３次元ビデオ技術のその後の展開と現状

延原 章平 *

1　３次元ビデオとは

３次元ビデオ [1] とは被写体の３次元形状と表面テクスチャをそのまま記録した映像メディアであり，1997 年の Kanade ら [2] と Moezzi ら [3] の先駆的研究を皮切りに，これまで多くの研究がなされてきた．本稿では３次元ビデオ生成の基本的な考え方をまず紹介し，３次元ビデオ生成に関わる現在の研究動向を概観するとともに，今後の展望について述べる．

まず３次元ビデオの生成の過程は，多視点映像撮影，３次元形状・表面テクスチャ推定，自由視点映像生成，圧縮・保存に大別される．ここで自由視点映像生成という観点で，Image-based Rendering のような関連技術と比較すると，３次元ビデオは被写体の３次元形状と表面テクスチャを明示的に推定する点にその特徴がある．また多視点映像を入力とした３次元形状・表面テクスチャ推定という観点では，CV および CG 分野における Image-based Modeling が関連深いが，特に３次元ビデオの場合は運動物体を対象とする点に特徴がある．

■多視点映像撮影　被写体の全周囲３次元形状を得るためには，本質的に被写体を全周囲から計測する必要が生じる．被写体が静止物体であると仮定できるならば，カメラを移動させながら撮影することによって多視点撮影を行うことができるが，自由に運動する被写体を対象とするならば，多視点カメラ環境を用意することが求められる．また技術的には多視点カメラ群の幾何的・光学的キャリブレーション，同期撮影の実現，撮影可能範囲を確保するためのカメラ配置設計などが課題となる [4][5].

■３次元形状・表面テクスチャ推定　多視点映像を入力として３次元形状・表面テクスチャ推定を推定する問題は CV 分野におけるもっとも基本的な課題のひとつとして数多くの研究がなされてきたが，その中でも特に３次元ビデオ生成の観点では，画像中における対象領域の輪郭情報（シルエット）を利用した視体積交差法と，画像間におけるテクスチャの一致度（photo-consistency）を利用した対応点推定に基づく多視点ステレオ法，およびそれらを統合したアルゴリズムが一般に用いられる．

これに加えて時間方向の情報，つまり Structure-from-Motion (SfM) を組み合わせる手法も提案されている [6]．これは一般に多数のカメラ群を用意することは容易ではないことから，多視点ステレオは疎なカメラ群による wide-baseline ステレオとなって photo-consistency の評価そのものが容易ではない一方で，SfM では narrow-baseline ステレオとなってより安定に対応付けを得ることができるという着眼点によるものである．

■自由視点映像生成　３次元形状と表面テクスチャを入力とした自由視点映像生成は，基本的には CG における 3D モデルのレンダリングと同じプロセスとなるが，３次元ビデオの場合は形状の表面テクスチャが実写であること，そしてそれが同一箇所に対して複数存在する点に特徴がある．これはもともとある表面形状を推定するにあたっては，その場所が複数の視点から撮影されていたはずであり，したがってその部分に対応付けられるテクスチャは複数台のカメラで撮影された画像となることに起因しており，レンダリング視点（仮想視点）に近い実カメラ画像をテクスチャとして採用するという視点依存レンダリングが一般的に採用されている [7]．これは複数画像を統合した単一テクスチャを生成するよりも，仮想視点に応じて実画像を切り替えたほうが視線方向に依存する光沢感などがより自然に再現できることによる．

■圧縮・保存　３次元形状と表面テクスチャを効率的にデータ圧縮する符号化法には，通常映像の圧縮符号化と同様に，フレームごとのデータ圧縮と，フレーム間差分によるデータ圧縮の２つの観点が存在する．

前者は主に Gu らによる Geometry Images と呼ばれる手法が知られており，これは３次元メッシュを２次元平面へと展開し，それを画像とみなして再サンプリング・圧縮する手法である．またこうして各フレーム独立に得られた画像の系列を，映像とみなして更に圧縮する手法も知られている．

一方後者は後述するように３次元ビデオから３次元運動フローを求め，これを利用してフレーム間圧縮を行う手法である．またその過程で全フレームで共通なテクスチャ画像を生成したり，またモーションフローをさらに解析して骨格運動として記述を単純化するなどの処理も行われることが多い．

前者はより信号処理的な側面が強く頑健な動作が期待

* 京都大学大学院情報学研究科

でき，後者では骨格運動や共通テクスチャを CG 的に編集することが可能となる．このようにどちらのアプローチにも長所が存在し，現時点で標準的な圧縮符号化法と呼べるものは広まっていない．

