

Un algoritmo de desentrelazado adaptativo con la repetición de imágenes basado en lógica difusa

Piedad Brox^{1,2}, Iluminada Baturone^{1,2}, and Santiago Sánchez-Solano^{1}

¹Instituto de Microelectrónica de Sevilla (CNM-CSIC)

²Departamento de Electrónica y Electromagnetismo. Universidad de Sevilla (España)

Email: {brox,lumi,santiago}@imse.cnm.es

Teléfono: (034) 954466666

Fax: (034) 954466669

Abstract—Esta comunicación presenta un interpolador temporal basado en lógica difusa que es utilizado para el desentrelazado de la señal de vídeo. El objetivo es que dicho interpolador sea capaz de aprovechar una característica que cada vez es más frecuente en las secuencias, debido a los estándares de conversión entre distintos formatos de transmisión, como es la repetición de imágenes. Este nuevo interpolador es incluido en un algoritmo adaptativo en función del movimiento, obteniendo una técnica de desentrelazado muy competitiva frente a otros algoritmos que son actualmente utilizados en ASICs de dispositivos comerciales de altas prestaciones. Los resultados de simulación obtenidos al desentrelazar secuencias de distintos materiales (film, vídeo e híbrido) muestran la superioridad del algoritmo propuesto.

I. INTRODUCCIÓN

Cada vez con más frecuencia existe la necesidad de conversión entre los distintos formatos de vídeo. La técnica simple y eficaz de aumentar la frecuencia de la secuencia, es decir, el número de fotogramas por segundo, es repetir fotogramas con la cadencia necesaria. Tras el proceso de conversión, la señal es transmitida pero la cadencia empleada en la conversión no es una información que se incluya en la transmisión. Por tanto, en la recepción de la señal no es inmediata la detección de imágenes repetidas.

El conocimiento de posibles patrones de repetición de fotogramas sería muy útil para facilitar algunas tareas de procesado de la señal de vídeo. Particularmente, el desentrelazado podría llevarse a cabo consiguiendo un resultado perfecto si la imagen repetida es correctamente detectada. Por este motivo, los ASICs de procesado de vídeo incorporados en los productos de gama alta, como pueden ser displays de LCD o plasma, DVDs o proyectores de alta definición, incluyen algoritmos de desentrelazado que son capaces de adaptar la estrategia de interpolación a la cadencia de repetición de imágenes, si es que ésta existe [1]-[3].

En el estado del arte de estos algoritmos capaces de detectar secuencias de repetición se encuentran dos grandes líneas de investigación. La primera de ellas es el desarrollo de algoritmos que son capaces de detectar cualquier patrón de repetición, sin necesidad de limitarse a la búsqueda de una cadencia definida por los estándares de conversión de un formato a otro [4]-[5]. La segunda se centra en un procesado de la imagen local, de modo que es posible detectar varios patrones de repetición dentro de una misma imagen. Esta

última línea de investigación es muy interesante para las secuencias de televisión debido al aumento de material híbrido, es decir material que proviene de distintas fuentes de grabación con distintos formatos [6]-[7].

El objetivo del trabajo de investigación que se presenta en esta comunicación es el desarrollo de un algoritmo de desentrelazado que sea adaptativo con la detección de imágenes o partes de la imagen repetidas. Para ello vamos a utilizar el algoritmo adaptativo con el grado de movimiento que desarrollamos en [8], y vamos a diseñar un interpolador temporal ' I_T ' que sea capaz de adaptar la interpolación del píxel en función de si existe o no una cadencia de repetición. Puesto que se realiza un procesado local de cada píxel de la imagen, se obtendrá un algoritmo muy competitivo frente a los implementados actualmente en dispositivos de altas prestaciones. Nuestro algoritmo realiza un desentrelazado adaptativo con la presencia local de movimiento, adaptativo con la presencia local de bordes, y adaptativo con la presencia de patrones de repetición de la imagen o partes de la imagen puesto que se realiza un procesado local de cada píxel de la imagen. Además la detección no se limita a una cadencia determinada de repetición sino que en principio puede ser detectado cualquier patrón de repetición.

