

多視点映像符号化 H.264/AVC Annex H (MVC) の国際標準化

木全英明[†]

3次元映像は臨場感の高い映像表現として注目されている。3次元映像をコンテンツとして普及させるためには映像フォーマットの標準化が大変重要である。本稿では、ITU-TとISO/IECの国際標準H.264/AVCで、3次元映像の一つの表現である多視点映像を扱えるように拡張した標準MVC(Multiview Video Coding)の標準化の経緯とその背景について概説する。

International Standardization of H.264/AVC Annex H (MVC) for Multi-view Video Coding

Hideaki Kimata[†]

3D video is expected to be a representation to enhance reality of visual media. International standardization of video format is very important to popularize 3D video. In this paper, the prehistory and background of standardization of H.264/AVC Annex H (MVC) in ITU-T and ISO/IEC standardization bodies, which focus on compression of multi-view video, are shown.

1. はじめに

多視点映像とは同期した複数のカメラ映像のまとまりのことを指す。近年、多視点映像は映画などの映像製作の場でも利用されており、またライブ・スポーツイベントの撮影にも利用されている。この多視点映像を圧縮する符号化方式およびフォーマットとして、初の国際標準H.264/AVC Annex H (MVC)が2009年に完成した。この標準化はISO/IECの中のMPEGで日本の提案から始まった、3次元映像のフォーマットの標準化に関する一大プロジェクトの結果の一つといえる。このMVC(Multiview Video Coding)は近年汎用的に利用されているH.264/AVC標準に対して、臨場感の高い映像表現を扱うための拡張であり、特に立体映像向けにはBlu-ray 3D標準にも採用されるなどで利用が期待されている。本稿では、標準化当時の時代背景を振り返りつつ国際標準化にいたった経緯と、MVC標準の主な技術内容を概説する。なお実際に標準化された技術の詳細については参考文献1) 2) 3) 4) も参照されたい。

2. MVC標準化の経緯

2.1 3次元映像を扱う3DAVプロジェクト

2001年からMPEG標準化団体の中で3次元映像のフォーマットと圧縮方式の検討が開始された。ちょうどMPEGではテレビ用のMPEG-2やモバイル向けのMPEG-4 Visualの標準化がひと段落して、次は3次元映像という流れがあった。映像アプリケーションとしてもナビゲーションや監視用途で全方位カメラ映像を使ったシステムが開発され始めており、またEyeVisionシステムのように被写体の周囲をグルリと見回すような映像効果を得られるような放送コンテンツ製作にも3次元映像が利用されつつあった。このような状況を踏まえつつ、電子情報技術産業協会(JEITA)からMPEGへ3次元映像の標準化の提案が行われた。提案内容はJEITAの2000年度から始まった基準認証研究開発事業「XMLをベースとした立体映像コンテンツ規格の標準化」がベースであった[5]。この提案を受けて、MPEGにおいて3次元映像のアプリケーションと要求条件の研究がはじまった。この活動を3DAV(3D Audio Visual)プロジェクトと呼ぶ。このプロジェクトを通じて、全方位映像、自由視点映像、立体映像などの主要な3次元映像のアプリケーションの表現フォーマットが研究された。図1に3DAVで研究対象とした3次元映像データの標準的な処理フローと標準化検討ブロックを示す。また図2に3DAVで想定するカメラ配置の例を示す。3DAVで扱う入力データは、同

[†] 日本電信電話株式会社, NTTサイバースペース研究所
NTT Cyber Space Laboratories, NTT Corporation

期した複数の2次元映像（多視点映像）で表現されることと、シーンの撮影時のカメラパラメータが得られていて実空間との対応ができることを共通的特徴とした。研究の結果、全方位映像については高解像度映像において部分的な領域を復号できる機能が必要であること、自由視点映像については光線を表現する機能ならびにユーザ視聴領域を合成して表示できる機能が必要であること、立体映像については奥行き情報を使うフォーマットが重要であることがわかった。そして3DAVプロジェクトの最終段階では、産業界から自由視点映像に注目が集められ標準化を進めることになった。自由視点映像のフォーマットには光線空間やCGベースなど大変多くの種類があるが、多視点映像というカメラ入力側で整理して、第一段階として多視点映像のフォーマットと圧縮符号化方式を検討することとなった。このような状況を踏まえてMVCの標準化が始まった。

