

3次元映像符号化の国際標準化 MPEG 3DAV の動向

木全 英明[†]

国際標準化団体 MPEG では自然画像をベースとした3次元映像について標準化の検討を進めている。3DAV とは3次元映像と3次元音声のアプリケーションや要求条件ならびに技術を研究し、標準化へ向けた検討を進める活動である。3DAV は2001年12月から MPEG で開始され、数々の検討結果を残してきた。本稿では、3DAV で検討されてきたアプリケーションを示し、3DAV で進めてきた検討内容を史実に沿って述べる。特に時間をかけてきた全方位映像と自由視点映像について実験内容を含めて技術動向を示す。また3DAV では、現在では自由視点テレビを最終ゴールとした多視点映像へ焦点を絞って標準化へ向けた検討を進めており、この多視点符号化について技術動向を紹介する。

Movement on MPEG 3DAV toward International Standardization of 3D Video

Hideaki Kimata[†]

MPEG has promoted 3D Video coding technologies that base on natural video as the 3DAV activity, in which applications and requirements on 3D Video and 3D Audio have been studied. MPEG initiated the 3DAV activity in December 2001, and it has demonstrated several evaluation results. I show applications that 3DAV will provide for us, and address what the 3DAV activity has studied along the historical events. Especially, I address the movement on omni-directional video and free-viewpoint video for which the 3DAV activity has spent many works, and I address MVC (Multi-view video coding), which is the current main task of 3DAV to start standardization, while FTV (free-viewpoint television) is set as the final goal.

1. MPEG 3DAV とは

マルチメディアの国際標準化団体 MPEG では3次元映像音声技術の国際標準化を見据えた3DAV 活動を行っている。3DAV は MPEG 会合中には Requirements グループや Video グループの中の小グループで活動を行っており、MPEG 会合間には AHG を開催するなど、会期を問わず活発に活動が続けられている。3DAV は 3D Audio Video の略であり、3D Audio と 3D Video 単体または組み合わせを指す。元々3D Video の

標準化提案が行われたことから開始された活動であり、現時点まで様々な活動を行い、MPEG 各パートで具体的な標準化活動を開始し、活動結果を報告としてまとめてきた。その活動の過程と結果は第2章以降で詳しく述べることで、本章では3DAV が目指しているサービスやアプリケーションを紹介する。

3D Video という言葉について、現在までに各人が自分の定義を持っていることもあり、映像の特徴を反映して明確に定義するのは難しい。私が3DAV の活動を通して感じたところでは、3D Video とはカメラの Geometry 情報を利用する映像表現を指すと考えれば良いようである。カメラの Geometry 情報とは、通常のピンホーム

[†] NTT アドバンステクノロジー株式会社
NTT Advanced Technology Corporation
MPEG 3DAV AHG co-chair

カメラモデルであれば SMPTE 315M [1]で定義される情報であり，3DAV 活動の中ではカメラパラメータと称している．例えばカメラの位置情報や焦点距離情報である．このカメラパラメータは，カメラ内部情報(intrinsic parameter)と外部情報(extrinsic parameter)に分類できる．私達の住んでいる世界は一般的に空間3次元と時間1次元の4次元空間と言われる．カメラは3次元空間の一部分を，カメラ方向に投影した情報を画像情報として記録する装置と考えられる．カメラパラメータが分かっていると，3次元空間のどの部分を投影したものなのかが分かる．3D Video は，この3次元空間のどの部分を撮影したのかという，実世界のスケール情報を利用した映像表現とも言える．

3D Video では実世界のスケール情報を利用するため，想定されるアプリケーションはカメラ配置に依存して異なる．3DAV では複数のカメラからの画像情報を組み合わせて利用することを想定している．想定しているカメラ配置は例えば図1に示したものである[2]．図1(a)はカメラがある直線方向あるいは平面に並んだ配置である．図1(b)はある点を中心に内向きにカメラを並べた配置であり，図1(c)はある点を中心に外向きにカメラを並べた配置である．図1(a)と1(b)を組み合わせると，カメラをアーク状に配置することも想定される．このように，これら3パターンの組み合わせで各種カメラ配置を表現できるため，3DAV 活動では，この3つに焦点を当てて議論を進めてきた．

