A survey of hybrid MC/DPCM/DCT video coding distortions

Michael Yuen[a], H.R. Wu[b],*

[a] *ESS Technology Inc, 48401 Fremont Blvd, Fremont, CA 94538, USA*
[b] *School of Computer Science and Software Engineering, Monash University, Clayton, Victoria 3168, Australia*

Abstract

The motion-compensated hybrid DCT/DPCM algorithm has been successfully adopted in various video coding standards, such as H.261, H.263, MPEG-1 and MPEG-2. However, its robustness is challenged in the face of an inadequate bit allocation, either globally for the whole video sequence, or locally as a result of an inappropriate distribution of the available bits. In either of these situations, the trade-off between quality and the availability of bits results in a deterioration in the quality of the decoded video sequence, both in terms of the loss of information and the introduction of coding artifacts. These distortions are an important factor in the fields of filtering, codec design, and the search for objective psychovisual-based quality metrics; therefore, this paper presents a comprehensive analysis and classification of the numerous coding artifacts which are introduced into the reconstructed video sequence through the use of the hybrid MC/DPCM/DCT video coding algorithm. Artifacts which have already been briefly described in the literature, such as the blocking effect, ringing, the mosquito effect, MC mismatch, blurring, and color bleeding, will be comprehensively analyzed. Additionally, we will present artifacts with unique properties which have not been previously identified in the literature. © 1998 Elsevier Science B.V. All rights reserved.

Zusammenfassung

Der bewegungskompensierte hybride DCT/DPCM-Algorithmus wurde mit Erfolg in diversen Videocodierungs-standards (z.B. H.261, H.263, MPEG-1 und MPEG-2) eingesetzt. Die Robustheit dieses Algorithmus wird jedoch durch eine inadäquate Bitzuteilung beeinträchtigt – entweder global für die gesamte Videosequenz oder lokal durch eine ungeeignete Verteilung der verfügbaren Bits. In beiden Situationen ergibt der Gegensatz zwischen Qualität und der Verfügbarkeit von Bits eine Verschlechterung der Qualität der decodierten Videosequenz, sowohl hinsichtlich eines Informationsverlustes als auch der Einfügung von Codierungsartefakten. Diese Verzerrungen sind ein wichtiger Faktor bei der Filterung und beim Codec-Entwurf sowie bei der Suche nach objektiven, psychovisuell basierten Qualitäts-metriken. In der vorliegenden Arbeit wird deshalb eine umfassende Analyse und Klassifikation der zahlreichen Codierungsartefakte durchgeführt, die durch die Verwendung des hybriden MC/DPCM/DCT-Videocodierungsalgorith-mus in der rekonstruierten Videosequenz verursacht werden. Artefakte, die bereits kurz in der Literatur beschrieben wurden – z.B. Blockeffekt, Klingeln, Moskitoeffekt, MC-Fehlanpassung, Verschmierung und Farbbluten – werden umfassend analysiert. Zusätzlich werden Artefakte mit speziellen Eigenschaften vorgestellt, die bisher noch nicht in der Literatur beschrieben wurden. © 1998 Elsevier Science B.V. All rights reserved.

---

*Corresponding author. Tel.: + 1 613 9905 3255; e-mail: hrw@dgs.monash.edu.au.

Résumé

L'algorithme hybride DCT/PCM à compensation de mouvement a été adopté avec succès dans plusieurs standards de codage vidéo tels que H.261, H.263, MPEG-1 et MPEG-2. Toutefois, sa robustesse est mise à l'épreuve en cas d'allocation inadéquate des bits, soit globalement pour la séquence entière, soit localement comme résultat d'une distribution inappropriée des bits disponibles. Dans l'une ou l'autre situation, le compromis entre qualité et disponibilité des bits a pour résultat une détérioration de la qualité de la séquence vidéo décodée, à la fois en termes de perte d'information et d'introduction d'artefacts de codage. Ces distortions sont un facteur important dans les domaines du filtrage, de la conception du codec, et de la recherche de métriques de qualité objectives basées sur des concepts psycho-visuels; de ce fait, cet article présente une analyse approfondie et une classification des nombreux artefacts de codage introduits dans la séquence vidéo reconstruite en utilisant l'algorithme de codage hybride MC/DPCM/DCT. Les artefacts qui ont déjà été brièvement décrits dans la littérature, tels que l'effet de bloc, le tremblement, l'effet moustique, le mésalignement MC, le flou, et le bavage des couleurs sont analysés en profondeur. Additionnellement, nous présentons des artefacts ayant des propriétés uniques et qui n'ont pas encore été identifiés dans la littérature. © 1998 Elsevier Science B.V. All rights reserved.

Keywords: Video coding distortions; Video coding; MPEG-1; Blocking effect; Ringing; Mosquito effect

1. Introduction

The hybrid video compression algorithm based on a motion compensated hybrid of temporal DPCM and the block DCT (hybrid MC/DPCM/DCT) has been adopted in many of the current international digital video coding standards, including H.261, H.263, MPEG-1 and MPEG-2 [13,14,29,30]. Evaluation and classification of coding artifacts, produced by the use of this hybrid video compression algorithm, become ever important in order to evaluate the performance of various video coding software and hardware products proliferating the telecommunications, entertainment, multimedia and consumer electronics markets. A comprehensive classification will also assist in the design of more effective adaptive quantization algorithms and coding mechanisms under the current constant bit-rate schemes, and, therefore, improve video codec performance. The classification of coding artifacts is of equal importance in the evaluation and minimization of these artifacts to achieve constant 'quality' video compression when the above hybrid video compression algorithm is applied.

Whilst there have been numerous discussions in the literature concerning the various image and video coding distortions related to the components in the hybrid MC/DPCM/DCT algorithm [17,22], a comprehensive study of the full range of coding artifacts remains to be seen. A brief summary of a number of the artifacts was provided in [25], and more recently, in an effort to standardize terms and definitions in the area of quality evaluation, the American National Standards Institute (ANSI) in [3] has also provided a brief descriptive list, and video examples, of impairments in reconstructed sequences. However, Plompen [24] has produced the best attempt with regard to describing the cause and effect of reconstruction distortions.

In this contribution, we present a comprehensive characterization of the numerous coding artifacts which are introduced into reconstructed video sequences through the use of the hybrid MC/DPCM/DCT video coding algorithm. The isolation of the individual artifacts, which exhibit consistent identifying attributes, will be conducted with the aim of obtaining descriptions of the artifacts' visual manifestations, causes and relationships. Additionally, the spatial and temporal characteristics of a video sequence that are susceptible to each artifact, and in which the artifacts are visually prominent, will be noted. To account for the adaptive quantization mechanisms and human visual system (HVS) masking properties utilized in practical coding applications, an MPEG-1 compliant codec, utilizing source content-based adaptive bit-allocation, will be employed as the origin of the coding distortions.

Due to the complexity of the HVS, which has not yet been satisfactorily modeled [6], the perceived distortion is not directly proportional to the

absolute quantization error, but is also subject to the local and global spatial, and temporal characteristics of the video sequence. As a consequence of these factors, at a local level it is not possible to provide a definitive value indicating the level of quantization that will induce any one artifact to a certain visual degree, or when an artifact is perceived to even exist (*just noticeable distortion*) [16]. Consequently, at a global level it is not possible to indicate a specific bit-rate at which any one artifact manifests. This is exacerbated by the different varieties of bit-allocation techniques that have been proposed which may, or may not, exploit the masking effects of the HVS.

Therefore, the following discussion will be limited to descriptions of the artifacts' visual manifestations, causes and relationships. We also describe the spatial and temporal characteristics of video sequences that are susceptible to each artifact and in which the artifacts are visually prominent. The characteristics of some of the artifacts may be directly attributable to aspects of the MPEG-1 or MPEG-2 coding processes (such as the use of bidirectional prediction); however, the overall discussion relates to the hybrid MC/DPCM/DCT algorithm in general. Where possible, frames affected by the artifacts have been extracted from MPEG-1 coded sequences and presented to illustrate the artifacts' visual effect; however, as noted by Clarke [6], a true visual appreciation of the artifacts can only be provided by viewing the affected sequences on a high quality video monitor.

2. Blocking effect

We define the blocking effect as the discontinuities found at the boundaries of adjacent blocks in a reconstructed frame [28,32], since it is a common practice to measure and reduce the blocking effect by only taking into account the pels at the block boundaries [2,19,35].[1] However, the term can be

generalized to mean an overall difference between adjacent blocks [24], but we have categorized these differences as individual artifacts. Examples of the blocking effect are shown in Fig. 1.

The cause of the blocking effect with respect to block-based coding is intuitively obvious, and has already been well documented [19,24,28]. In brief, due to the isolated nature in which individual blocks are coded, the level and characteristics of the coding error introduced into a block may differ from one block to another. This, in consequence, manifests as discontinuities between the boundaries of adjacent blocks: the blocking effect.

2.1. Intraframe coded blocks

The severity of the blocking effect is subject to the coarseness of the quantization of the DCT coefficients of either one or both of the adjacent 8×8-pel blocks. The threshold level of the quantization, in either of the blocks, above which the blocking effect is noticeable to the HVS, relies on the content of the blocks as well as the masking effects of the HVS. Generally, the effect is hidden in either the more spatially active areas, or the bright or very dark areas [10].

Since the blocking effect is more visible in the smoothly textured sections of a frame, the lower order DCT coefficients play the most significant role in determining the visibility of the blocking effect; this is especially true for the DC coefficient [22]. However, the blocking effect may occur in spatially active areas as a result of very coarse quantization. With the medium to higher order AC coefficients quantized to zero, an originally spatially active block will have a smoothly textured reconstruction which, in terms of the blocking effect, is subject to the same visibility concerns as originally smooth blocks.

2.2. Predictive coded blocks

The occurrence of the blocking effect in predictive coded macroblocks can be categorized into two different forms: one relating to the external boundary of the macroblock, and the other found

---

[1] Ramamurthy and Gersho [26] use the term 'grid noise' to describe these block-edge discontinuities. In [3], the blocking effect, the *DCT basis image artifact* (Section 3), and mosaic patterns (Section 8) are collectively defined under the term *block distortion/tiling*.

Fig. 1. Examples of the blocking effect; most evident in the smoothly textured regions which are of low-to-medium luminance.

between the four constituent luminance (8×8)-element blocks of the macroblock.

Since the MC prediction, generally, provides a good prediction of the lower-frequency information of a macroblock [21], the internal blocking effect, typically, does not occur for macroblocks with a smoothly textured content.[2] In this situation, the prediction error is minimal, so after quantization the prediction error is reduced to zero; if the prediction error for all the constituent blocks is quantized to zero, the internal blocking effect will not occur. The internal blocking effect mainly occurs in mildly textured areas where the prediction error is sufficiently large such that it is not quantized to zero. The combination of a relatively smooth prediction and

coarsely quantized prediction error results in visible discontinuities between the internal blocks. The areas within a frame which produce very large prediction errors usually contains high spatial activity, which masks the internal blocking effect.

Discontinuities induced between the internal blocks will result in discontinuities between adjacent macroblocks. However, the external blocking effect is most visibly significant around the borders of moving objects, and is a consequence of poor MC prediction, which is typical around moving areas in the recorded scene. This phenomenon will be discussed in more detail in Section 11. Suffice to say that the task of the MC prediction is made more difficult in these areas since it must produce a single motion vector for the whole macroblock where, in some situations, the constituent pels of a macroblock may have originated from separate objects within the scene with divergent motion. The result is a predicted macroblock that is ill-fitting

---

[2] The copying of the blocking effect to the current frame from the MC reference frames will be discussed in Section 10.
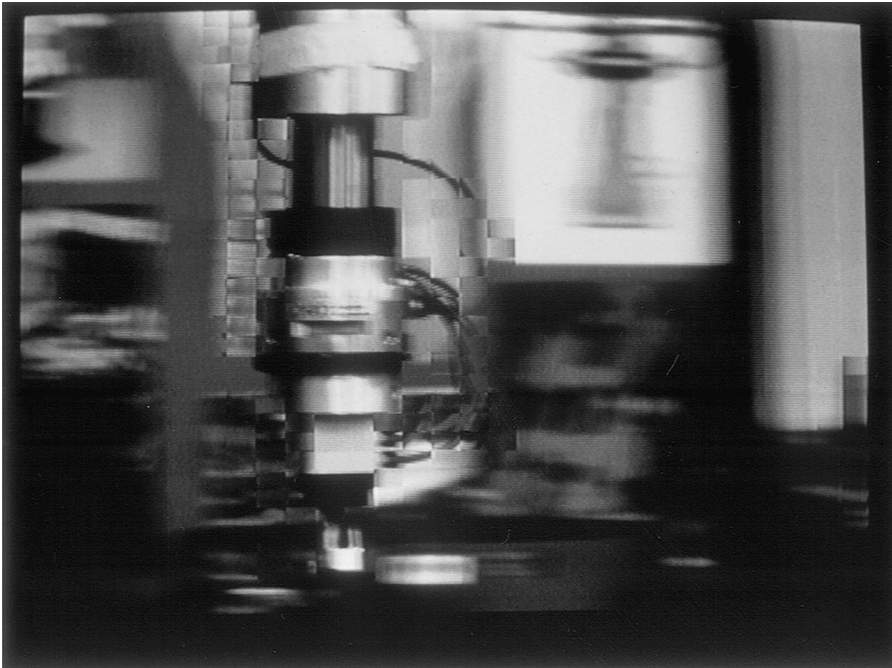
Fig. 2. Example of the blocking effect caused by block matching motion compensation, with a macroblock size of 16×16 pels; most evident in areas around the robot arm.

in one or more of the enclosed moving objects, with disparities occurring along the boundary of the macroblock. The contribution of the quantized prediction error to the reconstruction is, typically, a high-frequency noise-like effect, resulting from the presence of the edge(s) between the boundaries of the moving objects in the prediction, and the high energy of the prediction error needed to compensate for the poor prediction. This tends to mask the blocking effect, although it produces a temporal effect which will be discussed in Section 12.