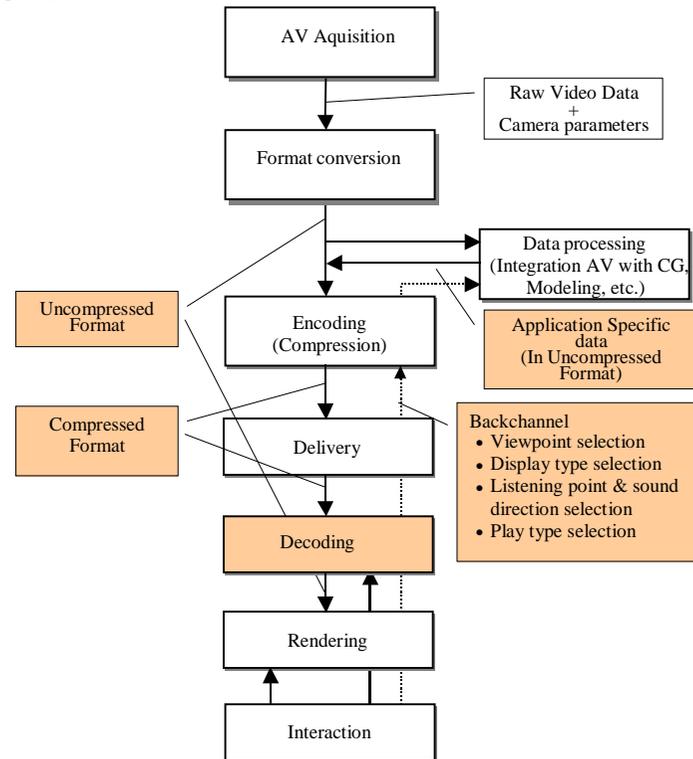


図1 3DAVで検討した3次元映像データ処理フロー

この自由視点映像のフォーマットの整理は非常に大切な分岐点であった。それまでは表示側すなわち仮想カメラ映像のレンダリングの処理を基準に整理を試みたため、3DAVの開始当初では共通化が非常に難しかった。これについて入力側で整理して解決できたことが大変有意義であった。またMPEGなどの標準化団体において、3DAVのような研究活動が行われたのは大変稀であるが、3次元映像産業には大変貴重であった。通常の2次元映像の標準化ではカメラやディスプレイ市場でデファクトになっている表現フォーマットがあるため、フォーマットについて特に大きな議論をすることはなく標準化が進められる。しかしながら3次元映像では、撮影と表示の方式が大変多岐にわたるため共通的な表現フォーマットの策定は難しく、これが3次元映像が現状広く使われていない要因の一つだと考えられた。このためMPEGという比較的開かれたな場で、フォーマットの策定へ向けて技術者が切磋琢磨して議論を交わしつつ研究し、それなりの結論を出せたことは大変有意義であった。

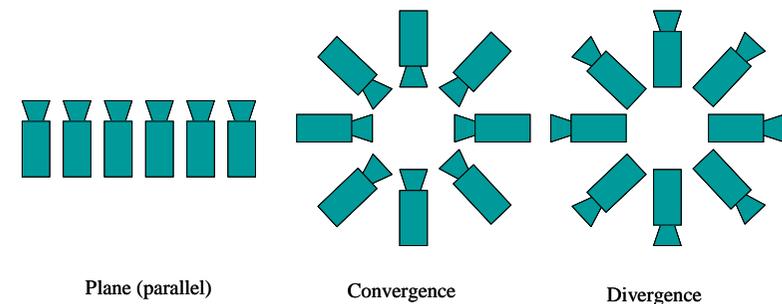


図2 3DAVで検討したカメラ配置の例

2.2 MVC標準化の流れ

3DAVプロジェクトと時を同じくして2001年頃よりH.264/AVC (MPEG-4 AVC)の標準化がMPEGとITU-Tの共同で開始された。技術審議は両標準化団体の共同チームJVT (Joint Video Team)で行われた。すなわち、ちょうどJVTでH.264/AVCの標準化が進められている間で、MPEGの3DAVプロジェクトでは3次元映像の研究が進められたことになる。そしてJVTで産業界にとって大変重要な規格H.264/AVC High Profileの標準化が完了した後で、MPEGではMVC標準化の提案募集が行われた。このHigh ProfileはBlu-rayやIPTVなど近年の主要な映像アプリケーションで共通して利用されている大変重要な規格である。このような状況もあり、MVC標準化の提案募集に対

