

CODIFICAÇÃO DE VÍDEO: A GUERRA DOS FORMATOS

Maurizio Pintus Nº 64286

Ivo Vieira Nº 53653

Instituto Superior Técnico

Av. Rovisco Pais, 1049-001 Lisboa, Portugal

E-mail: mauriziontus@tiscali.it, ivo.a.l.vieira@gmail.com

RESUMO

Este artigo retrata as principais normas de codificação de vídeo digital. Começa-se por fazer uma perspectiva histórica da codificação e da necessidade da normalização. As normas são apresentadas cronologicamente, com a indicação do ano em que foram publicadas e das respectivas entidades envolvidas. Houve uma preocupação em descrever as principais ferramentas de codificação que cada norma apresenta, como uma novidade em relação às suas antecessoras. É feita também uma breve referência aos formatos privados mais utilizados. Cada norma é concluída com a indicação das principais aplicações a que se destina e indicadores de desempenho. Há depois uma menção do problema das patentes. Conclui-se com uma avaliação das normas mais utilizadas actualmente e uma perspectiva do que se vai passar no futuro próximo.

Index Terms — H.261, H.263, MPEG-1, MPEG-2, MPEG-4, AVS, AVC, MJPEG, RealVideo, WMV, vídeo, norma.

1. INTRODUÇÃO

Nos finais dos anos 70 começou-se a usar vídeo digital, para um melhor processamento do vídeo analógico, em sistemas maioritariamente analógicos. No início dos anos 80 começa a difusão dos primeiros sistemas completamente digitais e, como consequência, a necessidade de normas internacionais para a regulamentação destes. Na ilustração 1 fica a cronologia das principais normas. Estas vêm regulamentar um conjunto de ferramentas de codificação, algumas das quais já existentes há alguns anos. Esta normalização foi, e continua a ser, de extrema importância para garantir a interoperabilidade entre os produtos de diferentes empresas. Uma norma, ao mesmo tempo, deve ser o mais flexível possível, especificando só o necessário para garantir a tal interoperabilidade. Esta flexibilidade é importante para garantir uma saudável competição entre empresas, bem como para aproveitar o aparecimento de novas tecnologias e resistir ao longo do tempo.

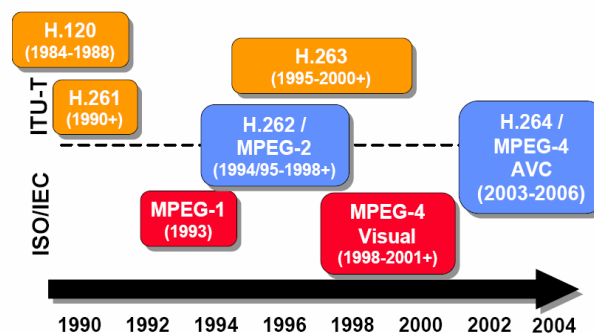


Ilustração 1. Cronologia das principais normas de codificação vídeo[1].

2. O INÍCIO

A primeira norma de codificação digital de vídeo foi CCIR (Comité Consultatif International pour la Radio) 601. Esta norma apareceu em 1982, pelo então CCIR, hoje ITU-R (International Telecommunication Union - Radiocommunication Sector) e usa PCM (Pulse Code Modulation) [2]. Isto significa que não há compressão e o débito binário é da ordem dos 216 Mbit/s. Este débito é extremamente elevado e faz com que esta norma apenas seja usada em estúdios de televisão. Em 1984 surgiu a primeira norma que introduziu compressão, designada ITU-T (ITU - Telecommunication Sector) H.120 [2]. A versão 1 desta norma implementava DPCM (Differential PCM), diferentes níveis de quantificação, codificação de comprimento variável, codificação inter-trama – manda apenas um novo bloco, se este for diferente do bloco da imagem anterior – e possibilidade de mudar para amostragem “quincux”. A versão 2 surgiu 4 anos depois e acrescentou compensação de movimento. Esta característica consiste basicamente em pesquisar blocos parecidos na imagem anterior ao bloco a codificar. Esta norma permite débitos binários de 1544 a 2048 kbit/s e destinava-se essencialmente a videoconferência.

3. H.261

Em 1990 surge a norma ITU-T H.261 que introduz, pela primeira vez, um esquema de codificação híbrida, com o uso da transformada DCT (*Discrete Cosine Transform*), da compensação de movimento e da codificação entrópica.

Cada trama é dividida em GOB (*Group Of Blocks*) e cada um deles constituído por 33 MB (MacroBlocos). Cada um destes MB é constituído por 4 blocos de luminância e 2 de crominância. Um bloco, por sua vez, é formado por 8 x 8 pixels e é terminado por uma palavra de EOB (*End Of Block*).

Existem 2 resoluções [3]:

- CIF (Common Intermediate Format): 352 x 288 pixels para a luminância (Y) e 176 x 144 pixels para cada uma das crominâncias (U, V). Isto significa que o formato é 4:2:0, posicionados em “quincunx”, progressivos, 30 tramas/s e com factor de forma 4/3. A trama é dividida em 12 GOB.
- QCIF (*Quarter CIF*): 176 x 144 pixels para a luminância e 88 x 72 pixels para cada uma das crominâncias. O ritmo das tramas e o factor de forma são os mesmos que no CIF. A trama é dividida em 3 GOB.