The external block effect may also occur where the boundary edge of a macroblock coincides with the boundary of a moving object; the straight well-defined edge of the macroblock produces an unnatural visual sharpness to the object's boundary. Examples of the blocking effect on 16×16-pel macroblock boundaries, caused by the block matching motion compensation algorithm, are shown in Fig. 2.

3. DCT basis image effect

The visual prominence of the blocking effect is primarily a consequence of the regularity of the size and spacing of the block-edge discontinuities. Each of the DCT basis images have a distinctive regular horizontally or vertically oriented pattern which make them visually conspicuous [34]. This leads to the *DCT basis image effect* which manifests as blocks bearing a distinct likeness to one of the 63 AC DCT basis images. The regular pattern of the basis images, and their fixed size, makes them visually prominent; examples are given in Fig. 3.

The effect is caused by coarse quantization of the AC DCT coefficients in areas of high spatial activity within a frame, resulting in the nullification of the low-magnitude DCT coefficients which are within the quantization *dead-zone* [29]. In a situation where a single AC basis image is prominent in the representation of a block, the result after coarse
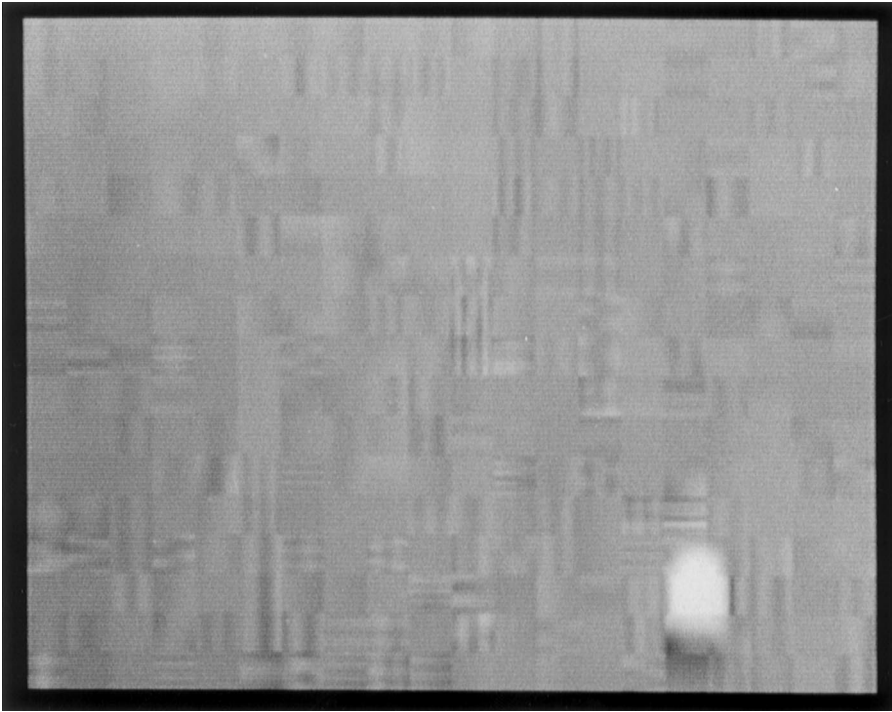
Fig. 3. Examples of the DCT basis image effect, extracted from the background of the *Table-tennis* sequence (see the example given for *ringing* in Section 7).

quantization is the reduction of all, except the most prominent, basis images to insignificance. This results in an emphasis of the pattern contributed by the prominent basis image, since the combined energy of the other AC basis images are insufficient to mute its contribution to the aggregated reconstruction.

The visual characteristics of the DCT basis images may also produce other coding artifacts when adjacent blocks are taken into consideration. For example, blocks suffering from the basis image effect invariably do not fit well, in terms of appearance, with the surrounding blocks; therefore, this results in the *mosaic pattern effect*, which will be described in Section 8. Additionally, for basis images which have a zero, or very low, frequency content in either the horizontal or vertical direction, the blocking effect may result along the sections of the block boundary which contain limited spatial activity. These basis images include those on

the top row and left column of Fig. 4, which have no vertical and horizontal activity, respectively.

3.1. Visual significance of each basis image

The MPEG-1 intra quantization weighting matrix was derived from experiments on the psychovisual thresholding of quantization distortion for each of the DCT basis images [12,29]. This thresholding relates to the ability of the HVS to discern changes of contrast in each of the basis images. The use of different weights for each of the basis images, and other research into the calculation of visually optimum quantization matrices [1,20,23,34], demonstrate that the HVS does not perceive each of the basis images with equal significance.

Therefore, the visual impact of the basis image effect is subject to the proportion of AC energy
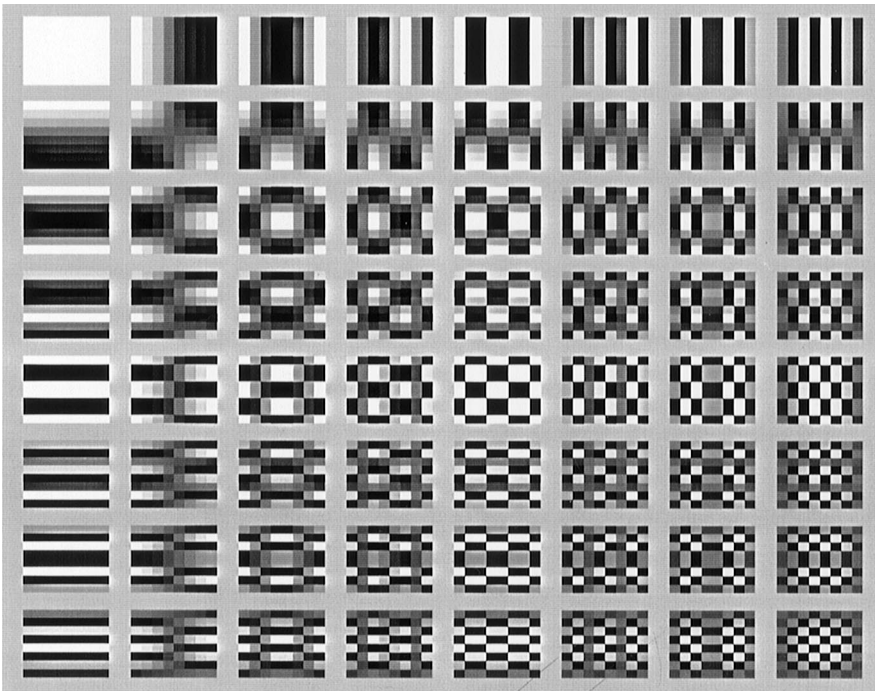
Fig. 4. 8 × 8 DCT basis images.

concentrated into any one coefficient, as well as the visual significance of the respective DCT basis image.

3.2. Predictive coded macroblocks

As with intraframe coded macroblocks, the basis image effect in predictive coded macroblocks occurs in high spatial activity areas. This is especially evident when the prediction offered by the MC reference contains little or no spatial detail. The attempt to compensate for the loss of high-frequency information with coarsely quantized prediction error may result in the emergence of a single AC basis image.

Similar to the mosaic pattern effect (see Section 8), if the same high spatial activity area is visible in a number of frames of an original sequence, the basis image effect will decrease over time in this area within the reconstruction. This is a consequence of the accumulation and refinement of higher-order AC component information for each succeeding P-type frame (MC reference frame) in the sequence.

3.3. Aggregation of major basis images

Although blocks containing the basis image effect resemble a single DCT basis image, it is possible that a significant proportion of the AC energy is concentrated into more than one AC coefficient. This is most evident when the primary basis images are of a similar directional orientation and frequency, resulting in an enhancement of parts of the contributed visual patterns. This will retain the frequency of the contributing basis images, as well as a perceived regularity in the aggregated pattern.

4. Blurring

Blurring manifests as a loss of spatial detail and a reduction in sharpness of edges in moderate to

Fig. 5. Example of blurring. *Table-tennis* sequence coded at 1 Mbps. Most evident around the net and the surface of the table.

high spatial activity regions of frames, such as in roughly textured areas or around scene object edges. Fig. 5 shows examples of blurring, taken from the *Table-tennis* sequence coded at 1 Mbps. The blurring is most evident in the area around the net as well as on the surface on the table.

For intraframe coded macroblocks, blurring is directly related to the suppression of the higher-order AC DCT coefficients through coarse quantization, leaving only the lower-order coefficients to represent the contents of a block; therefore, blurring can be directly associated with lowpass filtering. Similarly, in many respects, blurring, as a consequence of transform coding, can be considered as a specific case of the basis image effect where the prominent AC basis images after quantization are of a lower frequency, resulting in a reconstructed block with low spatial activity. Also similar to the basis image effect, the blurring of blocks in areas of high spatial activity may coincide with both the blocking effect and the mosaic pattern effect. The result of blurring of

the chrominance information will be discussed in Section 5.

For predictive coded macroblocks, blurring is mainly a consequence of the use of a predicted macroblock with a lack of spatial detail. However, blurring can also be induced in bidirectionally predicted macroblocks, where the interpolation of the backward and forward predictions results in an averaging of the contents of the final bidirectional prediction. In both these cases, the blurred details are supplemented by the prediction error which supplies some higher-frequency information to the reconstruction, thereby reducing the blurring effect.

5. Color bleeding

The blurring of the luminance information, as discussed in Section 4, results in the smoothing of spatial detail. The corresponding effect for the chrominance information results in a smearing of
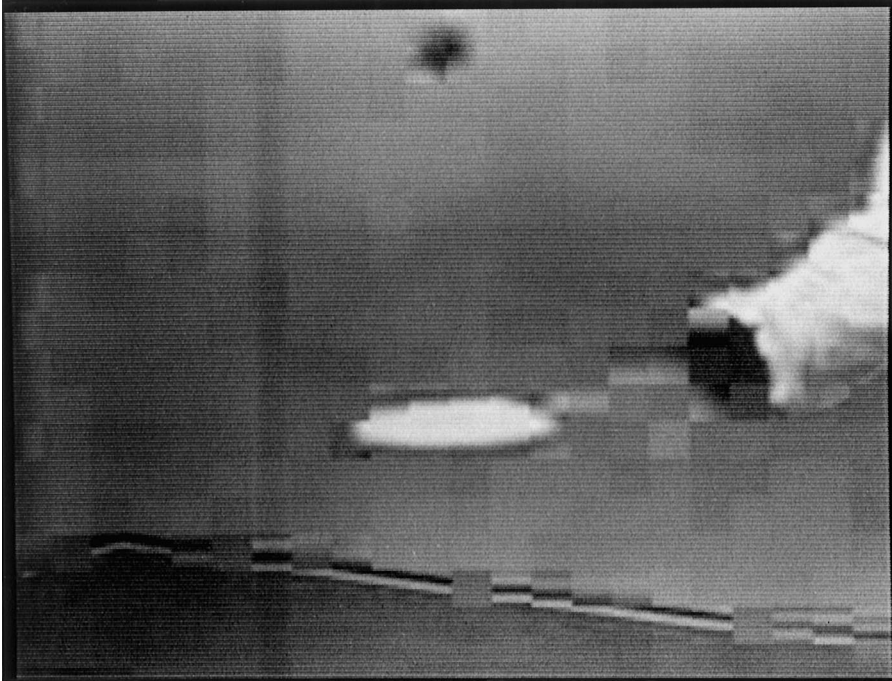
Fig. 6. The Cb chrominance component of a coded frame from the *Table-tennis* sequence.
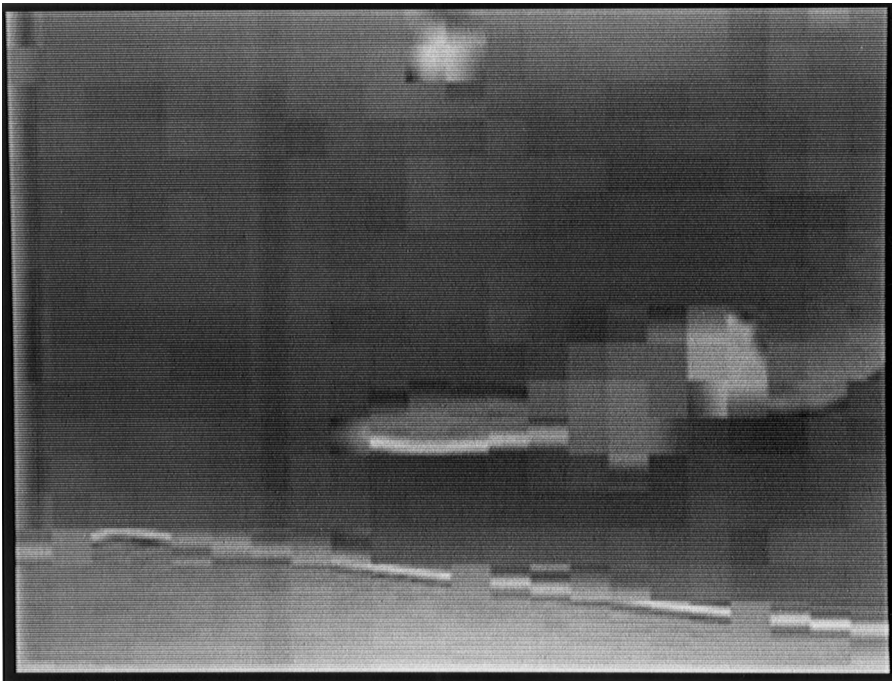

