

# MÉTODO PARA GENERACIÓN DE INFORMACIÓN LATERAL MULTI-CÁMARA PARA LA CODIFICACIÓN DISTRIBUIDA DE VÍDEO

Xavi Artigas, Francesc Tarrés, Luis Torres

{xavi, tarres, luis}@gps.tsc.upc.edu

Dpto. de Teoría de la Señal y Comunicaciones. Universidad Politécnica de Cataluña.

Campus Norte D-5, Jordi Girona 1-3, 08034 Barcelona

**Abstract-** This paper presents a side information generation method for multiview distributed video coding scenarios. Existing literature on this topic relies on pure temporal interpolation (by means of motion compensation), pure inter-camera interpolation (by means of disparity compensation) or a combination of both. In this work a different approach is used by calculating the motion vectors on an available conventionally-encoded camera (one that is not encoded using distributed principles) and then using the obtained motion vectors to generate the side information for the camera that requires it. This paper reports the status of the ongoing work and presents the results obtained so far after simulation on different multiview test sequences, alongside with a comparison with the simpler pure temporal and pure inter-camera methods. Finally, some conclusions and future work lines are given.

## I. INTRODUCCIÓN

La investigación y la estandarización en codificación de vídeo han adoptado hasta este momento un paradigma en el que la tarea de explorar las estadísticas de la fuente recae exclusivamente sobre el codificador, conduciendo a una distribución de la complejidad donde codificadores complejos interactúan con decodificadores más simples. La codificación distribuida de vídeo (DVC, un caso particular de la codificación distribuida de fuente o DSC) adopta el paradigma inverso, dando al decodificador la tarea de explorar las estadísticas de la fuente para obtener una compresión eficiente. Este paradigma es particularmente adecuado para escenarios de creciente interés tales como cámaras de vídeo inalámbricas, redes inalámbricas de vigilancia de bajo consumo, cámaras de vídeo no reutilizables, usos médicos, redes de sensores, adquisición multicámara de imagen, etc. Todos estos ejemplos se beneficiarían de codificadores de baja complejidad ya que disponen de recursos escasos de memoria, potencia de cálculo y energía para dedicar a la codificación.

Sin embargo, pese a que las bases teóricas para la codificación distribuida de fuente fueron fijadas hace treinta años con el trabajo de Slepian y Wolf [1] (para el caso sin pérdidas) y Wyner y Ziv [2] (para el caso con pérdidas), ha sido recientemente cuando su investigación ha tomado ímpetu. Esta investigación ha sido impulsada por el creciente interés en los escenarios antes mencionados, y conducida principalmente por Ramchandran [3] y Girod [4].

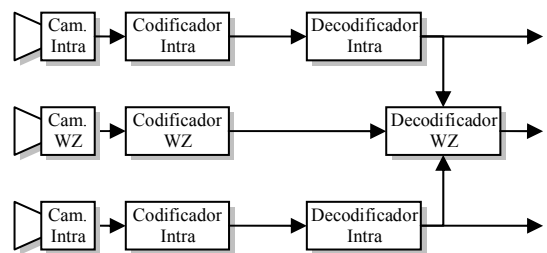


Fig. 1. Escenario físico.

Por otra parte, las técnicas multicámara se han investigado en el pasado, para la codificación [5] y para la interpolación de vistas, puesto que permiten, por ejemplo, crear vistas desde cámaras inexistentes, o la navegación libre por entornos físicos, dadas solamente algunas imágenes del escenario real [6].

El objetivo de la DVC multicámara es codificar eficientemente diversas señales de vídeo, pero explotar las posibles redundancias en el decodificador, obteniendo así las ventajas inherentes a la DVC como son una complejidad de codificación más baja, tolerancia a error incorporada o el hecho de que no es necesaria ninguna conexión entre las diversas cámaras, como sería el caso en una codificación conjunta tradicional. Esta comunicación presenta el estado actual de la investigación de un método para aplicar conceptos DVC al problema multicámara.

