次世代映像メディア ～自由視点テレビFTV～

講演日：2010年5月28日（金）

場所：機械振興会館

（正会員）谷 本 正 幸[†]

1. ま え が き

本日は、私どもが最も革新的な次世代映像メディアと考えています自由視点テレビについてご紹介をさせていただきます。

まず、映像メディアの代表であるテレビのこれまでの進歩を振り返ってみましょう。1926年に高柳健次郎博士が実現した世界で最初の電子式テレビ受像機と、NHKが2005年の愛知万博に出展したスーパーハイビジョンを比較してみますと、明らかに解像度が違います。これは走査線数が前者では40本で、後者では4,000本であるためです。すなわち、この80年間に走査線数は100倍、画素数では1万倍以上になったということで、非常に大きな進歩でございます。

では、今後映像メディアはどのように進歩していくのでしょうか。これについて、二つの方向、軸が考えられます（図1）。一つの方法は画素の軸で、これまでのようにテレビがハイビジョンになって、さらにスーパーハイビジョンになる。画素数が増える方向です。

もう一つの方法は視点の軸です。視点数が増える方向です。これまでのテレビは、ただ一つの視点の情報しか送っていませんので、視点数は1です。このことは、ハイビジョンであってもスーパーハイビジョンであっても、変わりません。しかし、5～6cm離れた左右両眼の情報を送る2眼式の立体テレビでは、視点数は2となります。

それでは、視点数をさらに増やしていったら、100とか、1,000とか、1万とかというふうになったら、一体それはどういうことになるのでしょうか。人間の目は二つだからそれを超える視点数はないのではと思われるかもしれませんが、ここに技術としての可能性があります。

視点数がどんどん増えると、どこにでも目があるような状況になります。それを超多視点と言っておりますが、そういう状況では、どこからでも自由にシーンが見える映像メディアが実現します。これが、自由視点テレビです。

すなわち、これまでの映像メディアは3次元空間情報のほんの一部である、1視点の情報しか伝達していませんでした。これからの映像情報メディアが挑戦すべき最重要課題は、3次元空間情報のすべて、すなわち全視点の情報の

図1 映像メディアのこれから

伝達です。それを実現するものを私どもは自由視点テレビ“FTV”（Free-viewpoint Television）と名付けました。

映像メディアの視点数は、2Dのテレビはもちろん1です。2眼のステレオですと2です。それから多眼のステレオでは、数視点から10数視点になります。こういう2眼のステレオや多眼のステレオを合わせて3Dテレビと言っております。

これに対して、FTVの視点数は無限大です。ここで無限大というのは、無限個の視点でシーンを一度に見るということではなくて、どこにでも視点を置くことができるという意味です。

それで、FTVとは、あたかもその場にいるかのように、自由に視点を变えて3次元シーンを見ることができる映像メディアであると定義しております。

このようなFTVが実現すれば、それは3次元空間情報のすべてを伝達して記録することができる究極の3次元テレビであり、時空間のアーカイブになります。そして臨場感の非常に高いヒューマンインタフェースになり、さらに、高い表現力を有する革新的なメディアアートやコンテンツの制作ツールにもなり得ると考えております。

2. FTVへの道

それでは、私どもがどのようにしてFTVに到達したかをお話します（図2）。15年ほど前に、3次元空間の光線情報を記述する光線空間の研究を始めました。そして、その情

[†] 名古屋大学 大学院工学研究科 電子情報システム専攻
"Next-Generation Visual Media: FTV (Free-viewpoint TV)" by Masayuki Tanimoto (Graduate School of Engineering, Nagoya Univ., Aichi)

報圧縮の仕組みや、光線空間を取得するときに、全体を一度に取ることができないので、飛び飛びの、少しの情報を取って、残りは補間で作って穴埋めする技術などの、理論的な検討を随分長い間やってきておりました。

理論検討だけでなく、実際に動くものを作りたいと考えていましたが、2000年になって、自由視点鳥瞰システムという、あたかも空を飛んでいる鳥が下を眺めるように、好きなところからシーンが見えるシステムを実現することができました。このときには、信号処理を行う能力が低かったので、パソコンをたくさん並べたPCクラスタを使ってリアルタイム処理を実現しました。

