Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG
(ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6)
15th Meeting: Busan, KR, 16-22 April, 2005

Document: JVT-O046
Filename: JVT-O046.doc

Title:	Improved FGS scheme for JSVM : some preliminary results[1]
Status:	Input Document to JVT
Purpose:	Proposal

Author(s) or Contact(s): Vincent Bottreau, Christine Guillemot, Hervé Jégou
IRISA-INRIA
Campus de Beaulieu
35042 Rennes cedex, France

+33 (0) 2 99 84 25 44
Tel: +33 (0) 2 99 84 71 71
Email: vbottrea@irisa.fr

Source: IRISA-INRIA

Abstract

JSVM [2] includes Fine Grain SNR (Quality) Scalability (FGS) for encoding the residue between the spatial transform coefficients of the original picture obtained after the MCTF of the corresponding spatial layer and their reconstructed base layer representation (or the subordinate enhancement layer representation). Each quality enhancement layer contains a refinement signal that corresponds to a bisection of the quantization step size. Thus transform coefficients are progressively refined by repeatedly decreasing the quantization step size and applying a modified CABAC entropy coding process akin to sub-bit-plane coding.

We describe here an alternative FGS coding scheme. It makes use of binarization codes adapted to the spatial (4x4) transform coefficient distributions. The second key feature of the approach relies on a rate-distortion optimized scanning order of the tree representing the binarization code. In addition, each residual frame is encoded using a rate-distortion optimization among transform coefficient subbands. The third novelty of the method is that the decoder uses an expectation-based decoding technique. In addition, the method avoids exhaustive quantization step search. Preliminary results show that the proposed method provides up to about 0.5 dB of improvement in coding efficiency performance for the QCIF15 layer of Scenario 2 test sequences when integrated into the SVM-3.0 software. However, these results may be seen as direct encoding for each target bit-rate since here we do not take into account the sub-optimality that may occur when decoding at a lower bit-rate.

Introduction

Built into the SVM-3.0 software, our improved FGS scheme functions broadly as follows:

- A base layer is encoded using a non-embedded scheme, possibly compatible with H.264-AVC, and the encoded version is reconstructed;
- This reconstructed base layer representation is subtracted from the original picture obtained after the MCTF of the corresponding spatial layer, thus yielding a residual signal;

[1] This work has been realized within IST European co-funded project DANAE [1] under contract IST-1-507113.

- This differential is transformed using 4x4 transform for all 3 color components, and the transform coefficients are separated into subbands by frequency. However, no additional transform is performed for luminance with intra16x16 prediction and chrominance as in AVC;
- Each 4x4 subband Probability Density Function is modeled as a Generalized Gaussian Density from which an adapted binarization of transform coefficients is derived;
- Transform coefficients are processed through a “bit transition” encoding engine that parses coefficient trees under a rate-distortion minimizing criterion. A CABAC entropy coding process is then applied on End Of Block markers, significance and sign maps;
- Rate-distortion optimization is performed across subbands for each residual signal in order to reach the target bit-rate;
- At any time of the decoding process, transform coefficients are optimally reconstructed to their expectation value.

Generation of transform coefficient subbands

A 4x4 transform is applied to each of the 3 color components of the differential between the original picture obtained after the MCTF of a spatial layer and their reconstructed base layer representation (or the subordinate enhancement layer representation). Here, no special attention is given to tuning the quantization step size. We make here the choice to arbitrarily set the quantization step size to its minimum authorized value following the idea that optimal encoding (in a rate-distortion sense) will be achieved during the subsequent steps. Note that it would be however always possible to adapt this quantization step size for instance on each component or per residual frame type.