Esta comunicación se estructura en tres secciones. La sección II dedicada a la descripción del algoritmo, se subdivide a su vez en tres subapartados. En el primero de ellos se detallan algunos patrones de repetición, que son utilizados en estándares de formatos de conversión. A continuación, se describe muy brevemente el algoritmo adaptativo con el grado de movimiento que se presentó en [8]. Finalmente, se describe el interpolador temporal, ' I_T ', propuesto en esta comunicación. En la sección III se muestra la eficiencia del algoritmo propuesto (comparándose frente a otras técnicas) al desentrelazar varias secuencias de distintos formatos, incluyendo secuencias híbridas. Finalmente la conclusiones obtenidas se presentan en la sección IV.

II. DESCRIPCIÓN DEL ALGORITMO

A. Transferencia entre los formatos de film y vídeo: el proceso pull-down

El formato de filmación de cine (film) no es directamente compatible con ningún formato televisivo estándar. Por lo

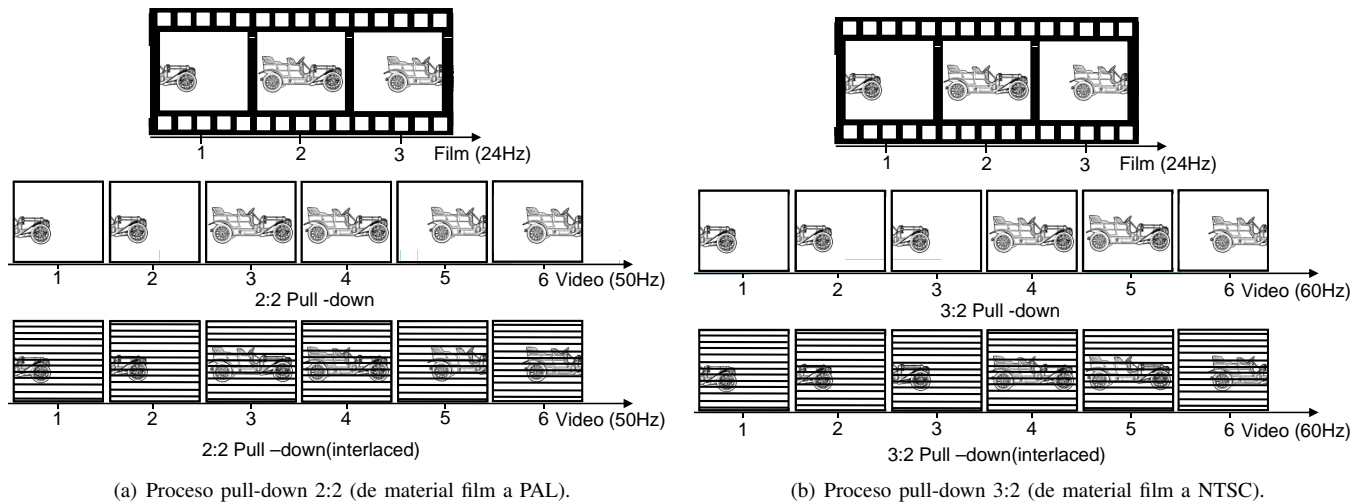


Fig. 1. Transferencia entre el formato film y los formatos NTSC y PAL.

tanto, el proceso de conversión a los estándares televisivos es un proceso necesario. El material film se caracteriza por tener 24 fotogramas progresivos que deben ser convertidos a 25-fotogramas/50-campos (PAL) ó 29.97-fotogramas/59.94-campos (NTSC), dependiendo del formato de emisión televisivo.

El cambio de frecuencia de 24 Hz en el material de film al de 50 Hz para el material de vídeo televisivo se denomina proceso pull-down 2:2 (ver Fig. 1(a)). Este proceso de conversión sería fácilmente implementable en caso de que la velocidad del material film fuera 25 fotogramas por segundo. Para lograr esto, la película originalmente fotografiada a 24 fotogramas/s tiene que pasarse un 4% más rápida para conseguir los 25 fotogramas por segundo. A continuación, cada fotograma de material film es escaneado dos veces, creando dos fotogramas. Si el estándar de transmisión requiere que la señal sea entrelazada, como ocurre en los estándares PAL y SECAM, cada fotograma se convierte en su correspondiente campo como puede observarse en la Fig. 1(a).

