

## 6

### Referências

AYGÜN, R. S. **Spatio-temporal browsing of multimedia presentations.** PhD dissertation. New York: USA, May 2003.

BOCCIGNONE, G., SANTO M. De, PERCANNELLA G. **An Algorithm for Video Cut Detection in MPEG Sequences.** In: IS&T SPIE Proceedings: Storage and Retrieval for Media Databases, 2000, San Jose, California.

BOCCIGNONE, G.; SANTO, M. De, PERCANNELLA, G. **A System for Parsing MPEG Videos.** In: IS&T SPIE Proceedings: Internet Imaging, January 2000, San Jose, California.

BOCCIGNONE, G.; SANTO, M. De; PERCANNELLA, G. **Automated Threshold Selection for the Detection of Dissolves in MPEG Video.** In: IEEE Proc. of ICME, July 2000, New York, USA.

CALIC J.; IZQUIERDO E. **Towards Real-Time Shot Detection in the MPEG-Compressed Domain. Image Analysis for Multimedia Interactive Services.** In: EURASIP Journal on Applied Signal Processing. 2001.

CALIC, J.; SAV, S.; IZQUIERDO, E.; MARLOW, S.; MURPHY, N.; O'CONNOR, N.E. **Temporal Video Segmentation For Real-Time Key Frame Extraction.** In: ICASSP, 2002.

CECCARELLI, M.; HANJALIC, A.; LAGENDIJK, R. L. **A sequence analysis system for video databases.** In: Time-Varying Image Processing and Moving Object Recognition, Vol. 4, pp 133-138, - V. Cappellini (Ed.), 1997 Elsevier Science. 1997

CHEN, A. L. P.; LIU, C.-C.; KUO, T. C. T. **Content-based Video Data Retrieval.** In: Proc. National Science Council (ROC), Vol. 23, No. 4, 1999.

CHEN, A. L. P.; KUO, T. C. T. **A Mask Matching Approach for Video Segmentation on Compressed Data.** In: Source Information

Sciences—Informatics and Computer Science: An International Journal archive, Volume 141 , Issue 1-2, March, 2002.

DREW, M. S.; LI, Z.-N.; ZHONG, X. **Video dissolve and wipe detection via spatio-temporal images of Chromatic histogram differences.** In Proceeding of IEEE Int. Conf. on Image Processing (ICIP 2000), volume 3, pages 909-932, 2000.

EKIN , A. **Sports Video Processing for Description, Summarization, and Search.** PhD dissertation. University of Rochester. Rochester: New York, 2003.

EWERTH, R.; FREISLEBEN, B. **Improving Cut Detection in MPEG Videos by GOP-Oriented Frame Difference Normalization.** Pattern Recognition, 17th International Conference on (ICPR'04). **Cambridge UK.** August 2004. Volume 2.

FARAG, W. E.; A-WAHAB, H. **A new paradigm for detecting scene changes on MPEG Compressed Videos.** In: Proc. Of IEEE International Symposium on Signal Processing and Information Technology, pp. 153-158, Cairo, Egypt, Dec. 2001.

HANJALIC, A.; LAGENDIJK, R.L.; BIEMOND, J. **Template-based Detection of Anchorperson Shots in News Programs.** IEEE International Conference on Image Processing (ICIP '98). Chicago : USA, 1998.

HOAD, T.; ZOBEL, J. **Video Similarity Detection for Digital Rights Management.** In: Proceedings of the Australasian Computer Science Conference, M. Oudshoorn (ed). Adelaide: Australia, February 2003. pp. 237-245.

IEC. **International Electrotechnical Commission Homepage.** Disponível em: <http://www.iec.org>. Acesso em: 1 dez. 2004.

ISO. **International Organization for Standardization Homepage.** Disponível em: <http://www.iso.org/> Acesso em : 1 dez. 2004.

ISO; IEC. **International Standard 11172-1**, System Encoding section. 1993

ISO; IEC. **International Standard 11172-2**, Video Encoding section. 1993

ISO; IEC. **International Standard 13818-1**, System Encoding section. 2000

ISO; IEC. **International Standard 13818-2**, Video Encoding section. 2000

ITU-T. **International Telecommunication Union - Telecommunication Standardization Sector (ITU-T) homepage**. Disponível em: <http://www.itu.int/home>. Acesso em: 1 jan. 2005.

IYENGAR, G.; NETI, C. **Speaker change detection using joint audio-visual statistics**. 6th Conference on Content-Based Multimedia Information Access (RIAO2000). Paris: France, April 2000.

JARINA, R.; DUFFNER, O.; MARLOW, S.; O'CONNOR, N.; MURPHY, N.

**MPEG-1 Bitstreams Processing for Audio Content Analysis**. In: Irish Signals and Systems Conference (ISSC 2002). Cork: Ireland, 25-26 June 2002.

JARINA, R.; DUFFNER, O.; MARLOW, S.; O'CONNOR, N.; MURPHY, N. **Rhythm Detection for Speech-music Discrimination in MPEG Compressed Domain**. In: 14th International Conference on Digital Signal Processing, 2002 (DSP 2002). 2002. Vol.1, pp 129- 132.

JUN S.-B.; YOON, K.; LEE, H.-Y. **Dissolve Transition Detection Algorithm Using Spatio-Temporal Distribution of MPEG Macro-Block Types**. Proceedings of the eighth ACM international conference on Multimedia. California: United States, 2000. International Multimedia Conference archive, ACM Press , pp 391 – 394.

KASTURI, R.; JAIN, R. **Dynamic vision**. In: R. Kasturi and R. Jain, editors, Computer Vision: Principles. Washington: IEEE Computer Society Press, 1991. pp 469 – 480.

KIM, Y.-M.; CHOI, S. W.; LEE, S.-W. **Fast Scene Change Detection Using Direct Feature Extraction from MPEG Compressed Videos**. In: 15th International Conference on Pattern Recognition, September, 2000, Barcelona. Vol. 3, pp. 178 – 181.

