Journal of Zhejiang University SCIENCE A
ISSN 1009-3095 (Print); ISSN 1862-1775 (Online)
www.zju.edu.cn/jzus; www.springerlink.com
E-mail: jzus@zju.edu.cn

A rate control scheme for H.264 video under low bandwidth channel

YIN Ming†[1,2], WANG Hong-yuan[1]
([1]Department of Electronics and Information Engineering, Huazhong University of Science and Technology, Wuhan 430074, China)
([2]College of Automation Engineering, Guangdong University of Technology, Guangzhou 510090, China)
†E-mail: yinmhn@126.com

Received Sept. 12, 2005; revision accepted Feb. 13, 2006

Abstract: The dilemma of the quantization parameter (QP) being involved in both rate control and rate-distortion optimization (RDO) prevents using the traditional rate control scheme. Although some rate control schemes are proposed to circumvent the dilemma, the inaccurate prediction model and improper bit allocation deter H.264 application on low bandwidth channel. To resolve this issue, this paper proposes a novel rate control scheme by considering the macroblock (MB) encoding complexity variation and buffer variation and by exploiting the spatio-temporal correlation sufficiently well. Simulations showed that this scheme improves the perceptual quality of the pictures with similar or smaller PSNR deviations when compared to that of rate control in JVT-O016.

Key words: Video coding, H.264, Rate control, HRD, Bit allocation
doi:10.1631/jzus.2006.A0990 **Document code:** A **CLC number:** TP37

INTRODUCTION

Rate control has been playing an increasing important role in real-time video communication. An encoder employs rate control to regulate output bit-streaming to meet certain channel bandwidth and buffer constraints while pursing high video quality. One fundamental problem in the encoder design is the selection of quantization parameter (QP) to maximize visual quality under constraints imposed by the computational complexity, delay, bandwidth and/or loss factors. At present, there exists common rate control schemes in standard, like MPEG-2 TM5, H.263 TMN8, MPEG-4 VM8.

H.264/AVC, jointly developed by the ITU-T VCEG (2003) and the ISO/IEC MPEG (2003) standards committees, is mainly intended for video transmission in all areas where bandwidth or storage capacity is limited, supplying enhanced coding efficiency and improved network adaptation. In implementing RDO for a macroblock (MB), a quantization parameter should be firstly determined for the MB by using the mean absolute difference (MAD) of the current frame and/or MB.

However, the MAD is only available after making mode decision by using the existing rate control scheme such as TM5, TMN8. To resolve this dilemma, Li *et al*.(2003a) presented a linear model to predict MAD and adopted a fluid flow traffic model to allocate target bit rate for current frame or MB. To meet the hypothetical reference decoder (HRD) requirements, the target bits are further bounded in (Li *et al*., 2003b).

However, to estimate the target bits for each frame, a common straightforward way is used, namely, an equal number of bits is allocated to each frame regardless of its complexity. Moreover, the linear MAD model is weak in predicting picture characteristics. As a result, there usually exists large fluctuation in the virtual buffer and non-uniform distortion over a video sequence. Meanwhile, Jiang *et al*.(2004) presented that the inaccurately predicted MAD value used in R-D model would result in wrong QP and consequently degrade RDO performance. Yuan *et al*.(2005) proposed an optimum bit allocation scheme to improve the rate control accuracy, though its large computational complexity deterred the application in real-time video transmission.

In addition, the encoding rate and frame rate are all low on low bandwidth channel so that the degree of correlation between frames is relatively low. Therefore, using B frame under low bit-rate cannot improve the image quality and could add to encoding complexity and delay reversely. Consequently, B frame is usually not applied on low bandwidth channel in (Song and Jay Kuo, 2001). Under low bit-rate, the bit consumption of coding H.264 header information varies so significantly that it could not be omitted.

To resolve this problem, a simple but accurate rate control scheme with H.264 HRD is proposed in this paper. We focus on the MAD prediction by effectively utilizing the spatio-temporal correlation. In addition, an optimum bit allocation is employed among MBs based on improved complexity estimation measure.

The organization of this paper is as follows. Section 2 briefly introduces preliminary knowledge for later section. In Section 3, HRD constrained rate control scheme is proposed. Numerous experimental results are presented in Section 4 for demonstrating the effectiveness of the proposed scheme under low bandwidth channel. Section 5 concludes this paper.