Para transferir 24 Hz de material film al formato televisivo NTSC, habría que pasar a 30 Hz pero esto no es deseable ya que causa una distorsión en la imagen y en el sonido que es considerado como inaceptable por el público en general. Una alternativa es reducir la frecuencia del material film de 24 Hz a 23.976 Hz, siendo este cambio de velocidad no apreciable por el espectador. Después de esto, los fotogramas son repetidos con una cadencia de repetición de tres y dos veces como se ilustra en la Fig. 1(b). Esto hace aumentar la frecuencia de la secuencia en un factor de 2.5, obteniéndose una secuencia de aproximadamente 60 campos por segundo. Finalmente y de acuerdo al estándar del formato de transmisión NTSC, la secuencia es entrelazada. Este método de conversión es denominado pull-down 3:2. La Fig. 1 ilustra estos dos procesos de conversión que son los más conocidos. Los modernos formatos de transmisión de televisión digitales también utilizan procesos de conversión similares.

En la Fig. 1 puede observarse cómo el proceso de desen-

trelazado sería perfecto en caso de que la imagen repetida sea correctamente detectada. Además su implementación es muy fácil ya que sólo consiste en ‘copiar’ las líneas no transmitidas del campo que es idéntico.

B. Algoritmo adaptativo con el grado de movimiento

La detección del grado de movimiento permite dirigir y optimizar el proceso de desentrelazado. Si no existe movimiento o es escaso, las líneas no transmitidas en un campo pueden obtenerse a partir de una técnica de interpolación temporal. Por el contrario, cuando existe movimiento la información temporal no es fiable, siendo preferible aplicar una interpolación espacial entre los píxeles del campo actual. La idea básica de los algoritmos adaptativos al grado de movimiento fue propuesta en [9], y se expresa matemáticamente mediante la expresión:

$$I_p(x, y, t) = (1 - \gamma)I_T(x, y, t) + \gamma I_S(x, y, t) \quad (1)$$

donde $I_p(x, y, t)$ representa la luminancia calculada para un determinado píxel, las variables x e y son las coordenadas espaciales del píxel en el fotograma, t indica el número de orden del fotograma en la secuencia, I_S e I_T representan los valores de luminancia obtenidos al utilizar técnicas de interpolación espacial y temporal, y γ indica el grado de movimiento mediante un valor comprendido entre 0 y 1.

El cálculo de γ y del interpolador espacial I_S fue descrito en [8], mientras que en esta comunicación se desarrolla un nuevo interpolador temporal, I_T , capaz de adaptar la interpolación temporal a la presencia de imágenes o partes de la imagen repetidas.

C. Descripción del interpolador temporal

La detección de los posibles patrones de repetición se realiza mediante el estudio de la evolución temporal del valor de luminancia del píxel analizado. Esta evolución temporal es analizada utilizando una ventana temporal de tres campos consecutivos. El uso de una ventana temporal más grande no

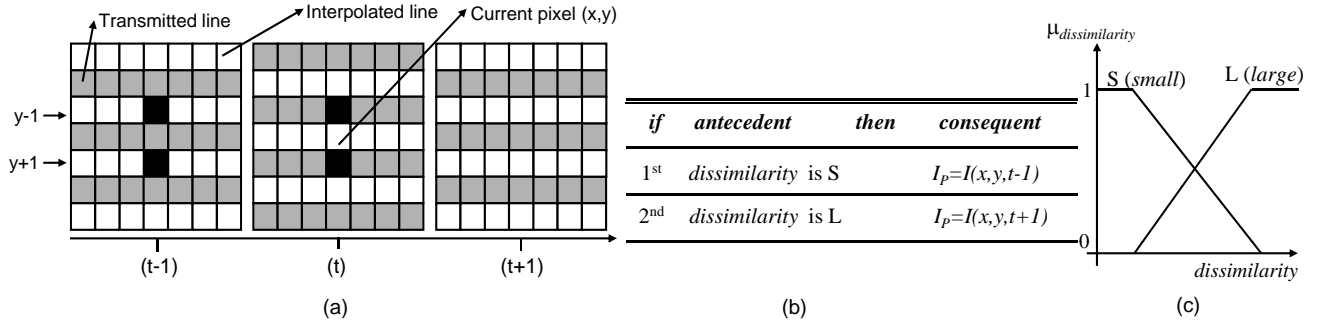


Fig. 2. (a) Ventana temporal utilizada para el cálculo del interpolador temporal. (b) Base de reglas del sistema difuso. (c) Conceptos difusos *SMALL*(S) y *LARGE*(L).

permitiría detectar correctamente algunos casos de conversión estándares. Ilustremos esto con el siguiente ejemplo. Imaginemos que queremos calcular las líneas no transmitidas (en negro) del campo 4 de la Fig. 1(b), y para ello estudiamos la evolución temporal del valor de luminancia de un píxel del campo 3 hasta el campo 6. Como puede observarse muchos de los píxeles que componen el perfil del coche han cambiado notablemente de posición espacial a lo largo de esta ventana temporal de cuatro campos. Esto conduciría al detector a pensar que en esta secuencia no existe repetición alguna de campos. Pero tal y como ilustra la Fig. 1(b) esta conclusión es totalmente errónea ya que los campos 4 y 5 son iguales.

El interpolador temporal evalúa la similitud entre dos de los tres campos que componen la ventana temporal del estudio mediante un simple sistema difuso, que es el encargado de determinar si el interpolador temporal debe utilizar el píxel del campo anterior, el píxel del campo posterior, o si debe realizar una interpolación entre ambos valores.

La entrada del sistema difuso es una medida de la disimilitud entre el primer y segundo campo de la ventana temporal (ver Fig. 2(a)). Matemáticamente se calcula como la convolución uni-dimensional de una matrix de pesos, C , y una matrix de diferencias de valores de luminancia, M :

$$dissimilarity(x, y, t) = \frac{\sum \sum M_{(i,j)} C_{(i,j)}}{\sum \sum C_{(i,j)}} \quad (2)$$

Las dimensiones y elementos de las matrices M y C se obtienen tras un estudio que se encuentra detallado en [10], siendo las siguientes:

$$C = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \quad (3)$$

$$M = \begin{pmatrix} M(x, y - 1, t) \\ M(x, y, t) \\ M(x, y + 1, t) \end{pmatrix} \quad (4)$$

Los valores de la matrix $M_{(i,j)}$ se calculan como:

$$M(x, y - 1, t) = \frac{|I(x, y - 1, t) - I_p(x, y - 1, t - 1)|}{2} \quad (5)$$

$$M(x, y, t) = \frac{|I(x, y, t + 1) - I(x, y, t - 1)|}{2} \quad (6)$$

$$M(x, y + 1, t) = \frac{|I(x, y + 1, t) - I_p(x, y + 1, t - 1)|}{2} \quad (7)$$

Para el cálculo de las expresiones (5) y (6) se necesita almacenar los valores de píxeles previamente calculados en el campo anterior (I_p). Esto implica que para poder calcular el segundo campo de la secuencia es necesario que el primero sea desentrelazado utilizando únicamente un método de interpolación espacial. Los valores de luminancia de los píxeles sombreados en negro en la Fig. 2(a) son los utilizados para evaluar la medida de disimilitud. La influencia de *dissimilarity* en la selección del interpolador temporal es evaluada considerando la siguiente base de reglas difusa:

- 1) Si la *dissimilarity* entre los campos $(t-1)$ y (t) es *SMALL* (S), el valor interpolado más adecuado se obtiene seleccionando el píxel en el campo anterior que tiene las mismas coordenadas espaciales ($I(x, y, t - 1)$).
- 2) Por el contrario, si la *dissimilarity* es *LARGE* (L), el píxel del campo anterior no es una buena opción y sería mejor apostar por el píxel del campo posterior ($I(x, y, t + 1)$).

La tabla de la Fig. 2(b) muestra de forma resumida la base de reglas de este simple sistema difuso. Los conceptos difusos de *SMALL* y *LARGE* son modelados con las funciones de pertenencia lineales a tramos que se muestran en la Fig. 2(c).

La salida del sistema difuso es la salida del interpolador temporal que viene dada por:

$$I_T = \chi_1 \cdot I(x, y, t - 1) + \chi_2 \cdot I(x, y, t + 1) \quad (8)$$

donde χ_i es el grado de activación de cada una de la reglas que se incluyen en la base de reglas del sistema difuso. La calidad de este interpolador es significativamente mejorada si se realiza un post-procesado de la señal *dissimilarity* utilizando operadores morfológicos tal y como se demuestra en [10].

III. RESULTADOS DE SIMULACIÓN

Las ventajas de incluir este nuevo interpolador temporal se han evaluado desentrelazando una serie de secuencias progresivas. Todas estas secuencias son en primer lugar entrelazadas 'artificialmente', eliminando las correspondientes líneas pares o impares de cada fotograma, y posteriormente, son desentrelazadas utilizando distintos algoritmos de desentrelazado. Para

TABLE I
PSNR RESULTS (IN DBS) FOR DIFFERENT DE-INTERLACING METHODS

	News	Mother	Paris	Salesman	Fargo 2	Bodyguard
Formato	QCIF	QCIF	CIF	CIF	PAL TV	PAL TV
Material	Video	Video	Video	Video	Film	Film
Duplicación de línea	25.18	31.81	23.61	29.75	28.79	31.45
Valor medio línea	29.25	35.94	26.67	35.04	34.31	35.61
ELA 3+3 [11]	26.63	35.39	25.53	32.11	33.66	35.08
ELA 5+5 [11]	25.92	34.2	24.64	30.17	32.16	33.93
Inserción campo	33.13	36.14	29.86	36.17	33.07	26.31
VT 2fields [11]	35.46	39.61	30.73	36.54	40.99	37.88
VT 3fields [11]	35.67	40.89	31.37	37.16	38.91	37.77
Median motion [11]	33.51	38.49	30.27	36.61	38.89	38.5
Van de Ville et al. [12]	34.73	38.49	33.12	37.62	40.11	39.05
MC Field insertion [11]	36.41	39.72	34.48	38.32	39.44	35.14
MC TBP [11]	36.2	39.49	34.29	38.16	39.48	34.89
MC TR [11]	36.16	39.51	34.21	38.01	38.97	33.19
Propuesta en [8]	38.25	41.36	35.68	38.49	42.29	40.61
Propuesta	39.91	42.37	36.91	38.94	43.65	41.11

evaluar la bondad de cada algoritmo se evalúa píxel a píxel el error cuadrático medio entre el valor de luminancia obtenido, I_p , y el valor original de la secuencia progresiva ($I_{original}$):

$$MSE = \frac{1}{M \times N} \sum_{i=1}^{M \times N} (I_p - I_{original})^2 \quad (9)$$

Otro valor cuantitativo que normalmente se utiliza es el PSNR, que se define a partir del MSE como sigue:

$$PSNR = 20 \log \frac{255}{\sqrt{MSE}} \quad (10)$$

Los resultados de los valores medios de PSNR obtenidos tras desentrelazar seis secuencias se encuentran en la Tabla I. Como puede observarse cada secuencia ha sido clasificada atendiendo al formato de la misma y a su procedencia. Las secuencias de material de vídeo no contienen secuencia de repetición alguna. Por el contrario, las secuencias de material film han sido convertidas al formato PAL utilizando el proceso pull-down 2:2.