Fig. 7. The Cr chrominance component of a coded frame from the *Table-tennis* sequence.

the color between areas of strongly contrasting chrominance.

As with blurring, color bleeding results from the quantization to zero of the higher-order AC coefficients, resulting in the representation of the chrominance components with only the lower-frequency basis images. Since the chrominance information is subsampled, the bleeding is not limited to an 8×8-pel area, as for the luminance information, but extends to the boundary of the macroblock.

For chrominance edges of very high contrast, or where the quantization of the higher order AC coefficients does not result in their truncation, the color artifact corresponding to the *ringing effect* occurs. This will be discussed in Section 7.

Figs. 6 and 7 show the Cb and Cr chrominance components, respectively, of an I-type frame from the *Table-tennis* sequence, coded at 0.6 Mbps. The blurring of the chrominance component, which leads to color bleeding, is most evident along the top edge of the arm in Fig. 6, as well as around the table-tennis paddle. The source of the chrominance ringing is seen along the edge of the table-tennis table in both Figs. 6 and 7. The photograph capturing the corresponding color frame is shown in Fig. 8.

It is interesting to note that strong chrominance edges are accompanied by strong luminance edges; however, the existence of a strong luminance edge does not necessarily coincide with a strong chrominance edge [22]. Therefore, color bleeding is not necessarily found at blurred edges in a reconstructed color frame.

Additionally, it should be noted that the chrominance subsampling of the source also causes
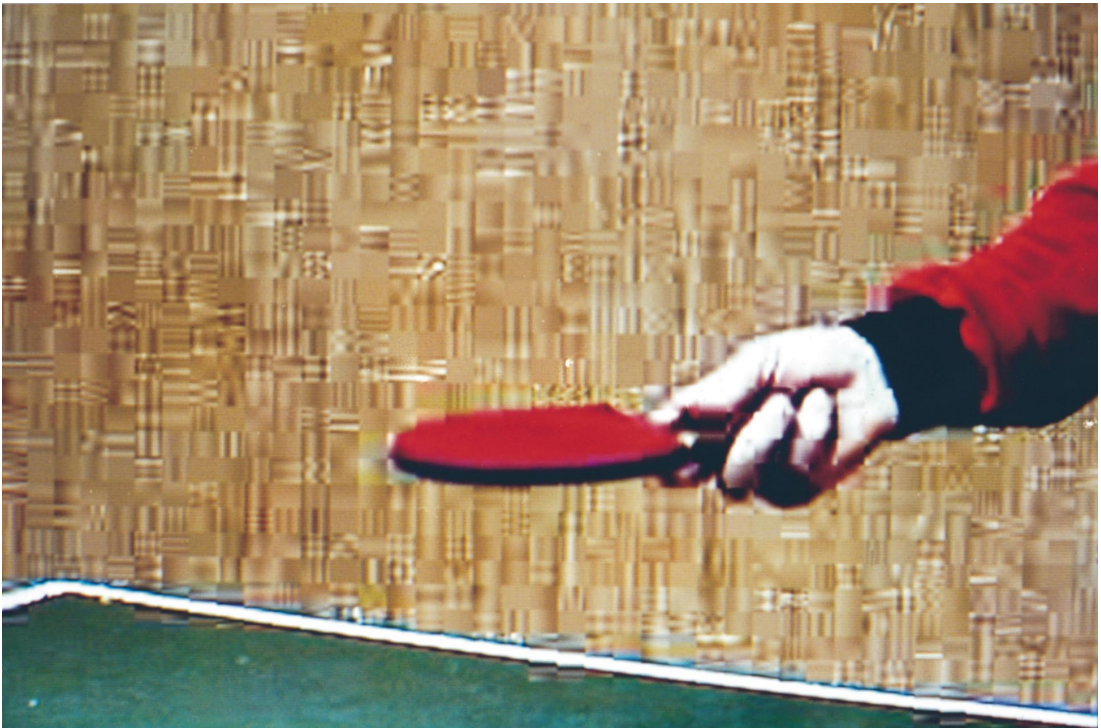

Fig. 8. Reproduction of a frame suffering from color bleeding and 'color ringing' from the *Table-tennis* sequence, corresponding with the Cb and Cr chrominance components shown in Figs. 6 and 7. Note the high-frequency changes in color around the table's edge, corresponding to the ringing, and the gradual bleeding around the arm.

color bleeding; for example, in addition to the 4:2:2 subsampling pattern of CCIR 601, when converting to SIF, MPEG-1 suggests the application of a further 4-tap subsampling filter [29]. Naturally, additional error is incurred when up-sampling back to CCIR 601. Therefore, color bleeding is an aggregate of both the subsampling of the original source as well as a result of the coding/compression processes.

6. Staircase effect

The DCT basis images are not attuned to the representation of diagonal edges and features [24]. Consequently, more of the higher activity basis images are required to satisfactorily represent diagonal edges or significant diagonally oriented features. Due to the typically low magnitude of the higher-order basis images, coarse quantization results in their truncation to zero. The contribution originally made by the higher-order basis images in forming the diagonal edge is diminished, resulting in a reconstruction exhibiting only the character-istics of the lower frequency basis images, which are generally either horizontally or vertically oriented. So, for a block containing a diagonal edge angled towards the horizontal, coarse quantization will result in a reconstruction with a horizontal orientation, and vice versa for blocks angled towards the vertical. The *ringing effect*, which is also found in blocks containing an edge, will be discussed in Section 7.

The staircase effect is related to both the blocking and mosaic pattern effects in terms of the manifestation of discontinuities between adjacent blocks. When a diagonal edge is represented within a string of consecutive blocks, the consequence of coarse quantization is the reconstruction of the diagonal edge as a number of horizontal or vertical steps. These individual steps do not merge smoothly at a block's boundary with the continuance of the edge in adjacent blocks. A number of horizontally oriented examples of the staircase effect are shown in Fig. 9; note that the step-wise discontinuities occur at block boundaries. Fig. 9 also exhibits significant *ringing*, especially around the dark diagonal edges in the top half of the image.
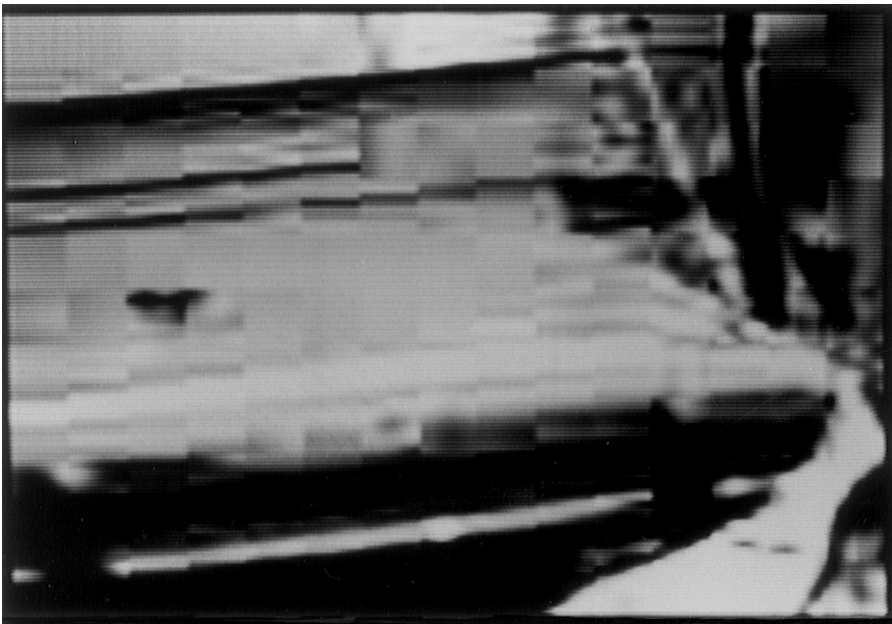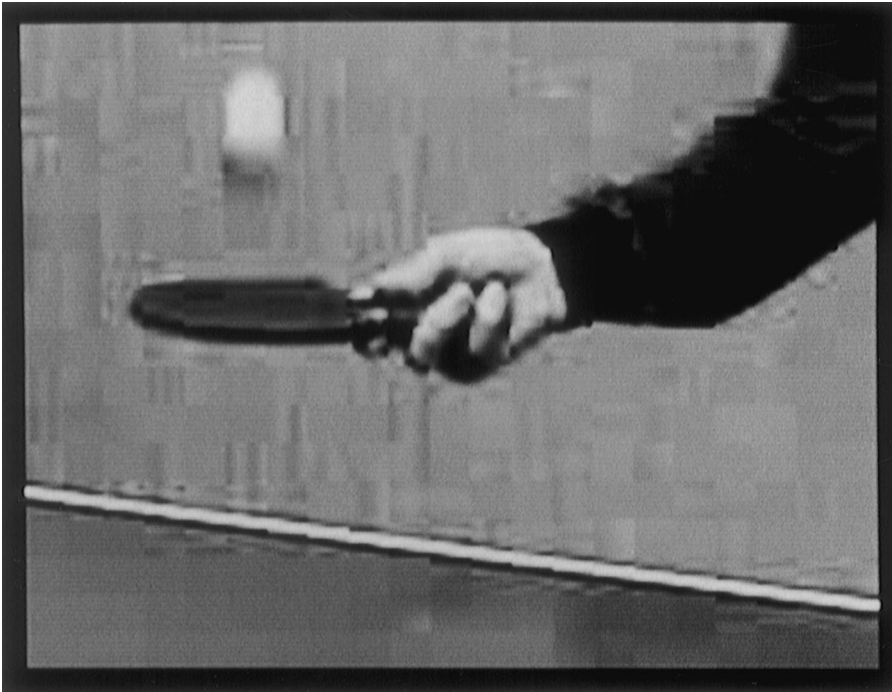

Fig. 9. Examples of the staircase effect.

Fig. 10. Example of the ringing effect, where it is most evident around the bright table-edge and the boundary of the arm.

The staircase effect is particularly noticeable with small block sizes (say, 6×6), and may not be visually discernible with the average block sizes used in block transform coders [26]. For larger block sizes, the discontinuity of the edge at block boundaries are spaced too widely for the eye to discern a regular, jagged, step-wise pattern. In this case, the staircase effect would be visualized as an occasional misalignment in an otherwise smooth edge.