Modeling of transform coefficient subbands

Each transform coefficient subband marginal distribution is modeled as a Generalized Gaussian Density (GGD) by adaptively varying two parameters α and β where α models the width of the Probability Density Function (PDF) peak (standard deviation), while β is inversely proportional to the decreasing rate of the peak. Sometimes, α is referred to as the scale parameter while is called the shape parameter. The GGD model contains the Gaussian and Laplacian PDFs as special cases, using β=1 and β=2, respectively.
In our implementation, this is realized through a gradient descent based algorithm using the Kullback-Leibler distance between the estimated GGD using (α,β) and the observed marginal distribution of transform coefficients. This way, every transform coefficient subband may be accurately represented using the couple (α,β). In the following preliminary results, α and β are both quantized on 5 bits for each of the 48 subbands (4x4 subbands for each component) of each differential, which obviously represents a significant overhead of about 480 bits (at the most) per residual image. More sophisticated quantization techniques may be investigated in order to reduce this (huge) information overhead. However, the distributions may not change a lot for a given type along the sequence, which would allow to have pre-defined parameters per frame type without requiring to model the distribution and to avoid transmitting the parameters.

Binarization of transform coefficients

Having estimated for each transform coefficient subband the parameters (α,β) of the modeled GGD, it is theoretically possible to design binarization codes that optimally, in a rate-distortion sense, map coefficient symbols into bin strings or codewords. To put it in a nutshell, such codes are built in such a way that every new transmitted bit gives the highest MSE decrease for a given rate constraint. The resulting codes have to be such that the symbol energy is mainly concentrated on the first bits of the symbol representation (i.e. on the first bit transitions of the corresponding codetree). A consequence of this is that the order according to which transform coefficient codewords are further parsed and processed is not necessarily the "classical" one, i.e. bit-plane by bit-plane for each coefficient at the same time, from the MSB to the LSB. Since to each bit transition one can associate an internal node of the codetree, this order is defined between internal nodes of the codetree. Bits are transmitted so that this order is verified as the bit clock increases. We will refer later to this by the term "bit transition" encoding. In our first experiments, this bit transition encoding strategy has been applied to FLC codewords, but it is expected that better results, both in terms of error-resilience and progressive decoding, may be obtained using special types of VLC codewords, especially over error-prone networks [3].

Bit transition encoding of transform coefficients

As previously mentioned, every transform coefficient codeword is progressively encoded by a bit transition encoding engine. Starting from the first bin of each codeword, the parsing process is done for every bit clock by bit transition levels taken into account an order (between internal nodes of the corresponding codetree) that maximizes the energy on the first transmitted bits. Then, for a particular subband, if the bit transition level of a particular codeword corresponds to the current bit clock order defined for the codetree of the subband it belongs to, then, significance and sign encoding primitives apply. This way, within the same subband, for a particular bit clock, some coefficients may be encoded with more bit transition levels than others.
Significance and sign bits are encoded by a CABAC coding engine. In our first experiments, intra-dependencies between bins of a particular codeword have presented good compression performances. However, since the first bins of every codeword present strong similarities, more sophisticated inter-dependencies might be exploited in order to design more efficient CABAC context models. In addition, a derivation of the End of Block (EOB) marker has been used so as to signal for a particular 4x4 block that there are no additional significant coefficients in the block.

Rate-distortion optimization across subbands

As previously seen, our method encodes every transform coefficient subband independently. For each residual frame a two-pass rate-distortion optimization between subbands is used to reach the target bit-rate defined as the percentage of the corresponding base layer frame bit-rate so that the sequence total target bit-rate is met. In our first experiments, only one NAL unit is generated. This only enhancement NAL unit includes all enhancement layers corresponding to a specific target bit-rate. Consequently, we did not incorporate for the time being the functionality of bit-stream truncation. Therefore, following results may be seen as direct encoding for each targeted bit-rate since here we do not take into account the sub-optimality that may occur when decoding at a lower bit-rate.

Expectation-based reconstruction of transform coefficients

Progressive decoding of the codewords is realized by considering the following expectation-based approach: codewords are decoded progressively by considering the bit generated by the bit transitions at a given level of the codetree as a bit-plane or a layer. The first bits received correspond to an internal node of the codetree from which a set of leaves and associated probabilities can be deduced. Then the optimal reconstruction value is given by the sum over deduced leaves of the values weighted by their corresponding probability.