その後、もっと広い空間の自由視点を実現したいということで、2005年にハイビジョンカメラ100台からなる100眼システムを作りました。

この頃になりますと、新しい視点を作り出す部分は結構うまくいくようになって、デスクトップのパソコンだとか、ノートパソコン上でも動くようになりました。最近では、iPodのようなモバイルプレーヤでも、リアルタイムで自由視点画像を作ることができるようになっております。

それから、実際に物が動き出した2001年頃から、私どもは、この新しいテレビを、もっと世界中の人に知ってもらいたいと考え、動画像の国際標準化を行っているMPEGに提案いたしました。標準化活動を続けてきております。

自由視点テレビに関連したものとして、多視点映像という範疇がございます。例えば、1999年の映画の「マトリックス」では、多数のカメラで撮影した画像を次々にスイッチして見せ、視点が飛んだように見える特殊映像を作りました。また、2001年に発表されたカーネギーメロン大学のEyeVisionというシステムでは、アメリカンフットボールのシーンを多数のテレビカメラで撮影し、カメラのスイッチによって、視点が移動するシーンを見せました。

しかし、これらは多視点の映像であって、自由視点ではありません。多視点映像では、撮影した映像を表示しますが、自由視点テレビでは、撮影していない視点の映像を作

り出して表示します。この点が両者の大きな違いです。

すなわち、FTVでは、たくさんカメラでシーンを撮影しますが、これはベースになる情報であって、これから自由視点の映像を作り出します。自由視点はどこにあってもよく、実カメラと実カメラの間でも、前でも後ろでも、どこでもよいということでございます。

このようなFTVを実現するための手法ですが、大きく分けて二つのアプローチがあります。一つは、モデルベースの手法で、3次元シーンのモデルを作り、そのモデルのある視点から見たらどう見えるかを計算します。しかし、実世界の正確なモデル化が困難であるため、実写品質の映像は得られていません。

もう一つは、私どもが採っている光線ベースの手法です。これは、限られた視点の画像をカメラで撮影します。例えば、実カメラはこのような飛び飛びの位置に置かれます。これらのカメラは、画像を撮影するというよりは、レンズの中に入ってくる光線を取得する装置であると、位置づけられます。

そして、実カメラとは別の位置に仮想カメラがあって、その画像を作りたいとき、仮想カメラのレンズの中心を通る光線群が必要です。そのうち、ある光線は実カメラでつかまえている。別の光線は別の実カメラでつかまえている。また、もう一つの光線はさらに別の実カメラでつかまえています。しかし、その他のほとんどの光線はつかまえていないので、実カメラが得た光線をもとにして、それらの光線を作り出します。つまり、自由視点画像は光線の統合と補間で作り出すということでございます。

図3に、私どもが2000年に世界で初めて実現したFTVである自由視点鳥瞰システムを示します。これは16台の小型CCDカメラをドーム状に置き、リモコンで走らせた車を撮影します。そして、撮影した映像をもとにして、3次元マウスで指定した視点の映像を、このPCクラスタによりリアルタイムで生成しました。

シーンも視点も動いていて、動いている実シーンの自由

図3 自由視点鳥瞰システム（世界初のFTV、2000年）

視点画像をリアルタイムで実現しています。だから、FTVの世界初の実現例ということでございます。

ただ、このシステムは、合成の仕方が簡単なため、映像はあまりきれいなものではありません。けれども、機能としては、動いているシーンを対象にして、仮想視点も自由にリアルタイムで与えて、それで自由視点画像をリアルタイムで作っているという点で、まさにFTVそのものでございます。

3. FTVの原理

そのあと、もっと品質の良いFTVを作ろうということで、光線ベースのFTVを実現しました。

撮影の仕組みについては、実空間のシーンをたくさんのカメラを使って撮影します。そして、1台目のカメラで撮った画像を、仮想空間の1の位置に衝立状に置きます。さらに、2台目のカメラで撮った画像を2の位置に置く。3台目のカメラで撮ったものを3の位置に置く。これを繰り返します。撮影する空間は実空間ですが、画像を衝立のように並べる空間は仮想空間で、頭の中で考えた空間です。これを光線空間Ray-spaceといっております。

このとき、カメラとカメラの間隔と、衝立と衝立の間隔は比例しておりますから、もしカメラがびっしり密に詰まっておれば、密な光線空間が得られます。カメラの間隔が広いと、光線空間も隙間が空いてスカスカになっています。だから、それは補間によって埋めるわけです。そうすると、密な光線空間が得られます。