7. Ringing

The representation of a block can be considered as a carefully balanced aggregate of each of the DCT basis images, such that the feature contributed by any one basis image is either enhanced or muted by the contribution of the other basis images [27]. Therefore, quantization of an individual coefficient results in the generation of an error in the contribution made by the corresponding basis image to a reconstructed block. Since the higher-frequency basis images play a significant role in the representation of an edge, the quantized reconstruction of the block will include high-frequency irregularities.[3]

The ringing effect is fundamentally associated with Gibb's phenomenon, and, as such, it is most evident along high contrast edges in areas of generally smooth texture in the reconstruction, and appears as a shimmering or rippling outwards from the edge up to the encompassing block's boundary. The higher the contrast of the edge, the greater the level of the peaks and troughs of the rippling. Examples of this are shown in Fig. 10, where it is most evident along the edge of the table-tennis table and the bottom of the player's arm. The generally smooth texture in the surrounding blocks

---

[3] If most of the higher-order AC coefficients are nullified as a result of very coarse quantization then, depending on the energy and distribution of the AC coefficients, the *basis image effect* or *blurring* may result instead.

results in a greater visibility of the ringing effect, where otherwise a masking of the ringing would be introduced.

It is worthwhile to note that associated with the parallel rippling away from diagonal edges is a less noticeable contouring whose direction is perpendicular to that of the rippling. This may be explained by examining the separable construction of the 2-D DCT [33]: the formulation of a single DCT basis image $b_{k,l}$ can be shown to be

$$b_{k,l}(m,n) = K \cos\left[\frac{(2m + 1)k\pi}{2N}\right]\cos\left[\frac{(2n + 1)l\pi}{2N}\right],$$
$$m,n \ = \ 0,1,\ldots,N - 1, \tag{1}$$

where (k,l) is the index of the associated DCT coefficient, and K is some constant. Using the trigonometric identities, Eq. (1) may be represented in vector notation as

$$b_{k,l}(m,n) = \frac{K}{2}\left\{\cos\left[\frac{\pi}{2N}\binom{2m + 1}{2n + 1}^{\mathrm{T}} \cdot \binom{k}{l}\right]\right.$$
$$\left. + \cos\left[\frac{\pi}{2N}\binom{2m + 1}{2n + 1}^{\mathrm{T}} \cdot \binom{k}{-l}\right]\right\},$$

where '·' denotes the dot product. This demonstrates that each DCT basis image may be considered to be the sum of two signals whose directions are conjugate: $(k,l)^{\mathrm{T}}$ and $(k, -l)^{\mathrm{T}}$. The exception to this is for $l = 0$, for which the two vectors are identical. So, given that the representation of a diagonal edge will be formed from basis images containing features in the same direction or angle as the edge, the aforementioned basis images will also contribute features which are perpendicular to the edge. Therefore, the imbalance in the aggregate caused by quantization error manifests features which are both parallel and perpendicular to the edge contained in the block.

The discussion and background of the ringing effect that has been presented so far also applies to the chrominance components; consequently, a similar effect to ringing occurs for the chrominance information at strong chrominance edges (see Figs. 6–8). The ringing of the chrominance information appears as wave-like transitions of color away from the chrominance edge up to the bound-

ary of the encompassing macroblock. The colors produced as a result of ringing often do not correspond to the colors of the surrounding area. Due to the subsampling of the chrominance information, the chrominance ringing spans a whole macroblock and is not localized to a single constituent block.

8. Mosaic patterns

The general nature of the *mosaic pattern* effect is the apparent mismatch between all, or part, of the contents of adjacent blocks; this has a similar effect to using visually ill-fitting square tiles in a mosaic. This may mean a block with a certain contour or texture dissimilar to the surrounding blocks, or a block used in the representation of an object which does not blend satisfactorily with the other constituent blocks.

The mosaic pattern, typically, coincides with the blocking effect; however, the existence of the blocking effect between two blocks does not necessarily imply the manifestation of the mosaic pattern between the same two blocks. For example, a smoothly textured area is highly susceptible to the blocking effect, but the smooth characteristics of the adjacent blocks would not induce a mosaic pattern. Measures to reduce the blocking effect in these low activity areas by smoothing the block boundaries have been shown to be effective [28,32], with no visually apparent mosaic pattern. Fig. 11 shows the result of such a filter proposed by Tzou [32] applied to the reconstructed frame depicted in Fig. 1. Compare the smoothly textured areas of the window and hood (bonnet) of the vehicle in Figs. 1 and 11. Examples of the mosaic pattern effect can be seen on the face of the human character in Fig. 11. The mosaic pattern may also be introduced by the various artifacts discussed in previous sections. This is most evident with blocks suffering from the basis image effect, as can be seen in Fig. 3.

8.1. Intraframe-coded macroblocks

For intraframe coded blocks, the mosaic pattern typically occurs in areas of the reconstructed frame

Fig. 11. Example of the mosaic pattern effect, where it is most evident around the character's face and adjacent to the horizontal edges near the van's window.

where high spatial activity existed in the original frame. Due to the typical quantization weightings and probability distributions associated with each of the AC coefficients, very coarse quantization will result in the truncation of a significant proportion of the higher-frequency AC coefficients to zero. Consequently, upon reconstruction, the blocks will contain textures constructed only from the lower-frequency AC basis images, which may be of a dissimilar texture or contour than their neighbors. It is important to note that even if the blocks contained a similar texture in the original frame, it cannot be guaranteed that the coarsely quantized reconstruction will be the same for all the blocks. There cannot even be a guarantee of a similarity of the general texture between the reconstructed blocks. An example of this can be seen by comparing Figs. 12 and 13, which show the background of the first frame of the *Table-tennis*

sequence, before and after transform coding,[4] respectively.

If an examination is made of the lower-order AC basis images in Fig. 4, it can be seen that their general orientations can be placed into one of three categories: horizontal, vertical and indeterminate.[5] One of the major factors affecting the visibility of the mosaic pattern effect is the general directional orientation of the adjacent blocks. For example, if the contents of two adjacent blocks are of dissimilar directional orientation, then the mosaic effect will be more pronounced; this is especially true for

---

[4] For this example, the same quantization weighting matrix and scaler were used for all the 8×8-pel blocks to code the frame in Fig. 13.

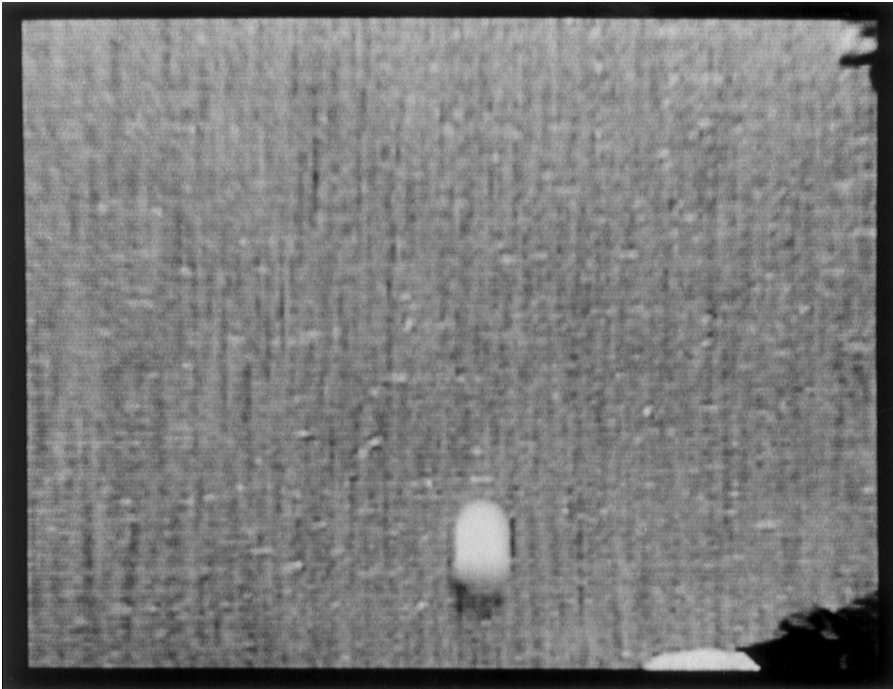[5] An attempt to define a set of diagonally oriented DCT basis images can be found in [24].

Fig. 12. A section of the background from the original (uncoded) *Table-tennis* sequence.
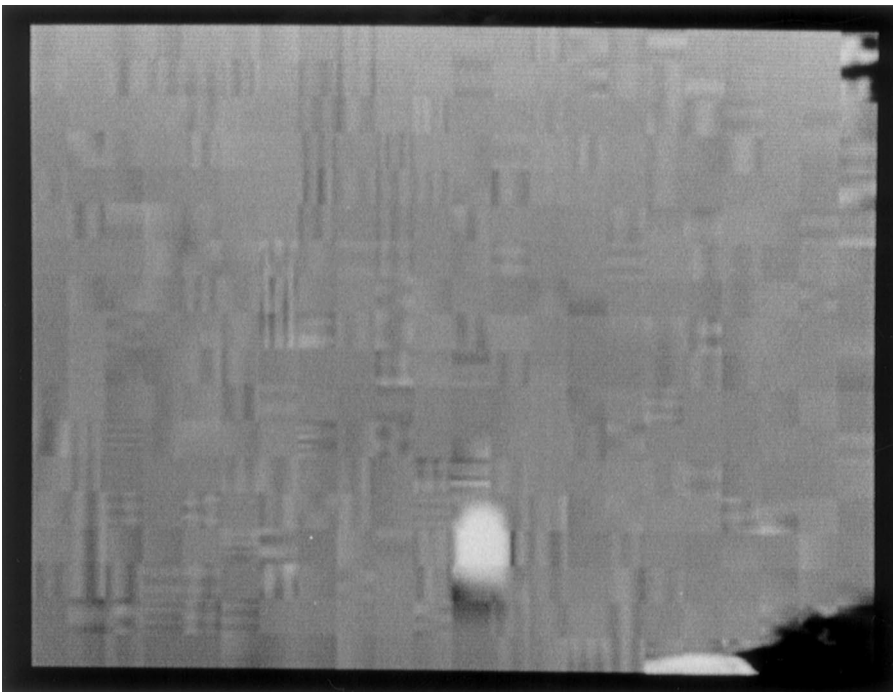

Fig. 13. The same section of the background from the *Table-tennis* sequence, shown in Fig. 12, after transform coding. An identical quantization matrix and scaler were used for all the 8×8-pel blocks.

a combination of horizontally and vertically oriented blocks.

Even if adjacent blocks do have a similar orientation, the visibility of the mosaic pattern will also be affected by the prominence of the major frequency component of each of the adjacent blocks. The worst case would be if a block is reconstructed using only a single AC basis image (see Section 3). A simple example of this can be seen by placing the lowest-order horizontal AC basis image adjacent to any of the higher-order horizontal basis images. Therefore, a similar major frequency component between the adjacent blocks would reduce the perceived mosaic pattern.

A number of aspects of the mosaic pattern effect for intraframe coded blocks have similarities to the *basis image effect*, which was described in Section 3. However, the coarseness of the quantization necessary for the mosaic pattern to occur is less than that for the basis image effect, and the resulting patterns of the adjacent blocks need only be sufficiently dissimilar: they need not resemble DCT basis images.

8.2. Predictive coded macroblocks

The typical consequence of the low bit-rate coding of I-type frames is the suppression of the higher order AC components in favour of the lower order components. For the predictive coded macroblocks of succeeding frames in a sequence, the MC prediction macroblocks originating from I-type frames typically contain a good prediction of the lower-order components [21]. Therefore, for areas of high spatial activity, the main task of the prediction error is to reconstruct the higher-frequency information.

As with intraframe coded macroblocks, the mosaic pattern effect in predictive coded macroblocks typically occurs in high spatial activity areas of a frame. The effect results from the attempt to compensate for the loss of high-frequency information in an MC predicted macroblock with coarsely quantized prediction error. This may produce differing higher-frequency AC components in the reconstruction of adjacent blocks within a macroblock, or in neighboring macroblocks. In smooth

areas, however, the prediction is of a reasonable quality[6] and the likeness between adjacent macroblocks preclude the mosaic pattern effect.

It is interesting to note that if the same high spatial activity area is visible in a number of consecutive frames of a sequence, the mosaic pattern gradually decreases over time. This is as a result of the accumulation and refinement of the higher-frequency AC component information for each successive MC reference frame. However, the visibility of the mosaic pattern in B-type frames is subject to the type of prediction used for each macroblock, as well as which reference frames were used for the MC prediction. Figs. 14–16 show the same section from the reconstruction of the *Table-tennis* sequence for an I-type frame and the following two P-type frames,[7] respectively. A comparison between these three reconstructions demonstrates the gradual refinement of high activity spatial information for each successive MC reference frame. However, it is important to note that the degree to which the high frequency information may be refined is subject to the coarseness of quantization of the related prediction error. Note that the I-type frame was filtered [36] for the blocking effect prior to it being used as an MC reference; this was done to separate out the effect of the propagation of the blocking effect to the following predictive coded frames (*false edges*, see Section 10).

9. False contouring

The simple artifact of false contouring often results from the direct quantization of pel values. It occurs in smoothly textured areas of frames containing a gradual transition in the value of the pels over a given area. The many-to-one mapping operation of quantization effectively restricts the allowable pel values to a subset of the original

---

[6] A poor prediction for a smooth area would typically result in the encoder intraframe coding the macroblock [29].