Experimental results

We have tested our method using the SVM-3.0 encoder. Experiments are conducted using Scenario 2 test sequences only for the first spatial layer, at 15Hz i.e. QCIF15. The encoder configurations are the one defined for MPEG Palma Meeting CE1.
RD curves for SVM-3.0 encoder using embedded quantization step sizes and SVM-3.0 encoder with our improved FGS scheme are shown in the following plots (where the abbreviation “FGS” is used to denote the use of our method). Up to 0.5 dB performance improvement has been achieved (particularly at middle interval bit-rate) by use of our improved FGS scheme. Again, please note that the following results may be seen as direct encoding for each targeted bit-rate since here we do not take into account the sub-optimality that may occur when decoding at a lower bit-rate.

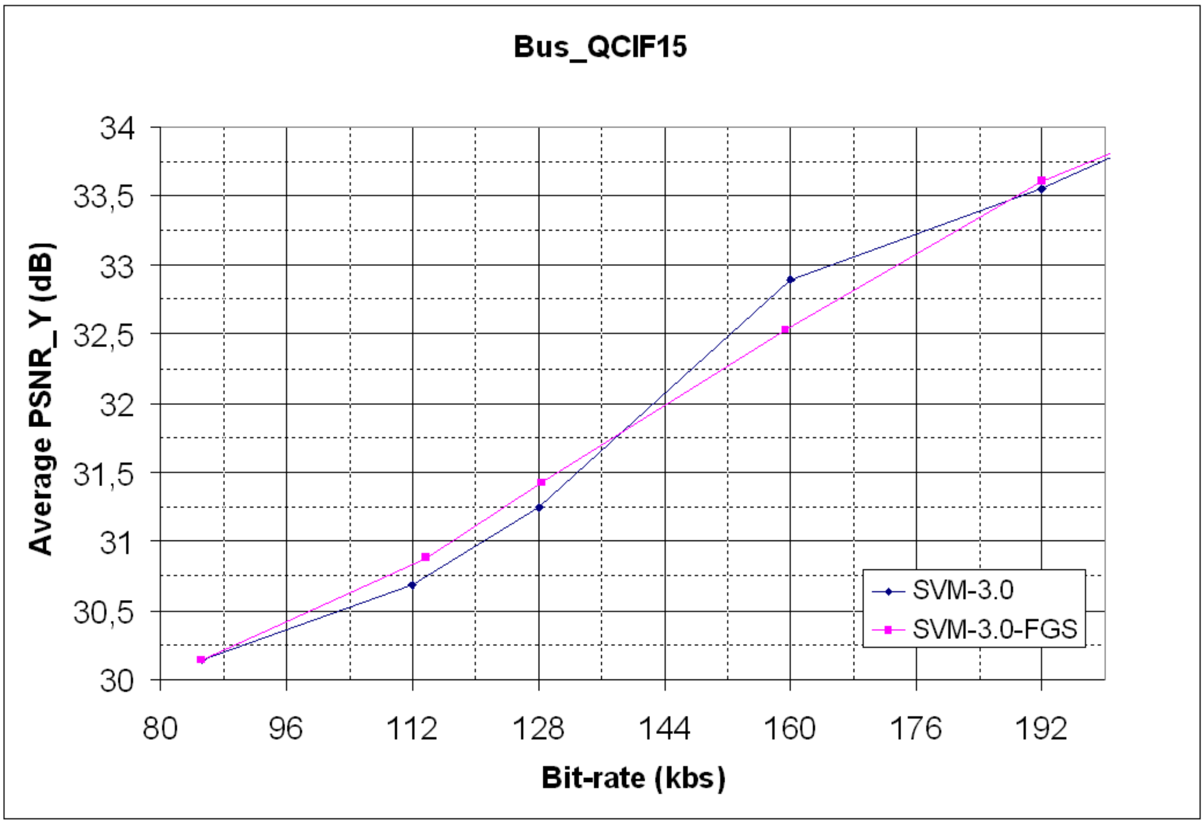

Figure 1: Performance comparison for Bus_QCIF15.