図4は、光線空間の構造を示しています。光線空間は、縦に切ると2次元像になりますが、水平に切ると直線構造の集まりになっています。この直線構造を使って補間をします。

このようにして、密な光線空間ができますと、自由視点画像の生成は非常に容易です。視点が与えられると、光線空間を切る位置がすぐわかって、そこで光線空間を垂直に切ると、その断面像がその視点から見た画像となります。別の視点だったら、別の位置で切ります。このようにして、いろんな視点の映像を瞬時に作ることができます。

もし映像を立体テレビに映したいのならば、並行した2

図4 直交光線空間の水平断面

本の包丁で二つの断面を作ります。それが右目画像と左目画像になります。それを2眼式3Dディスプレイに表示すれば、飛び出すテレビになります。もし、10本の3次元テレビに出す場合には、10本の並行した包丁で断面を作ります。

今ご説明した仕組みですと、カメラが直線上に配置されておりますので、実カメラ間とか、前進および後退した位置からシーンを見ることはできますが、横や後ろ側は見えません。情報が無いからです。そのようなことを実現したい場合には、円形にカメラを配置します。

円形カメラ配置の場合にも、撮影画像を順に衝立状に並べて光線空間を作ります。しかし、その構造は直線カメラ配置の場合とは異なります。例えば、0°のところで撮影した画像から始めて、1°のところで撮影した画像、2°のところで撮影した画像と順に置いていって、180°のところで撮影した画像では裏側が見えます。さらに回って360°のところで撮影した画像は0°の画像と同じですので、円形カメラ配置の光線空間の水平断面は、直線構造とはなりません。これはきちんと解析すると、正弦波構造になることがわかります。

このため、円形カメラ配置の場合には、直線状の物差しではなくて、正弦波状の物差しを使って光線空間を補間します。

4. FTVの技術

このような仕組みで作ったFTVシステムは、撮影、補正、エンコード、デコード、画像生成、表示の各部から構成されます。最初の撮影部では、たくさんのカメラでシーンを撮影します。カメラは例えば一直線上に正面に向けてきちんと配置します。

次は、補正です。この補正部では、位置の補正と色の補正を行います。カメラは、きちんと配置する必要があるのですが、実際にはそのようにはできません。位置も直線からずれたり、方向も少しガタガタしています。それをその

図5 FTVの撮影部と処理部

図6 さまざまなカメラ配置の100眼システム

まま並べたのでは、正しい光線空間が得られないので、これを画像処理で補正します。また、カメラに個体差があり、色特性がそれぞれ違うので、それも補正します。

次は、エンコード、デコードです。たくさんのビデオ、すなわちマルチビュービデオを符号化して送り、受信側でデコードします。

そして、受信側の画像生成部では、補間によって、足りない部分を埋めて、その後に包丁で切って断面を作ります。2Dのテレビだったら1本の包丁で1ビューを作ります。3Dのテレビだったら、その眼数に応じて、例えば5眼の3Dディスプレイだったら5ビューを生成します。

このように、FTVでは、撮像の眼数と表示の眼数が同じである必要はなく、撮影と表示を切り離して取り扱うことができます。

それでは各部をさらに具体的にご説明します。

まず撮影ですが、図5が鳥瞰システムの後に作ったもので、15台のカメラが円弧上に並べられています。これでこの水槽のシーンを撮影して、PCクラスタで自由視点画像をリアルタイムで生成しました。

しかし、これでは撮影できる範囲が狭いので、もっと大きな教室ぐらいの空間を撮影するために、ハイビジョンカメラを100台準備しました(図6)。鉄砲隊のような直線配置や円形配置など、いろんなカメラ配置が可能です。また、カメラを縦横の2次元アレイ状に配置すれば、水平方向だけでなく、上下方向の視点移動も実現できます。

それから補正です。位置補正前のカメラで撮影したままの画像はガタガタしていますので、画像処理で位置の補正を行います。画像間の対応関係がきちんと取れて、1画素以内に水平の位置が合います。

色に関しても、カメラにはそれぞれ個体差がありますので、補正前はカメラ画像を切替えると色がチラチラしますが、色合わせをすると揃います。光線の統合をするときに、複数のカメラの情報を使いますので、このことが必要です。