しては、各社から H.264/AVC High Profile を拡張した方式が提案された。標準化では標準のもとになる評価モデルを策定・改訂しながら進められるが、この評価モデルには High Profile をベースにしたものが選定された。また H.264/AVC を拡張した形で標準化をすることが合意され、技術審議は JVT で行うこととなった。以上のようなイベントに始まり、MVC 標準化は JVT の中で 2009 年まで行われた。図 3 に 3DAV と JVT でのイベントを整理する。

MVC 標準化では多視点映像の圧縮効率向上のための技術検討の他、自由視点映像などの想定アプリケーションに必要な機能を実現するための要求条件を議論し、それをもとにしたデータフォーマット (High level syntax) を審議した。MVC 標準の具体的な規格については、まず最も汎用的に複数の視点を扱うための Multiview High Profile が策定されたのち、ステレオ立体映像用の特化した形で Stereo High Profile が策定された。

なお MPEG では MVC 標準化の後半から、MVC 標準化と平行して、レンダリング側の処理も考慮して奥行き情報も含むフォーマットの標準化 3DV/FTV の検討が進められている。3DV/FTV の標準化は 2010 年中に開始が検討されている。

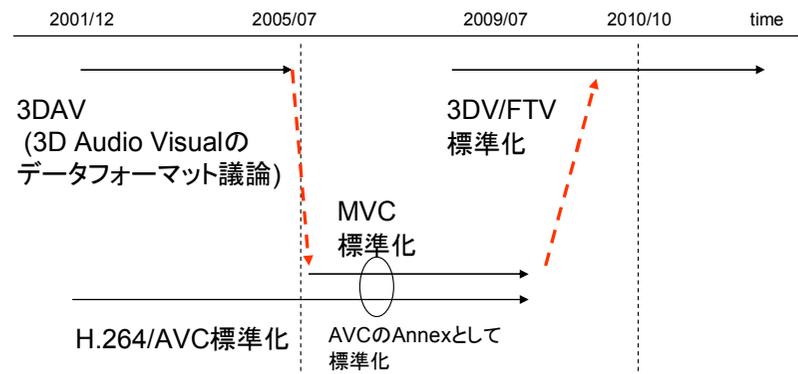


図3 3DAV から 3DV/FTV への流れ

3. MVC 標準の技術審議

3.1 圧縮効率向上のための技術

MVC 標準で審議された技術について紹介する。多視点映像は複数のカメラ映像を扱うため、カメラ台数分だけ情報量が増加する。そのため高い圧縮効率の実現が大変重要である。一方で、多視点映像は 2 次元映像の集合であるため、実装上では 2 次元映像の方式に近いことが望ましい。バランスよく両者を満たすため、圧縮効率向上のための技術としては、MVC 標準では元となる High Profile に対して視差補償を実現するための仕組みのみが採用された。標準化開始当初は、画像合成をして予測符号化する技術や、カメラ間の色の違いを補償する技術や、フォーカスの違いなどをフィルタリングする技術などが提案されていた、これらは 2 次元映像の実装からの改造が大きいとして採用には至らなかった。なお、画像合成をして予測符号化する技術については、大幅な圧縮効率向上が見込まれるため、MVC 標準化の中ではなく、3DV/FTV 標準化の中で引き続き検討が進められている。

3.2 複数のカメラ映像を扱うためのデータフォーマット

多視点映像はフレーム単位で複数のカメラ映像の同期がとれている映像である。ユーザ視聴側で多視点映像から仮想カメラ映像を合成するためには、同じ時刻の画像が遅延なく扱えることが重要である。MVC 標準のデコーダは同じ時刻の複数の画像を同じタイミングで復号するために、view という概念を導入し、view を単位に復号処理を繰り返し実行するマルチループ復号処理を採用した。このマルチループ復号処理の採用により、従来の動き補償を拡張する形で視差補償を実現する。図 4 に view の復号処理の流れを示す。1 つのデコーダで複数の view を復号するため、1 つの view に含まれるマクロブロックの最大数は view の数だけ減る。逆にいえば view に含まれるマクロブロック数を調整することで、1 つのデコーダで処理すべき総マクロブロック数を調整することができる。なお view はカメラで撮影される映像に限らない。例えば大きな映像を分割した領域を view として設定しても良い。