A norma disponibiliza dois modos de codificação dos MB:

- Inter: compara a imagem a ser codificada com a imagem anterior. Usado quando há bastante redundância temporal.
- Intra: assume-se que a imagem a ser codificada é independente da anterior. Usado quando não há, ou não se quer explorar, redundância temporal.

A primeira imagem, obviamente, é sempre codificada no modo Intra. As imagens seguintes, se codificadas no modo Inter, são comparadas apenas com a imagem anterior.

A estimativa de movimento é feita pesquisando, numa área de 32 x 32 pixels da imagem anterior, o MB mais parecido com aquele que se está a codificar, enviando-se o vector que representa a posição relativa entre os dois MB. Se estes dois MB forem significativamente diferentes envia-se também a diferença. Esta diferença é codificada usando a transformada DCT, que vai fazer a transformação do domínio espacial para o domínio frequencial. Os coeficientes DCT a transmitir são seleccionados e quantificados de maneira a poder-se explorar a irrelevância espacial. Estes coeficientes são varridos em zig-zag, começando nos de menor frequência, uma vez que o ser humano é mais sensível às baixas frequências. Na transmissão de um coeficiente não nulo é enviada o seu nível e a sua posição, sendo que esta é definida pelo número de coeficientes nulos que o precedeu. A norma usa 31 níveis de quantificação, sendo estes todos os números pares

entre 2 e 62. Por fim, é feita uma codificação entrópica, tendo como principal ferramenta a codificação de Huffman, para explorar a redundância estatística. O diagrama de blocos dum codificador H.261 é presente na ilustração 2.

Tendo em conta que a produção de bits pelo codificador é não uniforme, e o canal exige um fluxo constante, é usada uma memória de saída.

A correcção de erros é feita através de um código de blocos BCH (511, 493) – Bose-Chaudhuri-Hocquenghem – e o seu uso na descodificação é opcional.

A principais aplicações desta norma são a videotelefonia e a videoconferência, em canais síncronos a $p \times 64$ kbit/s, com $p = 1, \dots, 30$.

Embora esta norma tenha sido importante no estabelecimento de requisitos de compatibilidade, que influenciaram normas posteriores, ela já está ultrapassada.

4. MPEG-1

A versão final da norma MPEG-1 (*Moving Picture Experts Group*) foi aprovada em novembro de 1992 [4]. Esta norma mantém as seguintes características da H.261:

- Compensação de movimento aplicada aos MB 16 x 16;
- DCT aplicada aos blocos 8 x 8;
- Codificação de comprimento variável.

Novas características:

- Compensação de movimento bidireccional;
- Estimativa de movimento com resolução de $\frac{1}{2}$ pixel;
- Codificação de imagens só em componente DC;
- Matrizes de codificação ponderadas.

A primeira das novas características que a norma introduz, baseia-se na transmissão de tramas de 3 tipos: Intra (I), Preditas (P) e preditas Bidireccionalmente ou interpoladas (B). As tramas I são codificadas independentemente. As tramas P são codificadas tendo em conta as diferenças encontradas na trama I ou P anterior. Já as tramas B são codificadas tendo em conta a trama anterior e seguinte, também chamadas tramas âncoras, sendo estas I ou P, como se pode ver na ilustração 3. Uma sequência de tramas é organizada em GOP (*Group Of Pictures*) no qual há, pelo menos, uma trama I e um número variável de tramas P e B. Este tipo de codificação tem como principal diferença, em relação à norma H.261, a possibilidade de armazenar em memória tramas posteriores em relação à que está a ser codificada, para uma melhor predição. Esta característica é feita à custa de um atraso inicial que, para videotelefonia e videoconferência, não é aceitável.

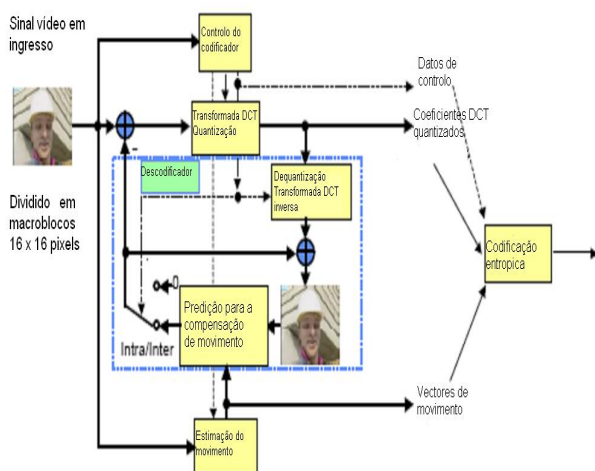


Ilustração 2. Diagrama de blocos dum codificador H.261 e seguintes.

Na MPEG-1 é introduzida uma estimativa de movimento mais. Esta estimativa é feita através de novos pixels, calculados pela média dos pixels adjacentes já existentes na trama, e que serão colocados entre estes. Desta maneira, consegue-se uma diminuição do erro de predição.

