H.R. Wu and K.R. Rao, (Eds.) Digital Video Image Quality and Perceptual Coding, CRC Press (ISBN: 0-8247-2777-0), 2005.

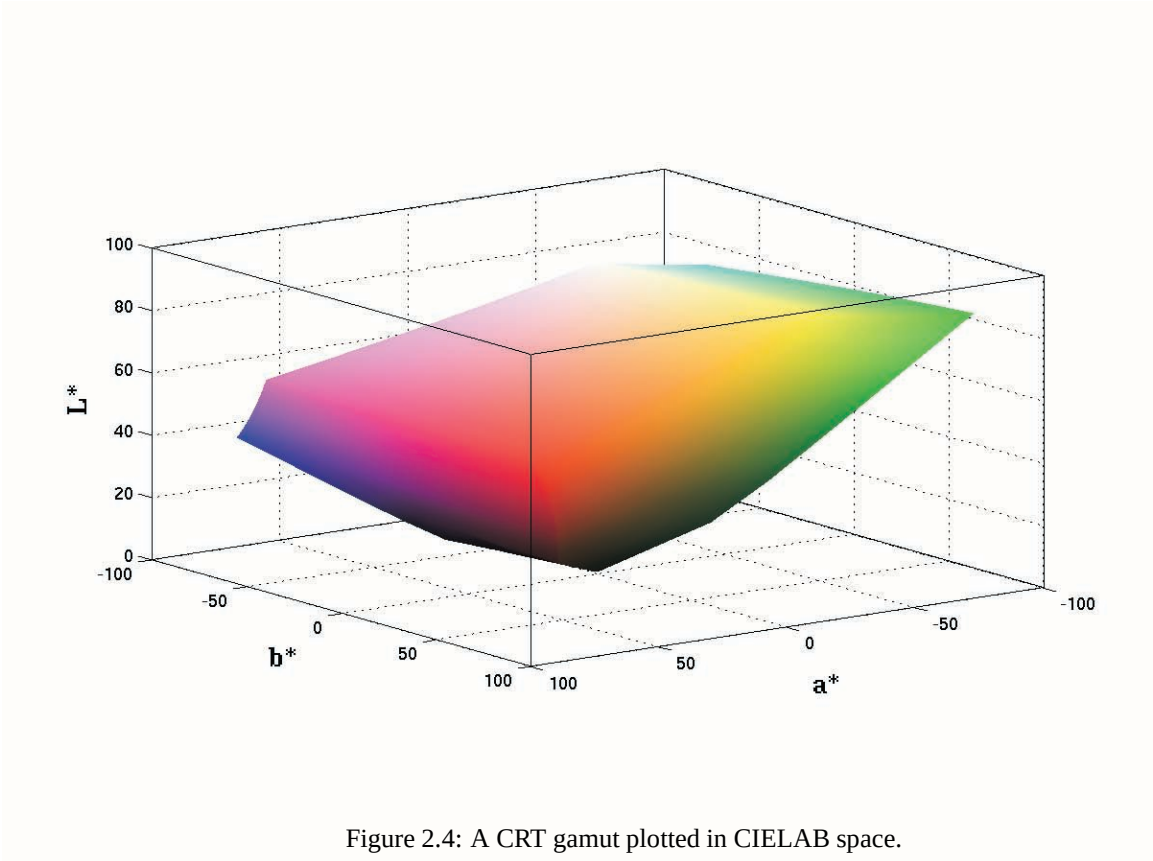


Figure 2.4: A CRT gamut plotted in CIELAB space.

pearance model that predicts the relative and absolute color appearance attributes based on specifying the surround conditions (average, dim, or dark), the luminance of the adapting field, the tristimulus values of the reference white point, and the tristimulus values of the sample.

2.4 Luminance and the Perception of Light Intensity

Luminance is a term that has taken on different meanings in different contexts and therefore the concept of luminance, its definition, and its application can lead to confusion. Luminance is a photometric measure that has loosely been described as the "apparent intensity" of a stimulus but is actually defined as the effectiveness of lights of different wavelengths in specific photometric matching tasks [SS99]. The term is also used to label the achromatic channel of visual processing.

2.4.1 Luminance

The CIE definition of luminance [WS82] is a quantity that is a constant times the integral of the product of radiance and $V(\lambda)$, the photopic spectral luminous efficiency function.

Figure 3.4: The C_r chrominance component of a coded picture from the *Table-Tennis* sequence.

Figure 3.5: Reproduction of a picture suffering from color bleeding and "color ringing" from the *Table-Tennis* sequence, corresponding with the C_b and C_r chrominance components shown in Figs. 3.3 and 3.4. Note the high frequency changes in color around the table's edge, corresponding to the ringing, and the gradual bleeding around the arm.

Figure 3.23: Macroblocks located centrally in the trunk of the tree in this picture provide examples of chrominance mismatch within the *Flowers* sequence.

tionally, with the huge popularity of DVDs, source material typically isolated within one geographic region (and TV standard, such as NTSC) is being distributed worldwide and are expected to be displayed on TVs which are not of the same TV standard or aspect ratio.

To allow for the correct rendition of various source formats on any given display, there is a need to perform correct scaling and frame-rate conversion. Following is a discussion of the various artifacts which are caused by both these tasks.

3.15.1 Video Scaling

The simplest method of scaling down is sample decimation, but such an approach introduces aliasing and severely distorts details in the direction of scaling especially with increasing scaling ratios. Similar artifacts are caused by sample repetition when scaling up. The crude nature of such approaches are sometimes necessitated by limitations in processing capacity.

More sophisticated approaches utilize convolution implemented using standard digital filter design techniques based on the scaling requirements. However, despite convolution providing an improved outcome compared to decimation/repetition, it is inevitable that some degree of blurring will result for large up-scaling factors. For down-scaling, the main issue is aliasing, with the level of aliasing generally a function of filter length

Figure 4.1: Standard test sequence — *Susie*

the test sequence *Susie* as shown in Figure 4.1 can be encoded at low rates with little apparent loss in quality.

There have been tests where the sequence *Susie* showed no visible quality impairment from such a bit rate reduction while other standard sequences, such as those shown in Figure 4.2, were significantly impaired and unacceptable to the viewer. *Waterfall* and *Mobile & Calendar* are exceptionally good test sequences because they highlight the loss in visual quality caused by reducing the video bit rate. *Tree* and *Tempete* stress other quality aspects of compression and transmission techniques.

It is very important that multiple types of video material be used in any image quality test involving motion video systems or equipment. Multiple sources should also be considered since film, studio and graphic material would also have different behaviors. In any given test, at least 8 to 16 different test material types should be used. Good test sequences can be obtained from various standards bodies such as SMPTE, IEEE and others. There are also open sourced sequences available from the Video Quality Experts Group (VQEG) at *www.vqeg.org*.