次に圧縮ですが、FTVはカメラをたくさん使いますので、

図7 さまざまな位置と時刻で生成した自由視点画像

カメラを10台使ったら、普通のビデオの10倍の情報になる。50台使ったら50倍で、100台使ったら100倍になるので、情報圧縮の必要性が高いのです。

多くのビデオの集まりをマルチビュービデオと言います。カメラがたくさんあると、そのビューの数だけ映像があります。それがマルチビュービデオです。

普通のMPEGの圧縮ですと、各々のカメラのフレーム間の冗長性を利用して情報圧縮します。しかし、FTVではすべてのカメラは同じシーンを撮影していますので、カメラ画像間にも冗長性があります。マルチビュービデオでは、この両方の冗長性を取ることが必要です。そのような方針で始めたのが、MVC、マルチビュービデオコーディングでございます。これについては、MPEGで検討が進められて、現在、標準化ができて上がっております。

それから、画像生成ですが、先ほどの15台のカメラのFTVの場合、実カメラ画像から作った自由視点画像は図7のようになります。仮想視点は実カメラと実カメラの間を動いたり、カメラ列から離れて前進したり後退することもできます。このシーンは、砂利や表面のガラスの反射、泡そして水草などがあったりして、非常に複雑なものですが、きちんと再現されております。

このときにはPCクラスタで画像生成を行いました。数

図8 ノートPCと携帯端末で実現したFTV

年前からノートPCでもFTVが実現できるようになっていますので、そのFTVプログラムをここで動かしてみましょう。

たくさんのカメラで撮影したマルチビュービデオの1フレーム分が、すでにパソコンの中に入っています。FTVプログラムは、それから自由視点画像をリアルタイムで作ります。画面内のカーソルをマウスで動かすことによって、視点を前後、左右に移動できます。

撮影したシーンの3次元構造は非常に複雑です。手前にも後方にも物体があります。後方にある物体は、視点の位置によっては手前の物体に隠れて見えませんが、視点を左右に移動させると見えるようになります。そのような物体の前後関係が非常に良く再現されています。

視点を前後に移動させると、単なるズームでないことがわかります。例えば、前進すると手前の物体が大きく見えるようになり、後方の物体が隠されてしまいます。

また、今はシーンが静止していますが、プログラムはビデオレートで動いていますので、もしマルチビューの動画がパソコンの中に入っていれば、もちろん動画のFTVになります。

図8は2年ほど前のiPodですが、その中にFTVのアルゴリズムを実装しました。タッチパネルで視点を与えますと、リアルタイムで自由視点画像を生成します。この場合もマルチビュービデオの1フレーム分しか入っていないので、動画ではありませんが、処理はビデオレートで行われています。もし、動画が入っていたら、リアルタイムのFTVになります。

次にFTVの表示として、二つの方法を示します。第一は視点追従型で、2Dや3Dのディスプレイに対して、視点を指示し、それに従った画像を表示します。2Dディスプレイだったら、もっと右に移ってくださいとか、左に行ってくださいとか、前進してくださいという指示を与えて見せます。3Dディスプレイの場合も、普通の3Dディスプレイだと視域が非常に限られていますけど、この仕組みを使うと、視域が非常に広がります。

第二は、視点追従をせずに、眼数の非常に多い3Dディスプレイに、ビューをたくさん作って表示するもので、多くの人が自由に視点を動かしてシーンを見ることができます。

第一の例として、2Dディスプレイの場合は、カメラでディスプレイを見ている人を撮影し、画像処理でこの人の頭をトラッキングしていて、その頭が下がれば下から見たシーンを表示し、上がれば上から見たシーンを表示します。トラッキングをせずに、リモコンやマウスで、視点移動の方向や大きさを直接指示することもできます。

視点の追従を行う4眼の3Dディスプレイでは、視点追従をしなければ、視域を超えるとクロストークが起こり、さらに進むと周期的にまた同じものが見えます。それに対して、FTVの技術を使いますと、クロストークなく非常に視域を広くすることができ、しかも両眼で見ると立体視ができています。

第二の例として、360°どこからでも見える光線再現型3Dディスプレイを示します。2Dディスプレイは、表面の1点からどんな方向にも等しい強度の光線を出しますが、この3Dディスプレイは、方向によって異なる強度の光線を出します。つまり、このディスプレイでは非常に細かく光線を制御し、円筒の外側では、あたかもこの人が中にいたと同じような光線が再現されます。このため、大勢の人がいろいろな方向から同時に立体視をすることができます。