[7] Due to the structure of the Group Of Pictures (GOP) used in the MPEG-1 coding process for this example, the P-type frames are the 5th and 9th frames in the sequence.
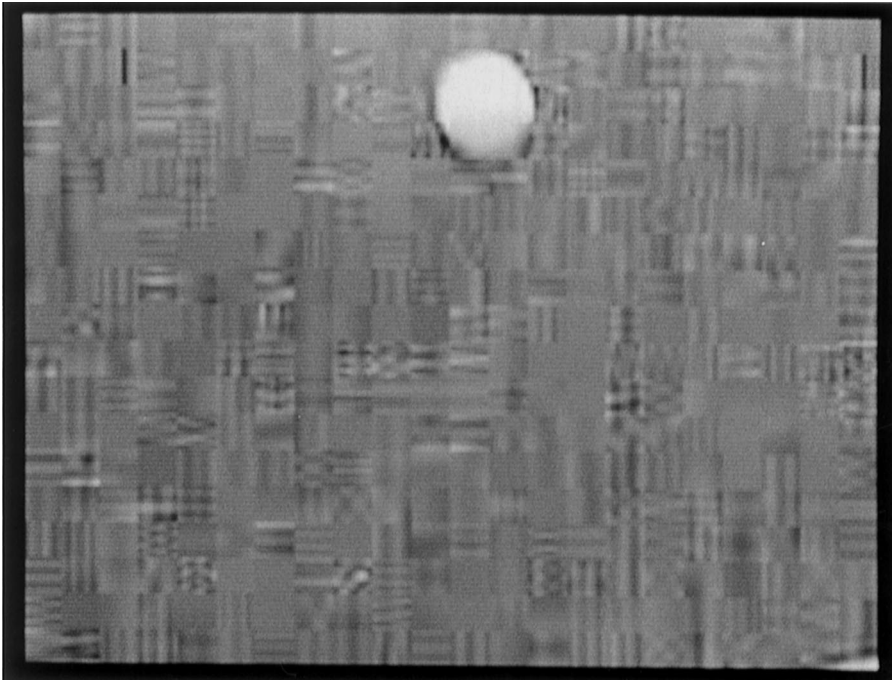
Fig. 14. Section of the reconstruction of an I-type frame from the *Table-tennis* sequence, coded at 0.6 Mbps, with the application of a filter for the blocking effect.
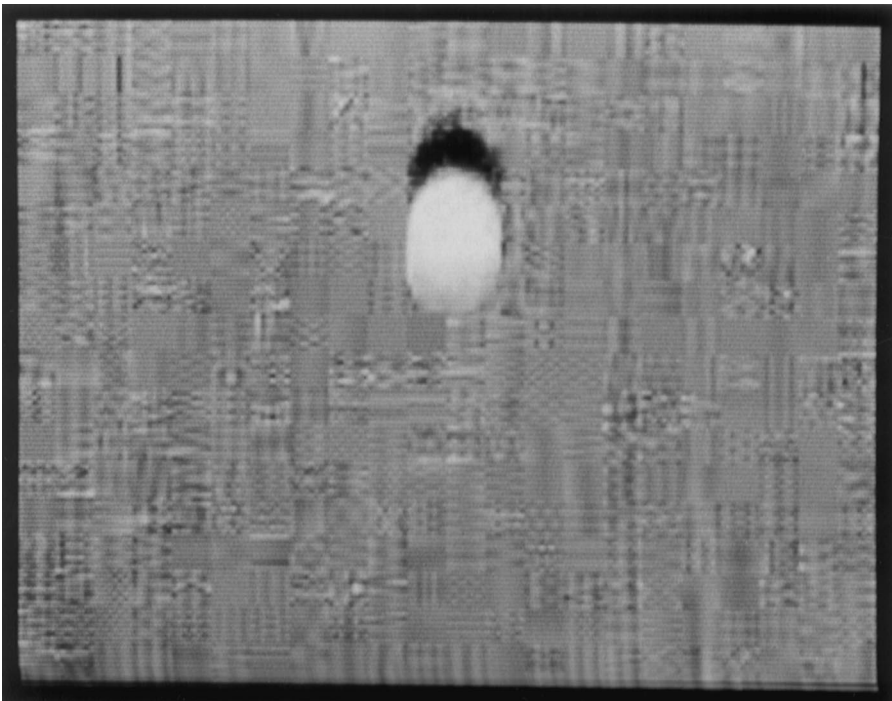

Fig. 15. Section of the reconstruction of the first P-type frame from the *Table-tennis* sequence, coded at 0.6 Mbps, using the I-type frame in Fig. 14 as the MC reference.
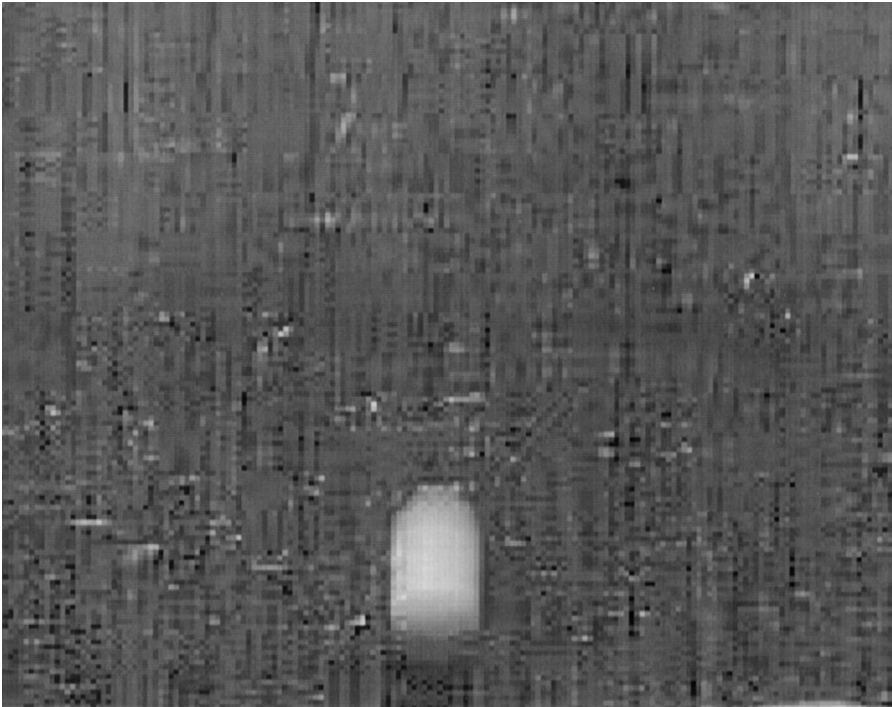
Fig. 16. Section of the reconstruction of the second P-type frame from the *Table-tennis* sequence, coded at 0.6 Mbps, using the P-type frame in Fig. 15 as the MC reference.

range, resulting in a series of step-like gradations in the reconstruction over the same given area. This artifact is directly attributed to either an inadequate number of quantization levels for the representation of the area, or their inappropriate distribution [11].

False contouring can also occur in block-based transform coding with similar circumstances and causes to that of direct quantization. The artifact is a consequence of the inadequate quantization of the DC coefficient and the lower-order AC coefficients in smoothly textured areas. The effect appears in the reconstruction as step-like gradations in areas of originally smooth transition, similar to the effect described above. However, it is important to note that the gradations tend to affect a whole block. Examples of this are shown in the background of Fig. 17, which contains a section of an I-type frame from the *Claire* sequence. The original frame had contained a gradual reduction of the luminance radially away from the speaker's head.

10. False edges

The exploitation of interframe redundancies relies on the transfer of previously coded information from MC reference frames to the current predictive coded frame. The transferred information, unfortunately, also includes the coding artifacts formed in the reconstruction of the MC reference. False edges are a consequence of the transfer of the block-edge discontinuities formed by the blocking effect into the current frame.

The generation of false edges is explained in Fig. 18, where the predicted macroblock copied from the MC reference frame contains the block boundary discontinuities formed by the blocking effect. These false edges may, or may not, coincide with the boundaries of macroblocks/blocks in the current predictive coded frame, but for skipped macroblocks the boundaries will always coincide if the MC reference is an I-type frame.
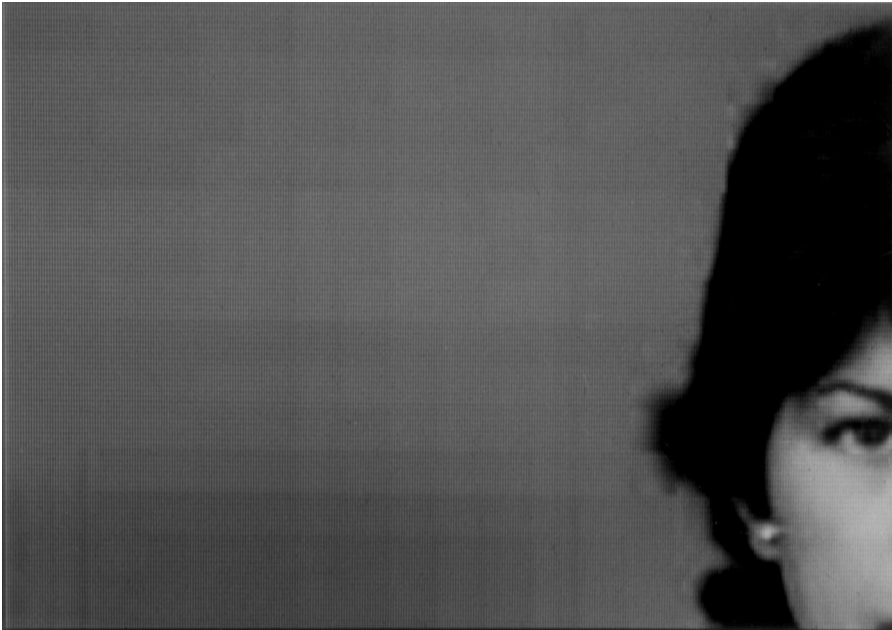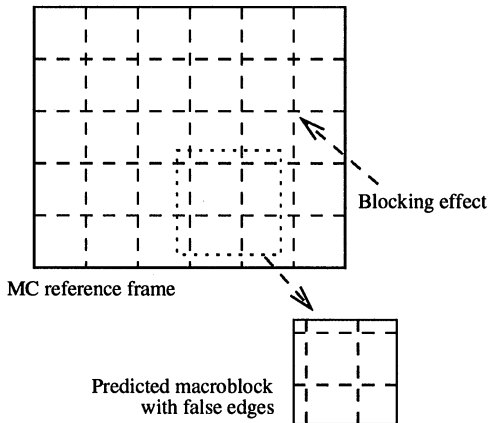
Fig. 17. Example of false contouring.

Fig. 18. Example of the propagation of the blocking effect to false edges.

As with the blocking effect, false edges are mainly visible in smooth areas of predictive coded frames. The prediction error in such areas would typically be minimal, or quantized to zero, and therefore the false edges would not be masked. Similar sections in an MC reference frame, which contain false edges, may be used as a prediction for high spatial activity areas; however, the false edges may not be perceivable in the reconstruction due to the addition of the high-spatial activity prediction error that is used to supplement the prediction. As a consequence of these masking effects, the manifestation of false edges typically do not propagate to subsequent predictive coded frames, especially in areas of high spatial activity.

11. MC mismatch

Motion compensation between frames of a sequence is commonly conducted on the luminance information using a full-search block-matching technique, where the contents of a macroblock to be coded is compared to all the possible macroblock-sized regions within a limited search window of the MC reference frames. The common metric used for the comparison is a disparity measure, such as the mean squared error or mean absolute error. The final product is a motion vector indicating

the spatial displacement of the current macroblock from its prediction.

This simple, yet computationally intensive, operation models the motion within a sequence as being composed of translations of rigid objects. This has, generally, been found to provide a close approximation of the true motion within a sequence [9], and the use of a disparity measure for the search criterion has the benefit of determining the prediction, within the search window, which produces the minimum prediction error energy. However, a problem arises as a result of assuming that the true motion of all the constituent pels of a macroblock are identical. This is most evident around the boundaries of moving scene objects, where a macroblock may encompass pels forming part of a moving object as well as pels representing sections of the scene not connected with the same object. In this situation, the motion of the pels within a macroblock would be better defined as a collection of multiple sub-macroblock motion vectors [9].