図1(a)と図1(b)のカメラ配置で実現されるアプリケーションには，例えばステレオ映像や自由視点テレビジョンが挙げられる．ステレオ映像は2カメラをアーク状に配置したものである．ステレオ映像は近年まで立体映像として扱われてきた映像表現である．映像を立体的に見せるところに特徴があり，映像情報自体は従来ある映像と同じものである．カメラパラメータは立体視に利用される．

自由視点テレビでは視聴者が見る視点を自由に選択できる映像表現であり，映像情報自体も従来にない全く新しいフォーマットを必要とする映像表現である．カメラパラメータは，この新しいフォーマットの定義において利用される．図1(a)を前提にしたカメラ配置では，従来のテレビジョンと同じ映像内容を対象としているが，従来のテレビジョンとは異なり視点位置

を変えることが出来る．図1(b)を前提にしたカメラ配置では，中心にある物体の外周を囲む映像内容となる．やはり視点位置を自由に変わることが出来る．アプリケーションとしては，テレビジョンの発展形にとどまらず，空間ナビゲーションや商品カタログなどのパーソナルユースでの利用や，有形/無形文化財のアーカイブが考えられる．例えば，インターネットで3D CGを使って表現されている商品イメージを，自由視点映像で表現することも可能である．実際に撮影して得られた映像情報を使うため，3D CGよりリアルな印象を視聴者に与えることが出来る．ステレオ映像は自由視点映像で2視点に限定したものに想定されるため，自由視点映像を立体視することも可能である．

図1(c)のカメラ配置で実現されるアプリケーションには全方位映像が挙げられる．全方位映像は視点位置を固定して複数の視線方向で撮影した映像内容となるため，自由視点映像の一例として捉えることも可能である．このため，全方位映像では，視点位置を視聴者が選択できない制限の元での，自由視点映像と同じアプリケーションを実現できる．例えば，空間ナビゲーションが挙げられる．視聴者は視線方向を自由に選択しつつ，ナビゲーション映像を見ることが出来る．近年では，カメラ光学系の進歩により，図1(c)のように複数のカメラを配置するのではなく，一つのカメラで全周囲の映像を撮影することも可能である(図2)[3]．