FTVを実現するには、高い多視点の撮影、表示技術が必要になります。そこで、私どもは多視点の撮影、表示の能力を示すものとして、画素視点積、ピクセルビュープロダクトという指標を定義いたしました(図9)。ここで画素とは、1ビューの中に画素がいくつあるかということで、これは光線の方向分解能を示しています。視点とは、そういうものが何ビューあるかということで、これは光線の空間分解能を示しています。この両方の積は多視点の撮影表示能力を表すとともに、光線の取得再生能力も表しています。

画素視点積は、撮影、表示とも、年とともに増加し、多視点の撮影表示能力が非常に高くなっていることがわかります。表示において、時分割の技術を使いますと、画素視点積をさらに1桁ぐらい上げることができます。

このように、多視点の撮影、表示の能力は着実に向上しておりまして、自由視点テレビを実現できる環境が整ってきております。

5. FTVの国際標準化

私どもは2001年から、MPEGにおきましてFTVの国際標準化活動を推進しております。FTVの第1フェーズとして2004年に始めたMVCの標準化が、2009年に成立しました。そして、ブルーレイの3DにMVCが採用されることが決定しております。

MVCによってたくさんのビデオの圧縮が可能になったわけですが、MVCだけではFTVは実現できません。それで、自由視点画像を作り出す技術なども含めた、FTVの新しい標準化を始めたいということを提案し、了承されました。それが、2007年から始めたFTVの第2フェーズです。現在、

図9 多視点の撮影表示能力を示す画素視点積の進歩

図10 MPEGにおけるFTV標準化の経緯

3Dディスプレイをターゲットにした3Dビデオの標準化を推進しています。

それでは、MPEGにおけるFTV標準化活動を、順を追ってもう少し詳しくご説明します(図10)。

まず、大きな流れとして、私どもは、2001年12月にFTVをMPEGに提案しました。そして3Dオーディオビジュアルグループで、3Dとしてこれからどういうものをターゲットにしていっていいかを議論いたしました。2004年3月に議論が収束し、ターゲットをFTVに絞ることが決定いたしました。

そして、FTVの第1フェーズとして、たくさんのビデオの圧縮、すなわち、マルチビュービデオコーディングを始

めました。そして2006年7月に、この活動をビデオの専門家が集まるジョイントビデオチームJVTに移し、2009年5月にMVCの標準化が完成いたしました。

MVCの完成を待たずに、私どもは2007年4月にMPEG-FTVグループを作り、FTVの第2フェーズをスタートさせました。現在このグループは、FTVの新しいビューを作り出す機能を利用する3Dディスプレイアプリケーションをターゲットにした3Dビデオの標準化活動をしております。

2009年10月に発行された“Applications and Requirements on 3D Video Coding”というMPEGの出力文書は、3Dビデオの標準化の目的と条件をまとめた重要な文書です。その中で、FTVは2001年に始まった新しいフレームワークである

図11 FTVレファレンスモデルと名古屋大学の貢献

こと、FTVの第1フェーズはMVCであり、FTVの第2フェーズは3Dビデオであることが、明記されております。

第1フェーズのMVC標準化につきましては、経済産業省の基準認証研究開発事業の支援を受けまして、JEITAの中にFTV委員会を立ち上げて、多くの企業にご支援をいただいでやってまいりました。

第2フェーズの3Dビデオにつきましては、総務省のSCOPEの支援を受けて研究活動を進めています。さらに、URCFの中にMPEG-FTVという作業班を作っていたいで、活動しているところでございます。

それでは、MVCと3Dビデオの内容をさらにご説明します。

まず、MVC標準化でございます。2001年から3Dに関連するさまざまな提案があつて、いろいろと審議いたしました。その中には、全方位ビデオ、FTV、モデルベースの3次元ビデオ、2眼のステレオ、1ビューにデプスをつけ加えた3Dなどがありました。そのようなものをいろいろ議論した結果、これからはFTVに絞るということを合意いたしまして、FTVの入力信号であるマルチビュービデオの圧縮の標準化を2004年にスタートさせました。

それが2009年5月に完成し、H.264/MPEG-4 AVCの拡張として、ISO/IEC 14496-10:2009 Annex Hという標準ができました。

MVCでは、ビュー間の冗長性を取らずに各ビューを独立に圧縮するサイマルキャストに比べて、ビットレートを20～30%削減できます。

それでは、今行っている3Dビデオの標準化についてご説明します。MVCと大きく異なる点はデプス情報の利用です。

自由視点画像を生成するとき、カメラ画像間の補間というプロセスがあります。このプロセスは、光線空間の水平断面の直線検出に相当する奥行き推定(デプスサーチ)と、その直線に沿った補間の二つのプロセスから成っています。直線に沿った補間は、足して2で割るとか、重みをつけて加