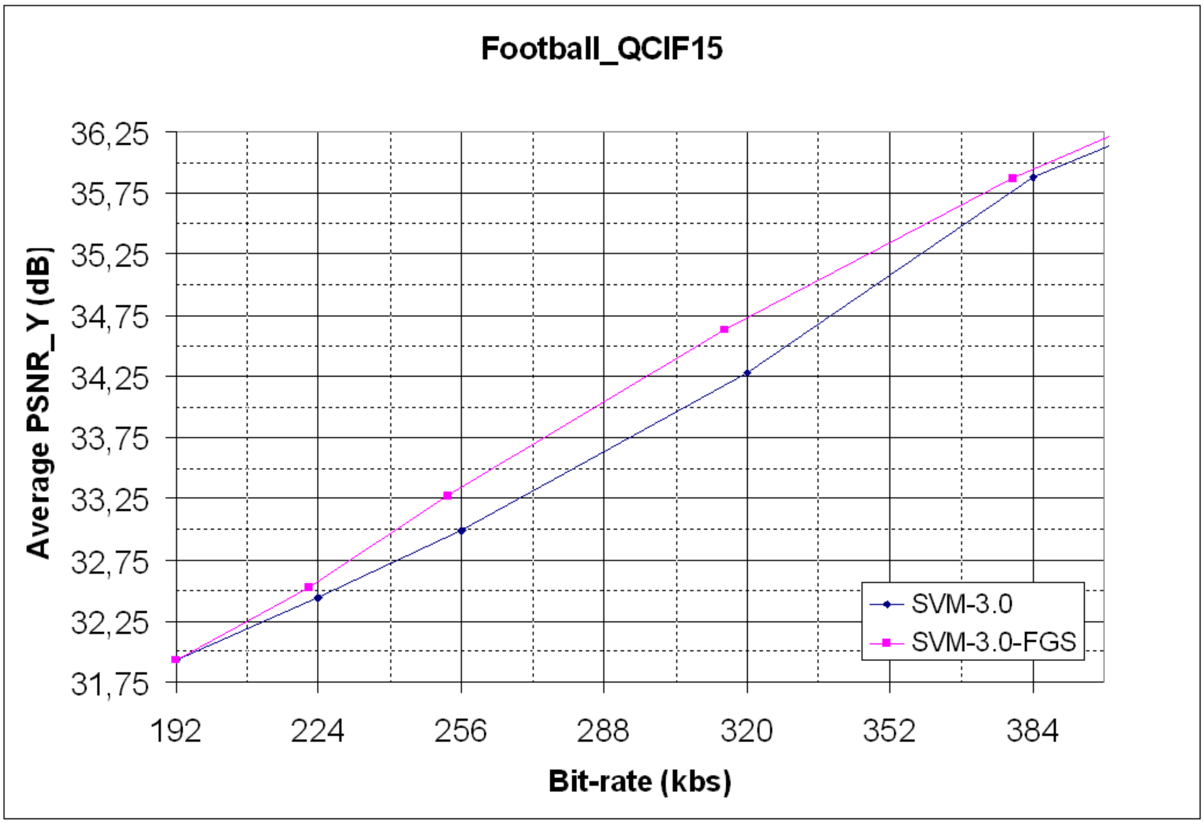

Figure 2: Performance comparison for Football_QCIF15.