3D Audio は現在 DVD で提供されている5.1チャンネルのように複数チャンネル音声を拡張したものであり，視聴者の実位置または仮想位置に

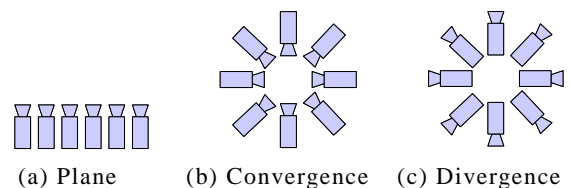


Fig.1. Examples of camera arrangements

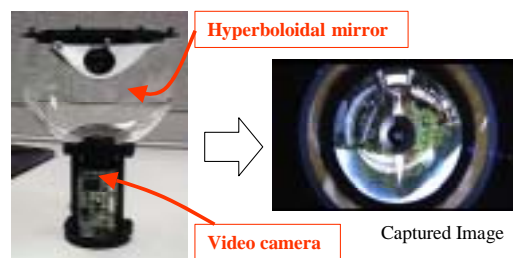


Fig.2. HyperOmni Vision camera

従って音声を変更するアプリケーションを想定している。

2. 3DAV の歴史と活動

元々 3D Video の標準化を MPEG へ提案したことから 3DAV 活動が始まった。2001 年 12 月に開催された MPEG 第 58 回 Pattya 会合で、(社) 電子情報技術産業協会 (JEITA) から提案された寄書 [4] を元に、3D Video の明確化と既存の MPEG 標準との差分を研究するグループ (AHG on 3D Video coding in MPEG) が Requirements グループの下に設立された (N4524)。JEITA では、当時、特定プロジェクト「XML をベースとした立体映像コンテンツ規格の標準化」(京都大学の松山教授が委員長) が進められており、その活動の一環として MPEG への 3D Video 標準化の提案活動が行われた。JEITA として参加された山下氏 (当時、株式会社アスキーに在籍) が、AHG の初代議長を務められた。この AHG の中で、3D Video を実現するスキームや、想定アプリケーションを実現するための要求条件の検討が進められた。2002 年 5 月の第 60 回 Fairfax 会合で 3D Audio も検討項目に加わり、グループ名が 3DAV になった (N4795)。この会合で Smolic 氏 (HHI に在籍) も AHG の co-chair になった。2002 年 12 月の第 63 回 淡路会合で山下氏に代わり私が co-chair になり (当時、NTT サイバースペース研究所に在籍) (N5341)、現在までこの体制が続いている。また、淡路会合では名古屋大学の谷本教授 (JEITA の特定プロジェクトの副委員長) のアレンジにより、3DAV セミナーが同時開催された [5]。3D Video 活動が開始されてから現在に至るまでの流れを図 3 に示す。

要求条件などの検討結果は Application and Requirements 文書としてまとめられており、2001 年 12 月の第 58 回 Pattaya 会合で N4559 が、MPEG からの最初の出力文書として作成された。以後、同文書は会合の度にアップデートされている。現時点での最新は N5877 [2] である。特に 2002 年 3 月の第 59 回 Jeju 会合の時点で、3DAV で検討するシステムのブロック図の概要が決定され、標準化すべき処理の明確化が行われた [6]。図 4 に 3DAV のブロック図を示す。従来の 2 次元映像を情報とするテレビジョンとは異なり、デコーダの出力となる "Uncompressed Format" も標準化の対象となっている。3 次元映像の信

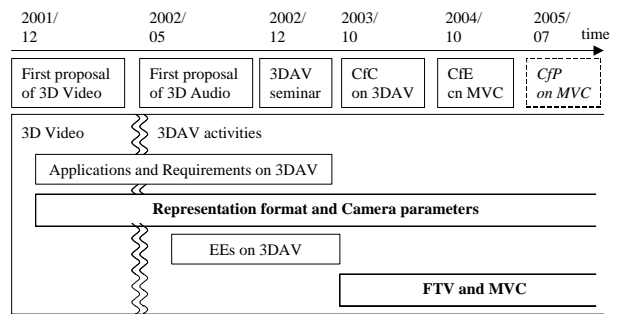


Fig.3. Chronological table on 3DAV

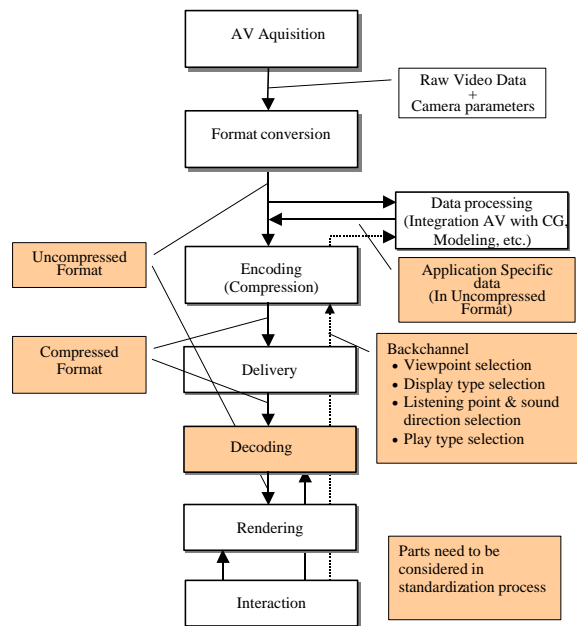


Fig.4. Block diagram of 3DAV system

号フォーマット (representation format) として現在世の中に標準的なものが無いためであり、このフォーマット自体から規格化する必要がある。この representation format は現在でも 3DAV では議論の中心にあり、想定したアプリケーションに対応するフォーマットは、未だ決定されていない。また、3DAV の特徴としてインタラクティブ性を備える映像メディアであることが挙げられる。representation format に依存してインタラクティブ性の内容は異なる。

