usa comúnmente **para comprimir secuencias de video** aprovechando redundancias temporales entre frames.
- La principal **diferencia con MPEG-4** es que las **técnicas** basadas en bloques usadas en otros estándares han sido **adaptadas a la estructura VOP**
- MPEG-4 provee de **tres modos** para codificar una entrada VOP

# NATURAL VIDEO CODING

## ● Motion Estimation and Compensation

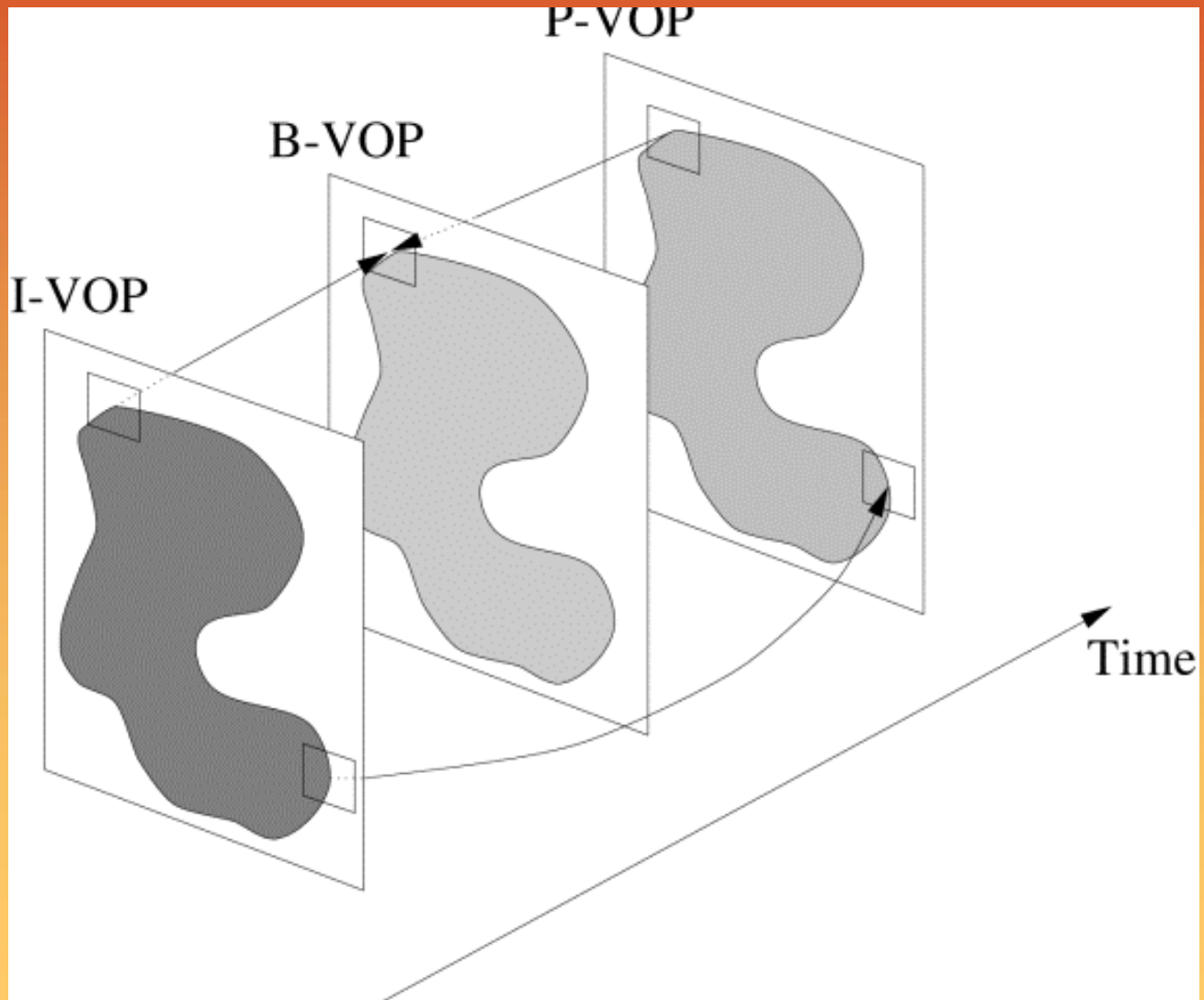
- **Modos** para codificar una entrada VOP