Outra novidade introduzida na norma MPEG-1 é a codificação de imagens só em componente DC. Esta característica codifica só a componente de mais baixa frequência de cada trama, para uma procura mais rápida de partes do vídeo.

São usados dois tipos de matrizes de quantificação: Intra e Inter. Nas matrizes Intra, os coeficientes mais altos traduzem altas frequências e, por sua vez, os coeficientes mais baixos, traduzem baixas frequências. Nas matrizes Inter todos os seus coeficientes têm o mesmo valor, porque estas representam uma diferença entre imagens e não uma imagem real.

O aparecimento da norma MPEG-1 tinha como principal objectivo o armazenamento digital de vídeo, com uma qualidade subjectiva igual ou superior ao VHS (*Video Home System*). Este armazenamento é feito, essencialmente em CD-ROM. Como a norma não tem preocupações de requisitos de tempo real, é possível esperar por tramas posteriores em relação à que se está a codificar. A norma permite o acesso aleatório, num tempo limitado (tipicamente 0.5 s), devido à presença de tramas Intra. É também possível o avanço e recuo rápido com um factor de velocidade entre 8 e 10, saltando a leitura das tramas Inter. O débito binário usado normalmente é, aproximadamente, 1.5 Mbit/s e para este débito a qualidade é superior à da H.261. Tipicamente são usadas imagens não interlaçadas, com resolução espacial CIF, a 25 Hz.

A norma não faz referência à correcção de erros, ficando esta característica ao critério de cada fabricante.

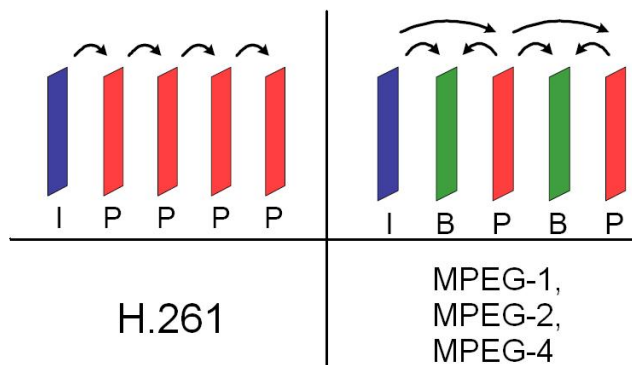


Ilustração 3. Imagens predição para a compensação de movimento em várias normas.

Isto torna a norma flexível, uma vez que não limita a criatividade e esforço dos fabricantes em fazer codificadores cada vez melhores. No final, sai a ganhar o consumidor dado que vai ter acesso a diferentes soluções que variam na sua qualidade e preço.

5. MPEG-2

Em Novembro de 1994, o ITU-T em conjunto com o ISO (*International Organization for Standardization*) estabeleceram a norma MPEG-2 (H.262 para o ITU-T). Esta norma mantém todas as características do MPEG-1, contudo, acrescenta uma codificação de conteúdos entrelaçados e várias formas de escalabilidade.

Para codificar em modo Intra imagens entrelaçadas, são possíveis duas estruturas: trama ou campo. Na primeira, o MB é tratado como se a imagem fosse progressiva. Este, para a aplicação da DCT, divide-se em quatro blocos em que cada um contém linhas pares e ímpares. Esta estrutura é mais indicada quando existe pouco movimento, porque existe pouca diferença entre dois quadros adjacentes. Na estrutura campo, o MB divide-se também em quatro blocos, porém, dois destes blocos contém linhas pares e os outros dois, linhas ímpares. Esta estrutura é mais indicada quando há muito movimento e pouco detalhe, uma vez que ao fazer a junção de linhas ímpares ou pares, vai fazer que dois pixels distantes sejam codificados como vizinhos. A ilustração 4 ilustra os dois modos. No caso de estimativa de movimento, são possíveis três modos. No modo trama, um MB dum campo ímpar é usado pela predição do MB correspondente no campo ímpar seguinte, e o mesmo pelos campos pares. O vector de movimento correspondente tem a duração de uma imagem (40 ms a 25 Hz). No modo campo, um MB é codificado em relação ao MB correspondente do campo anterior, e o vector de movimento correspondente tem a duração de um campo (20 ms a 25 Hz). No modo misto, a codificação é feita em relação ao MB de duas tramas.

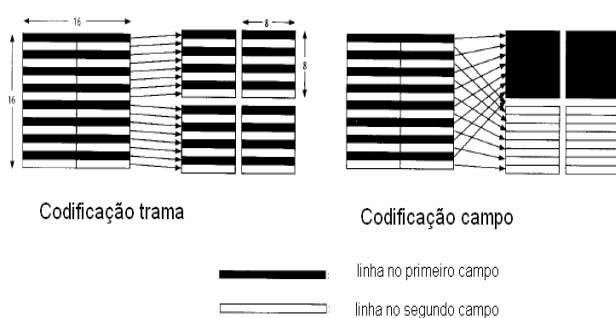


Ilustração 4. Os dois modos de codificação para imagens entrelaçadas na MPEG-2.