具体的な技術的進展については、2002 年 7 月の第 61 回 Klagenfurt 会合で、実際に MPEG の既存の標準との差分を明確にするための実験 EE (Exploration Experiments) が設定された [7]。以後、2003 年 7 月の第 65 回 Trondheim 会合の EE 最終報告書 [8] を作成するまで EE は継続された。EE は合計 4 つ設定された。

[EE1]: 全方位映像

[EE2]: 自由視点テレビ/自由視点映像

[EE3]: ステレオ映像の符号化

[EE4]: 3D TV 用 depth の符号化

これらのうち特に EE1 と EE2 が、最終報告書をまとめるまでメインに議論をされた。JEITA の特定プロジェクトは EE1 と EE2 に大きく貢献した。EE1 と EE2 の詳細については第 3 章と第 4 章にまとめる。EE3 と EE4 については早々に MPEG-4 として採用されている技術で実現可能であることが示され活動を終了した。3D Audio については、MPEG-4 BIFS で規格化されている Advanced Audio BIFS ノードを利用することで実現できることが確認された。

最終報告書を作成した後、2003 年 10 月の第 66 回 Brisbane 会合で CfC (Call for Comment) (N6051) を発行して、今後の 3DAV の活動内容について産業界からのコメントを募集した。EE2 で検討をした FTV(自由視点テレビ)に関心があるとのコメントが多数寄せられた[9]。そこで自由視点映像を最終ゴールと定め、まずは Video グループで審議しやすい多視点符号化(MVC)について、現在の MPEG 標準と比較をすることに焦点を絞った。2004 年 10 月の第 70 回 Palma 会合で、MVC について CfE (Call for Evidence) (N6720) が発行され、現在の MPEG 標準よりも優位性を示す実験結果を募集した。2005 年 1 月の第 71 回香港会合で優位性を示す十分な実験結果が示された[10]。MVC の標準化を進める次のステップとして CfP (Call for Proposal) の発行を予定しており、現在準備を進めている[11]。

他方、3DAV の元々の特徴であるカメラパラメータの規定については、未だ十分な検討がなされていない。SMPTE 315M をベースにした MPEG-7 descriptor の提案が 3DAV 以前にも行われたが[12]、MPEG-7 には採用されていない。

3. 全方位映像

全方位映像は視点方向のみ選択できる映像表現なので、ピンホールカメラモデルを想定すると、カメラ位置を中心とした球面に画像を投影することで表現することが出来る(図 5)。このため、全方位映像は、3 次元空間上の球面を定義する記述と、その球面上にマッピングする画像の組み合わせで表現できる。球面の定義は VRML や MPEG-4 BIFS で記述できる。従って既存の MPEG の技術で表現可能である。球面上の