- **Intra VOP (I-VOP):** el VOP se codifica independientemente de cualquier otro VOP

- **Predicted VOP (P-VOP):** un VOP puede ser predecido (usando compensación de movimiento) basándose en otro VOP previamente descifrado

- **Bidirectional Interpolated VOP (B-VOP):** un VOP puede ser predecido basandose tanto en pasados como futuros VOP's. Los B-VOP pueden ser solo interpolados basandose en I-VOP o P-VOP

# Modos de Codificación de VOP's

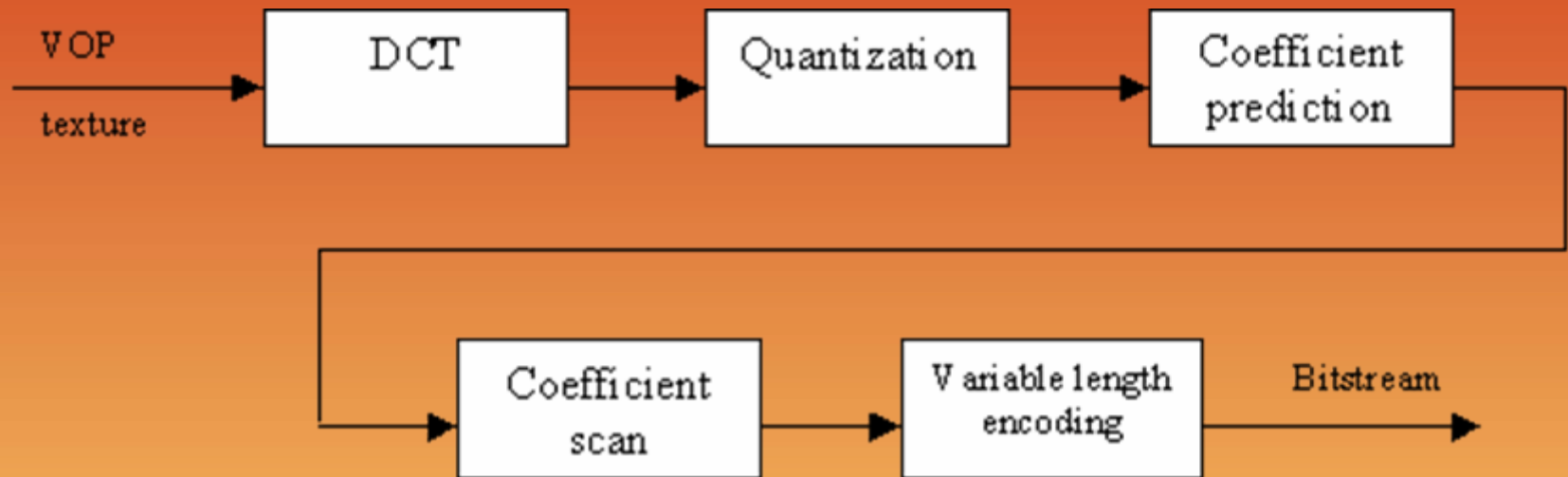


# NATURAL VIDEO CODING

## ● Texture Coding

- La información de textura de un VOP está presente en la **luminosidad Y** y las dos componentes de **crominancia Cb y Cr** de la señal de video
- En el caso de **I-VOP** la información reside directamente en los componentes de luminosidad y crominancia
- En los **P-VOP** y **B-VOP** la información de textura representa el error residual que queda después de la compensación del movimiento
- Para la codificación de la información de textura se usa el **estándar 8x8 block-based DCT**