La DVC multicámara no ha recibido atención de la comunidad científica hasta hace poco tiempo. Ramchandran [7] y Girod [8] han trabajado con imágenes estáticas. Guo [9] y Ouaret [10] han trabajado con secuencias vídeo usando homografía para realizar la interpolación entre las diferentes cámaras (Inter-cámara) y un proceso posterior de fusión con la interpolación convencional (Intra-cámara).

## II. DEFINICIÓN DEL PROBLEMA

Se considera el escenario representado en la Fig. 1. Este escenario puede aumentarse añadiendo más cámaras o cambiando su configuración, pero esta es la estructura mínima necesaria para describir las técnicas presentadas. Se usan tres cámaras, que no se comunican entre ellas. Dos de ellas son llamadas “Cámaras Intra” y trabajan de modo convencional, esto es, su señal de vídeo se codifica y decodifica independientemente de las otras cámaras. La

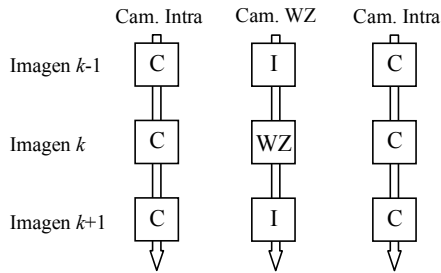


Fig. 2. Tipos de imágenes. C: Codificadas Convencionalmente, I: Intra, WZ: Wyner-Ziv, codificadas utilizando DVC.

tercera cámara, llamada “Cámara Wyner-Ziv” (WZ), codifica de forma individual, pero requiere las señales de las otras cámaras para ser decodificada. Esta decodificación conjunta permite a la cámara WZ transmitir a una tasa inferior a la que obtendría decodificando aisladamente.

La cámara Wyner-Ziv transmite algunas imágenes en modo Intra, es decir, codificadas independientemente de las demás imágenes. El resto, se llaman imágenes Wyner-Ziv y son las que aprovecharán la decodificación conjunta llevada a cabo en el receptor (Fig. 2). En el texto siguiente se asume que solo hay una imagen WZ entre cada dos imágenes Intra.

### III. CODIFICACIÓN DISTRIBUIDA DE VÍDEO

Se resume a continuación el proceso seguido para implementar los principios teóricos de Wyner y Ziv [2] en un codec real. La idea original puede encontrarse en [4] y el esquema general está dibujado en la Fig. 3.

Tal como se ha visto anteriormente, algunas de las imágenes de entrada (las llamadas Intra) se codifican y decodifican de forma independiente. El receptor, además, las utiliza para generar una estimación del resto de las imágenes, las llamadas WZ. Esta estimación se llama información lateral. Para las imágenes WZ el emisor solo tiene que transmitir los bits necesarios para corregir los posibles errores de esta estimación. Es por esto que cuanto mejor sea esta información lateral (mas correlada esté con la imagen a estimar) menos bits se requerirán para codificar dicha imagen (menos errores habrá que corregir).

En el codificador se utilizan turbocódigos sistemáticos para generar bits de paridad de las imágenes WZ. La parte sistemática se descarta y es substituida en el decodificador por la información lateral. El ratio de compresión deseado se obtiene perforando los bits de paridad, y si éste no fuera necesario, el decodificador puede solicitar más bits utilizando el canal de retorno.

Las técnicas descritas a continuación tratan solamente sobre la generación de la información lateral. Primero se resumen dos técnicas básicas y luego se describe la técnica que ha sido el principal objeto de este estudio.

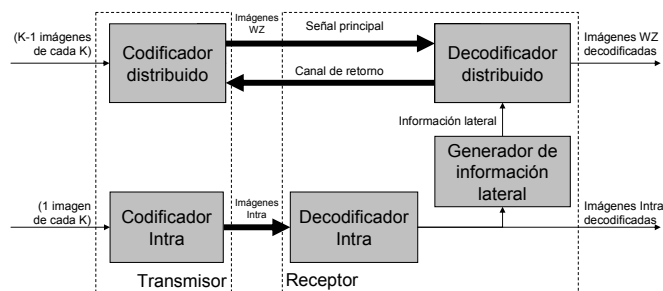


Fig. 3. Esquema general del sistema.