MC mismatch can be generally defined as the situation in which a satisfactory prediction cannot be found for a particular macroblock, resulting in a prediction whose spatial characteristics is mismatched with those of the current macroblock. The consequence of such a situation is a high level of prediction error which, if its integrity is compro-

mised after coding, results in a reconstruction with highly visible distortions or a faded replica of spatial features from the MC reference. In the extreme, i.e. the nullification of the coded prediction error, the propagated features will be reconstructed without augmentation, resulting in the presence of objects and spatial characteristics which are uncorrelated with the scene depicted in the current frame ('object retention' [3]).

For a better understanding of the MC mismatch effect, we examine a particular example where only two objects are involved, i.e. a macroblock straddling the boundary between two objects in a scene whose motions are divergent with relation to each other. Examples of such macroblocks are illustrated in 'Frame n' and 'Frame $n + k$' of Fig. 19. In both these situations, it is unlikely for there to exist the same congruence between the two objects within the MC reference frames as in the current frame; therefore, the chosen MC predicted macroblock would contain an unsatisfactory representation of one, or possibly even both, of the objects it is intended to represent. As depicted in Fig. 19, this results in a difference between the level and characteristics of the prediction error for the area surrounding an object in a predictive coded macroblock.

Since these macroblocks are situated on an object boundary, it will contain a significant amount
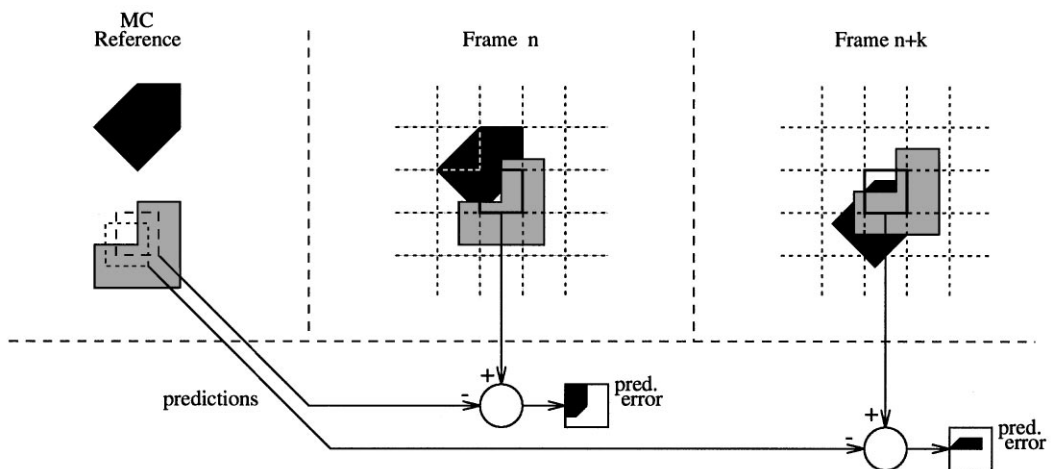

Fig. 19. Examples of the effect that changing background has on the characteristics and level of MC prediction error.

of high-frequency detail, such as a scene edge; there-fore, it would not be advantageous to intraframe code the macroblock, even though the poor predic-tion will result in a large prediction error energy. This corresponds with the intra/predictive coding decision mechanism used in MPEG-1 based on the variances of the macroblock and the prediction error.

Obviously, the above example is somewhat ideal-istic since the disparity measure used in the block-matching (prediction search) operation does not seek to retain the integrity of any one of the objects' contents, nor the boundary dividing the two ob-jects. The non-intuitive nature of the matching cri-terion hampers the between-frame consistency in the selection of predictions for corresponding sec-tions of scenes.

This is further exacerbated by the finite search window that restricts the spatial extent of the block-matching operation. As the camera pans across a scene, or the scene moves with relation to the camera, the areas of the scene covered by the prediction search windows of all macroblocks also change. As a result, a locally optimum prediction for a particular macroblock may move outside its search window, necessitating the selection of a new prediction. This is most noticeable in panning se-quences where the changes in prediction are seen as a steady panning motion interrupted by a regular series of abrupt changes in the spatial character-istics of the reconstructed scene content.

If a macroblock is reconstructed solely from a poor quality prediction, then, visually, the effect of MC mismatch would be seen as an ill-fitting square section of pels with relation to either one or more of the scene objects that the macroblock encompasses. Consequently, a relationship would exist with the previously discussed artifacts of the mosaic pattern effect and the blocking effect. How-ever, due to the changing spatial characteristics induced by the moving objects, a large prediction error energy will result. The presence of a boundary edge will also cause the prediction error to be impulse-like in content [24]. After quantization of

Fig. 20. Example of the high-frequency noise induced by MC mismatch around the boundaries of moving objects.

the DCT coefficients, the contribution of the prediction error to the reconstruction is typically a high-frequency noise-like effect. This helps to mask the mismatch of the prediction, although it also produces a temporal effect which will be discussed in Section 12. The effect of quantizing the higher order AC coefficients in blocks containing a high contrast edge was discussed in Section 7 with relation to ringing. Examples of the high-frequency noise induced by MC mismatch can be seen in Fig. 20 around the boundary between the arm and the background. This example was taken from a P-type frame, and it is important to note that all the macroblocks representing the arm and its boundary were forward predicted.

12. Mosquito effect

The mosquito effect is a temporal artifact seen mainly in smoothly textured regions as fluctuations of luminance/chrominance levels around high contrast edges, or moving objects, in a video sequence [18]. This effect is related to the high-frequency distortions introduced by both the ringing effect, and the prediction error produced by the MC mismatch artifact; both of which have been previously described. Generally, the degree and visibility of the fluctuations is less for the ringing-related effect than that resulting from the MC mismatch prediction error. The mosquito effect, regardless of the origin, is a consequence of the varied coding of the same area of a scene in consecutive frames of a sequence. This may mean a difference in the type of prediction (forward, backward, bidirectional or skipped), quantization level, MC prediction (motion vector), or a combination of these factors.

12.1. Ringing-related mosquito effect

In predictive coded frames, a macroblock containing a high contrast edge would most likely be predictively coded [29]. If available, the macroblock's MC prediction would also contain a high contrast edge corresponding to the contents of the current macroblock. Assuming that the I- or P-type MC reference frames were poorly quantized, this

prediction would also be suffering from the ringing effect due to the presence of the high contrast edge.

The mosquito effect occurs in such a situation, generally, as a result of varying attempts, from frame to frame, to correct the ringing effect in the MC prediction, or the use of a different prediction. This may be a consequence of the use of a different prediction type, a differently positioned MC prediction, or simply a change in the level of quantization of the prediction error. Any of these factors will cause a difference in the reconstruction of the same area of a frame. This assumes that the quantization is sufficiently coarse to prevent the satisfactory correction of the ringing effect, in which case under- or over-correction of the high frequency ringing noise may be introduced by the coarsely quantized prediction error.

As discussed in Section 7, the higher-frequency DCT basis images play a significant role in the representation of a high contrast edge, and the ringing effect results from the quantization error of the higher-order AC coefficients. To correct the ringing effect in the predicted macroblock, the prediction error will require the use of a significant level of the higher-frequency basis images; therefore, differences in the quantized prediction error, from frame to frame, will result in high frequency fluctuations in areas around high contrast edges during the display of the decoded video sequence.

Fig. 21 shows the difference between the same section of five consecutive coded frames from the *Claire* sequence, i.e. the difference between frames 0 and 1, 1 and 2, etc. The speaker's left arm remains relatively stationary over the duration of the five frames. High-frequency noise-like blocks can be seen around the high contrast edge formed by the speaker's arm and the background. Note that this noise is the difference between consecutive frames, therefore, it is equivalent to the temporal change in the area between consecutive frames.

As an indication of the different MC prediction types used by each of these predictive coded frames, each macroblock within the same cosited region is delineated in Fig. 22 for frames 1 through 4 of the coded *Claire* sequence, with the MC prediction method shown for each macroblock. These displayed regions correspond with those presented in Fig. 21.
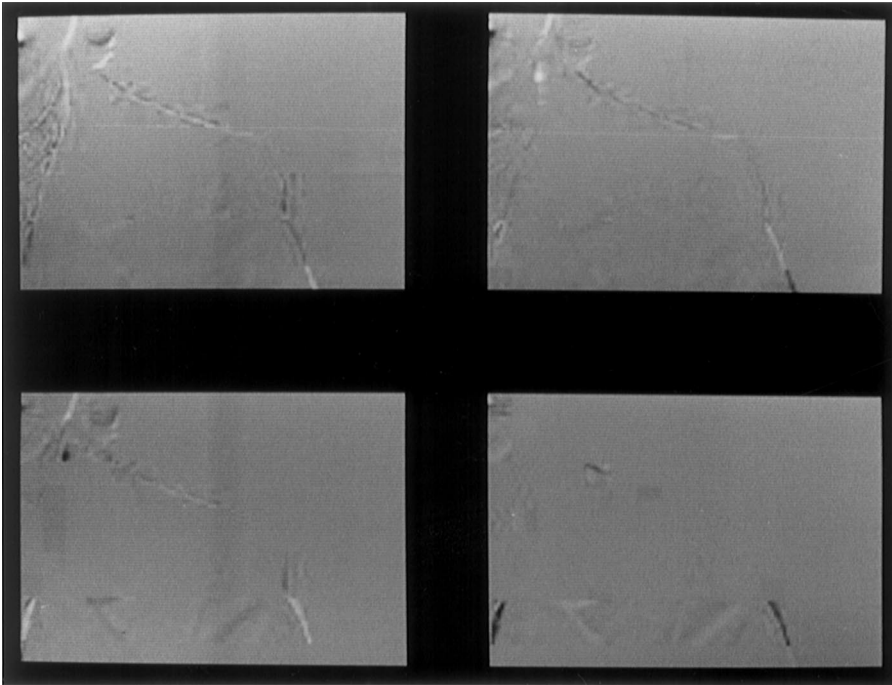
Fig. 21. Difference between the same section of five consecutive coded frames from the *Claire* sequence. The jaw of the speaker can be seen at the top-left of the difference images. To enable their display, the differences were offset by $+128$ and clipped to be within 0 and 255. Top left: 0 and 1; top right: 1 and 2; bottom left: 2 and 3; bottom right: 3 and 4.

12.2. Mismatch-related mosquito effect

As discussed in Section 11, the divergent motion around the boundaries of moving scene objects is a hindrance to the selection of a satisfactory MC prediction. The presence of the boundary edge, and the significant level of the prediction error energy, results in a high-frequency noise-like effect in the area around the boundary.

The causes of the mosquito effect in these areas are similar to those originating from the ringing effect, as discussed above. However, as a consequence of the changing spatial characteristics of either the moving object or its background, the differences would be more pronounced between the predicted macroblocks used in consecutive frames for the same section of the moving object. Also, if the object is moving with relation to the camera, the section of the object that is encompassed by any one macroblock will change from frame to frame. The different locations of the discontinuities caused by the changing position of the macroblock boundaries, with reference to the moving object, will result in an increased mosquito effect.

13. Stationary area temporal fluctuations

Similar fluctuations to that associated with the mosquito effect have also been seen away from object boundaries and high contrast edges, in stationary areas containing significant spatial activity. This is despite the masking effect provided by the high spatial detail. These fluctuations tend not to be perceived in similarly textured areas which are also undergoing significant motion, where it would be difficult to discern any minor difference between a section of a scene in one frame and the same section in its new position in the next frame of the sequence.