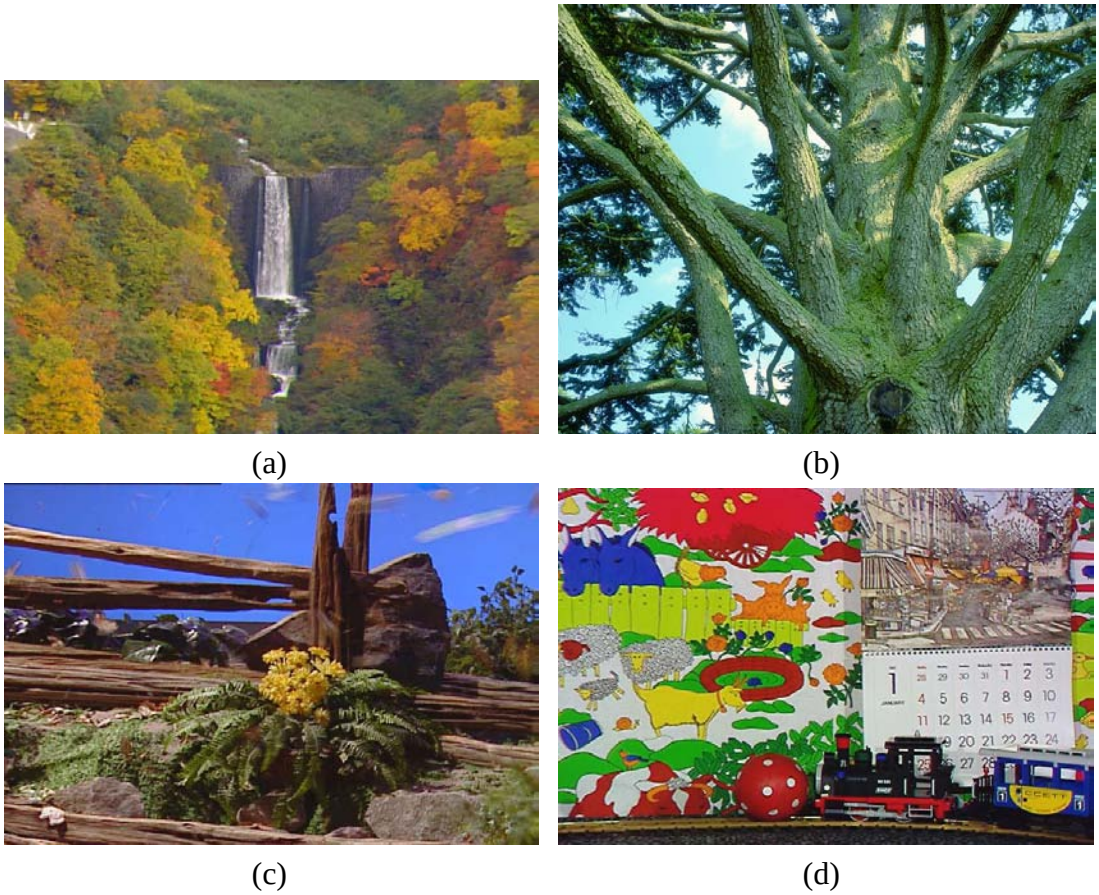


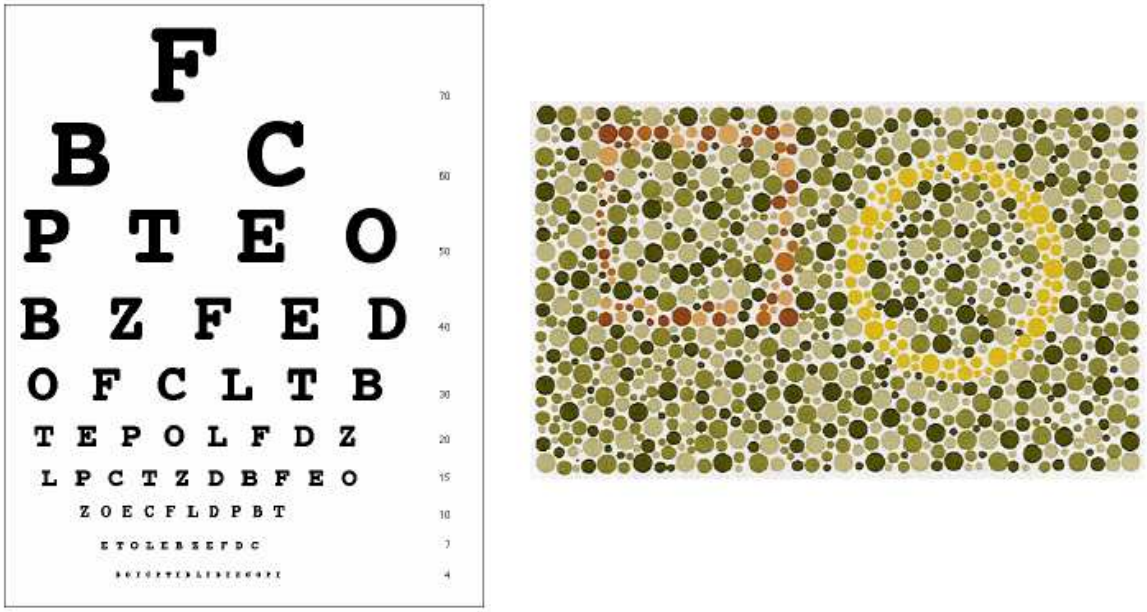
Figure 4.2: Examples of other Standard Test Sequences: (a) *Waterfall*, (b) *Tree*, (c) *Tempete*, and (d) *Mobile & Calendar*.

4.4 Selection of Participants — Subjects

Test subjects can be either experts or non-experts. All subjects should be screened for acuity, color blindness and other visual defects. To produce reliable results, a large number of participants should be used. It is generally accepted that between 16 and 24 subjects will provide a statistically valid result. The more participants the more statistically sound the results will be. Also ensuring that the subjects meet some minimum requirement by performing proper screening ensures that the data is valid and carries more weight in the research and consumer community.

4.4.1 Experts

Experts in the video field have extensive experience in evaluating, producing, distributing or designing systems that deal with video. These people have a specialized way



(a) (b)

Figure 4.3: (a) Snellen Eye Chart ©; (b) Ishihara Test plate ©

use a hidden reference; reference conditions are employed to judge the reliability and repeatability of any given test subject.

The data analysis associated with collecting and reporting subjective data can be found in journal articles or regular statistical textbooks and International Telecommunications Union Radiocommunications Broadcast Technology 500 (ITU-R Recommendation BT-500) [ITU98].

4.5.1 Test Chamber

The test environments in different facilities around the world vary from extremely complex to very simple. One of the most advanced video quality test facilities is the Communications Research Centre (CRC)/Industry Canada in Ottawa, Canada. Figure 4.4 shows a typical room layout at CRC for a dual monitor evaluation. This facility meets and exceeds the standards defined in the ITU-R Recommendation BT 500.7. Other facilities such as those at NTIA/ITS (National Telecommunications and Information Administration/Institute of Telecommunications Sciences) and Intel Corporation use soundproof chambers that are designed to meet industry standards, but accommodate much smaller displays than that at the CRC.

Three major factors that must be considered are lighting, ambient noise and the quality and calibration of the display. The most important information is contained in Recommendation BT-500, which one should read and understand before conducting

Figure 5.1: The same amount of noise has been added to these two images, such that their PSNR is identical. High-frequency noise was inserted into the bottom region of the left image, whereas band-pass filtered noise was inserted into the top region of the right image. The noise is hardly visible in the left image due to our low sensitivity to high-frequency stimuli and the strong masking by highly textured content in the bottom region. The smooth sky represents a much weaker masker, and the structured (low-frequency) noise is clearly visible. The PSNR is oblivious to both of these effects.

A number of additional pixel-based metrics have been proposed and tested [EF95]. It was found that although some of these metrics can predict subjective ratings quite successfully for a given compression technique or type of distortion, they are not reliable for evaluations across techniques. MSE was found to be a good metric for additive noise, but is outperformed by more complex HVS-related techniques for coding artifacts [ASS02]. Another study concluded that even perceptual weighting of MSE does not give consistently reliable predictions of visual quality for different pictures and scenes [Mar86]. These results indicate that pixel-based error measures are not accurate for quality evaluations across different scenes or distortion types.

5.5 The Psychophysical Approach

5.5.1 HVS Modeling Fundamentals

This section briefly introduces processes and properties of the human visual system to provide a common ground for the ensuing discussion of HVS-based metrics. More details on the fundamentals of human vision and modeling can be found in Chapter 2.

Models of the human visual system (HVS) account for a number of psychophysical effects [Win99a] which are typically implemented in a sequential process as shown in Figure 5.2.

where $\overline{R_L}(n)$, $\overline{R_M}(n)$ and $\overline{R_H}(n)$ denote the means of $R_L(n)$, $R_M(n)$ and $R_H(n)$, respectively.

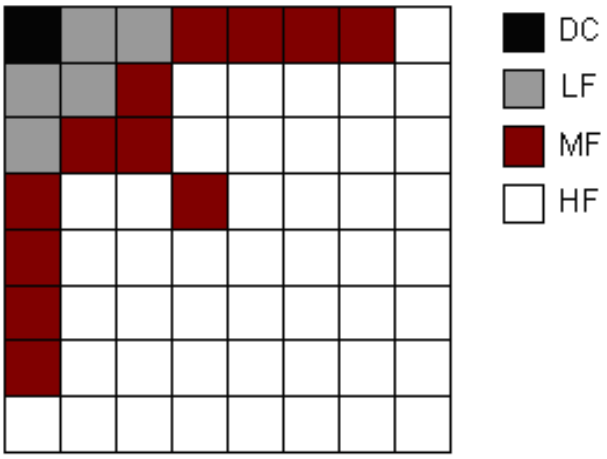


Figure 9.2: Coefficient Grouping for DCT Block Classification (N=8) (From [TV98], ©1998 IEEE).