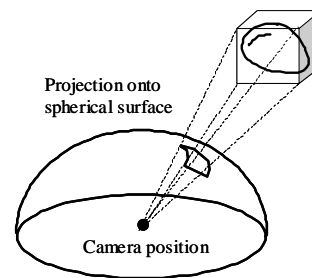


Fig.5. Geometry for omni-directional image

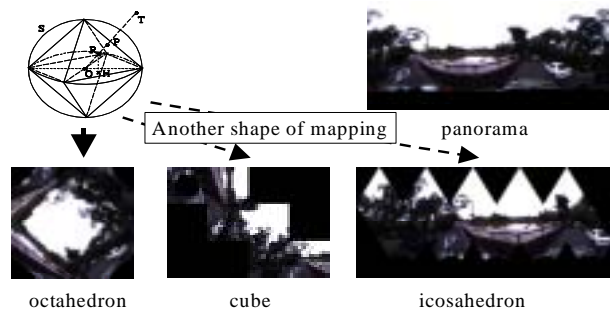


Fig.6. Patterns of geometrical mapping

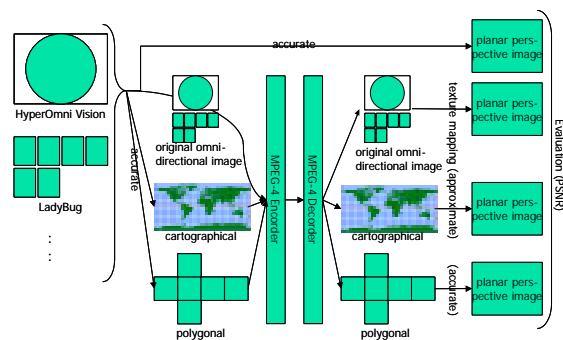


Fig.7. Evaluation procedure for omni-directional video

画像という点では、球面上の座標系を導入した画像符号化を検討する選択肢もあったが[13]、議論の焦点は既存の MPEG の技術で実現可能かどうかであったため検討から外れている。

既存の MPEG の技術で実現可能であることが明確になった後も、EE1 は球面への画像のマッピング方法を中心に議論が進められた([14]など)。符号化効率がマッピング方法に依存することを踏まえ、何か技術的な課題があるかどうかを見極める目的もあった。EE1 で検討したマッピング方法を図 6 に示す。符号化歪の無い透視画像に対する符号化歪のある透視画像の劣化度合いを比較することでマッピング方法の有効性を評価した(図 7)。評価の結果、20 面体マッピング法が最も劣化度合いが低いことが示された。しかしながら、球面を表現するポリゴン数が増

えると処理時間が大きくなる問題があり，結論としては，最も簡素な 8 面体でも十分実用的な符号化効率を実現できると考えられる．

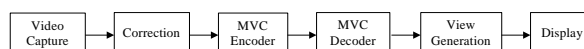


Fig.8. Basic components of FTV system

4. 自由視点テレビ/自由視点映像

自由視点映像は視点位置と視線方向を視聴者が選択できる映像表現であり，3DAV で検討されてきた映像メディアの中で最もインタラクティブ性の高いものである．自由視点映像について representation format の規定を含めて EE2 で議論を行った．EE2 では representation format の異なる複数の提案がなされている．提案されたものは，Ray-Space (光線空間)[15]をベースとした FTV (Free-viewpoint TV)と，モデルをベースとしたモデルベース FVV (Free-viewpoint video)に大別される．FTV と FVV は互いにアプローチの異なる手法として，別々に検討が進められた．FTV については，Ray-Space の構成方法から自由視点映像の生成までにおいて必要な技術の提案があり[16]，2次元のテレビジョンの自然な拡張であることが提案され，2次元のテレビジョンとの違いを中心に議論が進められた．特に自由視点映像を生成する view interpolation(視点補間)技術については，既存の MPEG には無く，新規に標準化を進めるべきとの提案が多くなされた．第 2 章でも示したように FTV は，CfC で強い賛同を受け，その後，MVC(多視点符号化)の CfE を発行するにいたる．自由視点映像の撮影から生成までの流れは図 8 のような構成が検討されており，自由視点テレビ映像デコーダと MVC との関係は，図 9 のようなブロック図が想定されている[17]．