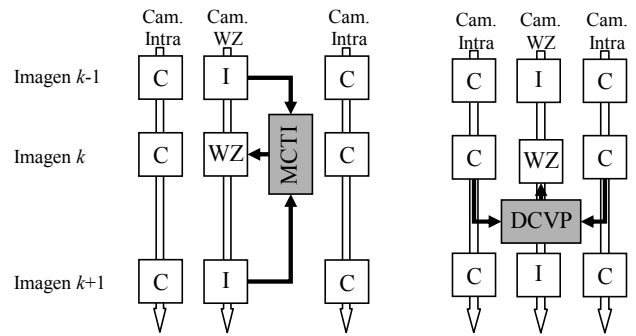


Fig. 4. Técnicas básicas.

### IV. GENERACIÓN DE LA INFORMACIÓN LATERAL

#### A. Técnicas básicas

Las dos técnicas más simples son la interpolación puramente temporal (MCTI, *Motion Compensated Temporal Interpolation*) y la puramente espacial o Inter-cámara (DCVP, *Disparity-Compensated Temporal Interpolation*), mostradas en la Fig. 4. MCTI utiliza interpolación por compensación de movimiento basada en bloques, y usa como imágenes de referencia las temporalmente adyacentes de la misma cámara WZ. DCVP emplea el mismo método, pero usa como imágenes de referencia las correspondientes al mismo instante temporal de las cámaras Intra. Cada una de estas técnicas es apropiada para un tipo de secuencias y no para otras. Por ejemplo, MCTI falla cuando hay mucho movimiento, y DCVP falla cuando hay oclusiones, reflejos o diferencias en la calibración de las cámaras.

Con tal de encontrar una técnica que funcionara en un mayor número de secuencias se realizó el trabajo descrito a continuación.

#### B. Estimación de movimiento multicámara (MVME)

La idea principal de la MVME (*Multi-View Motion Estimation*) se puede ver en la Fig. 5, y consiste en que los vectores de movimiento se encuentran en una de las cámaras Intra y se aplican luego en la cámara WZ para generar la estimación.

Los detalles se muestran en la Fig. 6. En primer lugar, se examina la relación entre las dos cámaras encontrando los vectores de disparidad. Estos vectores juegan el mismo papel que los vectores de movimiento: relacionan cada bloque de la cámara WZ con el bloque más similar en la cámara Intra utilizando *block matching*. Esto es parecido a lo que realiza DCVP.

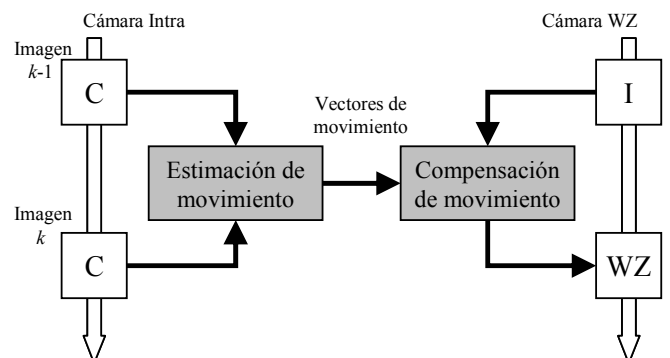


Fig. 5. Idea principal de la estimación de movimiento multicámara.