Based upon the similar methodology in [TV98], each DCT block is assigned to one of the three classes with descending order of the HVS sensitivity, namely, *Low-Masking*, *Medium-Masking* and *High-Masking* classes, according to $E_{mh}(n)$:

- $E_{mh}(n) \leq \mu_1$: the block is assigned to *Low-Masking* class.
- $\mu_1 \leq E_{mh}(n) \leq \mu_2$: if condition (9.28) or (9.29) is met, the block is assigned to *Medium-Masking* class; otherwise it is assigned to *Low-Masking* class.
- $\mu_2 \leq E_{mh}(n) \leq \mu_3$: if condition (9.28) or (9.29) is met, the block is assigned to *Medium-Masking* class; otherwise it is assigned to *High-Masking* class.
- $E_{mh}(n) > \mu_3$: if condition (9.28) or (9.29) is met for $\varphi_\tau = \tau \cdot \varphi$ and $\chi_\tau = \tau \cdot \chi$ (where $\tau < 1$), the block is assigned to *Medium-Masking* class; otherwise it is assigned to *High-Masking* class.

The condition used above for determining the *Medium-Masking* class is defined as:

$$\widetilde{E}_{dm}(n) \geq Q \tag{9.28}$$

$$max\{\widetilde{E}_d(n), \widetilde{E}_{dm}(n)\} \geq \varphi \quad and \quad min\{\widetilde{E}_d(n), \widetilde{E}_{dm}(n)\} \geq \chi \tag{9.29}$$

where $\varphi > \chi$.

The possible block classification according to $E_{mh}(n)$ is illustrated in Figure 9.3. The elevation with inter-band masking can be then determined as:

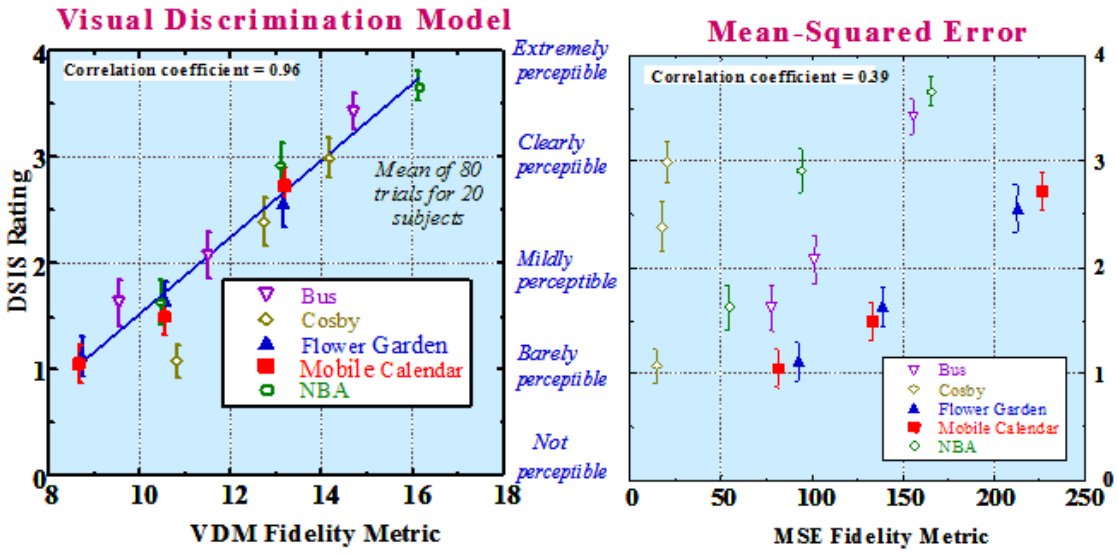


Figure 12.1: VDM vs. DSIS rating and MSE vs. DSIS.

directions in this area. Specifically, the following topics are covered in this chapter:

- Noise Visibility and Visual Masking – sensitivities of the human visual system to structured and unstructured noise, and their application to video encoding.
- Architectures that support perceptual based encoding – description of encoder architectures that specifically take advantage of human vision model. Architectures include macroblock, picture level control and look-ahead processing for a MPEG-2 encoder.
- Standards-specific features – origins of blockiness in standards, and in-loop filtering as a means to remove block visibility.
- Salience/masking pre-processing – determining in advance of encoding key areas to be preserved and the application of region based filtering to reduce bitrate requirements for a frame.
- Application to multi-channel encoding – using statistical variations in multiple video program streams to improve the efficiency of a transmission channel.
- Future Challenges – the potential of using advanced image interpolation techniques to compression.

12.2 Noise Visibility and Visual Masking

All lossy video coding systems add noise, or distortion, to the image, and the visibility of the noise is a key feature of the compression algorithm. Much research has been done

(a) (b)

Figure 12.3: (a) JPEG (block-based DCT) vs. (b) JPEG2000 [TM02](full-image wavelet) still image compression at the same normalized bit rate (0.25 bits/pixel).

cosine functions as shown in Figure 12.13. As discussed in more detail later, finely quantized DCT coefficients, upon inverse transformation back to pixels, have the desirable property of introducing noise at edges of objects, where the edge itself masks most of the distortion. However, coarsely quantized DCT coefficients introduce blockiness that is highly visible in near-uniform areas of the scene. Blockiness is a key distortion that needs to be minimized in block based compression approaches.

Alternative basis sets that do not induce structured noise can have a significant perceptual advantage over the DCT, but to realize this advantage, distortion measures must reflect the actual visibility of the distortion. One alternative basis set that is becoming more widely accepted is the DWT, or Discrete Wavelet Transform [TM02, SCE01, CSE00, RJ02, CP02]. For maximal efficiency, the DWT is often applied to large blocks, up to and including the full image size. In Figure 12.3, a comparison between 8×8 block-based DCT and a full-image DWT is shown. As can be seen, the appearance of the compression artifacts are quite different between the two transform methods.

12.3 Architectures for Perceptual Based Coding

For a particular scene, the bit rate can be used to predict quality but the variation from scene to scene can be quite significant. As shown in Figure 12.4, the subjective rating (DMOS[1]) of compressed content improves (the impairment difference becomes lower) with bit rate, as expected. However, there is a large variation in the shape of this rate-distortion curve as a function of scene content. Some content improves continuously as bit rate increases, while other content exhibits rapid quality saturation above a certain bit rate.

[1]DMOS, or Difference Mean Opinion Score, is a common subjective rating metric.

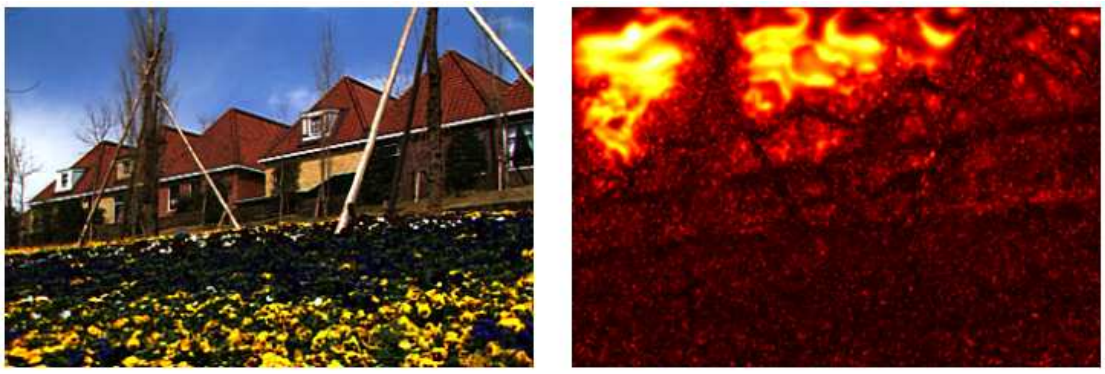


Figure 12.7: Masking map. Bright areas show areas of low masking. Perceptual based encoders encode the brighter areas with higher quality.