算するとかで、非常に簡単です。しかし、デプスサーチは非常に時間がかかる重い処理です。現在のFTVは奥行き推定と補間の両方合わせてリアルタイムで実現しています。

しかし、もっと簡単な仕組みはないかということで、別の仕組みを考えました。これまでは、新しいビューを作るときに、デプスサーチと補間をみんな受信側でやっていましたが、面倒なデプスサーチを送信側でやることにしますと、受信機のほうは非常に簡単にできます。この構成が2007年にMPEG-FTVで定めたFTVレファレンスモデルで、現在の3Dビデオはこの構成を採用しています(図11)。

もし、デプスサーチと補間をみんな受信側へ持ってくれば、これは光線空間を全部作ってしまうということになって、現在の技術では現実的なアプローチではありません。

FTVレファレンスモデルでは、送信側では、これまでと同様に、たくさんカメラでシーンを撮影し、マルチビューを得ますが、この他に、各視点点がデプス情報を持っている、つまり、ビューとデプスがペアになったものが視点点だけあるMultiview+depthのデータ構造を持つことが特徴です。

受信機側では、デプスとビューを利用して、3Dワーピングによって新しいビューを作ります。例えば10眼のディスプレイに表示するなら、10ビューを作ります。

FTVの第1フェーズでは、マルチビューだけ考えましたが、第2フェーズではデプスも同時に持つデータ構造を考えています。これを送って、受信側で2眼、または多眼の3Dディスプレイに表示します。

そのときに用いるディスプレイのタイプは決まっていまないので、一つのコンテンツを送って、どんなディスプレイにでも表示したい。そうすると、2眼でも見せたいし、5眼でも10眼でも見せたいとなったら、そこにFTVの新しいビューをつくるという機能がどうしても必要になります。

3Dビデオは、MVCより圧縮能率が高く、3次元画像を作

り出す能力も高いものの実現を目指しています。

3次元画像を作り出す能力とは具体的にどうということかといいますと、二つあって、一つはディスプレイアダプテーションで、同じコンテンツをいろんな眼数のディスプレイで見るために、新しいビューを作ることです。

もう一つはビューイングアダプテーションで、これは表示される3D画像の奥行きをユーザがコントロールできることです。つまり、見る人の好みによって、飛び出しの度合いを変えることを狙っています。ビューイングアダプテーションのためには、撮影と表示のベースラインを違える必要があります。これにもFTVの自由視点画像生成技術が使われます。

名古屋大学は、このMPEG-FTV活動に対して、FTVテスト画像、マルチビューからマルチデプスを作り出すデプス推定レファレンスソフトウェア、および、ビューとデプスから新しいビューを作り出す画像合成レファレンスソフトウェアを提供しています(図11)。

6. FTV研究開発の意義

FTVの研究開発の意義を最後に述べたいと思います。私どもはFTVの研究を長い間進めてきましたが、ある時期から世界的にマルチビューに対する関心が急速に高まってきました。これはMPEGでの活動がきっかけになっていると思います。

例えば、2007年の3DTVコンファレンスでは、14のオーラルセッション中の5セッションがマルチビュー関連でした。この年にはIEEEのトランザクションCSVTにマルチビュービデオコーディングの特集号が出たり、シグナルプロセッシングマガジンにマルチビューと3Dの特集号が出たりしました。特に、2007年11月に発行されたIEEEシグナルプロセッシングマガジンの特集号の表紙に、マルチビューと3Dをイメージする写真として載っていたものは、私どもの光線空間の写真でした。

最近では、いろんなコンファレンスで、マルチビューだけでなく、FTVがキーワードになって、セッションテーマになる例も出てきております。例えば、2010年12月に名古屋で開催されるピクチャーコーディングシンポジウムでは、FTVが特に重要なテーマとして挙げられています。

このように、FTVの研究が、多視点、自由視点という新しい研究分野の開拓に大きく貢献できたと考えています。

次に、多視点画像による新しい信号処理の可能性について述べます。たくさんの視点から見た画像というのは、ただ画像をたくさん集めたものに過ぎないと思われるかもしれませんが、実は、たくさん集めることによって、一つ一つではできないことが、できるようになる。そういう例をお示しします。