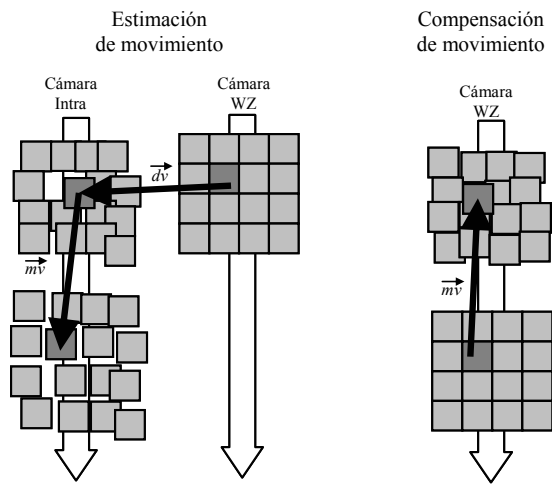


Fig. 6. Vista detallada de la técnica MVME.  $\vec{mv}$  y  $\vec{dv}$  son los vectores de movimiento y de disparidad respectivamente.

En segundo lugar cada bloque encontrado en la cámara Intra es de nuevo analizado para encontrar el más parecido, pero esta vez en una imagen temporalmente adyacente (tal como se haría en compensación de movimiento tradicional). En este paso se encuentran los vectores de movimiento. Finalmente estos vectores de movimiento encontrados en la cámara Intra se aplican a la imagen WZ para generar la información lateral.

Hay que remarcar que esta técnica funciona siempre que todas las cámaras se encuentren sobre el mismo plano y miren en la misma dirección. De otra forma, los vectores de movimiento encontrados en una cámara deberían ser transformados teniendo en cuenta las matrices de calibración antes de ser utilizados en otra cámara diferente.

Finalmente, la técnica descrita funciona para una sola cámara Intra y una sola imagen de referencia (temporalmente adyacente). En el escenario descrito en la sección II se dispone de 2 cámaras Intra, y cada una proporciona 2 imágenes de referencia, por lo tanto se pueden generar 4 estimaciones diferentes, tal como muestra la Fig. 7. Actualmente estas 4 estimaciones se están promediando, pero se están investigando fusiones más complejas.

## V. RESULTADOS

Para comprobar el rendimiento de las técnicas descritas se han simulado 3 secuencias, usando el escenario de la sección II. Los detalles de las secuencias están en la Tabla 1.

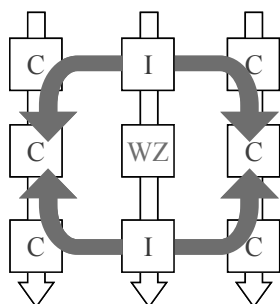


Fig. 7. 4 caminos diferentes, que producen 4 estimaciones diferentes.

Secuencia	Tamaño (píxeles)	Número de imágenes	Hz	Contenido
Breakdancers	256x192	100	15	Movimiento rápido
Exit	192x144	250	25	Movimiento lento
Uli	256x192	250	25	Gran disparidad

Tabla 1. Detalles de las secuencias simuladas.

En la Fig. 8 se muestran los resultados finales de la simulación de un codec DVC completo<sup>1</sup> tras utilizar las diferentes técnicas de generación de información lateral descritas. También se incluyen los resultados de utilizar el codec H.264 en modo Intra (sin codificación *Inter-frame*), para comparar con un codificador de complejidad parecida.

Se puede observar que solo para la secuencia con mucho movimiento la técnica estudiada (MVME) supera a la MCTI convencional. Para las otras dos secuencias la información de las cámaras Intra tiene que ser tratada de una forma más evolucionada para que sea útil a la cámara WZ.

En todos los casos se observa un rendimiento muy negativo de la técnica DCVP. Esto es debido a que en general las diferentes imágenes de una cámara están mucho más correladas entre ellas que con las de otras cámaras.

Finalmente, solo en las secuencias con movimiento lento los sistemas basados en DVC superan a una codificación tradicional simple (H.264 Intra). En estos casos, la interpolación temporal que los sistemas DVC utilizan en el decodificador les permite generar buenas estimaciones de las imágenes WZ. Cuando hay movimiento muy rápido (como en la secuencia *Breakdancers*, el mejor sistema es la codificación independiente de cada imagen (H.264 Intra).