- Avoiding motion paths of salient objects. It is often difficult to “not watch” an interesting object in the scene. Therefore, do not quantize heavily on an intersecting path associated with this object.
- Quantize more heavily in areas in which there is a large amount of local (but incoherent) motion energy.

Masking maps can be used in perceptual coding. A masking map shows where slight changes in image detail would be visible. Figure 12.7 shows an original image, and a masking map produced from this image. Note that the bright, or ”hot” areas are associated with the nearly uniform sky region. This means that special care must be taken to ensure that this region is coded with high fidelity, as any distortion, such as blockiness, will be readily noticed.

12.3.2 Perceptual Based Rate Control

Since quantization is a major factor contributing to the quality and coded bit rate of an image sequence, we now examine several ways of incorporating perceptual-based metrics into the rate control mechanism.

12.3.2.1 Macroblock level control

Setting the quantization level of individual macroblocks (16×16 blocks of luminance pixels) is the most difficult problem for any rate control scheme, including those that use perceptual models. Unfortunately, when applied to this level of granularity, most HVS models can only give relative estimates of visibility. However, this is still quite useful in making bit allocations for individual macroblocks. In Figure 12.5 the vertical hatch processing loop is used to control the macroblock quantization levels. Figure 12.8 demonstrates the results of this type of processing.

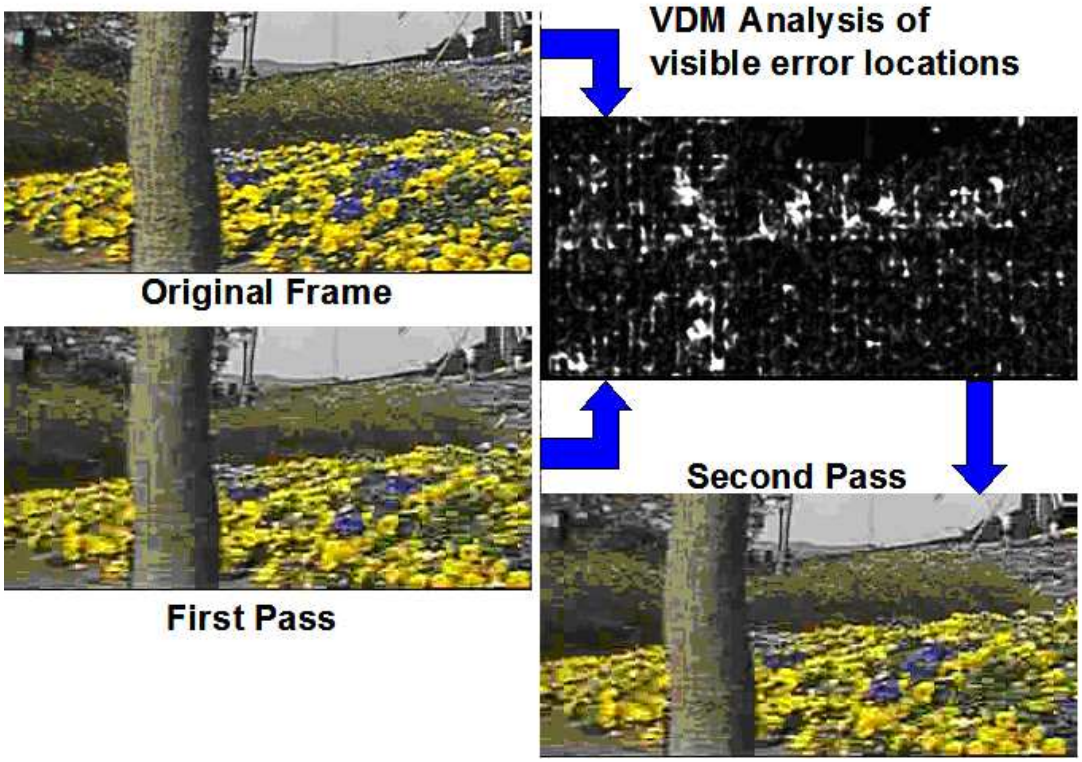
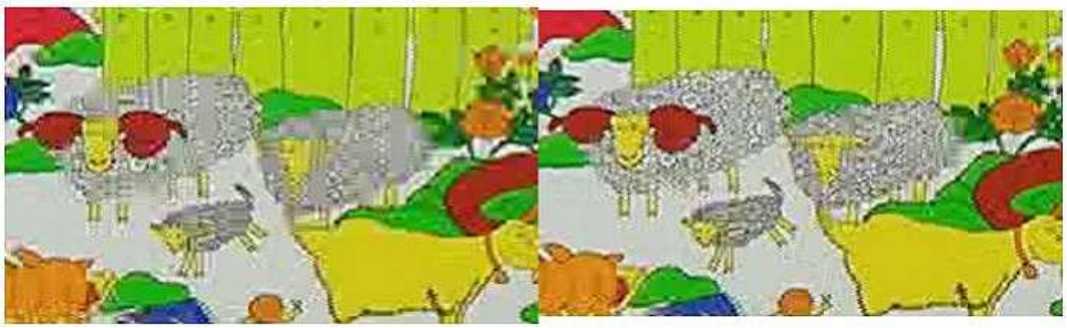


Figure 12.8: Demonstration of bit reallocation on the macroblock level using a visual discrimination model (VDM). This type of processing is shown in Figure 12.5 as the vertical hatch processing loop.

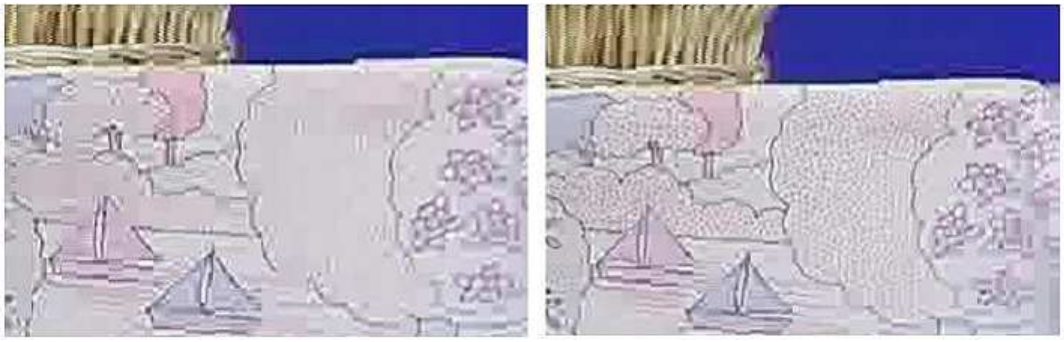
This algorithm is a variation on the MPEG Committee's TM5 rate control. During a first pass, the macroblock quantization is performed using a simple linear bit usage profile, the reconstructed field is then analyzed using a VDM that produces a distortion map as shown in Figure 12.8 (upper right quadrant). A second quantization is then performed using the distortion map to modify the original bit allocation. This process can be iterated until the distortion map meets a predetermined level of uniformity. The algorithm used is described in detail by Peterson and Lee [PL03]. To minimize the number of passes, a more sophisticated initial macroblock bit allocation should be performed (see [CZ97]). With a good initial start, a uniform distortion map can be achieved with only one additional pass.

12.3.2.2 Picture level control

Picture level control generally involves meeting a pre-assigned target bit rate (more precisely, a target bit count) for each picture type. Visual maskability of the source frame can give a starting point for the quantization profile to be used for the frame. After coding, the actual bit count can be measured, and the quantization profile can be scaled to reduce the difference. Figure 12.9 demonstrates the improvements that can be achieved



(a) Baseline Encoder-TM5 (b) DVC Encoder



(c) Baseline Encoder-TM5 (d) DVC Encoder

Figure 12.9: Examples of vision based encoding — Optimization of IPB bit allocations based on visibility of artifacts. The DVC encoder incorporates the VDM enhancements discussed in this section.

with this approach. The diagonal hatch processing loop shown in Figure 12.5 was used to produce the images in Figure 12.9. For this type of processing, a frame's complexity measure (as defined in TM5) is monitored for each frame type (IPB). Typically, the remaining bits for the GOP are apportioned according to the ratio of frame type complexity measures. To improve this performance, the results of the VDM analysis are used as a modifier to the frame type target bit allocations. For example, define the target bits for picture type j as TBj where j=I,P,B, the picture quality for a picture type as PQj and the average picture quality across all picture types (for the GOP) as <PQ>. Then the new target bits are computed as:

$$TB'_j \equiv TB_j \times \left(1 - k \times \frac{PQ_j - < PQ >}{< PQ >}\right) \tag{12.1}$$

where k is a reaction parameter that sets the rate of the compensation.. This approach typically takes one to two sub-GOPs (a sub-GOP is typically three frames long) after a scene change to settle into an optimal setting. For many scenes, this reallocation significantly improves image quality. (See [Lee03] for a detailed description of this approach.)

content, and therefore coded bit rate, in areas where the high frequencies would not be noticed.