KIM, N. W.; KANG, E. K.; IM, J. H.; KIM, T. Y. ; CHOI, J. S. **Scene change detection and classification algorithm on compressed video streams**. In: International Technical Conference on Circuits/Systems,

Computers and Communications (ITC-CSCC2001). Tokushima: Japan, July, 2001.

KOBLA, V.; DOERMANN, D. S. ; LIN, K. I ; FALOUTSOS, C.; **Compressed Domain Video Indexing Techniques Using DCT and Motion Vector Information in MPEG Video.** In: Proceedings of SPIE Conference on Storage and Retrieval for Image and Video Databases V. February 1997, San Jose, CA, USA. Vol.3022, pp. 200-211.

KOBLA, V.; DEMENTHON, D.; DOERMANN, D. **Detection of Slow-Motion Replay sequences for identifying Sports Video.** In: Proceedings of IEEE 1999 Workshop on Multimedia Signal Processing. Copenhagen: Denmark, september 1999. Pp 135-140.

KOPRINSKA, I. CARRATO, S. **Temporal Video Segmentation: A Survey.** In: Signal Processing: Image Communications, Vol. 16, pp 477-500. Elsevier Science, 2001.

LEW, M. S.; SEBE, N.; GARDNER, P. C. **Video Indexing and Understanding.** In: Principles of Visual Information Retrieval, Springer-Verlag, London, 2000.

LIENHART, R. **Reliable Dissolve Detection.** In Storage and Retrieval for Media Databases 2001, Proc. SPIE 4315. January 2001. pp. 219-230.

LIENHART, R. **Reliable Transition Detection in Videos: A Survey and Practitioner's Guide.** In: International Journal of Image and Graphics. July, 2001. Vol. 1, Number 3, pp. 469-486.

MARLOW, S.; SADLIER, D. A.; O'CONNOR, N.; MURPHY, N. **Voice Processing for Automatic TV Sports Program Highlights Detection.** Proceedings of 8th International Symposium on Social Communication, Santiago de Cuba, Cuba, 20 -24 January 2003.

MENG, J.; JUAN, Y.; CHANG, S.-F. **Scene Change Detection in a MPEG Compressed Video Sequence.** In: IS&T/SPIE Symposium Proceedings. San Jose, February 1995. Vol. 2419.

MICHAELIS. **Michaelis Dicionário Prático Língua Portuguesa.** Ed. Melhoramentos, 2001.

MPEG. **Moving Picture Experts Group Homepage**. Disponível em: <http://www.chiariglione.org/mpeg/> Acesso em: 15 dez. 2004

NGO, C.; PONG T.; CHIN, R.. **A Survey Of Video Parsing And Image Indexing Techniques In Compressed Domain**. In: Symposium on Image, Speech, Signal Processing, and Robotics (Workshop on Computer Vision) ISSPR. Hong Kong, Sept 1998.

NGO, C. W.; PONG, T. C.; CHIN, R. T. **Camera Breaks Detection by Partitioning of 2D Spatio-temporal Images in MPEG Domain**. In IEEE International Conference on Multimedia Computing and Systems, (ICMCS'99). France, June 1999.

O'TOOLE, C. **An MPEG-1 Shot Boundary Detector Using Colour Analysis of JPEG Images**. In: Technical report [98-03]. School of Computer Applications Dublin City University. September 1998.

PAN, H.; VAN, P.; BEEK, M; SEZAN, I. **Detection Of Slow-Motion Replay Segments In Sports Video For Highlights Generation**. in Proc. IEEE Int. Conf. On Acoustic, Speech and Signal Processing (ICASSP'02). 2002.

PICKERING, M. J. **Video Summarization and Retrieval**. In: MPhil to PhD Transfer Report. May 2002.

PYE, D.; HOLLING, N. J.; MILLS, T. J.; WOOD, K. R. **Audio visual segmentation for content-based retrieval**. In: 5<sup>th</sup> International Conference on Spoken language Processing. Sydney, December 1998.

SMEATON, A. F.; OVER, P.; TABAN, R. **The TREC-2001 Video Track Report**. In: NIST Special Publications 500-250: The Tenth Annual Text Retrieval Conference, Gaithersburg, MD, 52 – 60. 2001

STRACHAN, D. **Video Compression**. In: SMPTE Tutorial 105:68. February 1996.

SUN, H. **MPEG Standarts and its Applications**. MERL – Mitsubishi Eletronic Research Labs. 2002.

TAN, Y.-P.; SAUR, D. D.; KULKARNI, S. R.; RAMADGE, P. J. **Rapid Estimation of Camera Motion from Compressed Video with Application to Video Annotation**. In: IEEE Trans. on Circuits and

Systems for Video Technology. Local: Editora, Feb. 2000. Vol. 10, no. 1, pp. 133-146.

TEKALP, A. M. **Video Segmentation**. In: Bovik, A. (ed.), Handbook of Image and Video Processing, Academic Press, San Diego, p. 383-399, 2000.

TREC. **TREC Video Retrieval Evaluation homepage**. Disponível em <http://www-nlpir.nist.gov/projects/trecvid>. Acesso em : 10 jan. 2005.

TOBIN, B.; **Examination of MPEG Macroblock Types to Find Shot Cuts**. Technical Report, Dublin City University. Dublin, April, 1999.

TOTTERDELL, A. **An Algorithm for Detecting and Classifying Scene Breaks in MPEG Video Bit Streams**. School of Computer Applications Dublin City University. Technical Report [98 -3]. September, 1998.

VASCONCELOS, N.; LIPPMAN, A. **Content-Based Pre-Indexed Video**. In: Proceedings of International Conference on Image Processing. Santa Barbara, California: USA, 1997.