# Proceso de Codificación de Textura

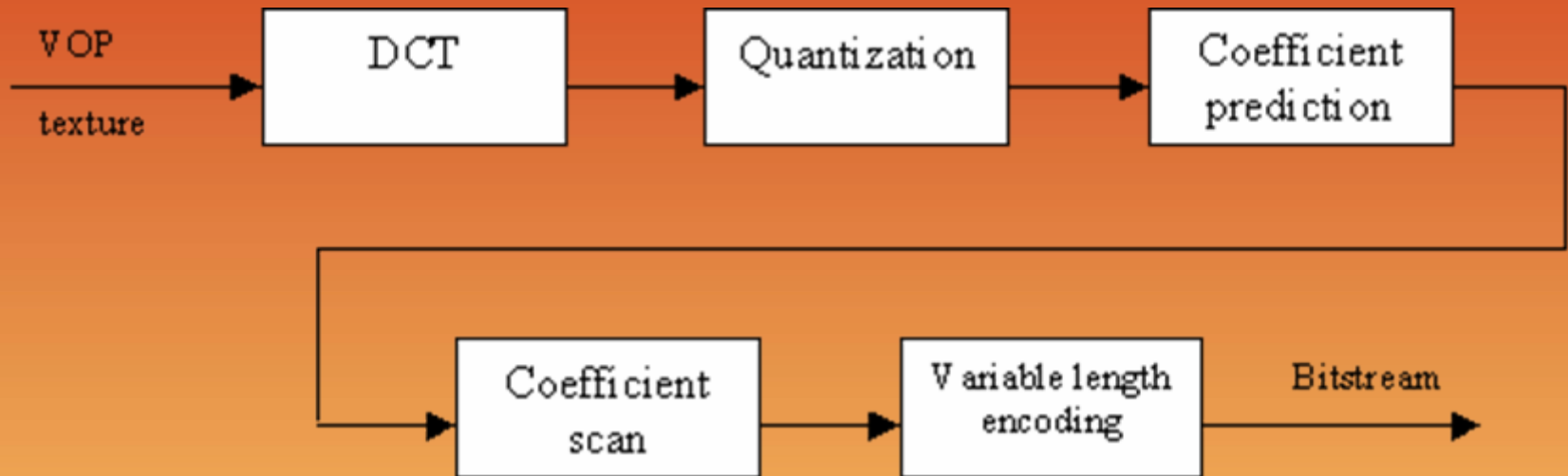


- **DCT:** los bloques (macroblocks) de textura de video se codifican usando 2-D 8x8 block-based DCT

- **Cuantificación:** los coeficientes DCT son cuantizados usando dos posibles algoritmos de cuantificación:

  - **Cuantificación MPEG:** derivado del estándar MPEG-2. Muy parecido al del estándar de imagen JPEG. Emplea las **tablas de cuantificación**, de forma que cada coeficiente recibe un tratamiento distinto según el índice espectral que lo representa

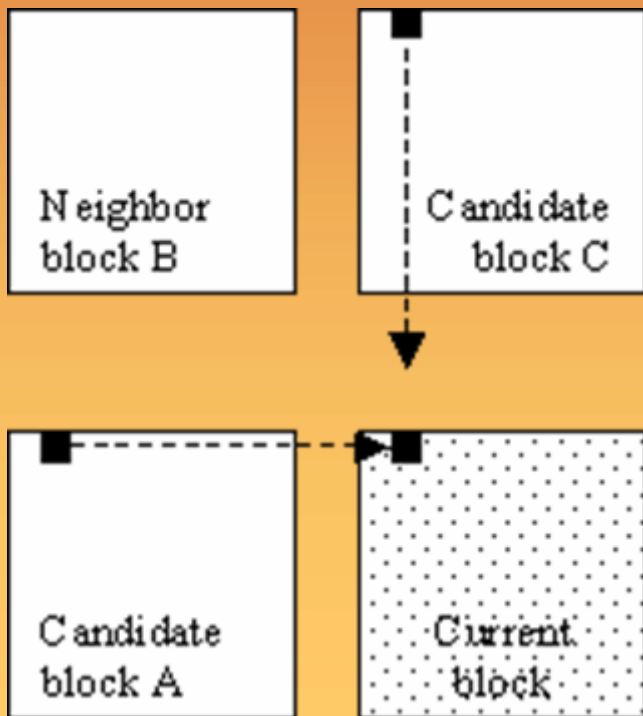
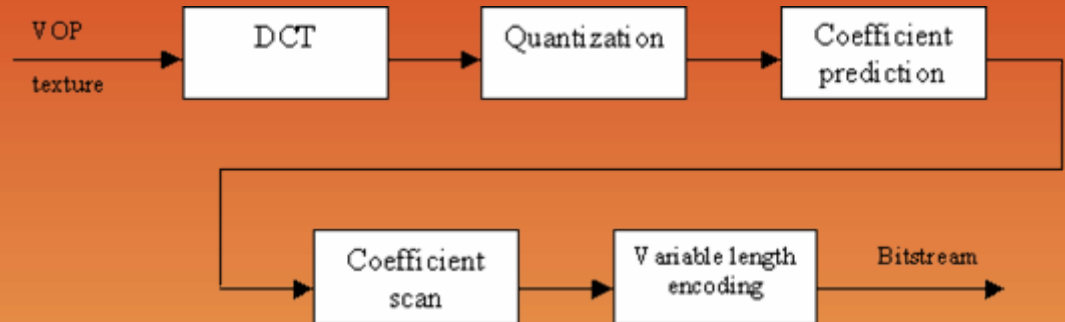
# Proceso de Codificación de Textura