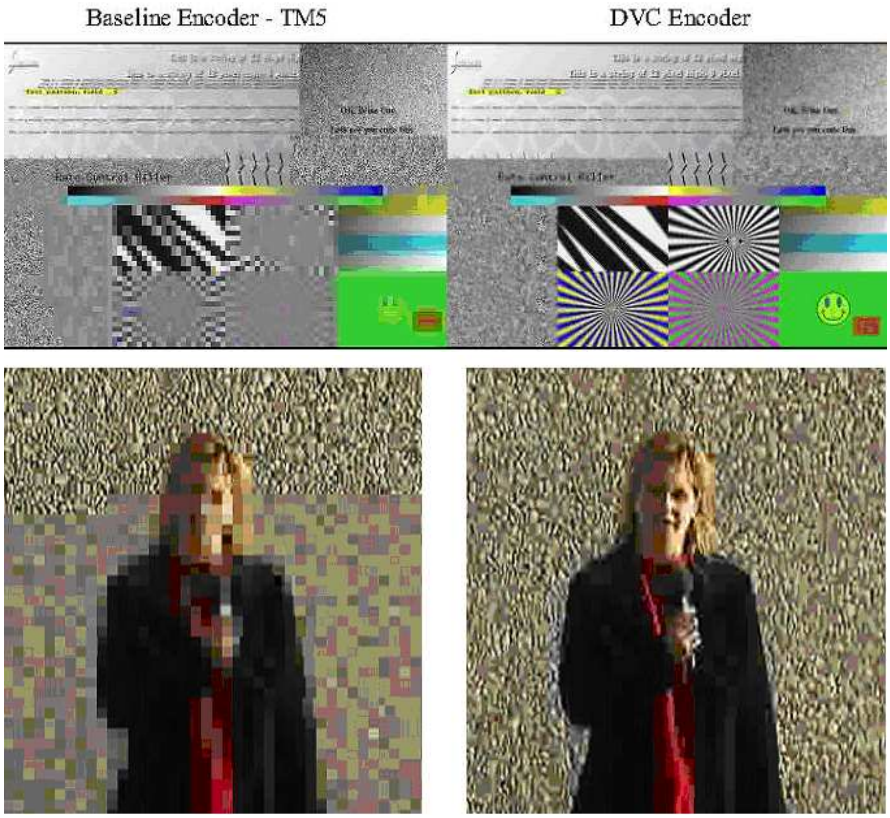


Figure 12.18: The left images are from an MPEG-2 baseline encoder using TM5 rate control. The right images are from an MPEG-2 encoder (DVC Encoder) that uses saliency pre-processing to filter out high frequencies before they are encoded.

12.6 Application to Multi-Channel Encoding

An additional requirement for many video distribution systems is the need to multiplex multiple video programs into a single distribution channel. It is not uncommon to see from 8-12 video program streams being multiplexed into a 30 Mbit/sec transport stream. Most products available today (statistical multiplexers) use MSE as part of their joint rate allocation scheme. The look-ahead approach (see Section 12.3.3) that estimates the rate-distortion behavior for the incoming video stream is ideally suited for use in a multi-channel multiplexer (Figure 12.19). With a sufficiently large look-ahead buffer (approximately a GOP), joint rate allocation can specify an optimal (i.e., minimized visual distortion) target bit rate for each encoding stream. The function of the joint rate allocation (JRA) module is to determine bit rates that equalize distortions across video

program streams while adhering to buffer constraints for the individual program streams and the transport stream. The JRA is well suited to linear programming techniques as well as approaches that are more heuristic.

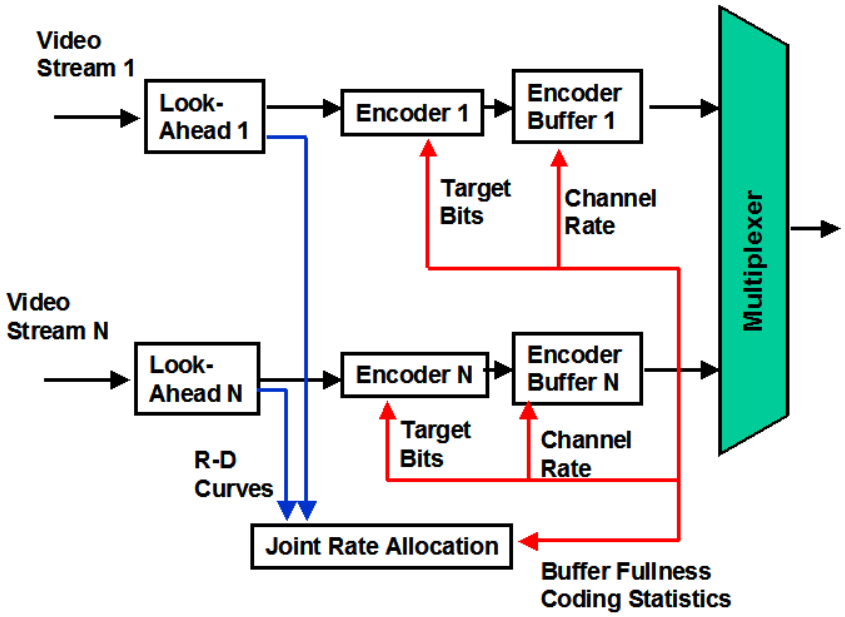


Figure 12.19: Schematic diagram of a multi-channel encoder.

References

[CP02] S. Cho and W. Pearlman. A Full-Featured, Error-Resilient, Scalable Wavelet Vdeo Codec Based on the Set Partitioning in Herarchical Trees (SPIHT) Algorithm. *IEEE Transactions on Circuits and Systems for Video Technology*, 12(3):157–171, March 2002.

[CSE00] C. Christopoulos, A. Skodras, and T. Ebrahimi. The JPEG2000 Still Image Coding System: An Overview. *IEEE Trans. Consumer Electronics*, 46(4):1103–1127, November 2000.

[CZ97] T. Chiang and Y. Zhang. A New Rate Control Scheme using Quadratic Rate Distortion Model. *IEEE Transactions on Circuits and Systems for Video Technology*, 7(1):246–250, 1997.

[Gha03] M. Ghanbari. *Standard Codecs: Image Compression to Advanced Video Coding*. U.K.: IEE, 2003.

[HW03] C. Ho and J. Wu. Toward User Oriented Scalable Video by Using Foveated FGS Bitstreams. In *ICCE 2003 Conference Proceedings*, 46–47, June 2003.

[HWW03] C. Ho, J. Wu, and S. Wang. A User Adaptive Perceptual Rate Control Scheme for FGS Videos. In *ICCE 2003 Conference Proceedings*, 42–43, June 2003.

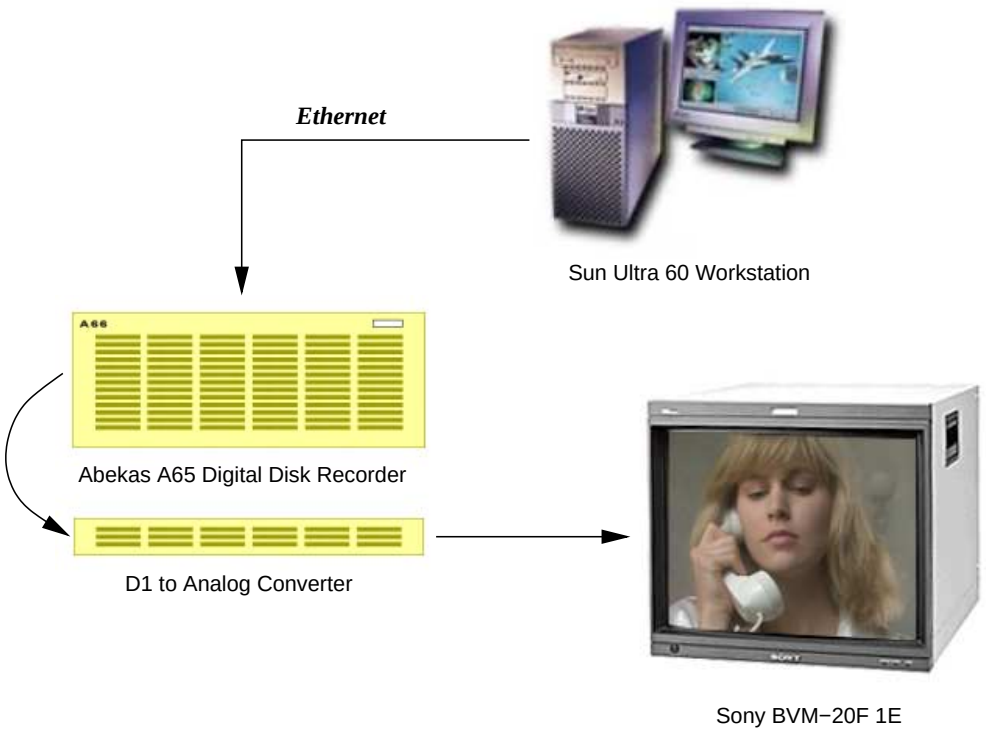


Figure 13.7: Equipment Setup for the DSCQS Test.

Table 13.2: Laboratory viewing conditions for the variant DSCQS test

Ratio of luminance of inactive screen to peak luminance:	≤ 0.02
Ratio of the luminance of the screen, when displaying only black level in a completely dark room, to that corresponding to peak white:	≤ 0.01
Display brightness and contrast setup via PLUGE - Rec. ITU-R BT.814 [ITU94a] and ITU-R BT.815 [ITU94b].	
Maximum viewing angle:	$\pm 30^o$
Ratio of luminance of background behind picture monitor to peak luminance of picture:	≤ 0.15
Other room illumination:	very low

13.3.5 Presentation of Material

The DSCQS test organizes the presentation of materials into sets. Each test set contains a pair of images, one for reference and the other for assessment. According to [ITU00], the order of presentation of the reference and test images should be randomized. However, in the interest of keeping the test simple so as to avoid systemic errors in the grading phase, this order has been fixed with the presentation of the reference image followed by the test image. This structure is given in Figure 13.8.

Figure 16.1: Illustration of color bleeding on one frame from the *Table-Tennis* sequence (see above the racket, as well as along the arm).

sual properties related to color image compression. Then, we give a thorough analysis of the color bleeding phenomenon, as a result of both quantization and decimation of chrominance data. Finally, an adaptive post-filtering algorithm is developed, based on the previous analysis. Simulation results for different color images show the improvement of the reconstructed video, both objectively and subjectively.

16.2 Analysis of the Color Bleeding Phenomenon

Color bleeding occurs when one color in the image bleeds into or overlaps into another color inappropriately. We demonstrate in this section that this distortion occurs because of both decimation and quantization of the chrominance components at the compression stage.

16.2.1 Digital Color Video Formats

Most current digital image compression standards handle colors as three separate components [Wan95]. Among the various existing colorimetric spaces, the YUV colorspace is widely used in digital video communications, where Y corresponds to the luminance part of the video signal, and U/V are the so-called chrominance components. Since the human eye is less sensible to colored details, the amount of chrominance information data can be reduced without decreasing the subjective image quality.

Several formats have been developed in recent years, in order to represent digital color video signals. The YUV 4:2:2 video format constitutes the reference for studio

Figure 16.3: One frame from the *Table-Tennis* sequence reconstructed from the Y coded component, and the C_b and C_r original components, respectively.

Figure 16.4: One frame from the *Table-Tennis* sequence: in this case, the image is reconstructed from the Y original component, and the C_b and C_r coded components, respectively.

component. Up to there, no sub-sampling is applied, so that the vector's length is equal to 16 pixels; the right half part is uniform, while a sharp transition is located in the left half part of the 1-D vector, corresponding to the presence of the color edge.

If the chrominance components are not sub-sampled, each of the 8-pixel half-vectors is transformed independently by means of DCT, and quantized. After decoding, coding errors are limited to the corresponding 8-pixel half-vector. In particular, spurious components due to ringing along the color edge are restricted to the left half-part, and do not bleed in the right one. However, in practice, chrominance data are sub-sampled prior to encoding (Figure 16.5c). After decimation (here, by a factor of 2), a single 8-pixel vector is assumed to represent the C_r color content of the whole 16-pixel length block. This 8-pixel vector is transformed by means of DCT, and then quantized. Due to quantization, the original balance of the DCT coefficients is corrupted, and distortions are introduced in the whole reconstructed signal (Figure 16.5d). Finally, the C_r chrominance component is interpolated in order to reconstruct the 16-pixel vector. Unfortunately,

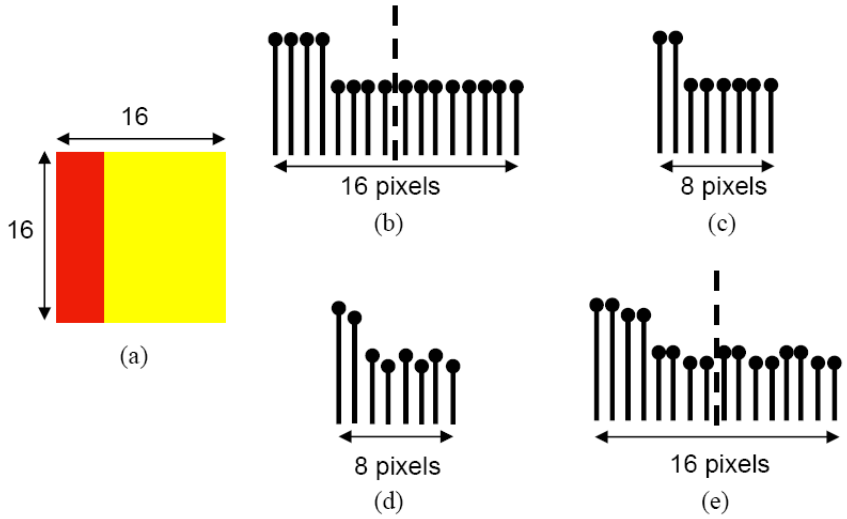


Figure 16.5: Schematic illustration of sub-sampling effect on the manifestation of color bleeding [CGC04]. (©2004 IEEE)

this process propagates errors to the right side of the signal, so that color bleeds locally beyond the image area where it should physically be restricted. This is clearly visible in Figure 16.5e where spurious chrominance variations have been artificially introduced in the right part of the 16-pixel decoded vector.

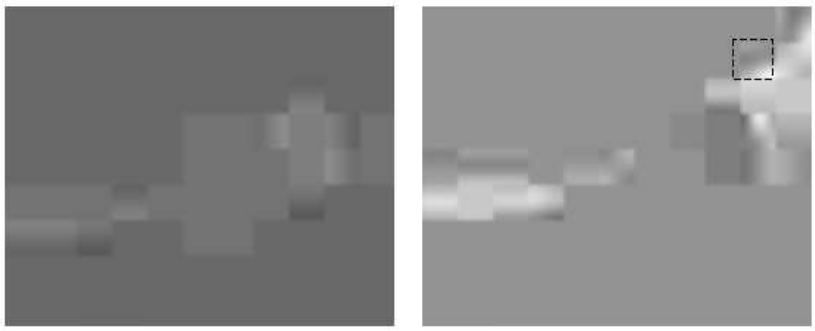


Figure 16.6: Enlarged parts of the C_b (left) and C_r (right) components of the coded frame shown in Figure 16.1; note that red color bleeds in the left part of the macro-block.

This analysis is easily extended to bi-dimensional signals. Figure 16.6 illustrates the C_b and C_r components of the coded frame presented in Figure 16.1. A 16×16 pixels macro-block, made from four adjacent blocks of 8×8 pixels each is selected (represented by dashed square lines). This macro-block contains a strong color edge, corresponding to the player's red shirt. Clearly, the original content of chrominance blocks has been widely modified: in particular, spurious red color components smear in the up-left part of the macro-block, leading to color bleeding distortion. In the case the same color image is compressed without chrominance decimation, no more color bleeding is

due to post-processing. Figure 16.9 shows the results of applying color bleeding removal to the color image shown in Figure 16.1. It is very noticeable that color defects have been efficiently reduced by adaptive neighborhood averaging. Note that the smearing of the color red has been efficiently cleaned up above the racket, as well as along the player's arm, and the ringing phenomenon has been removed, while preserving color edge sharpness. This leads to a great improvement in the visual appearance of the displayed color image.

Figure 16.9: Results of applying the de-color-bleeding post-processing to Figure 16.1.

Figure 16.10 shows another example of a color image subject to color bleeding. This image from the well-known *Foreman* sequence has been coded by means of DCT, using the quantization tables given in Table 16.1. Color bleeding is particularly noticeable on the foreman's chin (see the pink blob), as well as along the edges of the shoulder (middle left) and the crane (middle right). The right part of the figure illustrates the results of applying the post-filtering algorithm: clearly, the chrominance mismatches are greatly reduced, and the visual quality of reconstructed video is consequently better.

This chapter provides a thorough analysis of the color bleeding artifacts caused by standard digital image and video coding algorithms. A new post-filtering algorithm for the reduction of color bleeding artifacts is then proposed, that efficiently improves color fidelity. This algorithm is applied to the Y, C_r, C_b components of compressed video, and so can be implemented directly in classical decoders, without changing of colorimetric space. It can be integrated in a more general digital video post-processor addressing various image/video coding artifacts to improve rate-distortion optimization performance of the codecs.

Figure 16.10: One decoded frame from the Foreman sequence before (left) and after (right) post-processing.

References

[AFR01] A. Al-Fahoum and M. Reza. Combined Edge Crispiness and Statistical Differencing for Deblocking JPEG Compressed Images. *IEEE Transactions on Image Processing*, 10(9):1288–1298, September 2001.

[BK95] V. Bhaskaran and K. Konstantinides. *Image and Video Compression Standards*. Norwell, MA: Kluwer Academic Publishers, 1995.

[CGC04] F.-X. Coudoux, M. Gazalet, and P. Corlay. An adaptive Post-processing Technique for the Reduction of Color Bleeding in DCT Coded Images. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(1):114–121, January 2004.

[Gle93] W. Glenn. Digital Image Compression Based on Visual Perception and Scene Properties. *SMPTE Journal*, 102(5):392–397, 1993.

[ISO94] ISO/IEC JTC1 110918-1. ITU-T Rec. T.8.1. Information Technology–Digital Compression and Coding of Continuous-tone Still Images: Requirements and Guidelines. Technical report, ISO/IEC, 1994.

[MH98] M.Yuen and H.R.Wu. A Survey of Hybrid MC/DPCM/DCT Video Coding Distortions. *Signal Processing*, 70(3):247–278, November 1998.

[MPFL96] J. Mitchell, W. Pennebaker, C. Fogg, and D. LeGall. *MPEG Video Compression Standard*. New York: Chapman and Hall, 1996.

[NH95] A. Netravali and B. Haskell. *Digital Pictures: Representation and Compression, Second edition*. New York and London: Plenum Press, 1995.

[PL99] H. Park and Y. Lee. A Postprocessing Method for Reducing Quantization Effects in Low Bit-Rate Moving Picture Coding. *IEEE Transactions on Circuits and Systems for Video Technology*, 9(1):161–171, Febuary 1999.

[RY90] K. R. Rao and P. Yip. *Discrete Cosine Transform — Algorithms, Advantages, Applications*. Boston, MA: Academic Press, 1990.

Figure 17.15: Picture from the sequence *Bus* illustrating the positions of cell loss

Figure 17.16: Effect of concealment using macroblock replacement

Figure 17.17: Decoded *Flower Garden* image when motion compensated concealment using the motion vector in the macroblock above the lost macroblock is employed

Figure 17.18: Decoded *Bus* image when motion compensated concealment using the motion vector in the macroblock above the lost macroblock is employed

frame. It then calculates the total squared difference between these three lines and the corresponding three lines on the edge of a 16x16 block of data within a previous decoded frame. This is illustrated in Figure 17.19. The BMA estimates the lost motion vector as the one in which the squared difference between the surrounding lines (from the current decoded frame) and the block (from the previous decoded frame) is a minimum. Referring to Figure 17.19, this means that the total squared difference calculated by summing the following three square differences is minimized:

- squared difference between the pixels above the block and the pixels on the top line of the block (i.e. region A in Figure 17.19)
- the squared difference between the pixels to the left of the block and the pixels on the left edge of the block (i.e. region B in Figure 17.19)
- the squared difference between the pixels below the block and the pixels on the bottom line of the block (i.e. region C in Figure 17.19)

The search method employed to estimate the lost motion vector can be a full search over some area in the previous frame. Alternatively, the search process can be greatly speeded up if only a small number of candidate motion vectors are considered. These may include:

- the motion vector for the same macroblock in the previous frame
- the motion vectors associated with available neighboring macroblocks

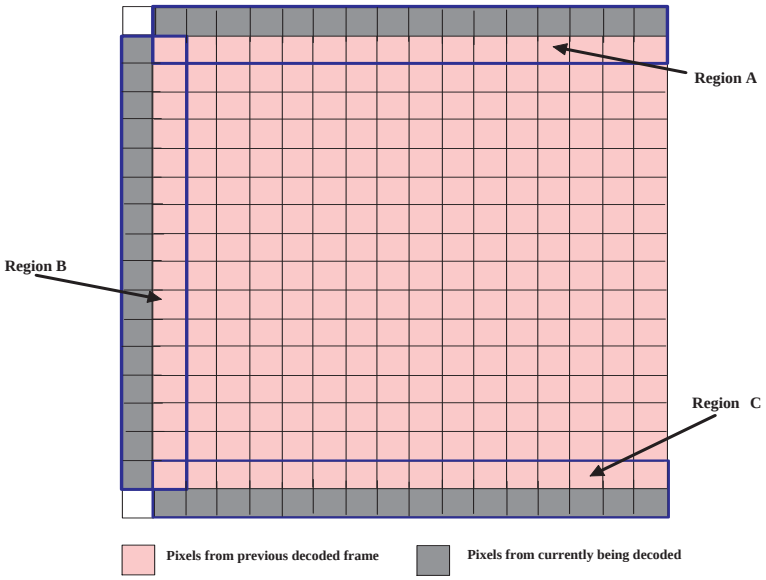


Figure 17.19: Matching technique employed in boundary matching algorithm [W. 93]

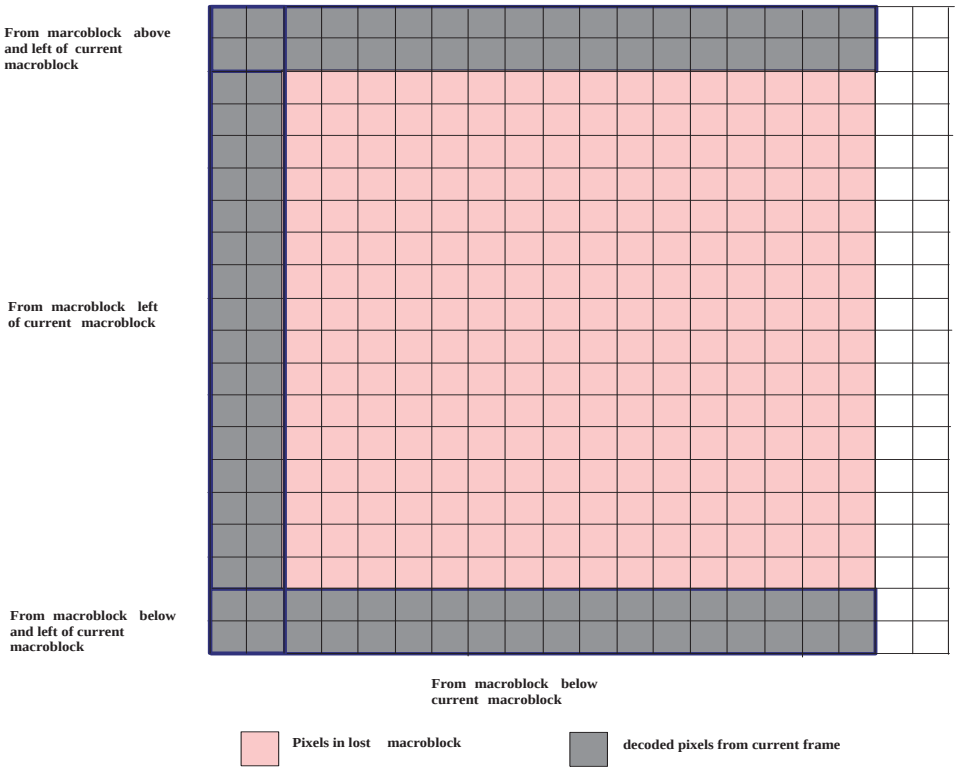


Figure 17.20: Matching technique employed in decoded motion vector estimation algorithm