PRELIMINARY KNOWLEDGE

Coding complexity estimation

Generally speaking, MAD as distortion measure is used for representing the complexity of coding unit. Larger MAD value shows higher coding complexity, possibly due to high motion or scene change, and vice versa.

To avoid the dilemma, a linear MAD model is proposed in JVT-G012. By using the model, the MAD of not yet coded blocks can be predicted by that in the co-located position of the previous P frame. The linear model has an effect on many cases, however, exhaustedly using temporal correlation cannot guarantee prediction accuracy, especially spatial correlation occurs without temporal correlation. Moreover, the accuracy of previous rate control schemes relies strongly on the linear regress model. Once large prediction error exists, the rate control performance is degraded significantly. Therefore, we propose a simple way to utilize spatio-temporal correlation sufficiently well.

Assume that the current basic unit is the mth MB in nth frame, (i, j) is its coordinate (Fig.1) and $MAD_n(m)$ denotes its complexity. The neighboring block is composed of a set $S=\{\psi(i-1, j-1), \psi(i-1, j), \psi(i, j-1), \psi(i', j')\}$, where $\psi(i-1, j-1)$, $\psi(i-1, j)$, $\psi(i, j-1)$, $\psi(i', j')$ respectively represent the upper-left, left, upper and temporal position block with their complexity being correspondingly $MAD_n(m_3)$, $MAD_n(m_2)$, $MAD_n(m_1)$ and $MAD_{n-1}(m)$. Considering the correlation of pictures in the temporal and spatial domain, $MAD_n(m)$ is predicted by using the adjacent position of the current frame and previous frame.

$$MAD_n^{\mathrm{Pred}}(m) = [MAD_n(m_1) + MAD_n(m_2) + MAD_n(m_3) + MAD_{n-1}(m)]/4. \tag{1}$$

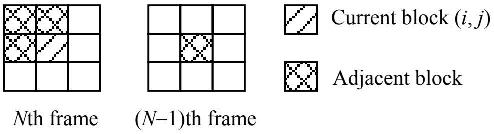


Fig.1 Position of the block

Further, we define the relative complexity measure as the MAD ratio of the current unit predicted MAD to the average MAD in all previously encoded P frames. The ratio can be easily calculated from Eq.(2):

$$MAD_{\mathrm{ratio},m} = \frac{MAD_n^{\mathrm{Pred}}(m)}{\frac{1}{n-1}\sum_{l=1}^{n-1} MAD_l}, \tag{2}$$

where MAD_l denotes the actual MAD of previously encoded P frame.

DCT coefficient statistical feature

Knowledge of the DCT coefficients' probability distribution is important in rate control for video coding, since the problems of bit allocation and quantization scale selection require information on the rate-distortion relationship as a function of the encoder parameters and the video source statistics. Statistics show that the AC coefficients of the DCT have approximately independent Laplacian distributions, which is the most popular model and widely used in practice (Netravail and Haskell, 1998).

Assume that $p(x)$ is the probability density function of the AC coefficients with parameter ω,

$$p(x)=\frac{\omega}{2}\exp(-\omega\,|\,x\,|), \tag{3}$$

where $\omega=1/E(|x|)$, $E(|x|)$ denotes the mean of the DCT coefficient absolute value in an MB.

In conventional rate control scheme, R-D model, such as the quadratic R-D model in (Li *et al.*, 2003b) or the linear model in (Li *et al.*, 2003c), is employed to estimate the target number of bits allocated to a frame or MB. In this paper, we shall adopt and enhance the model used in TMN8 in (Ribas-Corbera and Lei, 1999). So the coefficient rate model could have the following form:

$$R(Q)=\begin{cases}\frac{1}{2}\log_2(2\mathrm{e}^2\frac{\sigma^2}{Q^2}), & \frac{\sigma^2}{Q^2}>\frac{1}{2\mathrm{e}},\\ \frac{\mathrm{e}}{\ln 2}\frac{\sigma^2}{Q^2}, & \frac{\sigma^2}{Q^2}\le\frac{1}{2\mathrm{e}},\end{cases} \tag{4}$$

where σ^2 represents the variance of the displaced frame difference (DFD) in pixel domain, e is the base of natural logarithms.

For low bit rate application, the value of σ/Q is always small so that Eq.(4) is deduced to be:

$$R(Q)=\frac{\mathrm{e}}{\ln 2}\frac{\sigma^2}{Q^2}. \tag{5}$$

In general, a more easily computable form, in which MAD replaces σ^2, is developed:

$$R(Q)=\frac{\mathrm{e}}{\ln 2}\frac{MAD}{Q^2}. \tag{6}$$

In this paper, MB as basic unit is used for example. Let $R(m)$ denote the bits for coding MB m, $Q(m)$ represents the quantization step, α refers to the coefficient, and $MAD^{\mathrm{Pred}}(m)$ is the prediction of the current MB. Then the improved RD model formula is as follows:

$$R(m)=MAD^{\mathrm{Pred}}(m)\cdot\left(\frac{\mathrm{e}}{\ln 2}\frac{\alpha}{Q^2(m)}\right). \tag{7}$$

HRD-CONSTRAINED RATE CONTROL SCHEME

The scheme is composed of three layers: GOP layer, frame layer and MB layer target bit-allocation, which is determined based on target buffer, frame rate, the available channel bandwidth and the actual buffer occupancy. In GOP layer rate control, we need to compute the total number of remaining bits for all non-coded frames in each GOP. Then, the proposed algorithm uses frame-layer rate control to select the target number of bits for the current frame and an MB-layer rate control to select the value of QP for the MBs.

GOP layer rate control

In many applications, e.g. real-time encoders, the encoder does not know the total number of frames that need to be coded beforehand, or when scene changes will occur. Thus we adopted GOP layer rate control to allocate target bits for each picture. The H.264 standard does not actually contain GOP, but the terminology is used here to represent the distance between I pictures.

Assuming that each GOP consists of N frames, its structure is IPP…P, with I being an intra-coded frame, P being a forward predicted frame. We define $G_{k,n}$ as the nth frame in the kth GOP, and $B_{\mathrm{v}}(k,n)$ as the virtual buffer occupancy before coding the nth frame. Based on the fluid flow traffic model, $B_{\mathrm{v}}(k,n)$ can be shown as Eq.(8):

$$\begin{aligned}B_{\mathrm{v}}(k,n)=\min\big(&\max\big(0,B_{\mathrm{v}}(k,n-1)+R(k,n-1)\\&-u(k,n-1)/F\big),B_{\mathrm{s}}\big),\end{aligned} \tag{8}$$

where $B_{\mathrm{v}}(1,1)=B_{\mathrm{s}}/8$, $B_{\mathrm{v}}(k,n-1)$ is the buffer occupancy after encoding $G_{k,n-1}$, $R(k,n-1)$ denotes the actual generated bits in $G_{k,n-1}$, $u(k,n-1)$ is the available bandwidth, F stands for predefined frame rate and B_{s} is the buffer size whose maximum is determined by different level and profile (ISO/IEC, 2003).

The total number of remaining bits after the $(n-1)$th frame is shown in Eq.(9):

$$T_{\mathrm{GOP}}(k)=\begin{cases}(u(k,1)/F)\cdot N-B_{\mathrm{v}}(k-1,N+1)+B_{\mathrm{s}}/8, & k=1,\\ T_{\mathrm{GOP}}(k,n-1)+\dfrac{u(k,n)-u(k,n-1)}{F}\times(N-n)\\ \qquad -R(k,n-1), & k=2,3,\ldots,N.\end{cases} \tag{9}$$

The details of the GOP layer rate control are given in (Li *et al.*, 2003b; 2003c).

Frame-layer rate control

This layer consists of two stages: pre-encoding and post-encoding stages. The pre-encoding stage is aimed at computing target bits for the current frame. In the post-encoding stage, the model parameters are updated and where whether frame skipping is advisablely decided.

Pre-encoding stage: In this scheme, the starting QP of the first P frame in GOP is a predefined QP_0. Then the bit-allocation for the successive P frames is computed as follows: $T_f(k, n)=T_{GOP}(k, n)/(N-n+1)$. Taking the correlation of neighboring frames into account, the bit-rate of the previous frame has an impact on that of the current one, so the target bits are a weighted combination:

$$\begin{aligned}T_f(k,n)=&\frac{T_{GOP}(k,n)}{N-n+1}\cdot(1-\delta)\\&+\delta\cdot[T_f(k,n-1)-R(k,n-1)],\end{aligned}\tag{10}$$

where δ is a weighting factor with typical value of 0.05 in this scheme.

HRD requirements describing the status of the decoder buffer are incorporated to guarantee the buffer does not overflow or underflow in H.264. The obtained bit allocation from Eq.(10) may or may not be allowed by HRD requests. Hence, the target bit-rate for $T_f(k,n)$ should be further bounded by:

$$T_f(k,n)=\begin{cases}0.9U_n, & T_f(k,n)>0.9U_n,\\ L_n, & T_f(k,n)<L_n,\\ T_f(k,n), & \text{else,}\end{cases}\tag{11}$$

where the upper bound for picture n is $U_n=[t_r(n)-t_{ai}(n)]\cdot T_f(k,n)$, the lower bound is: $L_n(n)=\max((t_{bi}(n+1)-t_{ai}(n))\cdot T_f(k,n),0)$.$t_{ai}(n)$, called the initial arrival time, is the time when the first bit of frame n enters the coded picture buffers (CPB). $T_{bi}(n)$ is the earliest time when the coded picture data enter the CPB. $t_r(n)$, called the removal time, is the time when the frame n is removed from the CPB in (Wiegand, 2002; Ma *et al.*, 2003).

Post-encoding stage: After coding a frame, the prediction model parameters, such as in Eqs.(1), (2), are updated by linear regression method. Meanwhile, the actual bits generated are added to the buffer. If the updated buffer fullness has reached a safety margin, i.e. 85%, of the buffer size, the encoder will skip the upcoming frame. The modified buffer after frame skipping is:

$$B_v(k,n+1)=\max\left(\frac{B_s}{8},\ B_v(k,n-1)-\frac{u(k,n)}{F}\right).\tag{12}$$

MB-layer rate control

In this layer, an MB-layer adaptive scheme is developed to determine the QP for each MB in a frame, via 6 steps:

Step 1: Determine the number of texture bits, $T(k,n,m)$, for the current MB.

To get better bit estimation for each MB, we need to consider the unit complexity $MAD_{ratio,m}$ in the computation. If the MAD of the current MB is greater than the average MAD of the previous frame and the buffer fullness of the belonging frame is less than 75% level, then increase the target bits by 10%. We predict the MAD of the current MB in the current frame by using Eq.(1). The target bits for coding current MB is further adjusted in Eq.(13),

if ($MAD_{ratio,m}>1$ && $B_v(k,n)<0.75B_s$)

$$T(k,n,m)=1.1\times T(k,n)/N_{mb}\tag{13}$$

else

$$T(k,n,m)=T(k,n)/N_{mb}\tag{14}$$

where N_{mb} is the total number of MBs in a frame.

Since the coding structure in H.264 has become very complicated, a possible dramatic change may occur in the number of bits required to encode side information, such as MB mode, motion vector and quantization parameter. So the target bit for encoding the current MB is:

$$T(k,n,m)=T(k,n,m)-T_{n-1}^{\text{Head}},\tag{15}$$

where T_{n-1}^{Head} denotes the bits for motion vector and header in the $(n-1)$th picture.

Step 2: Compute the quantization step of current MB.

Substituting $T(k,n,m)$ into Eq.(7) yields Eq.(16), and $Q(m)$ is obtained by the picture spatial activity considering rate constraint.

$$T(k,n,m) = MAD^{\text{Pred}}(m) \cdot \left(\frac{\mathrm{e}}{\ln 2} \frac{\alpha}{Q^2(m)} \right). \quad (16)$$

Step 3: Adjust *QP* adaptively to enforce the CBP constraints.

Quantization step *Q*(*m*) is then converted to corresponding *QP*(*m*) based on the exponential formula from (ISO/IEC, 2003), as shown in Eq.(17):

$$Q(m) = 0.67 \times 2^{QP(m)/6}. \quad (17)$$

To maintain visual quality smoothness in successive frames, the computed *QP*(*m*) is adjusted so that the *QP* error between two neighboring MBs does not exceed 2.

$$QP(m) = \begin{cases} \max\left(QP(m-1)-2, QP(m)\right), & QP(m) < QP(m-1), \\ \min\left(QP(m-1)+2, QP(m)\right), & \text{else.} \end{cases} \quad (18)$$

In H.264/AVC standard, the quantization parameter is specified as 0~51. *QP*(*m*) is to be regulated,

$$QP(m) = \begin{cases} 0, & QP(m) < 0, \\ 51, & QP(m) > 51. \end{cases} \quad (19)$$

Step 4: Perform RDO for all MBs in the current frame.

Step 5: Update the number of remaining bits and the number of non-coded MB in the current frame.

Step 6: After coding all MBs in a frame, update the model parameter α in Eq.(7) by exploiting the MAD of error and actual bits generated for coding MB.

SIMULATION AND RESULTS

To evaluate the performance, we have implemented our proposed rate control scheme on the JM9.4 test model software from (JVT, 2005) and compared it with the JVT-O016. In the simulation, fast motion estimation and mode decision in (Yin *et al*., 2003) were applied. Many QCIF 4:2:0 sequences with various target bit rates or frame rates were tested. The structure of the GOP sequence is IPP...P, and motion estimation search range is ±32 with 5 reference frames. All other features such as Hadamard transform and RDO were enabled. When a frame is skipped, the *PSNR* calculation is decided by using the previous encoded frame instead of the skipped one.

Table 1 of the test results shows that our proposed scheme improves the average *PSNR*, and also significantly reduces the *PSNR* standard deviations of the sequences with high or low motion when compared with those of JVT-O016 and also shows that both algorithms can achieve accurate target bit rates.

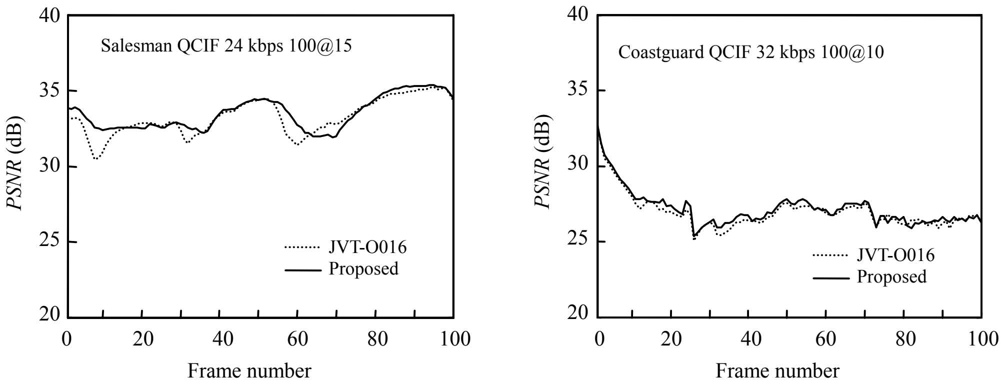
Additional comparisons are given in Fig.2, where the *PSNR* curves for Salesman and Coastguard with the proposed scheme and JVT-O016 are shown respectively. We can find that the proposed scheme achieves more stable reconstructed quality. Fig.3 shows the number of bits in each frame's buffer. From these figures, we can observe that the proposed scheme yields steadier buffer fullness, implying that our algorithm produces a more stable buffer status without overflow or underflow.

CONCLUSION

In this paper, a novel rate control strategy for H.264 on low bandwidth channel is proposed, which improves the prediction accuracy of MAD by using the spatio-temporal correlation. Meanwhile, an improvement measure of complexity estimation is applied for bit allocation among MBs. As demonstrated in our simulation, compared to H.264 and new proposal JVT-O016, our proposed scheme produces

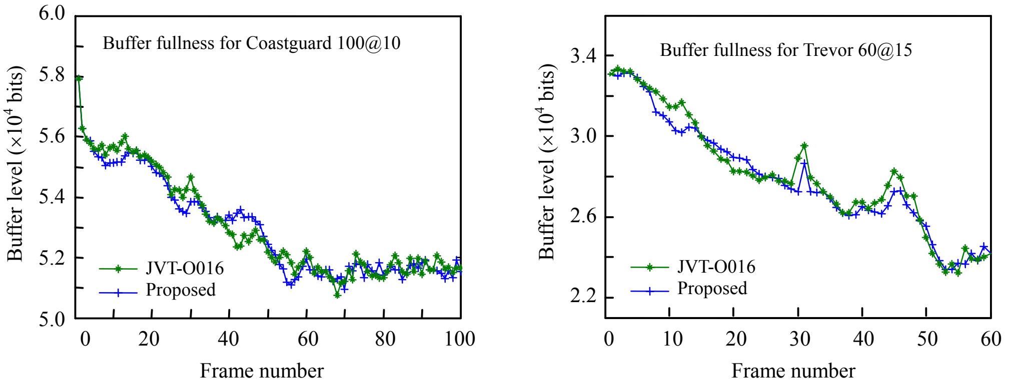
Table 1 Performance with the proposed rate control scheme