Los métodos de desentrelazado utilizados en la comparativa incluyen métodos muy simples lineales, que a su vez incluyen técnicas espaciales (duplicación de línea, valor medio de la línea, ELA 3+3 [11], ELA 5+5 [11]), técnicas temporales (inserción del campo anterior), y técnicas espacio-temporales (Vertico-Temporal (VT) con 2 y 3 campos [11]). También se incluyen dos métodos de movimiento adaptativo (median motion [11], y un algoritmo que también utiliza un sistema basado en lógica difusa para la evaluación de movimiento [12]). Finalmente, también se han evaluado tres métodos de compensación de movimiento descritos en [11] (MC de

inserción de campo o field insertion, MC Temporal Backward Projection (TBP), y MC Time Recursive (TR)). Todos estos métodos son ampliamente conocidos y algunos de ellos son actualmente utilizados en ASICs de dispositivos comerciales.

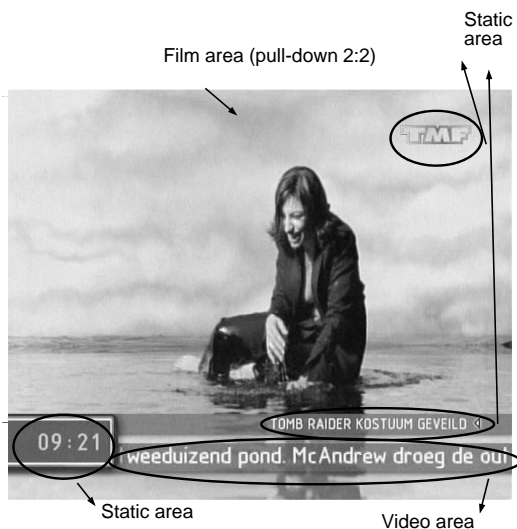
Como puede apreciarse en los valores de la Tabla I, el método propuesto mejora notablemente todas las técnicas de desentrelazado evaluadas. Las mejoras son casi siempre muy superiores a 0.5 dBs ¹, por lo que las mejoras deben de ser apreciadas cualitativamente al ser mostradas las imágenes desentrelazadas. Para ello hemos seleccionado un par de secuencias híbridas, es decir, secuencias que en un mismo fotograma contienen material de vídeo y film ² (ver Fig. 3). Las mejoras respecto a nuestra versión previa del algoritmo, presentado en [8], son también evidentes. La inclusión del nuevo interpolador temporal mejora todas las secuencias pero son particularmente significativas las mejoras en las secuencias de material film.

La imágenes desentrelazadas se muestran en las Fig. 4, 5, y 6. En Fig. 4 se muestra una zona de vídeo de la secuencia TMF donde nuestra propuesta se compara con algunas de la técnicas de desentrelazado incluidas en la Tabla I. La reconstrucción de las letras realizada por nuestra propuesta es bastante superior como puede verse en Fig. 4.

La secuencia híbrida Donna es mostrada en Fig. 5 y 6. En concreto, la Fig. 5 muestra una zona de la imagen que es de material film tras ser desentrelazada con nuestra propuesta y las técnicas de compensación de movimiento. Finalmente,

¹Este valor se establece como barrera por encima de la cual se obtiene una mejora de la imagen significativa a simple vista.

²Estas secuencias no se encuentran incluidas en la Tabla I porque sólo se dispone del material entrelazado



(a) Secuencia TMF.



(b) Secuencia Donna.

Fig. 3. Secuencias de material híbrido procesadas.



Fig. 4. Fotograma desentrelazado con distintos métodos de la secuencia híbrida TMF.

la Fig. 6 muestra otra zona del fotograma desentrelazado de la secuencia de Donna que corresponde a material de vídeo. Nuevamente nuestra propuesta es comparada con las técnicas