→ **Cuantificación H.263:** no utiliza tablas de cuantificación. Técnica menos compleja y fácil de implementar

- **Coefficient prediction:** para la mayoría de bloques vecinos existen relaciones de dependencia. Así, la energía media de los coeficientes cuantificados puede reducirse aún más utilizando predicción con los bloques vecinos

# Proceso de Codificación de Textura

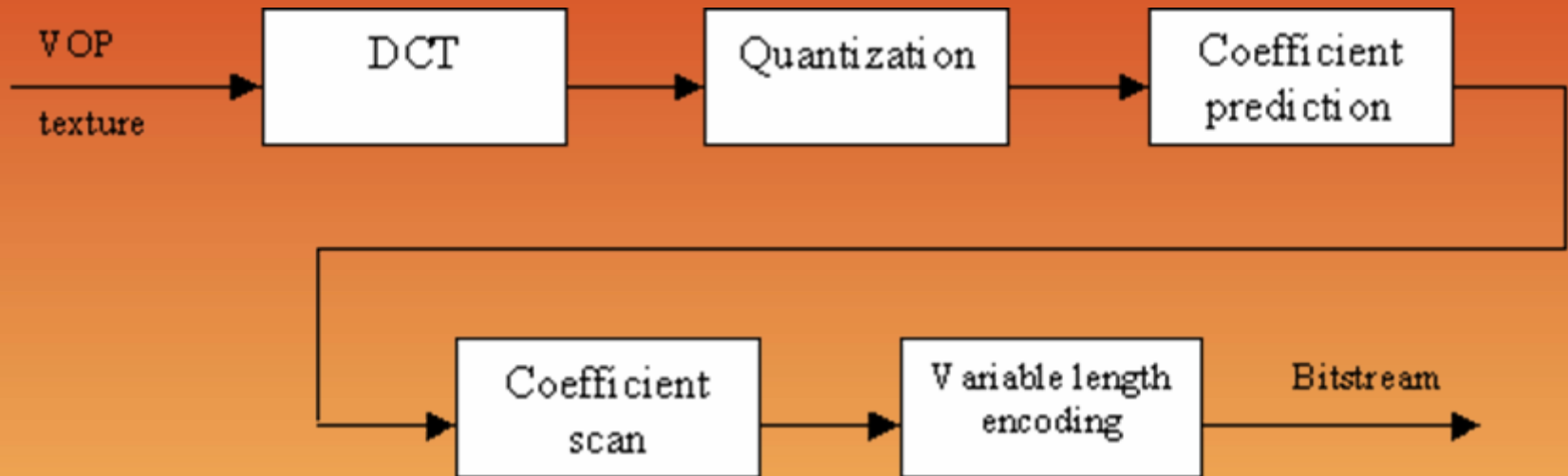


▪ **Coefficient prediction:** dos tipos de predicciones:

→ **Predicción DC:** la predicción se realiza para el coeficiente DC, escogiendo entre el DC de A o de C

→ **Predicción AC:** los coeficientes de la primera columna y primera fila del bloque actual se predicen con los coeficientes situados en la misma posición de los bloques candidatos (A y C)

# Proceso de Codificación de Textura



• **Coefficient Scan:** aplicado para transformar los datos en dos dimensiones en un vector unidimensional. Tres formas de hacer la transformación:

- **Recorrido en Zig-Zag:** los coeficientes son extraídos diagonalmente
- **Recorrido Horizontal:** extraídos horizontalmente
- **Recorrido Vertical:** extraídos verticalmente