3.3 部分復号のためのデータフォーマット

自由視点映像や全方位映像では、ユーザが視聴したい位置と方向の仮想カメラ映像が合成できれば良いため、多視点映像の圧縮データから全てのカメラ映像を復号する必要はない。そこでストリーム中のいくつかの view のみを復号する部分復号を実現する機能が採用されている。

復号する view の組み合わせを予め定義しておき、この組み合わせを切り替えることで、所望の (複数の) view のみを 1 つのデコーダで復号できるようにする。また部分復号が基本的な機能として採用されているため、逆に実際のデコーダでストリーム中の全ての view を復号する必要はない。

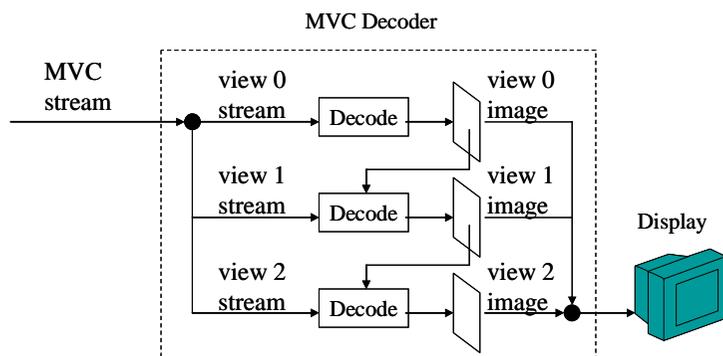


図4 MVCにおけるview単位の復号処理のフローの例

4. MVC 周辺技術の動向

符号化部分の標準化の他に、MVC のストリームをファイル化するファイルフォーマット標準化や MPEG-2 TS で伝送するためのシステム標準化が行われた。このようなシステム周りの標準化も行われたこともあり、今後は実サービスでの利用が進められると期待される。立体映像アプリケーションについては、立体映画用の Blu-ray 3D の他、携帯電話向けの開発例も報告がされている。また全方位映像（パノラマ映像）や自由視点映像向けのプレイヤーの開発例も報告されている状況である[6][7]。

5. おわりに

2001 年の 3DAV から始まった 3 次元映像に関する標準化から 2009 年の MVC 標準化完成までの経緯を概説した。標準化の進め方として、アプリケーションと要求条件を十分に研究することから始めたことにより、標準化が長期にわたり続いているデメリットもあるが、課題が整理された形でタイムリーに標準化を進められる大きなメリットを生むことができた。ITU-T の H.26L プロジェクトが H.264 として大きく花開いたように、今後も、標準化の場においても少し長いスパンで研究をするような土壌を持ち続けられる方がよさそうである。

参考文献

- 1) 改訂三版 H.264/AVC 教科書「H.264/AVC における MVC(多視点映像符号化)規格」, インプレス R&D, Jan., (2009).
- 2) 木全英明: 3 次元映像符号化の国際標準化 MPEG 3DAV の動向, 情報処理学会 AVM 研究会, 情報研報, 2005-AVM-48, (2005).
- 3) 木全英明: MPEG 3DAV 国際標準化の動向, 映像情報メディア学会誌 vol.60, no.2, pp.143-149, Feb., (2006).
- 4) 木全英明: 多視点映像符号化 MVC の国際標準化動向, 映像情報メディア学会誌 vol.61, no.4, pp.426-430, Apr., (2007).
- 5) XML をベースとした立体映像コンテンツ規格の標準化, 基準認証研究開発事業成果報告書, 電子情報技術産業協会 (2004).
- 6) Hideaki Kimata, Shinya Shimizu, Yutaka Kunita, Megumi Isogai, and Yoshimitsu Ohtani: Panorama video coding for user-driven interactive video application, IEEE International Symposium on Consumer Electronics (ISCE) 2009, (2009).
- 7) Shinya Shimizu, Hideaki Kimata, Yoshimitsu Ohtani: REAL-TIME FREE-VIEWPOINT VIEWER FROM MULTIVIEW VIDEO PLUS DEPTH REPRESENTATION CODED BY H.264/AVC MVC EXTENSION, IEEE 3DTV-CONFERENCE 2009, (2009).