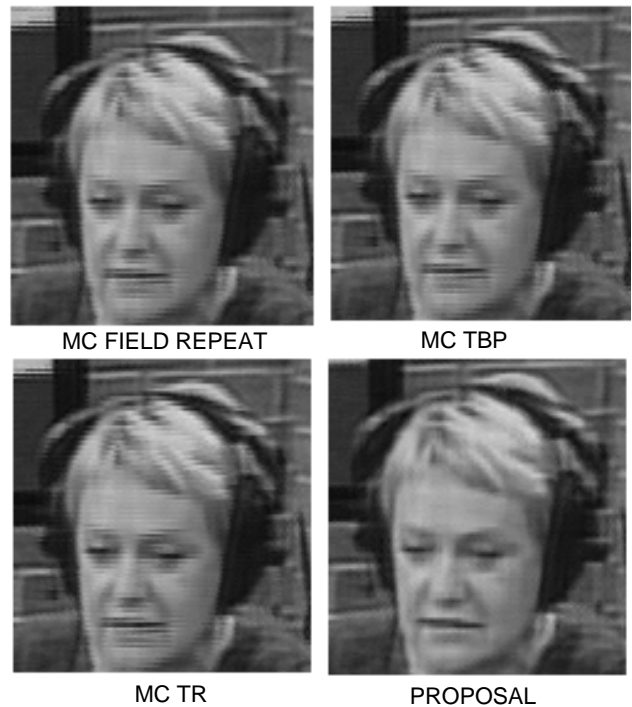


Fig. 5. Zoom de un área de material film de la secuencia híbrida Donna desentrelazada con métodos MC.

de compensación de movimiento.

IV. CONCLUSIONES

La repetición de imágenes en una secuencia es una técnica simple y eficaz utilizada para convertir un estándar de transmisión de una frecuencia determinada a otro estándar de frecuencia superior. Como resultado se obtiene que una misma imagen aparece repetida en la secuencia. Esto puede explotarse