# NATURAL VIDEO CODING

## ● Error Resilience

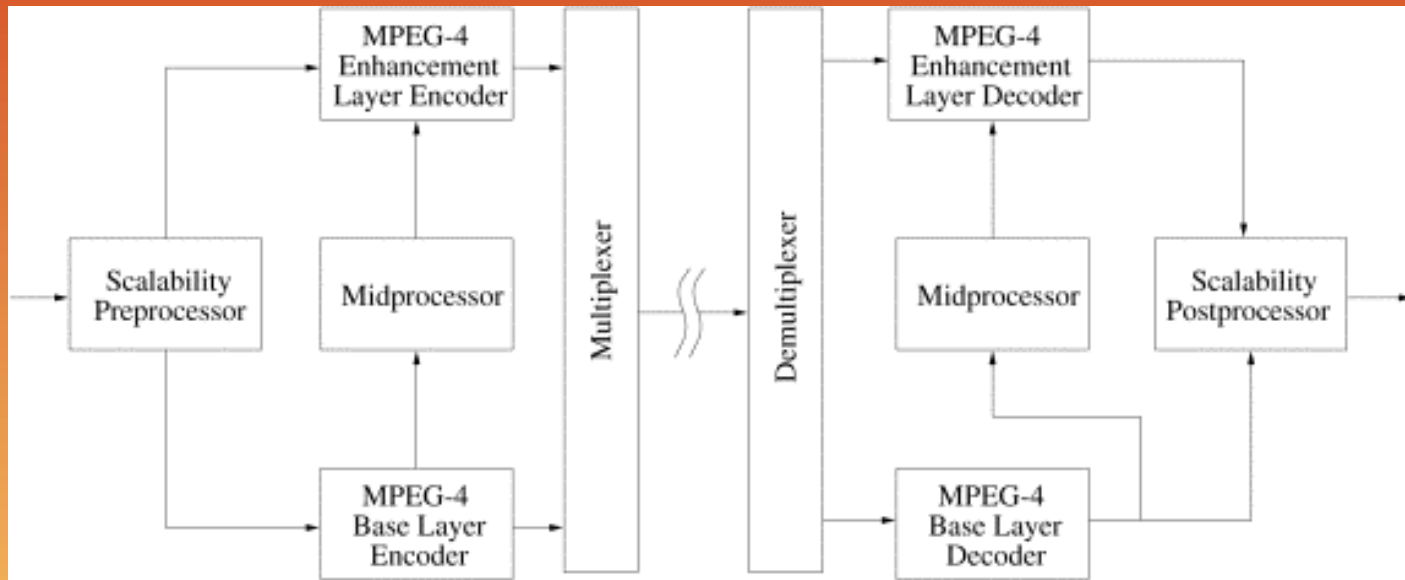
- Funcionalidad importante para el acceso universal através de entornos propensos a errores, como comunicaciones móviles
- MPEG-4 nos provee de varios mecanismos para permitir “elasticidad” de errores:
  - **Resynchronization**
  - **Data partitioning**
  - **Header extension code**
  - **Reversible VLCs**

# NATURAL VIDEO CODING

## ● Scalability

- La escalabilidad espacial y temporal se implementan usando múltiples VOL's. Consideremos dos VOL's, la capa base y la capa de mejora (*base-layer* y *enhancement-layer*)
- En la **escalabilidad espacial** la capa de mejora perfecciona la resolución espacial de un VOP dado por la capa base
- En la **escalabilidad temporal** la capa de mejora puede ser decodificada si el frame rate deseado es mayor que el ofrecido por la capa base.
  - Mejora la suavidad del movimiento en la secuencia