例えば、人形の前に柵があるシーンを多数のカメラで撮影します。そうすると、どの位置から撮っても柵が邪魔に

図12 画像生成誤差の多視点相関

なって、人形だけを見ることはできません。しかし、たくさんの視点から撮った画像を集めて、それで光線空間を作り、その光線空間に非線形のフィルタリングをかけます。そうすると、柵が消えて、人形だけのシーンが得られます。

もう一つの例として、多視点情報を利用した画像生成をご紹介します。ある視点から見たビューとデプスがあるとします。ビューとデプスがわかっているということは、対象物のテクスチャと形がわかっていることになりますので、他の視点から見たビューは3Dワーピングで容易に生成できます。このとき、デプスに誤差があると、生成したビューにも誤差が発生します。正しいデプスがわからないと、このことは避けられませんが、ビューがたくさんあれば、デプスの誤りを補正することができます。今、視点A、視点B、視点C、視点Dを考えます(図12)。この視点Aにビューとデプスがあるとしますと、これから視点B、視点C、視点Dのビューが作れます。それが、生成画像 V_B' 、 V_C' 、 V_D' です。だけど、これは視点Aのデプスが間違っていますから、本物ではないわけです。

しかし、視点B、視点C、視点Dにカメラを置いていたら、本当のビュー V_B 、 V_C 、 V_D がわかりますから、これと生成画像との差を取ると、画像生成誤差 $\Delta V_B = V_B - V_B'$ 、 $\Delta V_C = V_C - V_C'$ 、 $\Delta V_D = V_D - V_D'$ が得られます。これを求めてみると、各視点の画像生成誤差には強い相関があることがわかります(図12)。このことから、カメラDの画像が参照ビューとして与えられているならば、 ΔV_D が既知となり、これから ΔV_B や ΔV_C が推定できます。

このように画像生成誤差を推定し、デプスの誤りを補正すれば、生成画像の品質が向上します。まだ画像生成誤差を推定する仕組みの解明が充分でないため、実験結果では誤差が残っていますが、理屈の上からはゼロにできる可能性があります。

私どもは、多視点や自由視点の研究に役立てるため、FTVテストシーケンスを研究用に無償で提供しております(図13)。この中には、固定の多視点カメラで撮影したシー

図13 名古屋大学作成のFTVテストシーケンス

ケンスの他に、移動する多視点カメラで撮影したシーケンスも含まれています。名古屋大学谷本研究室のホームページ¹⁾に入ると、シーケンスにアクセスできます。ダウンロードにはIDとパスワードが必要ですが、ご連絡頂ければ開示をいたします。

7. む す び

現在、2022年のFIFAワールドカップを日本に招致する活動が進められています。2022FIFAワールドカップ日本

招致委員会は、サッカースタジアムの感動を再現して世界に配信する技術としてFTVに注目し、これを招致コンセプト²⁾の重要な柱としました。日本発の最先端映像メディアであるFTVがFIFAワールドカップを支える技術として役立つことを願っています。

これまで、FTVについて、高校生から社会人まで、さまざまな年齢層の方に紹介をさせて頂きましたが、いずれも大変興味を持って聞いて頂き、手ごたえを感じました。特に、これから進路を決める若い人たちの反応は大変良く、最先端映像メディアへの関心の高さがうかがわれました。このような関心が契機となって、多くの若者が映像情報メディアの分野に進み、本学会の未来を支える力となってくれることを確信しております。

本日は、長時間ご清聴頂きありがとうございました。

〔文 献〕

- 1) FTVテストシーケンス, <http://www.tanimoto.nuee.nagoya-u.ac.jp/>
- 2) 2022FIFAワールドカップ日本招致委員会コンセプトビデオ「サッカー観戦を革新する日本の技術力」, http://dream-2022.jp/jp/our_bid/bid_book/videos/jp02/index.html?keepThis=true&TB_iframe=true&width=711&height=450

谷本 正幸 1970年、東京大学工学部電気工学科卒業。1976年、同大学大学院博士課程修了。同年、名古屋大学工学部助手。1991年、同教授。現在、同大学大学院工学研究科電子情報システム専攻教授。通信方式、画像情報圧縮、画像処理の研究に従事。文部科学大臣表彰科学技術賞など受賞。当学会前会長。電子情報通信学会フェロー工学博士。正会員。