Para o varrimento dos coeficientes DCT, em vez da ordem em zig-zag, pode ser usado, para as imagens-trama, a ordem alternada, apresentado na ilustração 5, onde se codificam antes os coeficientes correspondentes a transições verticais. Isto porque a correlação vertical para imagens com muito movimento é pouca em relação à correlação horizontal (duas linhas adjacentes são adquiridas muito depois de que dois pixels adjacentes na mesma linha).

A escalabilidade é uma funcionalidade que visa uma descodificação útil do sinal comprimido, descodificando apenas algumas das suas partes. A norma prevê diferentes formas de escalabilidade. Com a escalabilidade na frequência (ou *data partitioning*) os diferentes níveis contêm um diferente subconjunto dos coeficientes DCT, correspondentes a frequências diferentes. A norma prevê dois níveis, o nível base e o nível de melhoramento. Ao primeiro atribui-se dados de endereçamento e controlo, bem como os coeficientes DCT de mais baixa frequência. Este nível é transmitido num canal de mais alta prioridade ou sujeito a um código de correcção de erros mais poderoso. O nível de melhoramento contém o restante da informação codificada. Com a escalabilidade espacial, o vídeo é codificado em vários níveis com resoluções espaciais diferentes. O nível de base é constituído por tramas I e P, enquanto os níveis de melhoramento são codificados como um conjunto de tramas P e B. A trama actual de um nível de melhoramento pode ser prevista fazendo uso da trama correspondente do nível anterior ou da trama anterior do mesmo nível. Com a escalabilidade de qualidade, ou de SNR (*Signal-to-Noise Ratio*), muda a precisão da quantificação nos níveis. O nível base é codificado com degraus de quantização mais largos; os níveis de melhoramento contêm a diferença entre o sinal do nível anterior e o sinal original, codificada com degraus de quantificação mais largos. A escalabilidade temporal permite ter níveis com resoluções temporais diferentes. No nível base, são enviadas apenas algumas imagens da

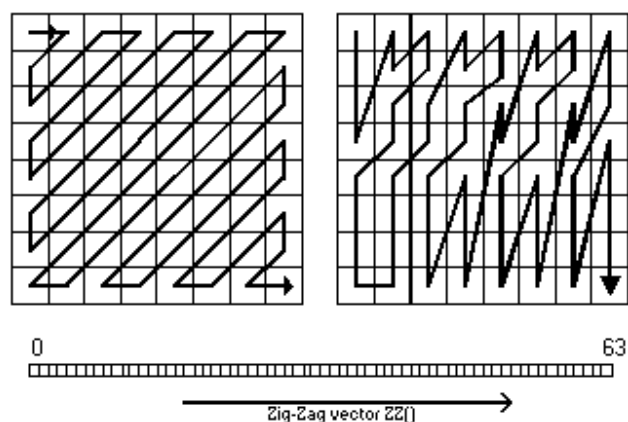


Ilustração 5. Diferença entre o varrimento a zig-zag (à esquerda) e o varrimento alternado (à direita)[5].

sequência codificada. Nos níveis de melhoramento envia-se as restantes. É também possível combinar as várias formas de escalabilidade para obter uma escalabilidade híbrida.

A escalabilidade é útil para, por exemplo, transmitir o mesmo vídeo a decodificadores normais, que o devem descodificar com a resolução básica, e a decodificadores HD (*High Definition*), que o podem descodificar com a resolução mais alta. Também é útil para garantir, no mesmo decodificador, uma qualidade básica em condições difíceis e uma melhor qualidade com melhores condições de recepção.

O MPEG-2 organiza o conjunto das ferramentas de codificação que disponibiliza por perfis. Cada perfil é um subconjunto de funcionalidades, definidos para satisfazer diferentes conjuntos de aplicações, e que podem ser suportadas de forma separada. São também definidos níveis de qualidade. Para cada nível é prevista uma gama de variação de parâmetros, como resolução espacial e temporal. A norma MPEG-2 prevê quatro níveis de resolução, desde a resolução SIF (até 360 x 288) até a resolução para HDTV em ecrã largo (até 1920 x 1152). Prevê, depois, vários perfis.

MPEG-2, pela possibilidade de codificar imagens entrelaçadas, como a compatibilidade com os aparelhos de televisão analógica requer, é usado na transmissão da televisão digital, também em definição normal e HD, em vários canais: satélite, cabo e antenas terrestres. Uma outra grande aplicação da norma é a gravação de vídeo em DVD (*Digital Video Disc*).

Inicialmente, eram previstas duas normas: uma para a resolução normal e uma para a alta resolução. A norma MPEG-2 acabou por incluir, também, a alta resolução e, como consequência, não surgiu a norma MPEG-3.