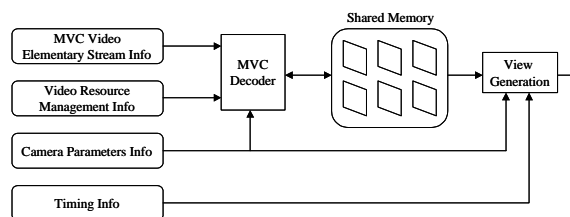


Fig.9. Architecture of an FTV decoder based on MVC

モデルベースの自由視点映像では，カメラで撮影した映像をモデルのテクスチャとすることに特徴がある．モデルベースにはモデルの作り方の違いにより複数の手法が提案されてきたが，テクスチャを低遅延でアップデートする機能が重要であることが認識された．このための手法を中心に議論が進められた．この手法は MPEG SNHC グループで規格作成が行われている AFX (Animation Framework eXtension)に採用された[18]．この採用によりモデルベースの自由視点映像は MPEG の技術で実現できることとなった．3DAV では形状の変化するモデルの表現と符号化効率改善提案があった([19]など)．MPEG の技術で実現できることが明確にはなったが，

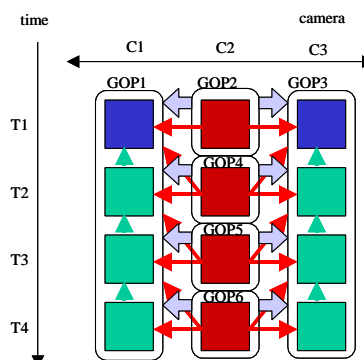


Fig.10. A MVC method for low delay random access

現在 SNHC グループでモデル符号化についての CE (Core Experiments)が行われている[20]．

5. 多視点符号化

FTV(自由視点テレビ)を最終ゴールとして，多視点映像符号化の標準化が検討されている．今年の 7 月に CfP (Call for Proposal)を発行する計画である．本章では CfE 以前または CfE の際に提案された技術を中心に紹介する．CfE では anchor に対する圧縮改善を主張する必要性があったため，anchor の方式に改良を加えた手法が主に提案された．Anchor は各カメラ映像を独立して MPEG-4 AVC で符号化する手法である[21]．

多視点符号化は MPEG-2 Multiview Profile でも実現できる．この手法では MPEG-2 が基本になっているため，利用できる参照画像には制限があり，また 1 つの視点映像をベースにして，階層的に他の視点映像を符号化する手法であるため，視点映像の低遅延のランダムアクセスは考慮されていない．3DAV で検討する MVC では，最終ゴールが自由視点映像であることを主眼にしているため，視点映像の低遅延ランダムアクセスは非常に重要な機能である[22]．MVC の CfE においても，視点映像の低遅延ランダムア

クセスを考慮した手法が提案された(図 10)[23] . また 3DAV に特徴的な手法 (カメラパラメータを利用する) として , CfE 以前より , view interpolation を予測符号化に適用する技術が提案されている [24] . この技術は視点映像のカメラパラメータを利用することで , 視差ベクトルの符号化に要するオーバーヘッドを低減する効果がある . 他にも , カメラパラメータを利用しないが , 視差ベクトルと動きベクトルにおけるオーバーヘッドを効果的に低減する手法が提案されている [23][25] .

CfE の評価では , 複数の提案方式の中でいずれかは , anchor に対して約 2 倍の圧縮効率を実現できることが示された [10] . CfP では , CfE で独立に提案された要素技術 , 特に view interpolation との組み合わせが提案されてくるものと期待される . View interpolation を取り入れるためにも , CfP 用のテスト画像を新たに募集している [26] .

6. まとめ

MPEG の 3DAV 活動について史実に沿って検討内容を紹介した . 今後は自由視点テレビを最終ゴールとする MVC の検討が精力的に進められると思われる . またカメラパラメータの規定方法についても今後の重要な検討課題である .