As before, the fluctuation's causes relate to the varied coding of the same area of a frame for

(a) Frame1

(b) Frame2
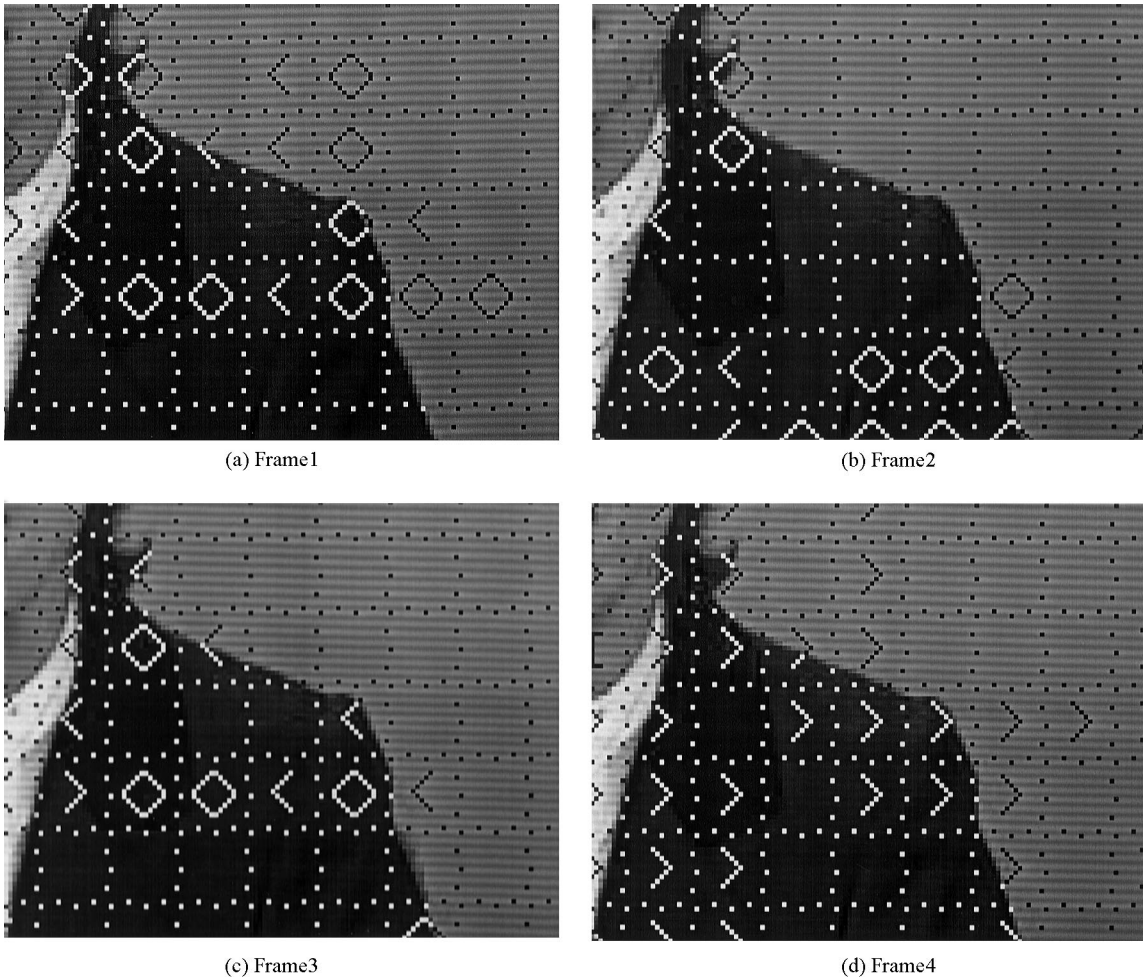

(c) Frame3

(d) Frame4

Fig. 22. The MC prediction types used to code each of the macroblocks in the same spatial region of frames 1 through 4 of the *Claire* sequence. The region corresponds to that used in the difference images in Fig. 21. Chevrons pointing right and left represent forward and backward prediction, respectively. Bidirectional (interpolated) prediction is represented by a diamond-like symbol, and the letter 'I' indicates that the macroblock was intraframe coded. The absence of a symbol indicates that the macroblock was skipped.

consecutive frame of a sequence. The high spatial activity in the area makes it unlikely that a macroblock will be intraframe coded, especially where there is no movement; therefore, the fluctuations result from the differing types of prediction, quantization levels, MC prediction, or a combination of these factors.

Since a relatively significant level of quantization error is introduced into the higher order AC coefficients in I-type frames, which propagates to the succeeding predictive coded frames of the sequence, the main task of the prediction error is to compensate for the loss of the higher-frequency information. Consequently, the difference in the prediction error from frame to frame causes high spatial frequency fluctuations.

Figs. 23 and 24 show two consecutive frames of the coded *Sons and Daughters* sequence where there is an absence of motion in the background, and slight motion by the two characters at the centre of

Fig. 23. Frame number 40 of the coded *Sons and Daughters* sequence.

Fig. 24. Frame number 41 of the coded *Sons and Daughters* sequence.

Fig. 25. Difference between frames 40 and 41 of the *Sons and Daughters* sequence. For the purpose of display, the difference is offset by + 128 and clipped to be within 0 and 255.

the scene. As an indication of the flickering in the background, Fig. 25 shows the difference between Figs. 23 and 24. Note that the difference in the background consists of a large number of blocks containing either high-frequency information, or no information at all (skipped).

14. Chrominance mismatch

As was mentioned in Section 11, motion compensation between frames of a sequence is commonly conducted using a full-search block-matching technique, where the contents of the macroblock to be coded is compared to all the possible macroblock-sized groups of pels within a limited search window of the MC reference frames. One important factor of this search is that it is conducted only with the luminance information, and the same motion vector is used by all the components (luminance and chrominance) of the macroblock.

Although a mismatch of the chrominance information occurs in the general MC mismatch artifact described in Section 11, the chrominance mismatch described in this section does not necessarily occur at object boundaries, and the quality of the associated luminance prediction is generally satisfactory. Chrominance mismatch appears as a misplacement of a macroblock with respect to its own general color and the color of the surrounding area within the frame.

The use of only the luminance information for the block-matching operation results in the selection of the macroblock-sized region within the search window which has the highest luminance correlation to the macroblock currently being coded. This correlation may not extend to the chrominance information, and may even be totally disparate to the chrominance information of the current macroblock. Fig. 26 shows a photograph of a predictive coded frame from the *Flowers* sequence. Examples of the manifestation of

Fig. 26. Macroblocks located centrally in the trunk of the tree in this frame provide examples of *chrominance mismatch* within the *Flowers* sequence.

chrominance mismatch can be seen within the tree-trunk, where sections of macroblocks containing a red-orange color (presumably, from the nearby roof tiles) are situated within the grey-brown trunk. It must be noted that the color disparity can be reduced by the coded prediction error, but its effectiveness is a function of the coarseness of the quantization.

15. Other temporal distortions

Although the aim of this paper is to illustrate the distortions introduced in a reconstructed sequence by the hybrid MC/DPCM/DCT algorithm, there are a number of additional distortions which manifest in video reconstructions as a result of supplemental processing. These include artifacts which stem from pre-processing of the source, supplementary processing within the encoder, post-processing of the decoded reconstruction, and distortions related to specific characteristics of the source. Here we will briefly discuss each of these distortions and their causes.

15.1. Jerkiness

Stilted and jerky motion – often found on occasions of high motion within video phone/conference sequences – is seen as time-discrete 'snapshots' of the original continuous scene strung together as a disjointed sequence. Motion jerkiness is associated with temporal aliasing, and is, in turn, a manifestation of an inadequate sampling/display rate necessary to accommodate the temporal bandwidth of the source [22].

Naturally, the appearance of jerkiness in a sequence will also be caused by transmission delays of the coded bitstream to the decoder. This is also subject to the decoder's ability to buffer against these intermittent fluctuations.

15.2. Scene changes

Due to the abrupt change in spatial characteristics of frames before and after a scene cut, there is little benefit in predictive coding across the cut; therefore, the first frame of a new scene is often intra-coded. Consequently, with this approach, the quality of the initial frames after a scene cut is generally poor, and usually associated with a gradual build up in quality as the finer spatial components are predictively accumulated.

However, the perceived spatial resolution of frames after a scene cut may be degraded by up to one-tenth of the norm without being detected, provided that the spatial resolution is gradually restored within approximately half a second [31]. Therefore, any loss of quality after a scene cut is generally masked, and is generally not perceivable except when the frames are viewed at low display rates or as individual images.

To assist in coding performance, a common practice is to remove the abrupt nature of a scene cut by preprocessing the source by alpha-mixing [15] the scenes before and after the cut (Zhang, S.T., 1997, pers. commun.). This provides some interframe correlation across the original cut, thereby allowing the formation of useful MC predictions, and also avoiding the need for intraframe coding, with its associated distortions.

15.3. Smearing

An artifact associated with the "exposure time" necessary for a camera to integrate the light projected onto it from a physical scene [22] is the smearing of spatial detail if moving objects in the direction of motion, resulting in a loss of spatial resolution and a blurring of features. The non-instantaneous nature of the "exposure" causes light from multiple points of the moving object to be integrated into a single point (pel) in the recorded image/frame. This occurs for all points of the image associated with the spatial extent of the recorded scene covered by the moving object during the integration (exposure) time.

The perceptibility of this smearing is heavily dependent on whether or not the viewer is visually tracking the moving object during the replay of the recorded sequence [22]. If the object is not tracked then it is unlikely that the smearing will be perceived; otherwise, the ability to detect the smearing depends on the speed of motion of the object, as well as the typical spatial masking effects.

15.4. Ghosting

Ghosting is an artifact of deliberate temporal low-pass filtering of the source sequence, and appears as a blurred remnant trailing behind fast moving objects ('object persistence' [3]). The original objective of these temporal filters is to eliminate additive signal noise within the source, utilizing the fact that this noise is uncorrelated between frames, whilst the image information is highly correlated. The basic method in which this is achieved is through the representation of the current picture as a weighted average of the current and previous pictures in a sequence, thereby averaging out the uncorrelated noise [4,5,7,8]. However, if the scene content, or part thereof, is uncorrelated between frames, as is the case with a fast motion object, the averaging process produces a 'ghost' of the uncorrelated scene content from the previous frame.

15.5. Down/up-sampling

Although the process of down- and up-sampling of the spatial resolution of a video sequence is not strictly a component of the hybrid MC/DPCM/DCT algorithm, the process may be a requirement when the resolution of the video sequence source material, or the display device, is different to that which is optimum for the coding algorithm. A common example of this is the need to down-sample CCIR Recommendation 601 (CCIR 601) format (720×576 or 720×480) source
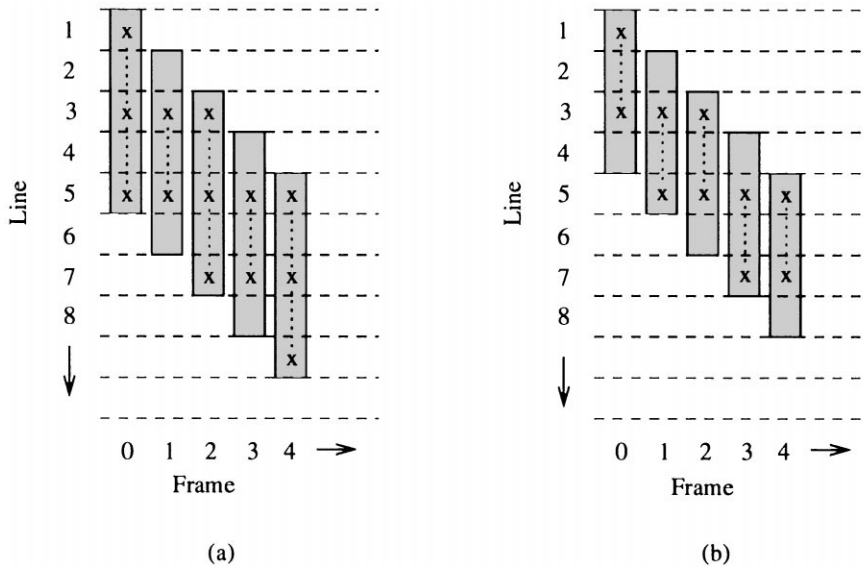
Fig. 27. The pels preserved in columns of an object, from frame to frame, as a result of discarding the even rows. (a) The case for columns of odd height and (b) for even height. 'x' marks the pels that are preserved in the down-sampled frame.

material [6] to MPEG-1 SIF (352×288) for MPEG-1 coding [29], followed by up-sampling of the coded reconstruction back to CCIR 601 for display.

The down-sampling process described in the MPEG-1 standard discards the even field to halve the vertical resolution. A horizontal decimation filter is then applied to the remaining odd field, which halves the horizontal resolution whilst reducing the spatial aliasing caused by the loss of the even field. This section describes a number of temporal artifacts related to the loss of the even field, affecting small, moving, spatial details which have a vertical component to their motion (the object's motion does not, necessarily, have to be solely vertical). The artifacts appear as alternating changes in vertical size, jittery movement and spatial fluctuations.

Fig. 27 demonstrates the loss of vertical information between frames for small objects undergoing vertical motion. Each row in both Fig. 27(a) and Fig. 27(b) represents a line in a CCIR 601 source frame, and each shaded column represents the same column of an object taken from consecutive frames in the sequence. Note that the same column is undergoing vertical downward move-

ment of a single pel between frames. For example, in Fig. 27(a), the column is spanning lines 1–5 in frame 0, and in frame 1 the column has shifted one pel downward to span lines 2–6.

Fig. 27(a) contains a column with an odd height of 5 pels, and the column in Fig. 27(b) is an even height of 4 pels. As a result of sampling the odd field (odd numbered lines), only the pels marked with an 'x' will be represented in the down-sampled SIF-sized frame. Naturally, an interpolation of only these 'x'-marked pels (represented in the figure by a line connecting the 'x's) will be used for the up-sampling back to CCIR 601.

If we concentrate on the odd-lengthed column in Fig. 27(a), notice that in frame number 0 three pels are carried over to the down-sampled frame and in frame number 1 only two pels are preserved. This alternating of vertical size of the object between frames causes a visual beacon-like effect in both the SIF-sized and CCIR 601-sized reconstructions, although it is more evident for the latter. Note that this only occurs if the object moves vertically an odd number of pels, otherwise, the vertical size of the object remains the same between frames.