A título ilustrativo se muestran ejemplos de informaciones laterales obtenidas con los diferentes métodos en la Fig. 9, junto con sus respectivas PSNR. Para esta secuencia se observa como la técnica estudiada (MVME) produce mejores estimaciones en términos de PSNR que las interpolaciones puras MCTI o DCVP. El emborronamiento que se observa en la imagen MVME se debe al promediado de los 4 caminos descritos anteriormente. A continuación se proponen algunas posibles soluciones junto con las conclusiones.

## VI. CONCLUSIONES

Se ha presentado una técnica para generar la información lateral que requieren los sistemas basados en codificación distribuida de vídeo, para el escenario multicámara. Al ser un método más evolucionado y que utiliza más información que las interpolaciones convencionales que se han venido utilizando hasta ahora es de esperar un mayor rendimiento en términos de *rate-distortion*. Esto solo ha sido así en una de las tres secuencias analizadas.

Esta técnica está aun en una fase muy temprana de su desarrollo y quedan aun muchas opciones por investigar, por lo que es de esperar que se obtengan mejores rendimientos en el futuro. A continuación se listan algunas de estas opciones.

En vez de encontrar los vectores de movimiento en la cámara Intra y aplicarlos a la cámara Wyner-Ziv, se pueden encontrar los vectores de disparidad en el instante de tiempo anterior y aplicarlos al actual. Esto proporcionaría 4 caminos

<sup>1</sup> Este software se llama DISCOVER-codec y está protegido por copyright por el proyecto "Distributed coding for video services" (DISCOVER), FP6-2002-IST-C contrato no.: 015314 de la Comisión Europea. No puede ser copiado, reproducido ni distribuido sin el consentimiento del consorcio. El software DISCOVER se originó del software llamado IST-WZ desarrollado en el Grupo de Imagen del Instituto Superior Técnico (IST), Lisboa-Portugal (<http://amalia.img.lx.it.pt/>), por Catarina Brites, João Ascenso y Fernando Pereira.

más (4 estimaciones más) con las que trabajar, haciendo un total de 8.

En vez de promediar las 4 estimaciones obtenidas se puede hacer una fusión mas evolucionada, como por ejemplo, desarrollar una medida de la confianza que se tiene en cada uno de los caminos y seleccionar el más fiable para cada bloque o píxel, o hacer un promedio ponderado por esta confianza.

#### AGRADECIMIENTOS

Este trabajo ha sido financiado por el proyecto TEC2005-07751-C02-02 del Ministerio de Educación y Ciencia y por el programa IST FP6 de la Comisión Europea.

#### REFERENCIAS

- [1] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources", IEEE Trans. Inform. Theory, vol. 19 pp. 471-480, July 1973.
- [2] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder", IEEE Trans. Inform. Theory, vol. 22, pp. 1-10, January 1976.
- [3] R. Puri and K. Ramchandran. "PRISM: A new robust video coding architecture based on distributed compression principles". Proc. of 40th Allerton Conf. on Comm., Control, and Computing, Allerton, IL, Oct. 2002.
- [4] B. Girod, A. Aaron, S. Rane and D. Rebollo-Monedero, "Distributed video coding", Proc. of the IEEE, vol. 93, no. 1, January 2005
- [5] Jens-Rainer Ohm, "Stereo/Multiview Video Encoding Using the MPEG Family of Standards", Invited Paper, Electronic Imaging '99, San Diego, Jan. 1999
- [6] Heung-Yeung Shum and Sing Bing Kang. "A Review of Image-based Rendering Techniques", IEEE/SPIE Visual Communications and Image Processing (VCIP) 2000, pp. 2-13, Perth, June 2000
- [7] G. Toffetti, M. Tagliasacchi, M. Marcon, A. Sarti, S. Tubaro, K. Ramchandran, "Image Compression in a Multi-camera System based on a Distributed Source Coding Approach". European Signal Processing Conference, Antalya, September 2005
- [8] X. Zhu, A. Aaron and B. Girod, "Distributed compression for large camera arrays", Proc. IEEE Workshop on Statistical Signal Processing, SSP-2003, St Louis, Missouri, Sept. 2003.
- [9] X. Guo, Y. Lu, F. Wu, W. Gao, S. Li, "Distributed multiview video coding", Proceedings of SPIE, January 2006, San Jose, California, USA. Vol. #6077.
- [10] M. Ouaret, F. Dufaux, T. Ebrahimi, "Fusion-based multiview distributed video coding", ACM International Workshop on Video Surveillance and Sensor Networks, Santa Barbara, CA, USA October 27, 2006