6. H.263

A norma H.263 começou a ser desenvolvida em 1993 pelo ITU-T *Study Group 15* e a norma foi publicada no Março de

1996. O algoritmo de codificação da H.263 é similar ao que foi usado na H.261, contudo, com alguns melhoramentos e mudanças para um melhor desempenho e resistência a erros. Destacam-se:

- Mais formatos de imagens e estruturas de GOB diferentes. Para além dos formatos CIF e QCIF, suportados pela H.261, a H.263 acrescenta os formatos Sub-QCIF, 4CIF e 16CIF. São mantidos os formatos para as crominâncias (4:2:0) e as resoluções para a luminância, em pixels, são, respectivamente, 128 x 96, 704 x 576 e 1408 x 1152. Na H.263, os GOB são estruturados numa ou mais linhas completas de MB.
- Codificação de vector de movimento e predição de $\frac{1}{2}$ pixel. É usada a média dos vectores de movimento de 3 MB, vizinhos (esquerda, cima e diagonal superior direita) do MB corrente, para a predição do vector de movimento deste último. O erro de predição é codificado e transmitido. Se um destes vizinhos estiver fora da fronteira duma imagem ou dum GOB, é usado um vector nulo em substituição do vector de movimento desse MB. Se dois vizinhos estiverem nessa situação, é usado, como predição, o vector de movimento do único MB do lado de dentro. A predição de $\frac{1}{2}$ pixel é feita como se descreveu na norma MPEG-1.
- 3D VLC. A norma H.263 acrescenta um elemento “último” ao par (posição, nível), para indicar se o coeficiente corrente é (último=1), ou não (último=0), o último coeficiente não nulo do bloco. Ficamos assim com um conjunto (último, posição, nível) e, desta maneira, não será necessário o envio da palavra de código EOB, usada na H.261.

A norma especifica 6 de modos de operação opcionais:

- Modo de Presença Contínua em Multiponto (Anexo C). Este modo permite enviar múltiplos fluxos de vídeo dentro de um único canal vídeo.
- Modo de Vector de Movimento Não Restrito (Anexo D). Os vectores de movimento podem apontar para pixels fora da imagem de predição. Isto significa que a predição para um MB a codificar pode ser um MB que fica em parte na imagem anterior e em parte fora, com a parte que falta enchida por uma repetição dos pixels no bordo da imagem. Isto é útil sobretudo se a resolução espacial for pequena. A segunda característica que este modo introduz é uma extensão do intervalo possível para os vectores de movimento, que agora podem ir desde -30 até 30, para as duas direcções da imagem, horizontal e vertical. Isto é útil quando há muito movimento entre duas imagens

consecutivas, por exemplo no caso duma resolução temporal baixa.

- Modo de Codificação Aritmética (Anexo E). A norma introduz a possibilidade de usar o código aritmético para explorar a redundância estatística, em vez do código de Huffman. Isto permite alcançar factores de compressão com um ganho de cerca de 10 % na codificação de imagens Intra e de cerca de 4 % na codificação de imagens Inter.
- Modo de Predição Avançado (Anexo F). É possível com este modo usar um ou quatro vectores de movimento por MB, ou seja, um para cada bloco. A predição para cada pixel do MB (ou bloco) é, depois, a soma pesada de três valores obtidos com três vectores de movimento: aquele do MB corrente, aquele do MB mais perto do pixel a codificar na direcção horizontal e aquele do MB mais perto do pixel a codificar na direcção vertical
- Modo de Tramas PB (Anexo G). Este modo introduz a codificação de duas imagens como uma só. A primeira, chamada trama P, é codificada em relação à imagem já codificada anterior. A segunda, chamada trama B, é codificada só através de vectores de movimento, em relação a uma interpolação linear entre a imagem anterior e a imagem P.
- Correção de Erros, como na norma H.261 (Anexo H).

O decodificador diz ao codificador quais as opções que consegue decodificar. Se o codificador suportar algumas dessas opções, pode activá-las. Estes modos de operação têm um significado semelhante aos perfis, descritos na norma MPEG-2.

Embora a H.263 tenha sido criada com o objectivo de desenhar uma norma capaz de funcionar bem em muito baixos débitos (cerca de 20 Kbit/s), depressa chegou-se à conclusão que a H.263 podia providenciar um significativo melhoramento da H.261 a qualquer débito. Mais especificamente, a H.263 pode poupar mais de 50% no débito necessário para representar vídeo a uma qualidade perceptual aceitável a muito baixos débitos, relativamente à H.261. Também a esses débitos, o H.263 consegue um ganho de 3 dB na SNR, em relação à H.261. De facto, a H.263 consegue uma codificação mais eficiente que a H.261 a todos os débitos. Comparando com a MPEG-1, a H.263 consegue também poupanças significativas a altos débitos (por ventura 30% a cerca de 1 Mbit/s) [6].

6.1. H.263+

Depois de estandardizada a versão 1 da H.263, um contínuo interesse nas capacidades do seu desenho básico, depressa tornou claro que era possível realizar futuros melhoramentos na norma. Assim, em Janeiro de 1998 a segunda versão da

norma, conhecida como H.263+, é aprovada. Esta manteve todo o conteúdo técnico da versão anterior, contudo, acrescentou mais 16 anexos, como modos opcionais. Estes vêm possibilitar que esta nova versão se destaque da sua precedente pelas seguintes características:

- Capacidade de providenciar uma alta resiliência a erros em redes móveis e de pacote.
- Melhoramentos na eficiência da codificação.
- Resolução dinâmica para uma adaptação da resolução de uma imagem codificada ao conteúdo.
- Suporte de uma grande variedade de resoluções de imagens de vídeo e débitos de tramas.
- Escalabilidade espacial, de qualidade e temporal para operações de multi-débitos simultâneas.