The alternation of size does not occur for even-length objects, as shown in Fig. 27(b). However, for
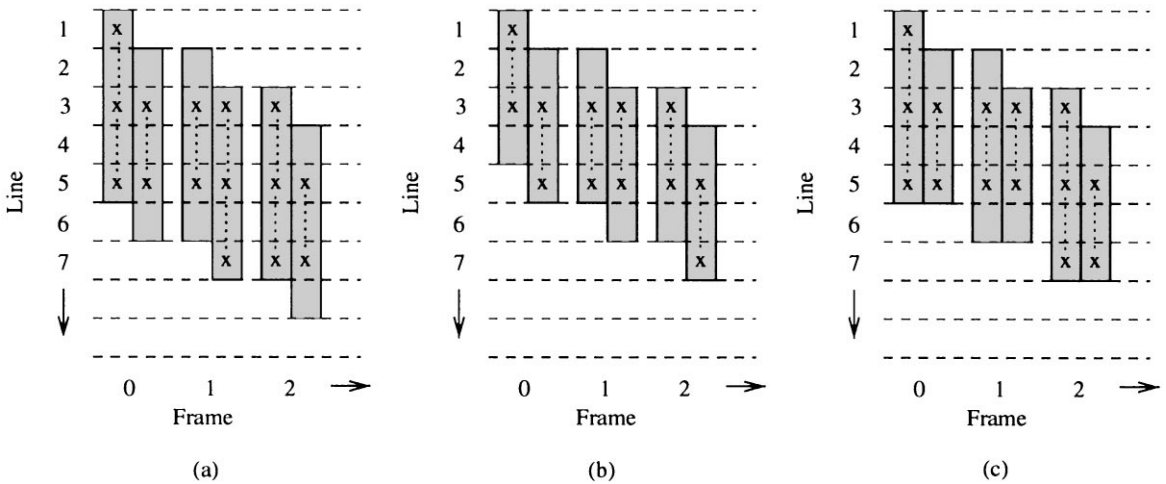
Fig. 28. The pels preserved in two adjacent, vertically misaligned columns of an object, from frame to frame, as a result of discarding the even rows. (a) The case for columns of odd height, (b) for even height and (c) for both odd and even. 'x' marks the pels that are preserved in the down-sampled frame.

single-pel vertical movement,[8] even-length objects are affected by jittery, stop-start motion. This can be seen by the pause in downward movement of the 'x's between frames 1 and 2 in Fig. 27(b), and then the jump between frames 2 and 3. This is more evident in the CCIR 601 reconstruction than for SIF, where there is a jump by two pels.

A visual busy-ness can be seen generally in small objects and details which are undergoing vertical motion. This temporal artifact relates to the differing effect that discarding a field has on the reconstruction of adjacent columns of an object that are not vertically aligned. Fig. 28 is set up similarly to Fig. 27 except that two adjacent columns of an object are considered together instead of just one for each frame. The sub-figures each show the situation where adjacent columns are vertically misaligned by a single pel. The busy-ness effect is most evident in Fig. 28(b) for even-lengthed columns, where it is seen that for frame 0 the 'x's in the two columns are vertically misaligned, whereas for

frame 1 they are aligned. This alternation continues for subsequent frames. Visually, this results in alternating vertical fluctuations of the columns forming the object. A similar effect occurs for columns of odd height in Fig. 28(a) and for adjacent columns of both odd and even height in Fig. 28(c).

Since these artifacts are only visually perceivable for very small objects or fine details, they would most likely be supplanted by the artifacts caused by the MC/DPCM/DCT algorithm which, especially for the DCT, typically results in the loss of fine spatial detail.

16. Summary

Due to the non-linearity of the quantization process, and the energy distributing effects of the inverse DCT, it is not possible to predict the distortionary outcomes of the quantization error; therefore, the aim of this paper was to identify the distortions which have a consistent identifying characteristic and, therefore, may be labeled as a consequence of the quantization process rather than a natural characteristic of the source.

---

[8] Actually, the stop-start motion will occur if the even-length object moves an odd number of pels; however, it is less noticeable for greater distances.

Table 1
Summary of the spatial and temporal coding distortions. The spatial distortions have been categorized according to their spatial extent within an image

Spatial			Temporal
8 × 8 block	16 × 16 macroblock	Global	
Blocking effect	Blocking effect	Staircase effect	False edges
Basis image effect	Color bleeding	Mosaic patterns	MC mismatch
Blurring		False contouring	Mosquito effect
Ringing			Chrominance mismatch
			Stationary area fluctuations

The quantization error introduced into the DCT coefficients can be distinguished into two basic processes. The first relates to fluctuations of a DCT coefficient around its original value, and the second results in the complete nullification of a DCT coefficient: the truncation effect.

The spatial distortions may be divided into two general categories, one relating to the actual distortions introduced into a block by the quantization process, and the other regarding the association and continuity between blocks. The latter distortion category is more a function of the independent coding of each of the blocks than the actual quantization error. The distortions in this group include the blocking effect, the mosaic pattern effect, the staircase effect, and false-contouring.

The distortions in the first general category can be further divided into two groups related to the two basic quantization error processes. The combination of fluctuations to the contribution of certain basis images by the quantization error, and the DCT's poor ability in representing high contrast edges and diagonal features, result in the distortions of ringing and the staircase effect. Although the quantization truncation effect also plays a role in the manifestation of these two distortions, truncation is mainly associated with blurring, color bleeding and the DCT basis image effect.

The temporal artifacts, in the main, result from the inconsistency in the coding of the same area of a scene in consecutive frames of a sequence. This includes variations in the type of predictions, quantization levels, MC predictions, or a combination of these factors. MC mismatch is associated with the inability to provide a satisfactory MC prediction for boundaries of moving scene objects. Whereas the mosquito effect is a general manifestation of the inability by the prediction error to satisfactorily correct for the loss of high frequency information in MC predictions.

As a final summary, Table 1 lists the spatial and temporal coding distortions. The spatial distortions have been divided according to their spatial extent within an image. The example images used in this paper can be found at the following ftp site: ftp.dgs.monash.edu.au/dgs/hrw/artifacts/.

References

[1] A.J. Ahumada, Luminance-model-based DCT quantization for color image compression, in: Proc. SPIE Human Vision, Visual Processing, and Digital Display, Vol. 1666, 1992, pp. 365–374.

[2] A.J. Ahumada, R. Horng, De-blocking DCT compressed images, in: Proc. SPIE Human Vision, Visual Processing, and Digital Display, Vol. 2179, May 1994, pp. 109–116.

[3] ANSI T1.801.02-1995, Digital Transport of Video Teleconferencing/Video Telephony Signals – Performance Terms, Definitions, and Examples, American National Standard for Telecommunications, ANSI, 1995.

[4] M.H. Chan, Motion adaptive noise filtering for video coding, in: Australian Broadband Switching and Services Symp. '91, Vol. 2, Sydney, Australia, July 1991, pp. 265–272.

[5] F.J. Chu, C.L. Yeh, Reduction of blocking artifacts using motion compensated spatial-temporal filtering, in: Proc. SPIE Image Processing Algorithms and Techniques II, Vol. 1452, 1991, pp. 38–45.

[6] R.J. Clarke, Digital Compression of Still Images and Video, Academic Press, London, 1995.

[7] T.J. Dennis, Nonlinear temporal filter for television picture noise reduction, Proc. IEE 127, part G (April 1980) 52–56.

[8] E. Dubois, S. Sabri, Noise reduction in image sequences using motion-compensated temporal filtering, IEEE Trans. Commun. COM-32 (July 1984) 826–831.

[9] F. Dufaux, F. Moscheni, Motion estimation techniques for digital TV: A review and a new contribution, Proc. IEEE 83 (June 1995) 858–876.

[10] B. Girod, The information theoretical significance of spatial and temporal masking in video signals, in: Proc. SPIE Human Vision, Visual Processing, and Digital Display, Vol. 1077, October 1989, pp. 178–187.

[11] R.C. Gonzales, R.E. Woods, Digital Image Processing, Addison-Wesley, Reading, MA, 1992.

[12] ISO/IEC JTC1/SC2/WG10, ISO IS 10918-1: Digital compression and coding of continuous-tone still images, (JPEG) ISO, 1990.

[13] ITU, ITU-T Recommendation H.261: Video Codec for Audiovisual Services at $p \times 64$ kbits, 1990.

[14] ITU, ITU-T Recommendation H.263: Video Coding for Low Bitrate Communication, 1996.

[15] K. Jack, Video Demystified: A Handbook for the Digital Engineer, HighText Pub., Solano Beach, CA, 1983.

[16] N. Jayant, Signal compression: Technology targets and research directions, IEEE J. Sel. Areas Commun. 10 (June 1992) 796–818.

[17] N.S. Jayant, P. Noll, Digital Coding of Waveforms: Principles and Applications to Speech and Video, Prentice-Hall, Englewood Cliffs, NJ, 1984.

[18] M. Kaneko, Y. Hatori, A. Koike, Improvements of transform coding algorithm for motion-compensated interframe prediction errors – DCT/SQ coding, IEEE J. Sel. Areas Commun. SAC-5 (August 1987) 1068–1078.

[19] S.A. Karunasekera, N.G. Kingsbury, A distortion measure for blocking artifacts in images based on human visual sensitivity, IEEE Trans. Image Process. 4 (June 1995) 713–724.

[20] S.A. Klein, A.D. Silverstein, T. Carney, Relevance of human vision to JPEG-DCT compression, in: Proc. SPIE Human Vision, Visual Processing, and Digital Display, Vol. 1666, 1992, pp. 200–215.

[21] D. Le Gall, MPEG: A video compression standard for multimedia applications, Commun. ACM 34 (April 1991) 47–57.

[22] A.N. Netravali, B.G. Haskell, Digital Pictures: Representation and Compression, Plenum Press, New York, 1988.

[23] H.A. Peterson, A.J. Ahumada, A.B. Watson, Improved detection model for DCT coefficient quantization, in: Proc. SPIE Human Vision, Visual Processing, and Digital Display, Vol. 1913, September 1993, pp. 191–201.

[24] R. Plompen, Motion video coding for visual telephony, PTT Research Neher Laboratories, 1989.

[25] A. Puri, R. Aravind, Motion-compensated video coding with adaptive perceptual quantization, IEEE Trans. Circuits Systems Video Technol. 1 (December 1991) 351–361.

[26] B. Ramamurthy, A. Gersho, Nonlinear space-variant post-processing of block coded images, IEEE Trans. Acoust. Speech Signal Process. ASSP-34 (October 1986) 1258–1268.

[27] K.R. Rao, P. Yip, Discrete Cosine Transform: Algorithms Advantages, Applications, Academic Press, San Diego, CA, 1990.

[28] H.C. Reeve, J.S. Lim, Reduction of blocking effects in image coding, in: Proc. IEEE ICASSP, Boston, MA, April 1983, pp. 1212–1215.

[29] Secretariat ISO/IEC JTC1/SC29, ISO CD11172-2: Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s, MPEG-1, Part 2: Video, ISO, November 1991.

[30] Secretariat ISO/IEC JTC1/SC29/WG11, ISO 13818.2: Information technology – Generic coding of moving pictures and associated audio information, MPEG-2, Part 2: Video, ISO, 1997.

[31] A.J. Seyler, Z. Budrikis, Detail perception after scene changes in television image presentations, IEEE Trans. Inform. Theory IT-11 (1) (January 1965) 31–43.

[32] K.-H. Tzou, Post-filtering of transform-coded images, in: Proc. SPIE Applications of Digital Image Processing XI, Vol. 974, 1988, pp. 121–126.

[33] C. van den Branden Lambrecht, Perceptual models and architectures for video coding applications, Ph.D. Thesis, Swiss Federal Institute of Technology, Lausanne, Switzerland, 1996.

[34] A.B. Watson, J.A. Solomon, A.J. Ahumada, DCT basis-function visibility: effects of viewing distance and contrast masking, in: Proc. SPIE Human Vision, Visual Processing, and Digital Display, Vol. 2178, May 1994, pp. 99–108.

[35] Y. Yang, N.P. Galatsanos, A.K. Katsaggelos, Projection-based spatially-adaptive reconstruction of block-transform compressed images, IEEE Trans. Image Process. 4 (July 1995) pp. 896–908.

[36] M. Yuen, H.R. Wu, K.R. Rao, Performance evaluation of POCS loop filtering in generic MC/DPCM/DCT video coding, in: Proc. SPIE Visual Commun. and Image Processing, Vol. 2501, Taipei, Taiwan, May 1995, pp. 65–75.