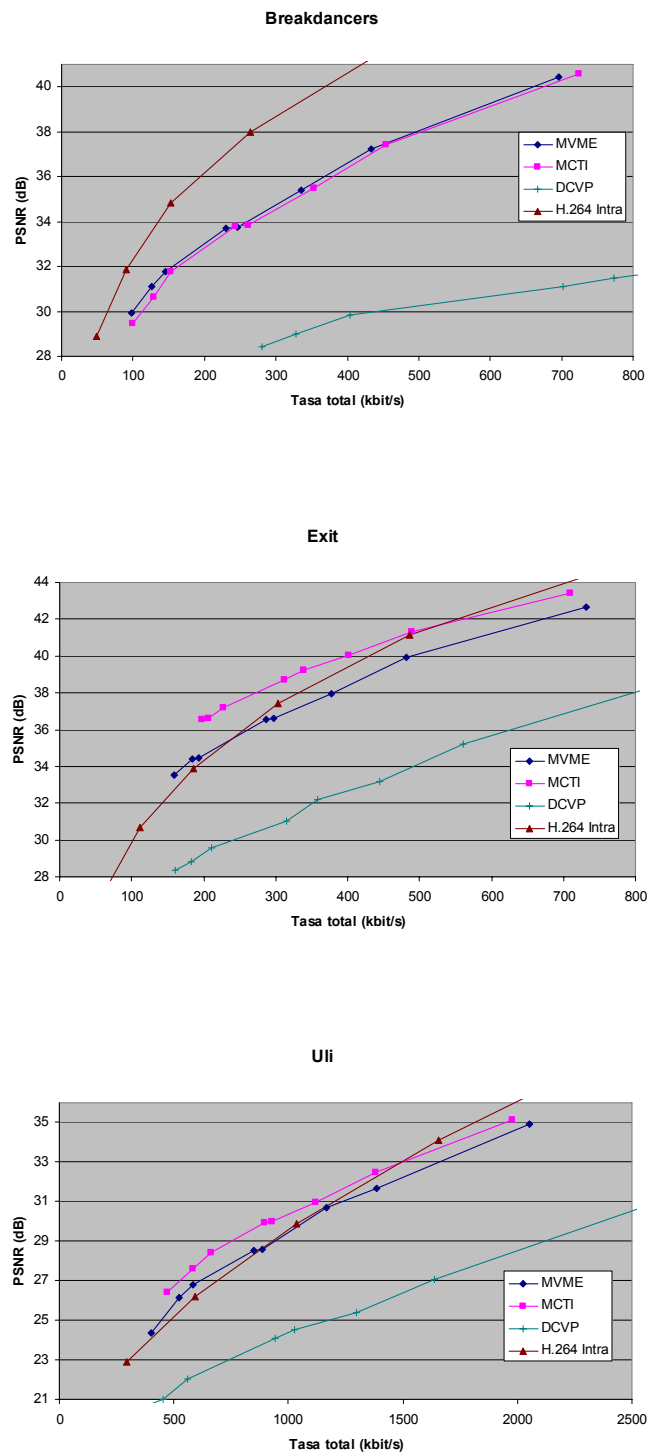


Fig. 8. Resultados para las 3 secuencias estudiadas.



Fig. 9. Informaciones laterales obtenidas con los diferentes métodos presentados para la primera imagen de la secuencia Breakdancers. Izquierda: DCVP (23.3 dB de PSNR), Centro: MCTI (26.1 dB de PSNR), Derecha: MVME (29.6 dB de PSNR).