# Framework generalizado de escalabilidad

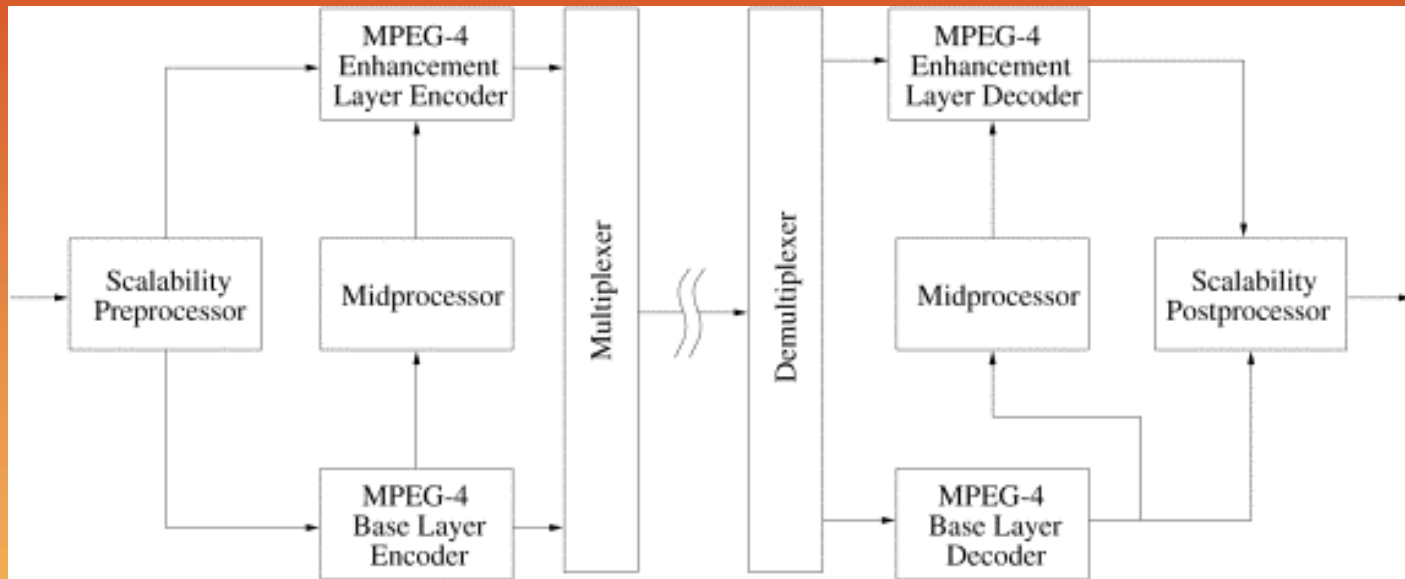


- MPEG-4 usa un framework generalizado de escalabilidad para permitir escalabilidades temporales y espaciales (mediante módulos separados)

→ **Preprocessor y Postprocessor:** usados para implementar la escalabilidad deseada. Trabajan con VOP's

→ **Midprocessor:** toma la capa capa base reconstruida por el preprocesador y aumenta las muestras. La diferencia entre el VOP original