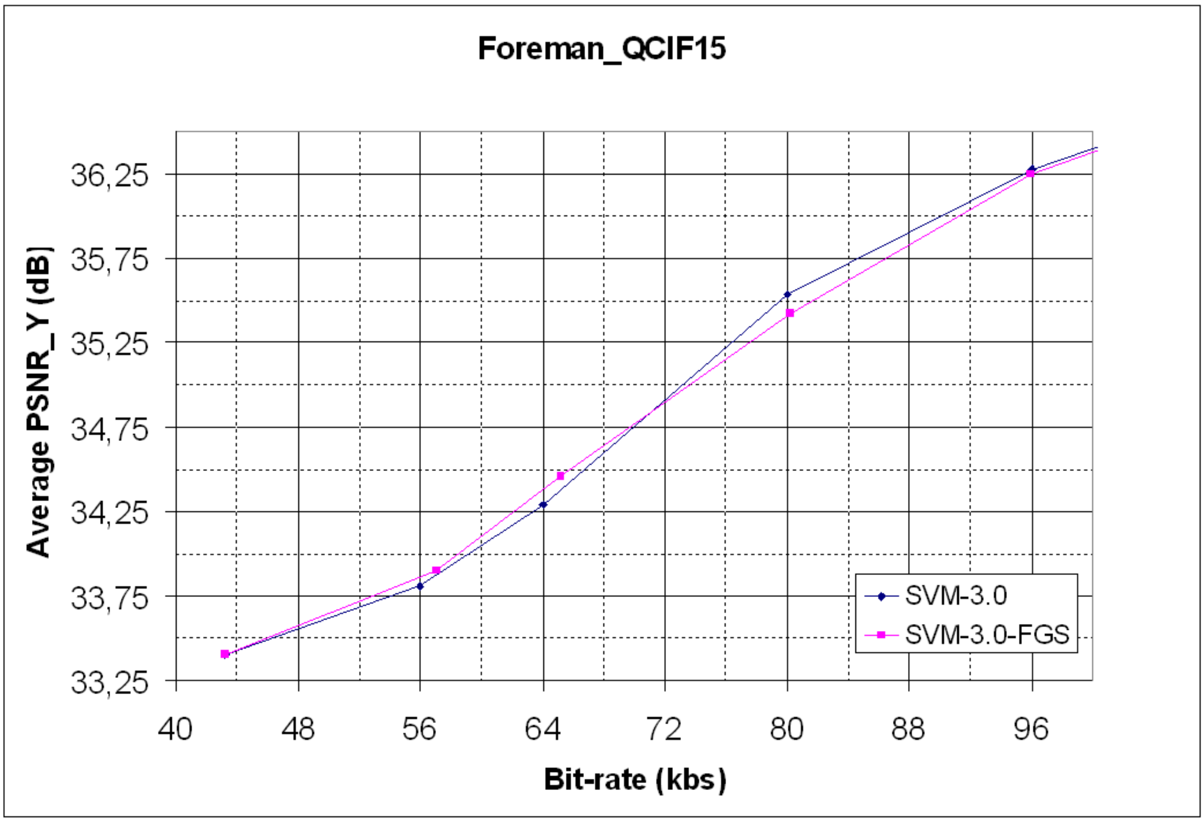

Figure 3: Performance comparison of Foreman_QCIF15.