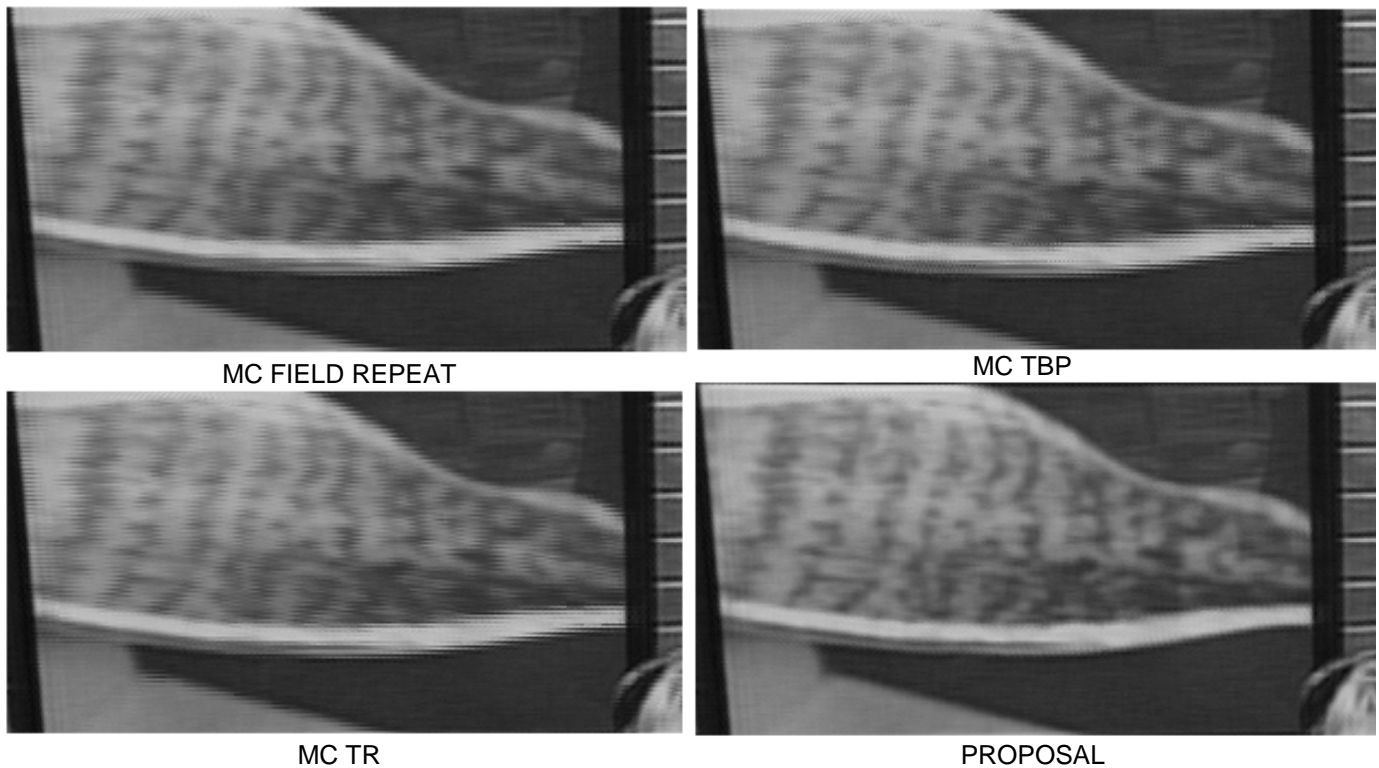


Fig. 6. Zoom de un área de material de vídeo de la secuencia híbrida Donna desentrelazada con métodos MC.

para mejorar el rendimiento de los algoritmos de desentrelazado, ya que la información de las líneas no transmitidas puede estar completamente disponible en el campo anterior o posterior de la secuencia. El interpolador temporal desarrollado, que se basa en esta idea, da lugar a un desentrelazado que es superior al que se obtiene con otras técnicas, algunas de ellas mucho más complejas.

AGRADECIMIENTOS

Este trabajo ha sido financiado parcialmente por los proyectos de investigación TEC2008-04920 del Ministerio de Ciencia e Innovación del Gobierno de España y P08-TIC-03674 de la Junta de Andalucía. Los autores quieren mostrar su agradecimiento a 'Philips Research Eindhoven' por proporcionar las secuencias híbridas analizadas.

REFERENCIAS

- [1] <http://www.anchorbaytech.com>
- [2] <http://www.hqv.com>
- [3] <http://www.tridentmicro.com>
- [4] P. Swan, *System and method for reconstructing noninterlaced captured content for display on a progressive screen*. Assignee: ATE Technologies, Inc. Thornhill, Canada, Apr. 25, 2000. United States Patent Office US 6,055,018.
- [5] G. de Haan, H. Huijgen, P. Biezen, and O. Ojo, *Method and apparatus for discriminating between movie film and non-movie film and generating a picture signal processing mode control signal*. Assignee: U.S. Philips Corporation, New York, USA, Nov. 15, 1994. United States Patent Office US 5,365,280.
- [6] G. de Haan, and R. B. Wittebrood, *Recognizing film and video object occurring in parallel in single television signal fields*. Assignee: Koninklijke Philips Electronics N. V., Eindhoven, NL, Aug. 2005. United States Patent Office US 6,937,655.

- [7] L. He, and H. Zhang, *Motion object video on film detection and adaptive de-interlace method based on fuzzy logic*. Assignee: nDSP Corporation, Campbell, CA, Sep. 258, 2004. United States Patent Office US 6,799,168.
- [8] P. Brox, I. Baturone, and S. Sánchez-Solano, *Aplicación de técnicas de interpolación basadas en lógica difusa al procesamiento de imágenes de vídeo*. Proc. of the Workshop IBERCHIP, no. 13, pp. 239-242, 2007.
- [9] A. M. Bock, *Motion-adaptive standards conversion between formats of similar field rates*. Signal Processing: Image Communication, vol.6, no.3, pp.275-280, 1994.
- [10] P. Brox, I. Baturone, and S. Sánchez-Solano, *A Fuzzy Motion Adaptive De-interlacing Algorithm Capable of Detecting Field Repetition Patterns*. Proc. Int. Symposium on Intelligent Signal Processing (WISP2007). Alcalá de Henares (Spain), Oct. 2007.
- [11] G. de Haan and E. B. Bellers, *De-interlacing - an overview*. Proc. of the IEEE, vol. 86, no. 9, pp. 1839-1857, Sep. 1998.
- [12] D. Van de Ville W. Philips, and I. Lemahieu, *Fuzzy-based motion detection and its application to de-interlacing*. Fuzzy techniques in image processing. Book Series of Studies in Fuzziness and Soft Computing, pp. 337-368, 2000.