文 献

- [1] "Camera Positioning Information Conveyed by Ancillary Data Packets," SMPTE 315M, 1999.
- [2] "Applications and Requirements for 3DAV," N5877 MPEG Trondheim Meeting, Jul., 2003.
- [3] K. Yamazawa, Y. Yagi, and M. Yachida, "Omnidirectional imaging with hyperboloidal projection," Proc. Int. Conf. on Intelligent Robots and Systems, vol.2, pp.1029-1034, Jul., 1993.
- [4] R. Yamashita, T. Matsuyama, K. Hasida, and K. Nomura, "Description of Omni-Directional and Multi-Viewpoint Video for 3D interaction with 2D Video", M7779 MPEG Pattaya Meeting, Dec., 2001.
- [5] <http://www.chiariglione.org/mpeg/events&tutorials/3dav/3dav.htm>.
- [6] "Preliminary Requirements for 3D Video support in MPEG," N4679 MPEG Jeju Meeting, Mar., 2002.
- [7] "Description of Exploration Experiments in 3DAV," N4929 MPEG Klagenfurt Meeting, Jul., 2002.
- [8] "Report on 3DAV Exploration," N5878 MPEG Trondheim Meeting, Jul., 2003.
- [9] A. Smolic, H. Kimata, "AHG on 3DAV Coding," M10218 MPEG Waikoloa Meeting, Dec., 2003.
- [10] "Report of the Subjective Quality Evaluation for Multi View Coding CfE," N6999 MPEG Hong Kong Meeting, Jan., 2005.
- [11] "Draft Call for Proposals on Multi-View Video Coding," N6910 MPEG Hong Kong Meeting, Jan., 2005.
- [12] S. Sethuraman, W. E. Sevrerson: "Proposal for 3D coordinate Systems and sensor parameters," M6866 MPEG Pisa Meeting, Jan., 2001.
- [13] T. Matsuyama and R. Yamashita, "Requirements for Standardisation of 3D Video", M8107 MPEG Jeju Meeting, Mar., 2002.
- [14] K. Yamazawa, H. Habe, T. Nomura, and T. Matsuyama, "Report of EE1 on Omni-directional Video", M9184 MPEG Awaji Meeting, Dec., 2002.
- [15] T. Fujii, T. Kimoto, and M. Tanimoto, "Ray Space Coding for 3D Visual Communication," PCS'96, 2, pp. 447-451. 1996.
- [16] M. Tanimoto and T. Fujii, "FTV -- Free Viewpoint Television", M8595 MPEG Klagenfurt Meeting, Jul., 2002.
- [17] "Requirements on Multi-view Video Coding v.2," N6834 MPEG Palma Meeting, Oct., 2004.
- [18] "Text of ISO/IEC 14496-16/FPDAM1 (AFX Extensions)," N6986 MPEG Hong Kong Meeting, Jan., 2005.
- [19] T. Matsuyama, Y. Katsura, M. Boehme, and H. Habe, "Compression Scheme for Model-based 3D Video", M9761 MPEG Trondheim Meeting, Jul., 2003.
- [20] "Animation Framework eXtension Core Experiments Description", N6988 MPEG Hong Kong Meeting, Jan., 2005.
- [21] H. Kimata and M. Kitahara, "Anchor bitstreams for preliminary Call for Evidence on multi-view video coding", M11231 MPEG Palma Meeting, Oct., 2004.
- [22] H. Kimata, M. Kitahara, K. Kamikura, Y. Yashima, T. Fujii, and M. Tanimoto, "System Design of Free Viewpoint Video Communication," CIT2004, Sep., 2004.
- [23] H. Kimata and M. Kitahara, "Multi-view video coding using shared reference picture memory and shared motion vector memory (Response to Call for Evidence on MVC)", M11570 MPEG Hong Kong Meeting, Jan., 2005.
- [24] M. Tanimoto and T. Fujii, "Ray-Space Coding Using Temporal and Spatial Prediction," M10178 MPEG Brisbane Meeting, Oct., 2003.
- [25] Y.-L. Lee and W.-C. Sung, "Multi-view video coding using 2D direct mode," M11266 MPEG Palma Meeting, Oct., 2004.
- [26] "Requirements on Test Material for Multi-View Video Coding Experiments", N6721 MPEG Palma Meeting, Oct., 2004.