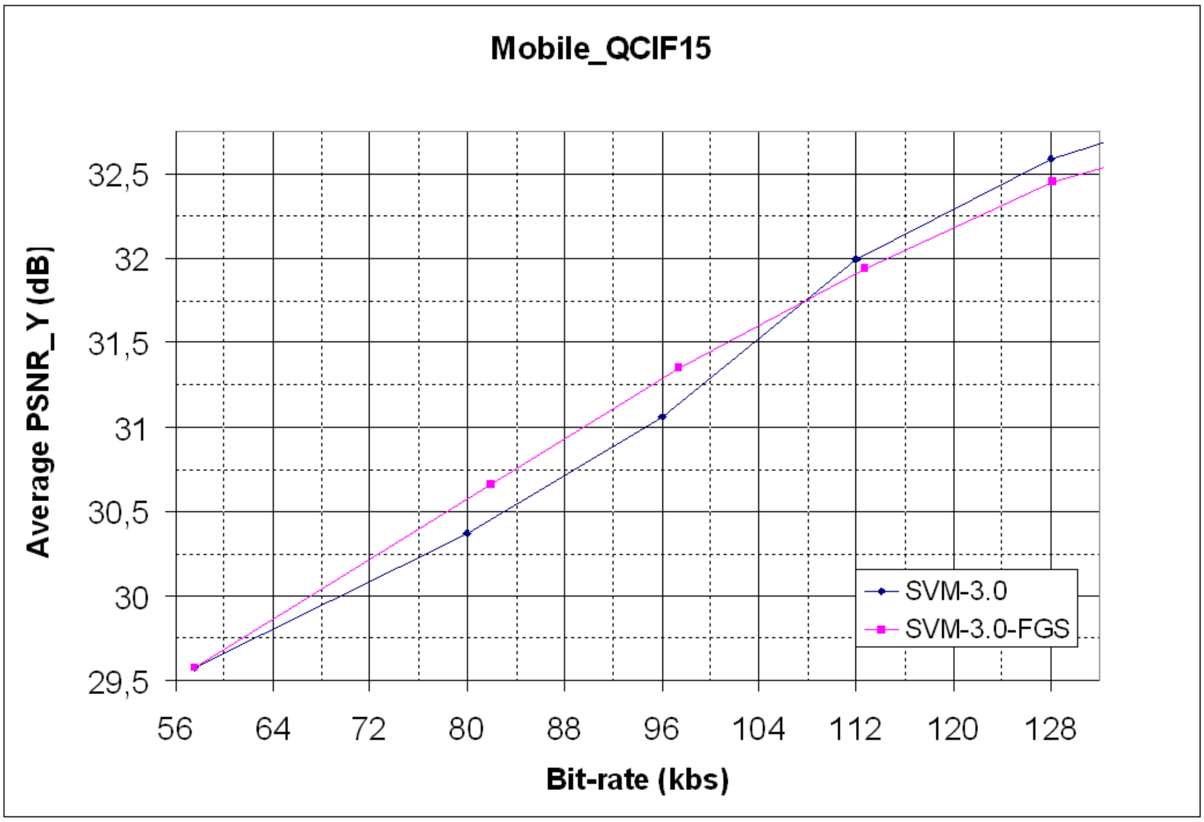

Figure 4: Performance comparison for Mobile_QCIF15.

Conclusion

An improved FGS scheme has been introduced, showing a significant improvement in coding performance for the targeted ranges of bit-rates. However, these preliminary results do not take into account the sub-optimality that may result from a bit-stream truncation at extraction. Future works would include, among other things, the integration of this method into JSVM 1, the encapsulation of enhancement layers in more NAL units, the support of bit-stream extraction so as to assess the performances of the scheme in a truly scalable mode, improvements of Gaussian Generalized Density parameters encoding, more sophisticated CABAC context modeling and introduction of VLC codewords. This latter point might be even more interesting in the context of error-prone channels in order to achieve good performances in terms of error-resilience and progressive decoding.

References

[1] DANAE website, http://danae.rd.francetelecom.com.
[2] “Joint Scalable Video Model JSVM 1”, Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG (ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6) Output Document JVT-N023, 14th Meeting: Hong Kong, CN, 17-21 January, 2005.
[3] Hervé Jégou and Christine Guillemot, “Progressive and Error-resilient transmission strategies for VLC encoded signals over noisy channels”, submitted to *Eurasip Journal of Applied Signal Processing (Eurasip JASP) special issue on video analysis and coding for robust transmission*.

(Append for Proposal Documents)

JVT Patent Disclosure Form

International Telecommunication Union
Telecommunication Standardization Sector

International Organization for Standardization

International Electrotechnical Commission

Joint Video Coding Experts Group - *Patent Disclosure Form*

(Typically one per contribution and one per Standard | Recommendation)

Please send to:

JVT Rapporteur Gary Sullivan, Microsoft Corp., One Microsoft Way, Bldg. 9, Redmond WA 98052-6399, USA
Email (preferred): Gary.Sullivan@itu.int Fax: +1 425 706 7329 (+1 425 70MSFAX)

This form provides the ITU-T | ISO/IEC Joint Video Coding Experts Group (JVT) with information about the patent status of techniques used in or proposed for incorporation in a Recommendation | Standard. JVT requires that all technical contributions be accompanied with this form. *Anyone* with knowledge of any patent affecting the use of JVT work, of their own or of any other entity ("third parties"), is strongly encouraged to submit this form as well.