2　現在の研究動向

上述のような３次元ビデオの基本的な考え方に対して，３次元形状復元など個別の要素技術の改善はもちろんのこと，近年の国際会議では下記のような新たな視点での研究が発表されている．

■マーカーレスモーションキャプチャ　３次元ビデオではもともと各時刻の形状を多視点映像から推定していたが，その際に形状のもととなる３次元モデルを事前知識として用意しておき，これを各時刻の多視点映像と一致するように変形させることができるならば，結果としてモーションキャプチャを特別なマーカー無しで実現できたことになる [8][9]．その際の３次元モデルは通常の３次元ビデオの１フレームを用いることもあれば，レーザスキャンなどによって別途用意される場合もある．

■単視点計測　特に Kinect のような深度センサーが普及することに伴い，単視点であれば容易に深度マップ，すなわち 2.5 次元形状を得ることができるようになった．一方で高精細な人体の３次元スキャンデータセットが公開されるようになり，パラメトリックな人体表面形状データを作成することが可能となった [10]．この２つの背景から，単視点で得られた深度情報に対してパラメトリックな人体表面形状データをフィッティングすることで３次元ビデオを生成する研究が提案されるようになっている [11]．

■非同期撮影　従来の３次元ビデオ撮影では，多視点カメラ群が同期撮影することを前提としており，特に屋外で同期撮影を実現することは容易ではなかった．一方でスマートフォンやアクションカメラなどの普及により，屋外での多視点カメラ環境そのものは容易に実現されるようになったため，同期撮影を前提としない３次元ビデオ生成法が研究されている [12][13]．

3　今後の展望

以上のように，多視点カメラ群を備えたスタジオでの同期撮影から各時刻の３次元形状と表面テクスチャの推定という基本アルゴリズムからスタートした３次元ビデオ研究は，より運動情報に着目した復元のような新たな応用の提案であったり，単視点での復元や非同期撮影を前提とした屋外撮影などのような撮影環境に対する制約の緩和などへと展開してきた．

しかしながら半透明物体のように復元が困難な被写体や，屋外や夜間のように極端な照明環境，あるいはスポーツのような広範囲もしくは顕微鏡環境下のようなミクロな環境など，未開拓な研究の方向性は数多く残されている．今後はこのような領域への展開が期待される．

■謝辞　本研究の一部は科研費 26240023 の補助を受けて行った．

参考文献

[1] Matsuyama, T. et al.: *3D Video and Its Applications*, Springer (2012).

[2] Kanade, T. et al.: Virtualized Reality: Constructing Virtual Worlds from Real Scenes, *IEEE Multimedia*, pp. 34–47 (1997).

[3] Moezzi, S. et al.: Virtual View Generation for 3D Digital Video, *IEEE Multimedia*, pp. 18–26 (1997).

[4] Starck, J. et al.: The Multiple-Camera 3-D Production Studio, *TCSVT*, Vol. 19, No. 6, pp. 856 –869 (2009).

[5] Joo, H. et al.: Panoptic Studio: A Massively Multiview System for Social Motion Capture, *Proc. ICCV*, pp. 3334–3342 (2015).

[6] Tung, T. et al.: Complete multi-view reconstruction of dynamic scenes from probabilistic fusion of narrow and wide baseline stereo, *Proc. ICCV*, pp. 1709–1716 (2009).

[7] Nobuhara, S. et al.: A Real-Time View-Dependent Shape Optimization for High Quality Free-Viewpoint Rendering of 3D Video, *Proc. 3DV*, pp. 665–672 (2014).

[8] Nobuhara, S. and Matsuyama, T.: Deformable Mesh Model for Complex Multi-Object 3D Motion Estimation from Multi-Viewpoint Video, *Proc. 3DPVT*, pp. 264–271 (2006).

[9] Liu, Y. et al.: Markerless motion capture of interacting characters using multi-view image segmentation, *Proc. CVPR*, pp. 1249–1256 (2011).

[10] Pishchulin, L. et al.: Building Statistical Shape Spaces for 3D Human Modeling, *ArXiv* (2015).

[11] Bogo, F. et al.: Detailed Full-Body Reconstructions of Moving People from Monocular RGB-D Sequences, *Proc. ICCV*, pp. 2300–2308 (2015).

[12] Zheng, E. et al.: Sparse Dynamic 3D Reconstruction from Unsynchronized Videos, *Proc. ICCV*, pp. 4435–4443 (2015).

[13] Mustafa, A. et al.: General Dynamic Scene Reconstruction from Multiple View Video, *Proc. ICCV*, pp. 900–908 (2015).