7. MPEG-4

7.1. Parte 2 (MPEG-4 Visual)

Esta norma é apresentada em 1998 pelo ISO e é similar às suas versões anteriores, MPEG-1 e MPEG-2, pois, tal como estas, baseia-se na compressão por DCT. A norma basicamente divide-se em 19 perfis [7], entre os quais se destacam o Perfil Simples – *Simple Profile* (SP) e o Perfil Simples Avançado – *Advanced Simple Profile* (ASP).

7.1.1. SP

Este perfil, o mais básico da norma MPEG-4, é, de certo modo, similar à norma H.263. Suporta só o varrimento progressivo e disponibiliza as seguintes novas ferramentas: compensação de movimento usando blocos de tamanho variável, predição dos coeficientes DCT em modo Intra, manuseamento de quatro fluxos na maioria dos níveis e algumas ferramentas de resiliência a erros. É usado normalmente quando certas condições de aplicações, como o tamanho de certos dispositivos ou a largura de banda de uma rede, exigem uma resolução e débito baixos. Como exemplos de aplicações que fazem uso deste perfil temos: câmaras de vídeo, telemóveis, sistemas de vigilância, etc.

7.1.2. ASP

O perfil suporta o entrelaçamento e faz uso de tramas B. Estas técnicas são desenhadas, basicamente, de uma maneira similar às já descritas aquando da norma MPEG-2. São suportadas as resoluções até à resolução da *Standard Definition Television* (720 × 576).

MPEG-4 ASP acrescenta uma compensação de movimento global e de $\frac{1}{4}$ de pixel. A primeira destas técnicas não é, actualmente, suportada na maioria das implementações, apesar de oficialmente a norma requerer

que os decodificadores a suportem. A maior parte dos codificadores também não a suporta e alguns especialistas dizem que, normalmente, a técnica não traz nenhum benefício em termos de compressão. Quando utilizada, a compensação de movimento global do ASP, apresenta um grande impacto negativo na velocidade e acrescenta uma considerável complexidade à aplicação. A compensação de movimento de $\frac{1}{4}$ de pixel é também omissa em algumas aplicações (embora não tanto como a técnica anterior), uma vez que apresenta, também, um certo prejuízo na velocidade nem sempre compensada em termos de qualidade de visualização.

Outros perfis disponibilizam codificação de vídeos de forma variável, representação integrada de sequências mono, stereo ou com múltiplas vistas (incluindo representações 3D), com possibilidade de novas perspectivas nomeadamente em termos do “realismo” das aplicações [8]. MPEG-4 parte Visual é cerca 1,4 vezes mais eficiente da MPEG-2.

7.2. Parte 10 (H.264/AVC)

Em 1998 o VCEG (*Video Coding Experts Group*) começou a desenvolver uma norma com um débito binário mais baixo do que o disponibilizado na MPEG-2. Em 2001 o VCEG juntou-se ao MPEG e esses formaram a *Joint Video Team*, que publicou a norma H.264/AVC em Março de 2003. A norma disponibiliza um novo nível na hierarquia duma sequência de vídeo: um grupo de *slices*, que pode conter desde um ou mais *slices*. A dimensão dos *slices* é flexível, como na norma MPEG-1 e ao contrário da MPEG-2. Em relação ao entrelaçamento, são definidos três tipos de codificação: um tipo de codificação campo e dois tipos de codificação trama. Com a codificação campo, cada campo é codificado como uma imagem separada, usando os campos para a compensação de movimento. Para a codificação trama do primeiro tipo, é a trama completa que vem codificada como uma imagem separada; para a do segundo tipo, em cada par de MB há uma mudança entre codificação trama e campo.

Existem várias diferenças com as normas anteriores em relação à compensação de movimento. Agora os blocos utilizados têm uma dimensão e uma forma variável, com uma dimensão mínima de 4 × 4. Como na MPEG-4 Visual, mas apresentando uma complexidade menor, existe uma compensação de movimento com precisão de $\frac{1}{4}$ de pixel, em vez da precisão de $\frac{1}{2}$ pixel presente na MPEG-2. Existem vectores de movimento que apontam para fora dos limiares da imagem, como na H.263. As tramas P podem ser codificadas em referência a mais de uma imagem anterior. As tramas B podem ser codificadas em referência a mais de uma imagem anterior e mais de uma imagem seguinte. Não existe uma dependência forte entre a ordem das imagens para compensação de movimento e a ordem das imagens para a visualização. Na AVC também as tramas B podem ser usadas como referência para outras imagens.

