Foreword

This body of work by Thomas Wiegand and Bernd Girod has already proved to have an exceptional degree of influence in the video technology community, and I have personally been in a position to proudly witness much of that influence.

I have been participating heavily in the video coding standardization community for some years – recently as the primary chairman ("rapporteur") of the video coding work in both of the major organizations in that area (the ITU-T VCEG and ISO/IEC MPEG organizations). The supporters of such efforts look for meritorious research ideas that can move smoothly from step to step in the process found there:

- generation of strong proposal descriptions,
- tests of effectiveness,
- adjustments for practicality and general flexibility, and
- precise description in a final approved design specification.

The ultimate hope in the standardization community is that the specifications written there and the other contributions developed there will prove to provide all the benefits of the best such efforts:

- enabling the growth of markets for products that work well together,
- maximizing the quality of these products in widespread use, and
- progressing the technical understanding of the general community.

The most well-known example of such a successful effort in the video coding community is the MPEG-2 video standard (formally identified as ITU-T Recommendation H.262 or as ISO/IEC International Standard 13818-2). MPEG-2 video is now used for DVD Video disks, direct-broadcast satellite services, terrestrial broadcast television for conventional and high-definition services, digital cable television, and more. The MPEG-2 story owes some of its success to lessons learned in earlier standardization efforts – including the first digital video coding standard known as ITU-T Recommendation H.120, the first truly practical success known as ITU-T Recommendation H.261 (a standard that enabled the growth of the new industry of videoconferencing), and the MPEG-1 video standard (formally ISO/IEC 11172-2, which enabled the storage of movies onto inexpensive compact disks). Each generation of technology has benefited from lessons learned in previous efforts.

The next generation of video coding standard after MPEG-2 is represented by ITU-T Recommendation H.263 (a standard primarily used today for videoconferencing, although showing strong potential for use in a variety of other applications), and it was the "H.263++" project for enhancing that standard that provided a key forum for Wiegand and Girod's work.

At the end of 1997, Thomas Wiegand, Xiaozheng Zhang, Bernd Girod, and Barry Andrews brought a fateful contribution (contribution Q15-C-11) to the Eibsee, Germany meeting of the ITU-T Video Coding Experts Group (VCEG). In it they proposed their design for using long-term memory motion-compensated prediction to improve the fidelity of compressed digital video. The use of long-term memory had already begun to appear in video coding with the recent adoption of the error/loss resilience feature known as reference picture selection or as "NEWPRED" (adopted into H.263 Annex N with final approval in January of 1998 and also adopted about two years later into the most recent ISO/IEC video standard, MPEG-4). But the demonstration of a way to use long-term memory as an effective means of improving coded video quality for reliable channels was clearly new and exciting.

Part of the analysis in that contribution was a discussion of the importance of using good rate-distortion optimization techniques in any video encoding process. The authors pointed out that the reference encoding method then in use by VCEG (called the group's *test model number 8*) could be significantly improved by incorporating better rate-distortion optimization. It was highly admirable that, in the interest of fairness, part of the proposal contribution was a description of a method to improve the quality of the reference *competition* against which their proposal would be evaluated. It was in this contribution that I first saw the simple equation

$$\lambda_{\mathrm{MOTION}} = \sqrt{\lambda_{\mathrm{MODE}}}\,.$$

A few months later (in VCEG contribution Q15-D-13), Wiegand and Andrews followed up with the extremely elegant simplification

$$\lambda_{\mathrm{MODE}} = 0.85 \cdot Q^2\,.$$

For years (starting with the publication of a paper by Yair Shoham and Allen Gersho in 1988), the principles of rate-distortion optimization had become an increasingly-familiar concept in the video compression community. Many members of the community (myself included, starting in 1991) had published work on the topic – work that was all governed by a frustrating little parameter known as λ. But figuring out what value to use for λ had long been a serious annoyance. It took keen insight and strong analysis to sort out the proper relationship between a good choice for λ and Q, the parameter governing the coarseness of the quantization. Wiegand, working under the tutelage of Girod and in collaboration with others at the University of Erlangen-Nuremberg and at 8x8, Inc. (now Netergy Networks), demonstrated that insight and analytical strength.

The ITU-T VCEG adopted the rate-distortion optimization method into its test model immediately (in April of 1998), and has used that method ever since. It is now preparing to adopt a description of it as an appendix to the H.263 standard to aid those interested in using the standard. I personally liked the technique so much that I persuaded Thomas Wiegand to co-author a paper with me for the November, 1998 issue of the IEEE Signal Processing Magazine and include a description of the method. And at the time of this writing, the ISO/IEC Moving Picture Experts Group (MPEG) is preparing to to conduct some tests against a reference level of quality produced by its recent MPEG-4 video standard (ISO/IEC International Standard 14496-2) – and it appears very likely that MPEG will also join the movement by choosing a reference that operates using that same rate-distortion optimization method.

But long-term memory motion compensation was the real subject of that 1997 contribution, while the rate-distortion optimization was only a side note. The main topic has fared even better than the aside. The initial reaction in the community was not one of unanimous enthusiasm – in fact some thought that the idea of increasing the memory and search requirements of video encoders and decoders was highly ill-advised. But diligence, strong demonstrations of results, and perhaps more iteration of Moore's Law soon persuaded the ITU-T VCEG to adopt the long-term memory feature as Annex U to Recommendation H.263. After good cross-verified *core experiment* results were shown in February of 1999, the proposal was adopted as draft Annex U. Additional good work described in this text in regard to fast search methods helped in convincing the skeptics of the practicality of using long-term memory. Ultimately, draft Annex U was adopted as a work item and evolved to preliminary approval in February of 2000 and then final approval in November of 2000.

A remarkable event took place in Osaka in May of 2000, when Michael Horowitz of Polycom, Inc. demonstrated an actual real-time implementation of Annex U in a prototype of a full videoconferencing product (VCEG contribution Q15-J-11). Real-time efficacy demonstrations of preliminary draft video coding standards has been an exceedingly rare thing in recent years. The obvious improvement in quality that was demonstrated by Horowitz's system was sufficient to squelch even the smallest grumblings of criticism over the relatively small cost increases for memory capacity and processing power.

In only three years, the long-term memory proposal that started as a new idea in a university research lab has moved all the way to an approved international standard and real market-ready products with obvious performance benefits. That is the sort of rapid success that research engineers and standards chairmen dream about at night.

Even newer ways of using long-term memory (such as some error resilience purposes also described in this work) have begun to appear and mature. Other concepts described in this work (such as affine multi-frame motion compensation) may one day also be seen as the initial forays into the designs for a new future.

As the community has grown to appreciate the long-term memory feature, it has become an embraced part of the conventional wisdom. When the ITU-T launched an initial design in August of 1999 for a next-generation "H.26L" video coding algorithm beyond the capabilities of today's standards, Wiegand's long-term memory idea was in it from the very beginning. The tide has turned. What once seemed like the strange and wasteful idea of requiring storage and searching of extra old pictures is becoming the accepted practice – indeed it is the previous practice of throwing away the old decoded pictures that has started to seem wasteful.

GARY J. SULLIVAN, PH.D.
Rapporteur of ITU-T VCEG (ITU-T Q.6/SG16 Video Coding Experts Group),
Rapporteur of ISO/IEC MPEG Video (ISO/IEC JTC1/SC29/WG11
Moving Picture Experts Group Video Subgroup),
Microsoft Corporation Software Design Engineer
May, 2001