tomado y la salida del midprocessor forma la entrada para el codificador de la capa de mejora

# Framework generalizado de escalabilidad



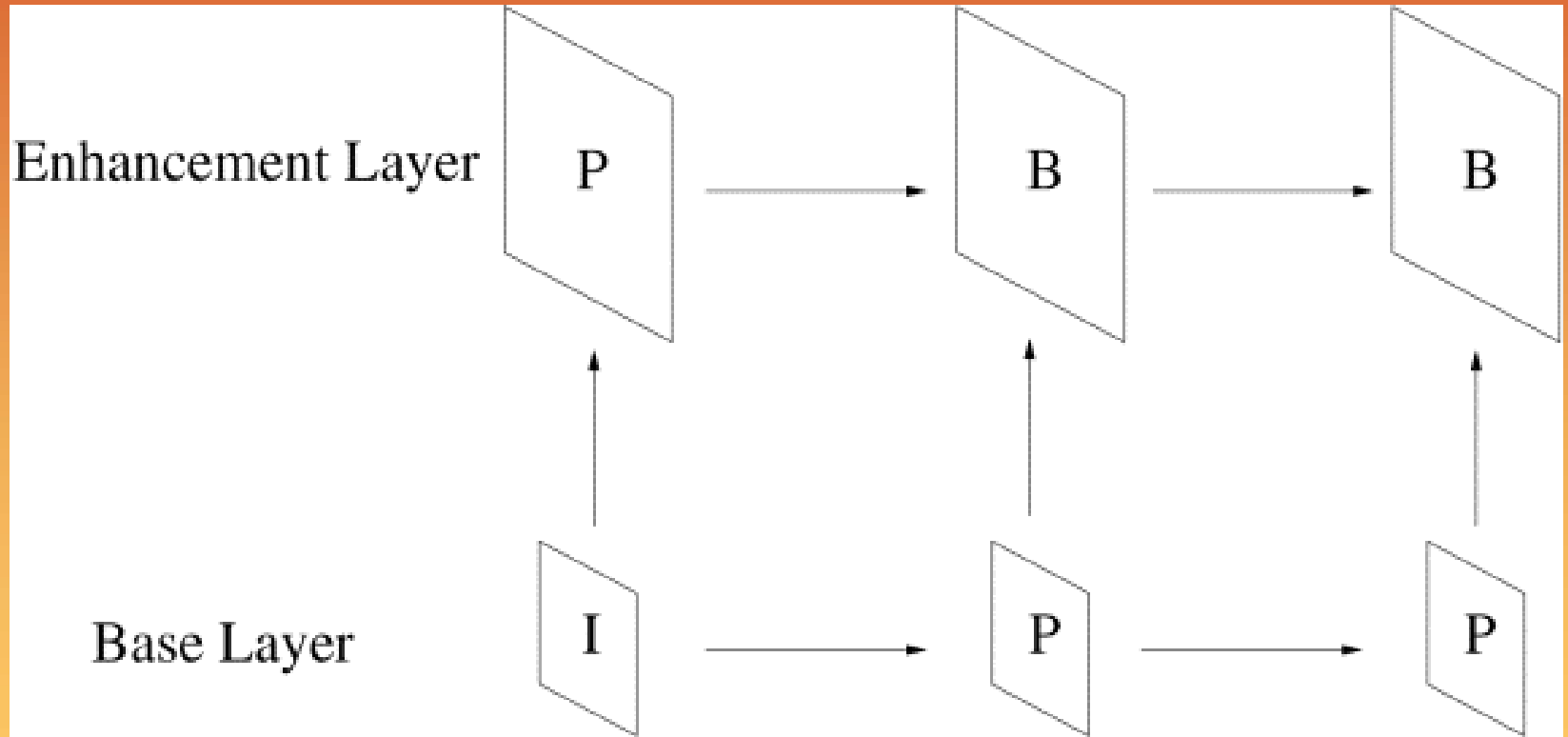
- Para implementar **escalabilidad temporal**, el preprocesador separa los frames en dos flujos.
  - Un flujo forma la entrada para el codificador de la capa base
  - Mientras que el otro flujo es procesado por el codificador de la capa de mejora
  - En este caso se elude el midprocessor

# NATURAL VIDEO CODING

## ● Spatial Scalability

- Los VOP's con capas de mejora se codifican como P-VOP o B-VOP
- Si un VOP en la capa de mejora coincide temporalmente con un I-VOP en la capa base, éste se trata como un P-VOP
- Si un VOP en la capa de mejora coincide con un P-VOP en la capa base, éste se trata como un B-VOP

# Comportamiento de capa base y capa de mejora



# NATURAL VIDEO CODING

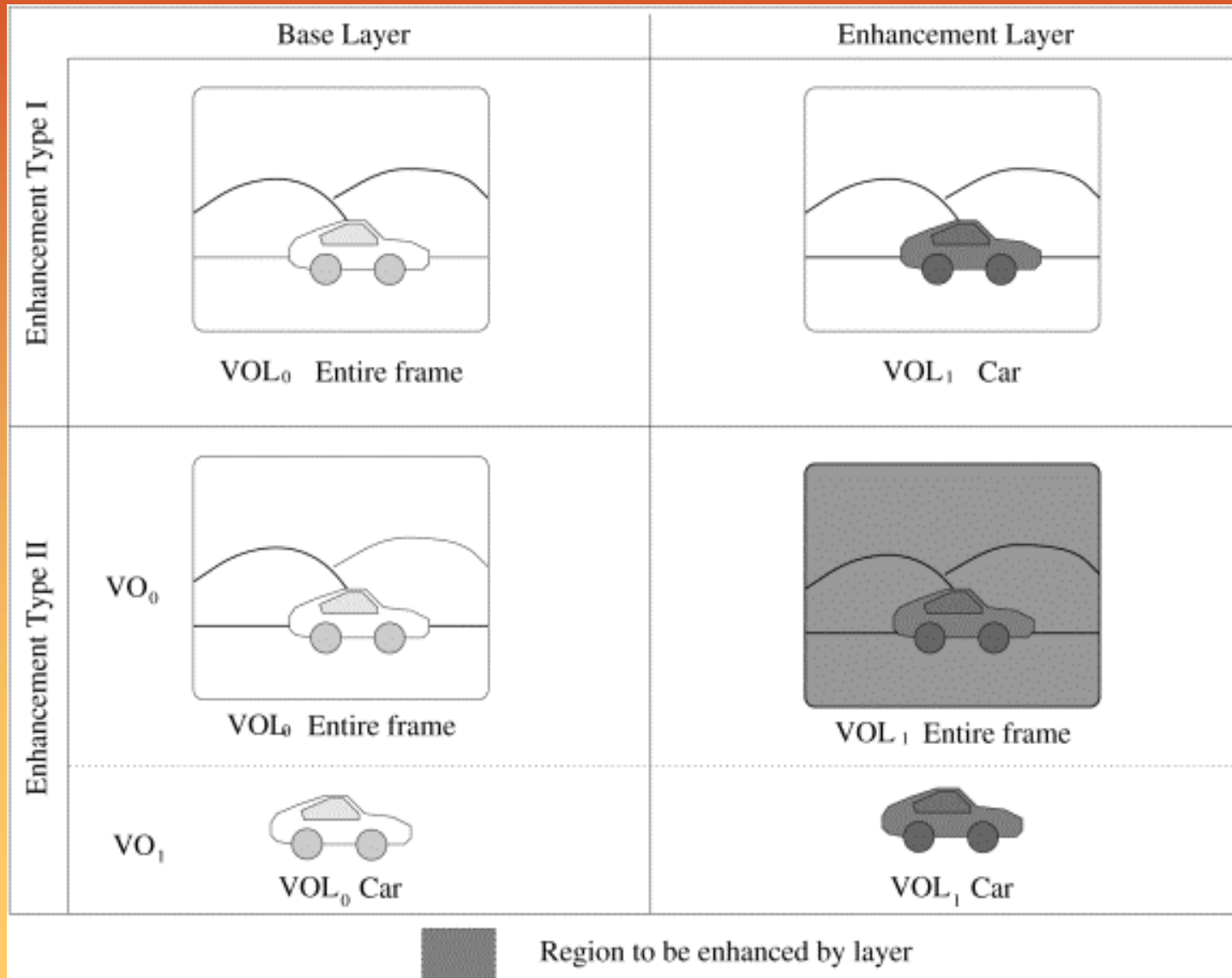
## ● Temporal Scalability

- Las capas de mejora transportan información para ser visualizada entre frames de la capa base. La capa de mejora actuará de una de estas formas:

- **Tipo 1:** la capa de mejora mejora la resolución de una porción de la capa base

- **Tipo 2:** la capa de mejora mejor la resolución de la totalidad de la capa base

# Tipos de mejora para escalabilidad temporal



# Comparativa entre formatos

	<b>MPEG-4</b>	<b>Windows Media</b>	<b>Real</b>	<b>Flash</b>
Audio/Video Codec	Standards based; multi-vendor support.	Proprietary	Proprietary, but supports automatic download of MPEG-4 plug-in.	Proprietary + proprietary Real and QuickTime formats.
Interactivity	Highly interactive.	Limited	Yes, via SMIL.	Highly interactive.
Digital Rights Management	Interfaces to proprietary DRM. More interoperable DRM under development in MPEG-4 and MPEG-21	Microsoft DRM	Content access control	No
Real-time stream control	Yes	Yes	Yes	No
Synchronization	Audio, video and all other objects can be tightly synchronized with high accuracy	Tight synchronization between audio and video	Tight synchronization between audio and video	No synchronization between scene and streams
Broadcast capable	Yes, including interactive features	A/V only	Scene must be unicast	No
Object model support	Video/audio and rich 2D/3D mixed media, synthetic graphics. DRM on separate streams.	Audio/Video only	Video/audio and mixed media through SMIL based protocol. No streaming of mixed media.	Video/audio and mixed media through proprietary protocol.
Graphic Objects	Yes	No	No	Yes
Transport	Support exists for HTTP, UDP, RTP/RTSP, MPEG-2TS, mobile	HTTP, UDP, RTP/RTSP, mobile	HTTP, RTP/RTSP, mobile	HTTP
PC, Set Top Box, Wireless	Yes	Yes	Yes	No

# BIBLIOGRAFÍA

- *The MPEG-4 Book [IMSC PRESS Multimedia Series]*
- *<http://es.wikipedia.org/wiki/MPEG>*
- *<http://www.chiariglione.com/mpeg>*

# MPEG-4

**Introducción a la  
codificación natural**

**Jose Antonio Cantón Claro**

**[jack.grim@gmail.com](mailto:jack.grim@gmail.com)**