É presente uma predição pesada para a compensação de movimento, com a qual o sinal de predição pode ser pesado e deslocado, segundo quantidades especificadas pelo codificador. A H.264 AVC deduz o movimento nas áreas saltadas, uma vez que estas contêm pouco movimento, em vez de deixar a área fixa. A norma melhora o método de inferência de movimento para *slices* B, conhecido como “compensação de movimento directa”, já presente na H.263+ e na MPEG-4 visual. O filtro de *deblocking*, útil para eliminar o efeito de bloco e melhorar a qualidade subjectiva do vídeo, é usado agora na malha de predição para a compensação de movimento, com melhores resultados do que um filtro usado em processamento posterior, melhorando também a capacidade de predição.

Existem outras diferenças, relativamente às normas anteriores, na exploração da redundância espacial. Na predição Intra, se forem do mesmo *slice* (para garantir a independência dos *slices*), os MB podem ser codificados em relação aos blocos ou MB adjacentes na mesma imagem. É também possível codificar um MB Intra em relação a um MB Inter; com o modo *Constrained Intra Coding Mode*, só os MB Intra são usados como predição.

A norma disponibiliza três tipos de transformada, usadas em relação ao tipo de resíduo a codificar:

- A transformada de Hadamard (mais eficiente mas também muito mais complexa que a transformada DCT) 4 x 4 é usada para codificar os coeficientes DC da luminância num MB codificado no modo Intra 16 x 16, portanto, sem subdivisão do MB;
- A transformada de Hadamard 2 x 2 é usada para codificar os coeficientes DC da crominância em todos os MB;
- A transformada DCT inteira (mais fácil de implementar do que a DCT normal) 4 x 4 é usada para codificar todos os outros blocos.

A transformada DCT é aplicada não só aos blocos de dimensão 8 x 8, mas também aos blocos de dimensão 4 x 4 e 16 x 16. Para a computação dos coeficientes DCT, os codificadores e decodificadores que usam esta norma precisam de uma aritmética a 16-bit e não a 32-bit, como naqueles que usam as normas anteriores. O uso da transformada DCT inteira garante que todos os decodificadores decodifiquem o mesmo vídeo com a mesma qualidade.

Na exploração da redundância estatística, a novidade é o uso duma codificação entrópica *Context-Adaptive Binary Arithmetic Coding*, uma técnica de codificação aritmética que se adapta ao conteúdo e, por isso, mais eficiente do que a usada pela norma H.263.

A uma qualidade fixada, o débito binário dum vídeo codificado H.264/AVC é metade da dum vídeo codificado

em MPEG-2. A norma foi produzida para muitas aplicações, como:

- Televisão;
- Gravação em DVD e similares;
- Vídeotelefonia e videoconferência em LAN, DSL, redes sem fios e móveis;
- *Video-on-demand* e *streaming* em ISDN, DSL, redes sem fios.

A sua difusão ainda é limitada, principalmente, à Internet e dispositivos móveis, que oferecem uma banda mais pequena daquela disponível em canais *broadcast*. Nas outras aplicações, como a televisão e a gravação em DVD, ainda existe uma inércia em aderir à norma MPEG-4, em parte devido ao bom desempenho que a MPEG-2 ainda oferece.

Existem depois algumas extensões da AVC, como a SVC (*Scalable Video Coding*), a MVC (*Multiview Video Coding*). A primeira tem como objectivo fornecer várias formas de escalabilidade, com um débito binário e uma complexidade parecidas a da norma AVC básica. Já as normas MPEG-2 e MPEG-4 garantiam escalabilidade, mas com a necessidade dum débito binário e duma complexidade do decodificador demasiado altos. A MVC quer oferecer a possibilidade de ver o mesmo vídeo com diferentes pontos de vista.

8. OUTROS FORMATOS

8.1. MJPEG

Este formato livre, aprovado pela ISO no Novembro de 2001, é baseado na norma MJPEG, baseada pela sua vez na norma JPEG, usada pela codificação de imagens. É suportado pela *Digital Cinema Initiatives*, um consórcio que junta muitos grandes estúdios cinematográficos, que a utiliza para cinema digital. Ao contrário de todas as outras grandes normas, não explora a redundância temporal; cada trama é codificada em modo Intra com uma variante da norma para imagens JPEG2000 [9].

8.2. AVS

Existem outras normas privadas, entre as quais se destaca a *Audio and Video Coding Standard* (AVS). Esta norma foi desenvolvida desde Junho de 2002, pelo departamento de ciência e tecnologia da China. Em Abril de 2005 a norma foi aprovada e passou a ser a norma nacional [10]. É utilizada, na China, para *broadcast* de alta resolução, gravação em DVD, comunicação sem fios e para multimédia na Internet. A propriedade intelectual deste formato é possuída pela China.

Os autores dizem que AVS tem uma eficiência de codificação parecida à da AVC e 2-3 maior da MPEG-2.

8.3. RealVideo

RealVideo é um formato propriedade da RealNetworks. A primeira versão deste formato surgiu em 1997 e era baseada na H.263 até que, a sua versão 8, passa a usar uma codificação própria. Em 2006 surgiu a versão 10, a mais recente. A principal aplicação deste formato é o *streaming* de vídeo pela Internet. Os proprietários reclamam, para vídeos com a mesma qualidade, um débito binário 80% menor do da MPEG-2, 15 % menor do da H.264 e 30 % menor do da WMV 9 [11].

8.4. WMV

Windows Media Video (WMV) é propriedade da Microsoft e foi originalmente desenhado para aplicações de *streaming* na internet, como um competidor do RealVideo. A primeira versão do *codec* foi introduzida em 1999 e construída a partir da implementação da MPEG-4 parte 2, feita pela Microsoft. Em Março de 2006 o WMV 9 ganhou a adopção, atribuída pela *Society of Motion Picture and Television Engineers* para os formatos de distribuição física de HD DVD e Blu-ray Disc, com o nome de VC-1. A Microsoft afirma que a eficiência da compressão é 3 vezes maior da MPEG-2 e 2 vezes maior da MPEG-4 [12].

9. PATENTES

As normas MPEG-2, MPEG-4 Parte 2 e H.264/AVC são normas abertas, no sentido que podem ser conhecidas por todos, mas prevêem que os fornecedores de produtos e serviços que as utilizam tenham de pagar pelas tecnologias usadas. A maior entidade envolvida no *licensing* para as normas MPEG é uma organização privada, MPEG LA. A duração das patentes é, normalmente, 20 anos. Como exemplo, uma empresa tem de pagar 0.20 \$ cada ano para cada codificador ou decodificador H.264, depois das primeiras 100 000 unidades. Sobre as 5 milhões de unidades por ano, os direitos de autores são 0.10 \$ para cada unidade.

As patentes para MPEG-2 são possuídas por 25 empresas, as de MPEG-4 Visual são possuídas por 28 empresas, as de AVC/H.264 por 23 empresas e as de VC-1 por 18 empresas [13].

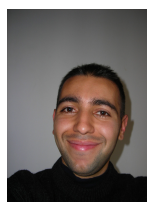
O facto de uma norma conter tecnologias patenteadas ou não, determina muitas vezes a escolha de um formato ou outro.

10. CONCLUSÃO

O MPEG-2 é ainda o formato de codificação vídeo mais utilizado mas, provavelmente, irá ser substituído em poucos anos pelo MPEG-4 AVC, que garante uma codificação muito mais eficiente. Entre os formatos não MPEG, o AVS é agora o mais importante e o único que pode lutar contra o AVC, sobretudo porque é suportado por uma grande potência como a China.

11. REFERÊNCIAS

- [1] Gary J. Sullivan, "Overview of International Video Coding Standards", *ITU-T VICA Workshop 22-23 July 2005*, ITU Headquarter, Genova.
- [2] Fernando Pereira, "A Digitalização e a Compressão na Era da Convergência" in *Convergências XXI*, APDC, Março 1996.
- [3] Fernando Pereira, cadeira de "Comunicação de Áudio e Vídeo", IST, ano 2007/2008 (Secção Material de Estudo). Disponível: <http://www.img.lx.it.pt/~fp/cav/com.htm>
- [4] Sítio oficial da MPEG: <http://www.chiariglione.org/mpeg>
- [5] Chad Fogg, "MPEG-2 FAQ - Questions that Should Be Frequently Asked About MPEG", version 3.8 (April 2, 1996) by Chad Fogg. Disponível: http://bmrc.berkeley.edu/research/mpeg/faq/mpeg2-v38/faq_v38.html
- [6] Tsuhan Chen, Atul Puri, *Multimedia Systems, Standards, and Networks*, Chapter 3, "H.263 (Including H.263+) and Other ITU-T Video Coding Standards". Disponível: <http://books.google.com/books?id=2YhiCRZNUfwC&printsec=frontcover>
- [7] Fernando Pereira, Paulo Nunes, "Levels for MPEG-4 Visual Profiles", no sítio da MPEG Industry Forum. Disponível: <http://www.m4if.org/resources/profiles/index.php#MPEG42>
- [8] P. Kauff, J.-R. Ohm, S. Rauthenberg and T. Sikora, "The MPEG-4 Standard and its Applications in Virtual 3D Environment".
- [9] Sítio oficial da MJPEG. Disponível: <http://www.jpeg.org/jpeg2000/j2kpart3.html>
- [10] Sítio oficial da AVS. Disponível: <http://www.avs.org.cn/en>
- [11] Sítio oficial da RealNetworks. Disponível: <http://www.realnetworks.com>
- [12] Sítio oficial da Windows Media. Disponível: <http://www.microsoft.com/windows/windowsmedia/>
- [13] Sítio oficial da MPEG - LA. Disponível: <http://www.mpegla.com/>



Maurizio Pintus, aluno da universidade Università degli studi di Cagliari, Italia, encontra-se a frequentar, como ERASMUS, o 5º ano do Mestrado em Engenharia Electrotécnica e de Computadores, no Instituto Superior Técnico, com o nº 64286.



Ivo Vieira, aluno do Instituto Superior Técnico, encontra-se a frequentar o 4º ano do Mestrado em Engenharia Electrotécnica e de Computadores, com o nº 53653.