WEE, S. **Manipulating temporal dependences in compressed video data with applications to compressed-domain processing of MPEG video** In: IEEE ICASSP. Arizona: Phoenix, March, 1999.

WEE, S.; SHEN ,B.; APOSTOLOPOULOS, J. **Compressed-Domain Video Processing**. In: *HP Labs Tech Report, 2002-282*. HP Laboratories Palo Alto. October 6th , 2002

WILLS, C. **Video segmentation for post-production**. In: Proc. SPIE Storage and Retrieval for Media Databases. San Jose: California, 2002. pp 158-167.

WU, M.; WOLF, W.; LIU, B. **An Algorithm for wipe detection**. In Proc. of IEEE International Conference on Image Processing (ICIP'98), vol.1, pp.893-897, Chicago, IL, Oct. 1998.

ZHANG, H. J.; KANKANHALLI, A.; SMOLIAR, W. **Automatic Partition of Full-Motion Video**, Multimedia Systems 1. Vol. 1,no.1, pp. 10-28. 1993.

ZHANG, H. J. **Content-based Video Browsing and Retrieval**. Handbook of Multimedia Computing. Boca Raton: CRC Press, 1999.

ZHANG, H. **Content-based video analysis, retrieval and browsing**. Multimedia Information Retrieval and Management - Technological Fundamentals and Applications. Springer (Ed.), 2003. Series: Signals and Communication Technology. 2003. Chapter 2, XVII, p. 476.

## Apêndice I

### Os Padrões MPEG-1 e MPEG-2

Este apêndice é uma introdução aos padrões de vídeo MPEG-1 e MPEG-2. É dirigido a iniciantes no estudo do tema, sem a intenção de explorar em detalhes os mencionados padrões.

#### Aspectos históricos dos padrões MPEG-1 e MPEG-2

As comissões International Organization for Standardization (ISO, 2004) e International Electrotechnical Commission (IEC, 2004) trabalham na elaboração de padrões de referência. Os padrões estabelecem uma linguagem comum entre grupos de desenvolvedores e consumidores de tecnologia, possibilitando a compatibilização e manutenção de produtos. Juntas, ISO e IEC, formaram em 1988 um comitê para o desenvolvimento de padrões de vídeo e áudio digital codificados, chamado Moving Picture Experts Group (MPEG, 2004).

O primeiro padrão de vídeo lançado por esse comitê foi o MPEG-1 (ISO & IEC, 93a; ISO & IEC, 93b), formalmente denominado ISO/IEC 11172. Esse padrão fornece suporte à codificação de vídeo associado a áudio em taxas de aproximadamente 1,5 Mbit/s. Seu projeto de desenvolvimento tem como objetivo atender aplicações como vídeo em CD-ROM (VCD), Digital Audio Tapes (DATs), codificadores de câmeras digitais (CamCoders) e aplicações de transmissão de vídeo através de redes de telecomunicação ou pela Internet.

O padrão MPEG-1 é formado por uma sintaxe para a representação codificada dos dados dos vídeos. Suas definições estabelecem os resultados das etapas do processo de decodificação, sem a descrição exata de algoritmos a serem implementados em codificadores e decodificadores. Por essa razão, distintas arquiteturas de codificadores e decodificadores podem apresentar diferenças consideráveis no fluxo de dados produzido (*stream*) para representar o vídeo codificado. A flexibilidade desse padrão é limitada apenas por um conjunto de

restrições, especificadas para garantir a adaptação de produtos de hardware e suas aplicações.

A sintaxe do padrão permite codificar vídeos em resoluções e taxas de codificação variadas, sendo projetada para obter um comportamento otimizado em vídeos não entrelaçados, no formato Standard Interchange Format (SIF), a taxas de 24 a 30 quadros por segundo, com resolução de 352 por 240 pixels para NTSC e 352 por 288 para PAL/SECAM. O uso de vídeo entrelaçado pode ser simulado pelo decodificador durante a exibição.

Uma característica importante do padrão MPEG-1 é o suporte às funcionalidades comuns em aparelhos de videocassete (VCR), chamadas *Trick Modes*. Entre essas funcionalidades encontram-se: o avanço desacelerado (*slow motion*), o avanço rápido (*fast forward*), o retrocesso (*reverse*) e a pausa (*pause*) (Figura 30) Além das funções convencionais, esse padrão também permite o acesso aleatório ao conteúdo do vídeo.



Figura 30 - *Trick Modes*

O desenvolvimento de um segundo padrão de codificação de vídeo pelo comitê MPEG teve como principal objetivo a adaptação do padrão MPEG-1 para dar suporte à *High Definition Television* (HDTV). Entre os novos requisitos incluídos no novo padrão, denominado MPEG-2 (ISSO, IEC 2000a; ISO, IEC 2000b), destacam-se:

- o suporte a vídeo entrelaçado, estabelecendo a compatibilidade com a transmissão de TV analógica;
- inclusão de canais multiplexados de informações, permitindo canais diferentes de áudio em diversas línguas, legendas, etc.;
- critérios de sincronismo mais rigorosos;
- maior controle de erros para transmissão em redes de baixa confiabilidade, como redes sem fio;
- otimização projetada para a resolução de HDTV, com a possibilidade de usar o formato SIF progressivo ou entrelaçado.

A primeira versão do MPEG-2 foi aprovada em 1994 pelo comitê MPEG como ISO/IEC 13818. Essa versão também foi aprovada pela International Telecommunication Union - Telecommunication Standardization Sector (ITU-T, 2005), com o nome de H.262.

O padrão MPEG-2 foi amplamente incorporado pela indústria, sendo adotado tanto pelo sistema de transmissão terrestre de televisão digital europeu, Digital Video Broadcasting (DVB), quanto pelo norte-americano, Advanced Television (ATV). Sua ampla aceitação no mercado de vídeo digital pode ser ilustrada pelas aplicações que utilizam esse padrão, como por exemplo: DVD, *Personal Video Recorder* (PVR), cinema digital, transmissão terrestre de vídeo, TV a cabo, Direct Broadcast Satellite TV (DBS), distribuição de vídeo por redes de banda larga em redes por fibra ótica (ATM), Digital Video Tape Recorder (VTR) e entrega de programas direct-to-home (DTH).

## Apêndice II

### Estrutura hierárquica MPEG-1 e MPEG-2

A informação contida nos vídeos nos formatos MPEG-1 e MPEG-2 é organizada em uma estrutura de camadas hierárquicas, ordenadas a partir da camada mais global para a camada mais interna, a saber: Video Sequence, Group of Pictures (GOP), Picture, Slice, Macroblock e Block (ISO & IEC 1993b; ISO & IEC 2000b). A Figura 31 apresenta uma ilustração dessa hierarquia.

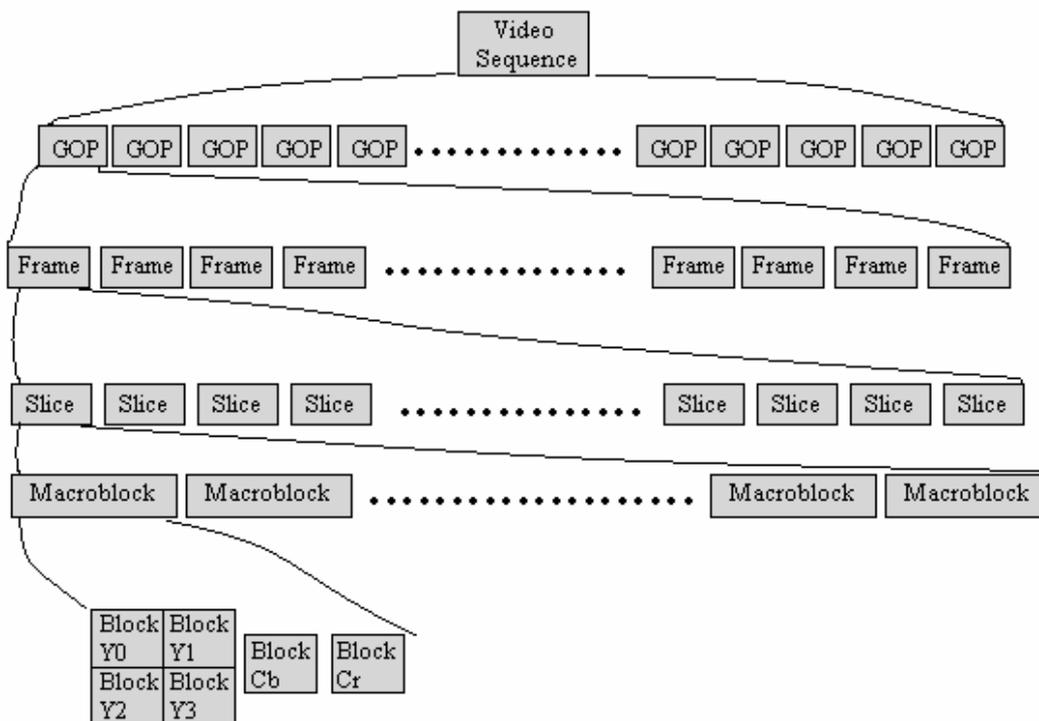


Figura 31 – Hierarquia MPEG

A camada Video Sequence é o topo da estrutura MPEG. Essa camada é formada pelo agrupamento de GOPs. Em seu cabeçalho, possui as informações necessárias para inicializar o estado do decodificador, permitindo decodificar qualquer seqüência, sem ser afetada por eventos anteriores.

A camada GOP é formada pelo agrupamento de Frames. A sintaxe MPEG permite que um GOP contenha qualquer número de quadros. O tamanho dos GOPs tipicamente varia de 9 a 18 quadros (Frames). O cabeçalho da camada GOP

possui informações temporais e de edição do vídeo. Cada GOP atualiza a referência temporal entre os quadros do vídeo. São usadas duas ordenações para a seqüência de Frames (ver Figura 32):

- a ordem de armazenamento ou transmissão (*transmitted order*): posiciona os quadros âncoras antes dos quadros a eles dependentes; Essa ordenação é usada no fluxo de dados (*stream*) codificado e na transmissão de vídeos MPEG; e
- a ordem de exibição (*display order*): posiciona os quadros na seqüência lógica para exibição do vídeo. Durante a exibição dos vídeos deve ser feita a reordenação dos quadros para a seqüência natural relacionada à ordem cronológica dos quadros.



Figura 32 – Ordem de exibição e ordem de transmissão

O primeiro quadro em ordem de transmissão (não é necessariamente o primeiro em ordem de exibição), é codificado sem usar predição, no modo intraframe. Neste modo o quadro pode ser decodificado sem nenhuma dependência em relação a outro quadro. As demais figuras interiores ao GOP podem usar ou não predição. Um GOP é um ponto de acesso aleatório, ou seja, é a menor unidade de codificação interior à seqüência que pode ser decodificada independentemente.

A camada Frame representa uma imagem, ou quadro, da seqüência do vídeo. Cada imagem é formada por um ou mais Slices. O cabeçalho dessa camada armazena uma informação temporal relativa ao interior do GOP e informações sobre o tipo da figura. Existem quatro tipos de figuras, classificadas de acordo com sua codificação: intracoded frame (I), predictive coded frame (P), backward-predicted coded frame (B) e dc-coded frame (D).

Existe uma classificação abstrata na literatura para ressaltar a independência dos quadros I. Nessa classificação, os quadros I também são denominados intracodificados, mas os quadros tipos P e B são ditos intercodificados.

Os quadros I não utilizam nenhum tipo de predição. Desta forma, não possuem dependência a outros quadros. Esse tipo de quadro não fornece taxas eficientes de compressão por representar toda a informação contida na figura. Em contrapartida, por serem independentes, fornecem pontos para acesso aleatório rápido ao vídeo. Os padrões MPEG-1/2 exigem a presença de, pelo menos, dois quadros I a cada segundo de vídeo, com o intuito de assegurar o acesso aleatório ao vídeo.

Os quadros P são codificados referenciando uma figura anteriormente codificada – predição forward –, usando compensação de movimento. O objetivo da predição é reutilizar áreas de imagens anteriormente já codificadas em outros quadros. O tipo P pode referenciar áreas em outras imagens I ou P. Tais quadros referenciados são chamados âncoras. Ao reutilizar informações, os quadros P conseguem melhores taxas compressão do que os quadros I.

O terceiro tipo mais comum de imagem, classificada como B, é codificada usando predição com compensação de movimento em relação a dois outros quadros: um anterior e outro posterior no tempo, fazendo uma predição bidirecional. Esse tipo de quadro utiliza como âncoras para sua predição os dois quadros I ou P mais próximos em sua vizinhança. Durante os processos de codificação e de decodificação, os quadros âncoras devem ser processados antes do quadro dependente. Por essa razão, existe uma ordenação para codificação e transmissão dos quadros diferente da ordem de exibição do vídeo. Na ordem de codificação ou transmissão cada quadro âncora deve ser posicionado antes do quadro dele dependente. Não é permitido o uso de quadros B como referência para a predição de outros quadros. Normalmente, se comparados aos quadros I e P, os do tipo B possuem as melhores taxas de compressão.

Os quadros classificados como D são raramente usados, possuindo o propósito de facilitar o acesso ao vídeo em taxas de bit muito baixas e de visualizar o vídeo em modo de aceleração, *fast-forward search mode*. Esse tipo de quadro armazena apenas a informação dos coeficientes DC da imagem.

Não existe uma ordem padrão para a codificação dos quadros nos diversos tipos. Fica como encargo do codificador decidir a disposição dos tipos de quadros,

de forma a atender os requerimentos do padrão MPEG-1/2 e de serviços específicos da aplicação em desenvolvimento. Durante a escolha entre qual tipo de quadro utilizar, o codificador pondera entre diminuir o espaço de armazenamento ocupado pelo vídeo e o tempo do processo de descompressão. Visando aumentar a compressão em um ambiente de recursos de armazenamento escassos, devem ser codificados mais quadros do tipo B, causando menor agilidade para a descompressão. No extremo oposto, relaxa-se a compressão em troca de maior agilidade para a descompressão, resultando na inclusão de mais quadros do tipo I e aumento do tamanho total do vídeo.

A camada Slice é formada pelo agrupamento de Macroblocks percorridos na direção horizontal, da esquerda para a direita, de cima para baixo. Um Slice pode conter no mínimo um Macroblock, ou até várias linhas de Macroblocks. Essa camada é responsável pela recuperação de erros nas figuras. Ao encontrar dados corrompidos o decodificador pode esperar o próximo Slice ao invés de descartar toda a figura. Ao iniciar um novo Slice, todos os dados necessários para reconstruir os valores dos coeficientes DC e os *motion vectors*, usados na compressão do vídeo, são reinicializados com as informações de sua posição na figura e escala de quantização, tornando-os independentes a informações de outros Slices. Desta forma, as informações da camada Slices são suficientes para recuperação de corrupções de dados locais. Essa característica permite que o decodificador possa avançar para o próximo Slice, ao encontrar um erro.

A camada Macroblock é a unidade básica usada para o cálculo da compensação de movimento e para alterações na escala de quantização. Cada Macroblock é composto por componentes Blocks e um cabeçalho. Dentro da hierarquia MPEG essa é a menor camada a selecionar um método de compressão, sendo considerada a menor unidade codificada na figura.

Um Macroblock é uma região codificada a partir de 16 por 16 *pixels* da imagem original. O número total de Macroblocks em uma figura depende do tamanho em *pixels* da figura, não sendo admitido um número fracionário de Macroblocks. Desta forma, as dimensões horizontal e vertical da imagem devem ser múltiplas de 16. O codificador é responsável por assegurar essa proporcionalidade. Quando for necessário adaptar um vídeo com alguma das dimensões não proporcional a 16, o codificador adiciona *pixels* para preencher o quadro na parte inferior direita das imagens de todo o vídeo.

Os Macroblocks podem ser classificados em cinco tipos abstratos: I (intra), P (predicted), B (backward), Bi (bi-directionally-predicted) e S (skipped). A distribuição nos diferentes tipos de Macroblocks é feita durante a codificação, considerando a confiabilidade da estimativa de movimento e a eficiência de codificação. Diferentes tipos de Macroblocks podem estar presentes em diferentes tipos de quadros, não exclusivamente nos quadros que levam nome de classificação correspondente à classificação do Macroblock.

Os Macroblocks do tipo I são intracodificados, não referenciam, portanto, outros Macroblocks. Os quadros I possuem apenas Macroblocks I. O contrário já não é verdadeiro (os Macroblocks I também podem ser codificados em quadros P e B).

Os Macroblocks P fazem predição forward e podem estar presentes em quadros P e B (Figura 33). Essa forma de predição utiliza outro Macroblock como referência para sua codificação. O Macroblock usado como referência pertence a uma imagem já computada no processo de compactação, anterior na ordem de exibição, que terá sua informação passada para frente, daí a denominação forward (em frente). Cada Macroblock P possui um vetor de movimento indicando o deslocamento espacial entre o bloco atual a ser decodificado e o bloco referência em um quadro âncora anterior. Tal vetor representa a compensação de movimento.

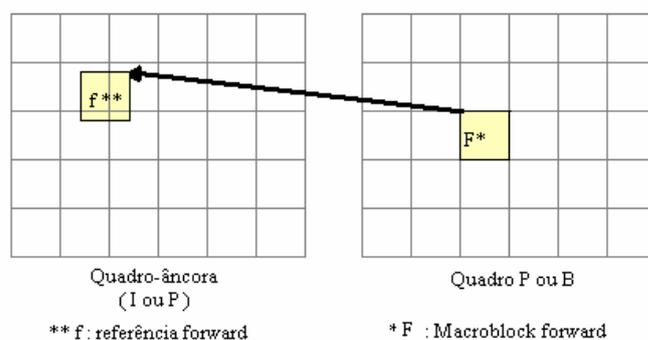


Figura 33 – Macroblock de referência Forward

Dois tipos abstratos de Macroblocks são utilizados apenas em quadros B. São eles o tipo B e o Bi. Um Macroblock B pode fazer apenas predição *backward* (Figura 34), enquanto um Macroblock Bi faz os dois tipos de predição (Figura 35), *forward* e *backward*. A predição *backward* possui o comportamento inverso da predição *forward*. A constituição de Macroblocks Bi é encontrada pela média

dos dois Macroblocks referenciados nos quadros âncoras mais próximos (em ambas as direções).

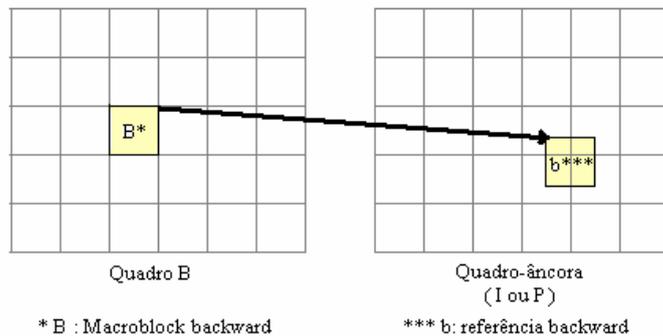


Figura 34 – Macroblock de referência Backward

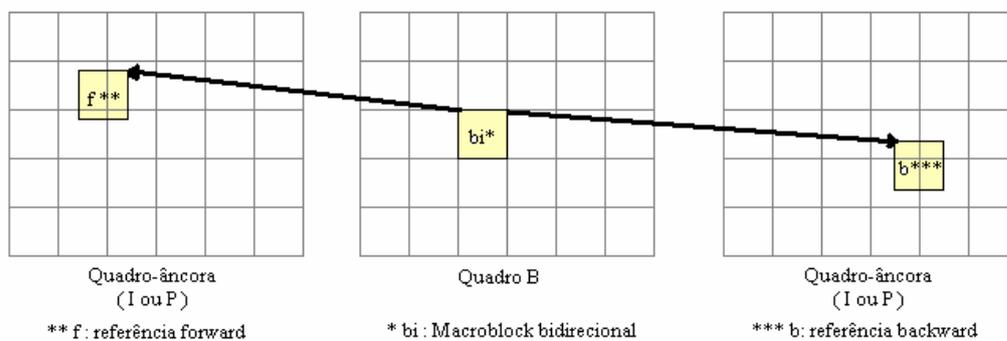


Figura 35– Macroblock de referência Bidirecional

Os quadros P e B podem apresentar mais um tipo de Macroblock importante, denominado Skipped. Os Macroblocks S são utilizados quando o Macroblock em processo possui alto grau de similaridade em relação ao Macroblock referência no quadro âncora. Essa similaridade permite que o Macroblock no quadro âncora seja copiado na posição do Macroblock atual, sem nenhuma modificação, nem perda de qualidade perceptível. Os Macroblocks S são responsáveis por áreas no interior da imagem integralmente reaproveitadas dos quadros âncoras, ou seja, não ocupam espaço de armazenamento. Normalmente, esse tipo de Macroblock aparece em áreas da imagem internas a objetos parados na cena.

A visão humana consegue distinguir a informação de luminância, informação de brilho, com maior precisão do que a informação de crominância, informação de cor. Para reproduzir essa maior sensibilidade, os valores da luminância dos *pixels* são codificados sem perda de resolução, enquanto os valores de crominância são sub-amostrados, sem perda perceptível de qualidade.

A informação de luminância em um Macroblock no padrão MPEG-1 é encontrada a partir de uma área de  $16 \times 16$  *pixels*, repartida em quatro blocos  $8 \times 8$ , mantendo a resolução original. Entretanto, a informação contida nos dois canais de crominância, Cb e Cr, é sub-amostrada em apenas um bloco  $8 \times 8$ , para cada canal, para representar a área de  $16 \times 16$ .

Os canais de crominância dos Macroblocks são sub-amostrados nas direções vertical e horizontal, na razão de 2 por 1. Desta forma, o padrão MPEG-1 estabelece o uso de quatro blocos para armazenar a luminância e dois para a crominância. Esse padrão limita o número total de Macroblocks a 396.

O padrão MPEG-2 possui outras distribuições de blocos, conhecidas como 4:2:0, 4:2:2 e 4:4:4 (Figura 36). O formato 4:2:0 possui a mesma distribuição do padrão MPEG-1. O formato 4:2:2 possui 4 blocos para luminância e dois para cada canal de crominância, ou seja, dois blocos Cr e dois Cb. O formato 4:4:4 possui 4 blocos para cada componente cromática e quatro para a componente de luminância.

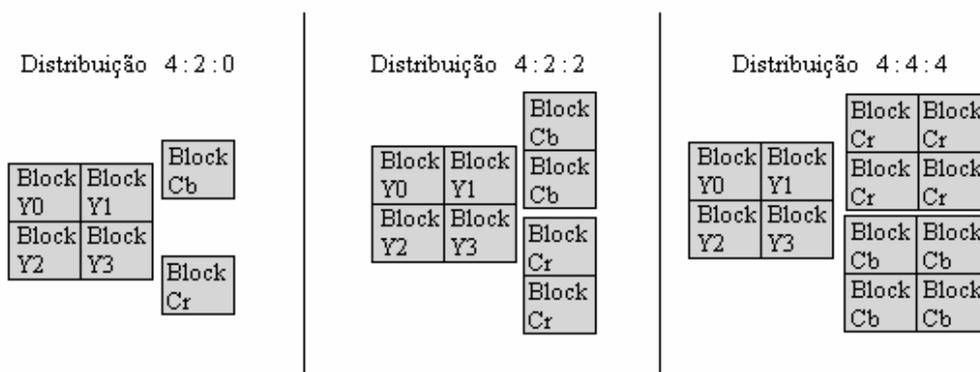


Figura 36 – Distribuição dos blocos no padrão MPEG-2

A camada mais interna, camada Block, é unidade básica para aplicação da Transformada Discreta de Cosseno, DCT (Figura 37). Cada Block é formado por uma matriz de 8 por 8 coeficientes. Desta forma, os 64 valores armazenados nos blocos representam a informação dos *pixels*, transformados para o domínio de frequências.

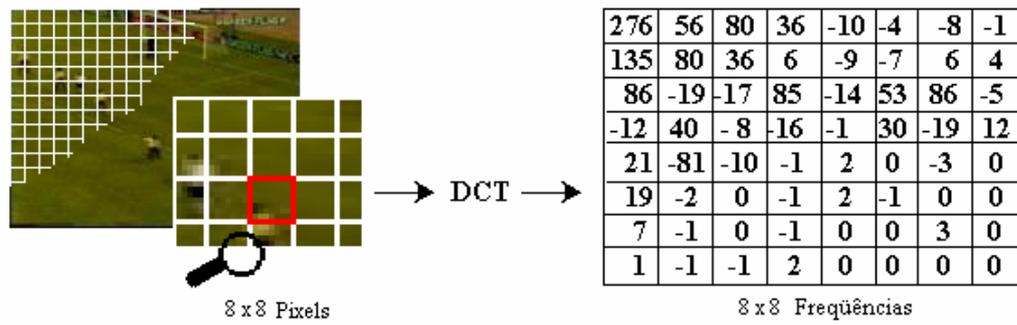


Figura 37 – Camada Block

## Apêndice III

### Compressão MPEG

A escolha de algoritmos para implementação de um codificador MPEG-1 ou MPEG-2 não é estabelecida por esses padrões (ISO & IEC 1993b; ISO & IEC 2000b). Entretanto, estão estabelecidas as etapas a serem percorridas por uma arquitetura de codificação. A Figura 38 ilustra uma representação abstrata dessa arquitetura.

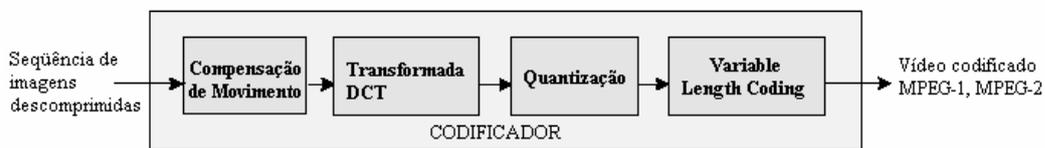


Figura 38 – Codificador MPEG Abstrato

Os padrões MPEG-1/2 especificam a implementação de técnicas de compressão de imagens e dados. Uma introdução a essas técnicas é abordada neste apêndice.

### Compressão MPEG-1/2

As técnicas para compressão almejam diminuir o tamanho de armazenamento do vídeo sem perda perceptível de qualidade. Para atingir esse objetivo, os padrões MPEG-1/2 procuram eliminar informações redundantes das imagens da seqüência do vídeo.

A compressão MPEG está baseada em duas premissas. A primeira está na observação de que a visão humana é incapaz de perceber toda a informação presente em uma imagem complexa em movimento. Graças a essa limitação da visão, é possível desenvolver processos de compressão que descartem informações imperceptíveis. A segunda premissa está baseada no fato de que as imagens em seqüência possuem muitas áreas repetidas a serem eliminadas pelo processo de compressão.

Durante a compressão, a busca por informações redundantes é feita em relação a duas dimensões da seqüência de imagens. Uma primeira abordagem de compressão procura áreas de informação constante entre quadros seqüenciais no vídeo, analisando a redundância através da dimensão do tempo. A segunda dimensão da compressão analisa internamente cada quadro, adotando critérios espaciais.

As próximas seções deste apêndice aprofundam o estudo da Compressão Temporal e Espacial respectivamente.

## **Compressão Temporal**

A compressão temporal foi desenvolvida para descartar informações semelhantes entre quadros sucessivos. Em uma seqüência contínua de imagens pertencentes à mesma cena, muitos objetos permanecem nas mesmas posições e alguns se deslocam em pequenas distâncias, de forma que imagens consecutivas apresentem áreas semelhantes. O padrão MPEG-1/2 usa a técnica de predição com compensação de movimento por Macroblocks para reduzir essas redundâncias temporais inerentes ao vídeo.

Em uma primeira etapa do processamento da compressão temporal, cada imagem é dividida em unidades de 16x16 *pixels* chamadas de Macroblocks. Os Macroblocks são comparados individualmente a outros Macroblocks em imagens anteriormente codificadas que lhes servirão como referência. O movimento é considerado como sendo a diferença entre o posicionamento do Macroblock codificado e o de referência. Um vetor bidimensional, denominado vetor de movimento (*motion vectors* - MV), é usado para representar o deslocamento entre a posição do Macroblock e a referência encontrada pelo pareamento, ou seja, especifica a posição do bloco previamente codificado em relação à posição do bloco atual. São codificados no fluxo de dados resultante apenas o erro de predição (chamado de residual e encontrado pela diferença entre os dois Macroblocks) e os vetores de movimento. Os Macroblocks com predição exclusivamente forward ou backward possuem apenas um MV, enquanto dois MVs são especificados para cada Macroblock com predição bidirecional.

Com o objetivo de encontrar o pareamento do Block a ser codificado, no quadro âncora é feita uma busca ao redor da área do Macroblock, limitada pelas

extensões máximas do vetor de movimento. O pareamento procura um Block similar ao Block sendo codificado.

A compensação de movimento considera que localmente, em relação à área do Macroblock, a figura pode ser representada pela translação de um Macroblock em uma outra figura na vizinhança.

A divisão da imagem em Macroblocks para a compensação de movimento permite que áreas distintas da imagem apresentem vetores de movimento distintos, de forma a flexibilizar a técnica de compensação de movimento a comportamentos distintos entre os objetos presentes nas imagens.

### Compressão Espacial

A compressão espacial procura diminuir a informação redundante no interior de cada imagem, a partir da análise da correlação existente entre os *pixels* e sua vizinhança. No padrão MPEG-1/2 esse tipo de compressão é aplicada em unidades de blocos de *pixels*. A Figura 39 ilustra as etapas da compressão espacial a serem abordadas nesta seção.

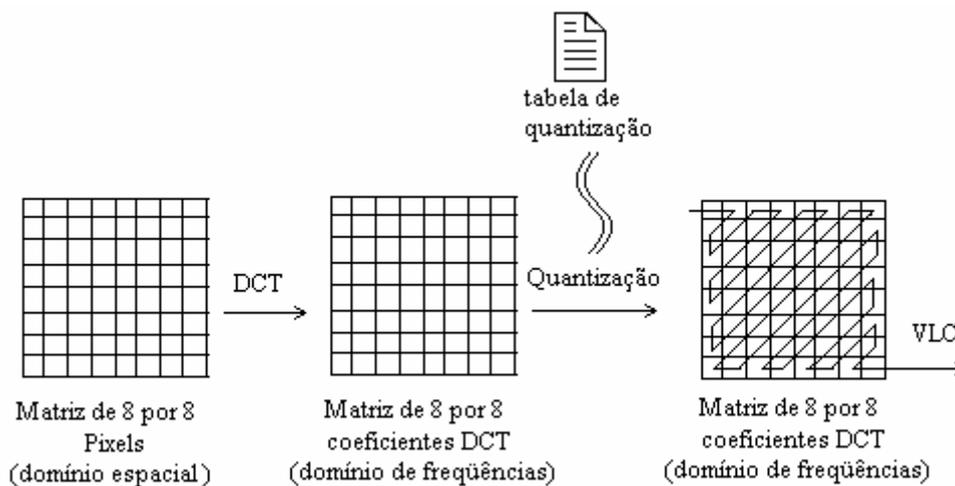


Figura 39 - Etapas da compressão espacial

O processo de compressão espacial se inicia pela aplicação da Transformada de Cosseno Discreta (DCT) a cada bloco de 8x8 *pixels*. Dessa forma, a informação contida no bloco passa do domínio espacial para o domínio de frequências. Nesse estágio os valores armazenados no bloco representam os coeficientes das frequências horizontais e verticais da região representada pelo

bloco. Os valores dos *pixels* podem ser reconstituídos pela aplicação da Inversa da Transformada de Cosseno Discreta (IDCT) nos coeficientes de frequência.

O coeficiente DCT na posição (0,0) é chamado de coeficiente DC e representa as frequências zero horizontal e zero vertical. Os demais coeficientes do bloco são chamados de coeficientes AC. O valor do coeficiente DC possui uma interpretação especial, por ser proporcional à média dos valores dos *pixels* do bloco 8x8. Os coeficientes AC representam uma ou mais frequências verticais e horizontais diferentes de zero. Depois de aplicar a transformada DCT ao bloco, ou seja, codificar o bloco no domínio das frequências, os dados ainda possuem o mesmo volume, já que não foi feito nenhum descarte de informação. A matriz de 8x8 *pixels* passa a ser representada por uma matriz de 8x8 coeficientes.

No domínio da frequência ocorre uma concentração da energia do sinal da imagem, equivalente às informações mais perceptíveis ao olho humano, em torno das baixas frequências. Essa informação de maior relevância visual é armazenada na parte superior esquerda dos blocos. As informações menos perceptíveis, representadas pelas altas frequências, são armazenadas no extremo oposto do bloco, na parte inferior direita.

A segunda etapa da compressão espacial consiste na aplicação de técnicas de quantização para reduzir a informação contida nas frequências altas. A escolha dos níveis para a quantização tende à criação de coeficientes AC com valor zero. Com esse objetivo, são feitos passos de quantização que favorecem à não percepção, pelo sistema visual humano, da perda de frequências altas, superiores a um nível determinado. A codificação de valores zero para coeficientes AC consecutivos aumenta o ganho da compressão dos dados. Como as altas frequências aparecem como zero, essas podem ser descritas apenas com o número total de zeros do bloco. Desta forma, já que o olho humano é menos sensível a essas frequências, são usados menos bits para representá-las.

O resultado esperado do processo de quantização é a representação esparsa dos dados de forma que na maioria das amplitudes os coeficientes DCT sejam iguais a zero. Com essa distribuição, apenas as amplitudes e localizações dos coeficientes diferentes de zero precisam ser armazenados na *stream* de dados codificada.

O terceiro estágio da compressão espacial usa a técnica de *Variable Length Coding* (VLC). A técnica de VLC consiste na codificação estatística que associa

palavras-código aos valores a serem codificados. Valores muito freqüentes são associados a palavras-código de comprimento pequeno, enquanto valores menos freqüentes são associados a palavras-código de comprimentos maiores. Em média, a associação de palavras-código aos valores a serem codificados usando essa análise de freqüência gera um código total menor do que a representação dos dados originais.