addition, pixels from the above-left macroblock (even if this macroblock is itself a concealed macroblock) and below-left macroblock (if received correctly) are used to complete the encirclement of the lost macroblock. If it is assumed that only two surrounding lines are used and that macroblocks above, below and right of the lost macroblock are available then the lines used are as shown in Figure 17.20.

The algorithm then performs a full search within the previous frame for the best match to the available lines of decoded pixels from the current frame. The macroblock of data which is surrounded by the lines which best matches the lines from the current frame is assumed to be the best match to the lost macroblock. In the experiments performed for this work, a search area of (-16, +16) pixels is used both horizontally and vertically. However, this is a decoder option and can be used to trade performance against computational complexity. Motion estimation is performed to half pixel accuracy since it was found during the experiments that the filtering associated with half pixel accuracy prediction tends to smooth any blocking artifacts which might otherwise be apparent between concealed macroblocks and normally decoded macroblocks within a frame. Hence, to speed up processing, a search to single pixel accuracy was first performed and then the best match within (-0.5, + 0.5) pixels surrounding that point was chosen for concealment.

Figure 17.23: Sampled picture (No.18) from original sequence

Figure 17.24: Sampled picture (No.18) decoded by Above-MVs Tech. (CLP=0.05)

Figure 17.25: Sampled picture (No.18) decoded by 8-around-line with bi-directional search (CLP=0.05)

Figure 17.26: Sampled picture from (No.57) from original sequence

Figure 17.27: Sampled picture (No.57) decoded by Above-MVs Tech. (CLP=0.05)

Figure 17.28: Sampled picture (No.57) decoded by 8-around-line with bi-directional search (CLP=0.05)

Figure 17.29: Sampled picture from (No.60) from original sequence

Figure 17.30: Sampled picture (No.60) decoded by Above-MVs Tech. (CLP=0.05)

Figure 17.31: Sampled picture (No.60) decoded by 8-around-line with bi-directional search (CLP=0.05)

Figure 17.32: Sampled picture from (No.18) decoded by Above-MVs Tech. (CLP=0.01)

Figure 17.33: Sampled picture (No.18) decoded by 8-around-line with bi-directional search (CLP=0.01)

Figure 17.34: Sampled picture (No.57) decoded by Above-MVs Tech. (CLP=0.01)

Figure 17.35: Sampled picture (No.57) decoded by 8-around-line with bi-directional search (CLP=0.01)

Figure 17.36: Sampled picture (No.60) decoded by Above-MVs Tech. (CLP=0.01)

Figure 17.37: Sampled picture (No.57) decoded by 8-around-line with bi-directional search (CLP=0.01)

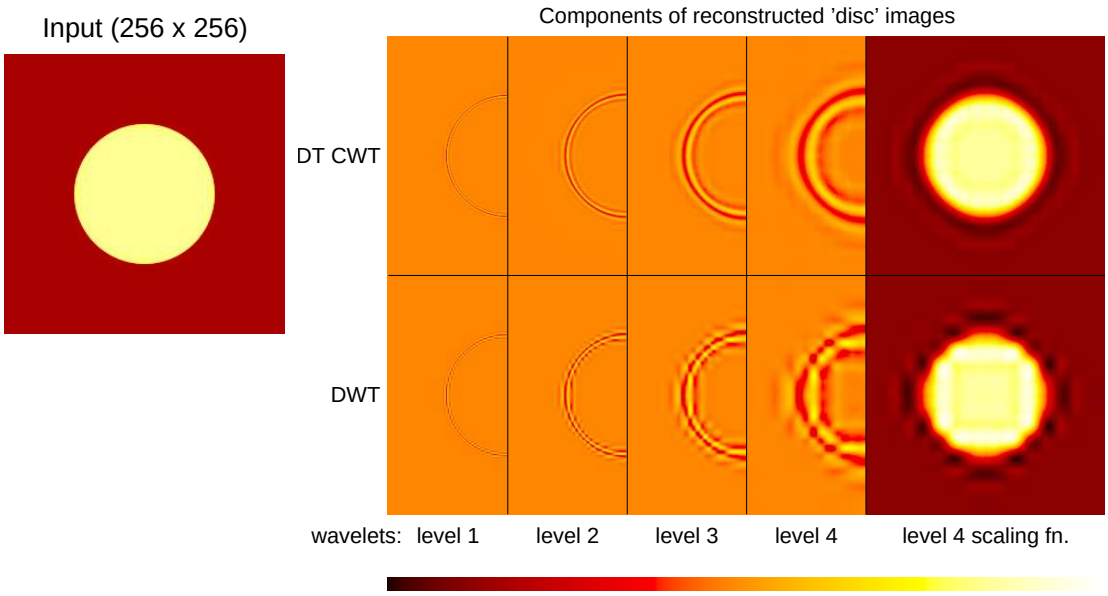


Figure 18.2: Wavelet and scaling function components at levels 1 to 4 of an image of a light circular disc on a dark background, using the 2-D DT CWT (upper row) and 2-D DWT (lower row). Only half of each wavelet image is shown in order to save space. Courtesy of N. G. Kingsbury.

To achieve further performance improvements for digital picture coding in a perceptual quality and bit rate optimization sense, selection of a better coding framework and structure may hold a vital key. In addition to vision modeling inspired transforms such as the Gabor transform [Dau88], the cortex transform [Wat87] and the steerable pyramid decomposition [SFAH92], a number of alternative (complete and over-complete) mathematical frameworks and structures have been made available such as directional wavelet transforms [Vet01, DV03, DVV02, DV01, DV02] and the DT CWT [Kin01], offering solutions to eliminating some hard to solve problems associated with classical complete transforms such as the DCT and the DWT (which have dominated picture coding applications) from their roots. A key issue is what performance criteria should be used in selecting the most suitable framework for the next generation picture compression technology.

18.1.3 Decisions Decisions

When going forward and searching for a better mathematical framework or coding structure for digital picture compression, it all comes down to how much one is prepared to expand or decompose a given picture in an alternative representation (or representations) in order to achieve maximum picture compression. It is one thing to acknowledge that the time honoured philosophy for thousands of years, "What needs be shrunk must first

Figure 18.6: Examples of JPEG2000 coding using different distortion metrics at 0.125bpp: (a) Original *Lena* image, (b) the MSE, (c) CVIS, and (d) a vision model based perceptual metrics [TTW04]. ©IEEE.

in various chapters seems enough a reason for them to be ruled out for consideration as a serious contender for picture compression. This, however, does not prevent introductions of plenty of HVS inspired picture coders over the years, such as model- or object-based coding techniques [TK96], and foveated coders as discussed in Chapter 14. Neither incomplete knowledge of human vision nor engineering/technological difficulties could deter further research and investigations along this direction. From discussions and examples presented in subsections 18.1.2 and 18.5.1, HVS inspired coders such as vision model based coders do offer a more concerted approach to minimization or elimination or balancing of perceptual picture coding artifacts. Significant computational overhead existing in more sophisticated vision model based picture coders will remain as an obstacle in practical applications, notwithstanding that an ever increasing