Sequence	Frame rate (fps)	Average *PSNR* (dB)			*PSNR* Std. deviation		Bit rate (kbps)	
		JVT-O016	Proposed	Gain	JVT-O016	Proposed	JVT-O016	Proposed
Foreman (48 kbps)	100@15	33.62	33.39	−0.23	1.19	1.04	48.07	48.20
Salesman (24 kbps)	100@15	33.22	33.47	0.25	1.19	1.08	24.13	24.15
News (24 kbps)	75@7.5	34.25	34.56	0.31	0.63	0.67	24.09	24.07
Trevor (64 kbps)	60@15	35.10	35.53	0.43	2.17	2.01	64.38	64.22
Coastguard (32 kbps)	100@10	26.96	27.05	0.09	1.17	1.17	32.06	32.00



(a) (b)

Fig.2 *PSNR* comparison for Salesman (a) and Coastguard (b)



(a) (b)

Fig.3 Buffer fullness comparison for Coastguard (a) and Trevor (b)

smoother visual quality variations during low or high motions and results in similar or higher visual qualities with lower computational complexity. In short, our scheme is very useful in low bit rate video streaming over Internet or wireless networks.

References

ISO/IEC 14496-10, 2003. Coding of Audiovisual Objects—Part 10: Advanced Video Coding.

ITU-T Recommendation H.264, 2003. Advanced Video Coding for Generic Audiovisual Services.

Jiang, M., Yi, X., Ling, N., 2004. Frame-Layer Bit Allocation Scheme for Constant Quality Video. IEEE International Conference on Multimedia and Expo. Taipei.

JVT Reference Software Version 9.6, 2005. Http://ftp3.itu.ch/av-arch/jvt-site/reference_software.

Li, Z.G., Pan, F., Keng, P., *et al.*, 2003a. Adaptive Basic Unit Layer Rate Control for JVT. Joint Video Team of ISO/IEC and ITU 7th Meeting, JVT-G012.

Li, Z.G., Gao, W., Pan, F., 2003b. Adaptive Rate Control with HRD Consideration. Joint Video Team of ISO/IEC and ITU 8th Meeting, JVT-H014.

Li, Z.G., Zhu, C., Ling, N., Yang, X.K., Feng, G.N., Wu, S., 2003c. A unified architecture for real-time video coding systems. *IEEE Trans. Circuits Syst. Tech.*, **13**(6):472-487. [doi:10.1109/TCSVT.2003.813420]

Ma, S.W., Gao, W., Wu, F., Lu, Y., 2003. Rate Control for JVT Video Coding Scheme with HRD Considerations. Proceedings International Conference on Image and Processing. Spain, p.793-796.

Netravail, A.N., Haskell, B.G., 1998. Digital Picture-Representation and Communication. Plenum Press.

Ribas-Corbera, J., Lei, S., 1999. Rate control in DCT video coding for low-delay communication. *IEEE Transaction on CSVT*, **9**(1):172-185.

Song, H., Jay Kuo, C.C., 2001. Rate control for low-bit-rate video via variable-encoding frame rates. *IEEE Transaction on Circuits and Systems for Video Technology*, **11**(4): 512-521. [doi:10.1109/76.915357]

Wiegand, T., 2002. Editor's Proposed Draft Text Modifications for Joint Video Specification. Ftp://ftp.imtc-files.org/jvt-experts/2002-10Geneva/JVT-E146d7ncm. zip.

Yin, P., Tourapis, H.Y.C., Tourapis, A.M., Boyce, J., 2003. Fast Mode Decision and Motion Estimation for H.264. ICIP2003.

Yuan, W., Lin, S.X., Zhang, Y.D., *et al.*, 2005. Optimum Bit Allocation and Rate Control for H.264/AVC. Joint Video Team of ISO/IEC and ITU 15th Meeting, JVT-O016.