This information will be maintained in a "living list" by JVT during the progress of their work, on a best effort basis. If a given technical proposal is not incorporated in a Recommendation | Standard, the relevant patent information will be removed from the "living list". The intent is that the JVT experts should know in advance of any patent issues with particular proposals or techniques, so that these may be addressed well before final approval.

This is not a binding legal document; it is provided to JVT for information only, on a best effort, good faith basis. Please submit corrected or updated forms if your knowledge or situation changes.

This form is *not* a substitute for the *ITU ISO IEC Patent Statement and Licensing Declaration*, which should be submitted by Patent Holders to the ITU TSB Director and ISO Secretary General before final approval.

<u>Submitting Organization or Person:</u>

Organization name ____________________

Mailing address ____________________

Country ____________________

Contact person ____________________

Telephone ____________________

Fax ____________________

Email ____________________

Place and date of submission ____________________

<u>Relevant Recommendation | Standard and, if applicable, Contribution:</u>

Name (ex: "JVT") ____________________

Title ____________________

Contribution number ____________________

(Form continues on next page)

Disclosure information – Submitting Organization/Person (choose one box)

☐ 2.0 The submitter is not aware of having any granted, pending, or planned patents associated with the technical content of the Recommendation | Standard or Contribution.

or,

The submitter (Patent Holder) has granted, pending, or planned patents associated with the technical content of the Recommendation | Standard or Contribution. In which case,

☐ 2.1 The Patent Holder is prepared to grant – on the basis of reciprocity for the above Recommendation | Standard – a free license to an unrestricted number of applicants on a worldwide, non-discriminatory basis to manufacture, use and/or sell implementations of the above Recommendation | Standard.

☐ 2.2 The Patent Holder is prepared to grant – on the basis of reciprocity for the above Recommendation | Standard – a license to an unrestricted number of applicants on a worldwide, non-discriminatory basis and on reasonable terms and conditions to manufacture, use and/ or sell implementations of the above Recommendation | Standard.

Such negotiations are left to the parties concerned and are performed outside the ITU | ISO/IEC.

☐ 2.2.1 The same as box 2.2 above, but in addition the Patent Holder is prepared to grant a “royalty-free” license to anyone on condition that all other patent holders do the same.

☐ 2.3 The Patent Holder is unwilling to grant licenses according to the provisions of either 2.1, 2.2, or 2.2.1 above. In this case, the following information must be provided as part of this declaration:

- patent registration/application number;
- an indication of which portions of the Recommendation | Standard are affected.
- a description of the patent claims covering the Recommendation | Standard;

In the case of any box ***other than 2.0*** *above, please provide the following:*

Patent number(s)/status ____________________

Inventor(s)/Assignee(s) ____________________

Relevance to JVT ____________________

Any other remarks: ____________________

(please provide attachments if more space is needed)

(form continues on next page)

Third party patent information – fill in based on your best knowledge of relevant patents granted, pending, or planned by other people or by organizations other than your own.

Disclosure information – Third Party Patents (choose one box)

☐ 3.1 The submitter is not aware of any granted, pending, or planned patents *held by third parties* associated with the technical content of the Recommendation | Standard or Contribution.

☐ 3.2 The submitter believes third parties may have granted, pending, or planned patents associated with the technical content of the Recommendation | Standard or Contribution.

For box 3.2, please provide as much information as is known (provide attachments if more space needed) - JVT will attempt to contact third parties to obtain more information:

3rd party name(s) ______________________

Mailing address ______________________

Country ______________________

Contact person ______________________

Telephone ______________________

Fax ______________________

Email ______________________

Patent number/status ______________________

Inventor/Assignee ______________________

Relevance to JVT ______________________

Any other comments or remarks: