

■2 群 (画像・音・言語) - 4 編 (3 次元画像)

4 章 処 理

【本章の構成】

本章では以下について解説する.

- 4-1 画像生成/レンダリング
- 4-2 処 理
- 4-3 圧縮・符号化
- 4-4 2D-3D 変換

■2 群-4 編-4 章

4-1 画像生成/レンダリング

■2 群-4 編-4 章

4-2 処 理

(執筆著者：山崎俊彦) [2012年8月 受領]

4-2-1 3次元シーン解析・加工，奥行き解析

(1) 3次元シーン解析

3D デジタルカメラ¹⁾や 3D カメラレコーダの発売により，3次元画像・映像は身近なものになった．一方で，これまで撮りためられた大量の画像・映像資産には，基本的に3次元情報は付与されていない．映画など，ごく一部のコンテンツについては手作業で3次元化がなされる場合もあるが，すべての2次元画像・映像に対して手作業で3次元情報を付与することは現実的でない．単視点の画像や映像化の奥行き情報を推定し，仮想的に3次元化する技術が幾つか提案されている．なお，カメラを動かしながら映像を撮影している場合には後述の Multi-View Stereo などの手法により3次元化が可能であるが，カメラの動きが一定の条件を満たしていないといけない，動物体の3次元化はできない，などの問題がある．

1枚の2次元の画像から3次元の奥行きを推定・表現する試みとしては Hoiem らによる“Automatic Photo Pop-up”が有名である²⁾．この研究では，スーパーピクセル（色や質感が似ているピクセルを集めた小領域）ごとに「地面」，「地面に対して垂直な物体」，「空」のいずれかの判定をし，幾何学的制約によりおおまかな3次元構造を推定することで3次元モデル化を行っている．また，文献3)のようにスーパーピクセル同士の奥行き関係を Markov Random Field (MRF) によって推定するものや，文献4)のように物体検出，領域分割，奥行き推定などの処理結果を統合して更に多段階学習を行うことでより高精度な奥行き推定を可能とした研究例もある．

■参考文献

- 1) 富士フイルム株式会社「FinePix REAL 3D W1」，
http://www.fujifilm.co.jp/corporate/news/articleffnr_0305.html
- 2) D. Hoiem, A.A. Efros, and M. Hebert: “Automatic Photo Pop-up,” ACM SIGGRAPH 2005.
- 3) A. Saxena, M. Sun, and A.Y. Ng: “Learning 3-d scene structure from a single still image,” IEEE ICCV, pp.1-8, 2007.
- 4) G. Heitz, S. Gould, A. Saxena, and D. Koller: “Cascaded Classification Models: Combining Models for Holistic Scene Understanding,” Neural Information Processing Systems (NIPS), 2008.

(2) 3次元シーン加工

正確な3次元情報を計測・表示するばかりではなく，より高い視覚効果を与えるために3次元情報に対して加工を行う研究も行われている．3次元シーンの加工は，2次元シーンの加工とは異なり，視差や奥行きなどの情報が3次元的に破綻しないように処理しなくてはならない．3次元映像の加工には，更に時間的な整合性も求められる．

(a) 奥行き変更・強調

生成された3次元モデルを別の映像に重畳したり，シーン中の3次元オブジェクトを違和感なく消去したり，奥行きを考慮しながら別の場所に貼り付けたり³⁾するオブジェクト編集がこれまで提案されている．

また、正確な3次元表示にこだわらず、Multi-Perspectiveの方法論を取り入れて、モデルの関節部分ごとに異なるカメラパラメータを用いて映像を合成することで、誇張表現を行うレンダリング手法⁶⁾も提案されている。

■参考文献

- 5) J. Xiao, X. Cao, and H. Foroosh : “3D Object Transfer Between Non-overlapping Videos,” IEEE Virtual Reality Conference, pp.127-134, 2006.
- 6) K. Utsugi, T. Naemura, T. Koike, and M. Oikawa : “Multiperspective Rendering for Anime-like Exaggeration of Joint Models,” ACM SIGGRAPH, Talks, Article no.29, 2011.

(b) 3次元リタargeティング (非線形リサイズ)

リタargeティング (非線形リサイズ) とは、主観の品質を保ちながら非線形に画像のサイズを変更する技術である。携帯端末やタブレット端末の普及により様々な解像度を持つディスプレイが登場しており、多様な解像度でも違和感なく表示するために必要な技術である。2次元画像・映像に対しては2000年台中頃から活発に研究が行われている。それに対し、視差情報を考慮した3次元リタargeティングも幾つか提案されている^{7)~9)}。

■参考文献

- 7) K. Utsugi, T. Shibahara, T. Koike, K. Takahashi, and T. Naemura : “Seam Carving for Stereo Images,” 3DTV Conference, 2010.
- 8) C.-H. Chang, C.-K. Liang, Y.-Y. Chuang : “Content-Aware Display Adaptation and Interactive Editing for Stereoscopic Images,” IEEE Transactions on Multimedia, vol.13, no.4, pp.589-601, Aug. 2011.
- 9) C. Birklbauer and O. Bimber : “Light-Field Retargeting,” Eurographics, 2012.

(c) その他の3次元シーン加工

ほかにも、推定した奥行き情報に対して高解像度で撮影された画像を適切にレンダリングすることでよりリアルな3次元シーンを表現する高品質・高精細化¹⁰⁾、映像中の3次元構造を理解することで時空間的に違和感のない視覚効果を与える研究¹¹⁾などがある。これらの技術を実現するためには、3次元物体の形状を正しく認識できるだけでなく、レンダリング視点位置を正確に把握していることが求められる。

■参考文献

- 10) P. Bhat, C.L. Zitnick, N. Snavely, A. Agarwala, M. Agrawala, B. Curless, M. Cohen, and S.B. Kang : “Using photographs to enhance videos of a static scene,” Eurographics Symposium on Rendering, pp.327-338, 2007.
- 11) G. Zhang, Z. Dong, J. Jia, L. Wan, T.-T. Wong, and H. Bao : “Refilming with depth-inferred videos,” IEEE Transactions on Visualization and Computer Graphics, vol.15, no.5, pp.828-840, 2009.

(3) 3次元奥行き・形状解析

本章では、物体に接触することなく、カメラやセンサなどを用いて3次元形状を計測する方法について述べる。なお、画像や映像に奥行き情報のみが付与された場合は、2.5Dデータ、もしくはRGB+D(depth)などと表現されることもあり、全周囲の情報を含んでいる3次元データと区別されることもある。

奥行きを計測するには、計測器側から被計測物体へ光・超音波など何らかのエネルギーを照

射して3次元物体があることによる反応の違いを計測する能動的な手法と、撮影した画像のみから得られる情報を用いて計算する受動的な手法とがある。

(a) レーザーレンジデータによる計測

レーザーを照射し、レーザーが被写体で反射されて戻ってくるまでの時間を計測することで距離を計測する能動的な方式である。レーザー光の高い指向性により非常に高精度（数 mm～数十 mm の誤差）、かつ広範囲な計測（数十 cm～数百 m）が可能である。ただし、ポイント計測を多数回繰り返して計測するため、計測に時間がかかる。そのため、動物体の計測には向かない。一般的に大型なものが多いが、近年では計測範囲を数 m に限った USB タイプのセンサも登場している。

(b) Time-of-Flight カメラによる計測

光などを照射し、被写体で反射されて戻ってくるまでの時間を計測することで距離を計測する能動的な方式である。原理的には上記レーザーレンジセンサと同じである。レーザーレンジセンサと比較して計測範囲が短い（数 m～数十 m）などの問題がある反面、面状に計測できるため動物体のリアルタイム計測が可能である。ただし、現状では解像度は QVGA 程度のもが多い。また、フェムト秒のパルス光を照射し、反射光をピコ秒の時間分解能を持つ超高速センサを用いて反射時間を計測することで、見通し範囲内にはない物体を計測する研究などもある¹²⁾。

赤外線を使うタイプのもは、原理上屋外の太陽光のもとでは安定した動作が難しい。超音波を用いたものもあるが、計測範囲や精度は光を用いたものに比べて劣ることが多い。

■参考文献

- 12) A. Velten, T. Willwacher, O. Gupta, A. Veeraraghavan, M.G. Bawendi, and R. Raskar : “Recovering Three Dimensional Shape around a Corner using Ultra-Fast Time-of-Flight Imaging,” Nature Communications, Mar. 2012.

(c) Structured Light による計測

被写体にある既知のパターンを照射し、それをカメラで撮影したときのパターンのゆがみから奥行きを計測する能動的な方式である^{13),14)}。複数のコードパターンを照射して計測する空間コード法、線状のパターンを照射しながらスキャンする光切断法などがある。これらの手法は計測に時間がかかる、またそのために動物体の計測に向かないという問題がある。高速に計測するための手法もいくつか提案されている¹⁵⁾。Microsoft 社の Kinect¹⁶⁾ は、赤外線を用いてあるランダムなパターンを照射することで高速な奥行き計測を実現している。

■参考文献

- 13) J. Batlle, E. Mouaddib, and J. Salvi : “Recent progress in coded structured light as a technique to solve the correspondence problem: a survey,” Pattern Recognition, 31(7), pp.963-982, 1998.
- 14) F. David, T. Sliwa, and Y. Voisin : “A Comparative Survey on Invisible Structured Light,” SPIE Electronic Imaging—Machine Vision Applications in Industrial Inspection XII. pp.90-97, 2004.
- 15) R. Sagawa, H. Kawasaki, R. Furukawa, and S. Kiyota : “Dense One-shot 3D Reconstruction by Detecting Continuous Regions with Parallel Line Projection,” IEEE ICCV, pp.1911-1918, 2011.
- 16) <http://www.xbox.com/ja-JP/kinect>

(d) ステレオ法による計測

ステレオ法とは、複数の視点から同一シーンを撮影し、視差情報から奥行きを推定する方法である。Middlebury Stereo Vision Page¹⁷⁾には標準データセットと、これまで提案されてきた数々の手法のリスト及び性能比較が掲載されているので参考にされたい。

また、照明を制御して明暗情報を測定することで3次元形状を計測する照度差ステレオ¹⁸⁾や、複数枚の画像を用いて物体の3次元形状をモデリングする多視点ステレオ法¹⁹⁾などの研究も進んでいる。多視点ステレオ法では、画像に含まれる特徴点同士のマッチングからカメラパラメータを推定することも可能で、大量の画像群を使って街全体を3次元モデル化する試みなども行われている²⁰⁾。

■参考文献

- 17) <http://vision.middlebury.edu/stereo/>
- 18) Y. Matsushita : ICIP 2010 Tutorial on Photometric Methods for 3-D Modeling, Sep. 26, 2010. Slide available at <http://research.microsoft.com/en-us/people/yasumat/>
- 19) S.M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski : “A Comparison and Evaluation of Multi-View Stereo Reconstruction Algorithms,” CVPR, 2006.
- 20) S. Agarwal, Y. Furukawa, N. Snavely, B. Curless, S.M. Seitz, and R. Szeliski : “Building Rome in a Day,” Communications of the ACM, vol.54, no.14, pp.105-112, Oct. 2011.

4-2-2 多眼・多焦点画像処理, 多眼・多焦点画像からの再構成

多眼・多焦点画像から新しい視点・焦点の画像を再構成する手法はイメージベーストレンダリング (Image-Based Rendering) と呼ばれている。多眼・多焦点画像処理の基本的な考え方は空間を満たす光線をすべて正しくサンプリングできれば、任意の視点の画像を再構成できるというものである。最も一般性の高い表現方法は、視点の座標 (x, y, z) 、光線の方向 (θ, ϕ) 、波長 λ 、時刻 t の7次元で空間を表現した Plenoptic Function²¹⁾ である。この7次元のデータを用いることで空間中のすべての光線を記述することが可能となる。しかし、この Plenoptic Function は膨大な情報を持ち、扱いが困難である。そのため、幾つかの制約条件をつけることで次元を縮退させ、現実的に解くことのできる問題に落とし込むのが一般的である。どのような制約条件をつけるかによって多種多様な手法が提案されている。詳細についてはサーベイ論文^{21)~24)}を参照されたい。

Image-Based Rendering の手法のなかでも、静的シーンのみを扱い (時刻 t を省略)、波長成分と光の減衰や干渉を無視することで4次元の形式として光線を表現した Light-Field Rendering^{25),26)} や Lumigraph²⁷⁾ は実用性の高さで特に有名である。Light-Field Rendering^{25),26)} は、撮影後に焦点を変えることのできるカメラとして2012年に米国 Lytro 社によって製品化されている。また、この4次元の表現形式は国内では藤井ら^{29),30)}によって文献^{25)~27)}に先だって発表されている。

■参考文献

- 21) E.H. Adelson and J.R. Bergen : “The plenoptic function and the elements of early vision,” Computation Models of Visual Processing, M. Landy and J.A. Movshon, eds., MIT Press, Cambridge, pp.3-20, 1991.
- 22) H.-Y. Shum, S.B. Kang, and S.-C. Chan : “Survey of image-based representations and compression techniques,” IEEE Transactions on Circuits and Systems for Video Technology, vol.13, no.11, pp.1020-1037, Nov. 2003.

- 23) C. Zhang and T. Chen : “A Survey on Image-Based Rendering - Representation, Sampling and Compression,” EURASIP Signal Processing: Image Communication, pp.1-28, vol.19, no.1, Jan. 2004.
- 24) S.C. Chan, H.-Y. Shum, and K.-T. Ng : “Image-Based Rendering and Synthesis,” IEEE Signal Processing Magazine, vol.24, no.6, pp.22-33, Nov. 2007.
- 25) M. Levoy and P. Hanrahan, “Light Field Rendering,” ACM SIGGRAPH, pp.31-42, 1996.
- 26) R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan : “Light Field Photography with a Hand-Held Plenoptic Camera,” Stanford University Computer Science Tech Report CSTR 2005-02.
- 27) S.J. Gortler, R. Grzeszczuk, R. Szeliski, and M.F. Cohen : “The Lumigraph,” Proc. ACM SIGGRAPH, pp.43-54, 1996.
- 28) <http://www.lytro.com/>
- 29) T. Fujii and H. Harashima : “Coding of an Autostereoscopic 3-D Image Sequence,” SPIE Visual Communications and Image Processing (VCIP), vol.2308, pp.930-941, 1994.
- 30) 藤井俊彰 : “3次元統合画像符号化の基礎検討,” 東京大学工学系研究科博士論文, 1994. Available at: <http://repository.dl.itc.u-tokyo.ac.jp/dspace/bitstream/2261/1866/1/111140.pdf>

4-2-3 メッシュポリゴン生成・圧縮・符号化・検索

(1) メッシュポリゴン生成

静止3次元物体については4-2-1節に述べた手法で計測・モデリングが可能なので、ここでは動的3次元物体のモデリングについて述べる。多視点映像から動物体を3次元モデル化する試みは1990年代半ばから研究が活発になされている^{31),32)}。一般的には、多視点映像から動的3次元モデルを生成するためには、すべてのカメラ視点から見えている3次元空間上の点(ボクセル)のみを残すという手法がとられる。これは、Shape from Silhouette, Volume Intersection, 視体積交差法などと呼ばれている。この手法は、原理上凸形状しかモデリングすることができない。凹部分は多視点ステレオ法やグラフカットなど別の手法で追加的に計算しなくてはならない。

このようにして生成された3次元ボクセルモデルは、表現とレンダリングの効率化のためにマーチングキューブ法³³⁾などを用いて3次元メッシュモデルとして表現されることが多い。生成される動的3次元映像のメッシュの結線情報はできるだけ時間的に一定であるほうが符号化効率の観点から都合がよい。また、結線情報が時間的に一定であれば各頂点の動き追跡ができていくことになり、頂点テクスチャレンダリングなどにも都合がよい。そこで、前フレームのメッシュを変形し、現フレームのシルエットと一致させるようにすることで時間的に一貫性のある結線情報を持った3次元メッシュモデルを生成する研究も盛んに行われている^{34),35)}。

■参考文献

- 31) T. Kanade, P. Rander, and P. Narayanan : “Virtualized reality: constructing virtual worlds from real scenes,” IEEE Multimedia, vol.4, no.1, pp.34-47, Jan./Mar. 1997.
- 32) E. Stoykova, A.A. Alatan, P. Benzie, N. Grammalidis, S. Malassiotis, J. Ostermann, S. Piekh, V. Sainov, C. Theobalt, T. Thevar, and X. Zabulis : “3-D Time-Varying Scene Capture Technologies – A Survey,” IEEE Transactions on Circuits and Systems for Video Technology, vol.17, no.11, pp.1568-1586, Nov. 2007.
- 33) W.E. Lorensen and H.E. Cline : “Marching cubes: A high resolution 3D surface construction algorithm,” ACM SIGGRAPH, pp.163-169, 1987.
- 34) J. Starck and A. Hilton : “Surface capture for performance-based animation,” IEEE CGA, vol.27, no.3, pp.21-31, 2007.
- 35) E. de Aguiar, C. Stoll, C. Theobalt, N. Ahmed, H.P. Seidel, and S. Thrum : “Performance capture from sparse multi-view stereo,” ACM SIGGRAPH 2008, #98, 2008.

(2) メッシュポリゴンの圧縮・符号化

(a) 静止3次元モデル(実写3次元映像)の圧縮・符号化

静止3次元モデルの符号化は1990年代頃から研究されている。3次元モデルのほとんどはポリゴンメッシュモデルとして記述されており、ポリゴンメッシュモデルの最も基本となるデータは頂点の幾何情報と頂点同士の結線情報である。

また、ポリゴンが規則的に記述されている場合、結線情報から幾何情報はある程度予測できるため、予測との誤差情報だけ符号化しておけば幾何情報の復号は比較的簡単とされている。そのため、初期の圧縮・符号化の研究では主に結線情報の圧縮に重点を置き、幾何情報は付加情報として扱われていた。ポリゴンをリンゴの皮をむいていくように展開して可逆圧縮する方法³⁶⁾、Progressive Mesh³⁷⁾のように段階的にメッシュ表現を簡素化して圧縮する方法などがある。

また、幾何情報の圧縮に関する研究例もある。文献39)では、3次元モデルを2次元平面に展開して頂点の幾何情報を2次元画像のRGB値として表現することで従来の2次元画像符号化技術を応用して幾何情報を圧縮している。2次元画像・映像圧縮で用いられているDCTと同様にラプラシアン行列を用いてメッシュの周波数主成分分析を行い、高周波成分を切り捨てることで効率的な符号化を行う試みもなされている⁴⁰⁾。

静止3次元モデルの圧縮については既に幾つものサーベイ論文41)~45)も発表されているので参照されたい。

■参考文献

- 36) G. Taubin and J. Rossignac : "Geometric compression through topological surgery," ACM Transactions on Graphics, 17(2), pp.84-115, 1998.
- 37) H. Hoppe : "Progressive Meshes," ACM SIGGRAPH, pp.99-108, 1996.
- 38) X. Gu, S. Gortler, and H. Hoppe : "Geometry images," ACM Trans. Graphics (SIGGRAPH), 21(3), 2002.
- 39) Z. Karni and C. Gotsman : "Spectral compression of mesh geometry," ACM SIGGRAPH, pp.279-286, 2000.
- 40) H. Briceño, P. Sander, L. McMillan, S. Gortler, and H. Hoppe : "Geometry videos: A new representation for 3D animations," Symposium on Computer Animation 2003, 136-146, 2003.
- 41) C. Gotsman, S. Gumhold, and L. Kobbelt : "Simplification and Compression of 3-D Meshes," Tutorials on Multiresolution in Geometric Modelling, Springer, 2002.
- 42) P. Alliez and C. Gotsman : "Recent Advances in Compression of 3-D Meshes," Symp. Multiresolution in Geometric Modeling, Sep. 2003.
- 43) J.L. Peng, C.S. Kim, and C.-C.J. Kuo : "Technologies for 3D Mesh Compression: A Survey," Journal of Visual Communication and Image Representation, vol.16, pp.688-733, 2005.
- 44) A. Smolic, K. Müller, N. Stefanoski, J. Ostermann, A. Gotchev, G.B. Akar, G. Triantafyllidis, and A. Koz : "Coding algorithms for 3DTV - A survey," IEEE Transactions on Circuits and Systems for Video Technology, vol.17, 1606-1621, 2007.
- 45) A. Smolic, R. Sondershaus, N. Stefanoski, L. Váša, K. Müller, J. Ostermann, and T. Wiegand : "A Survey on Coding of Static and Dynamic 3D Meshes," Three-Dimensional Television Signals and Communication Technology, pp.239-311, 2008.

(b) 動的3次元モデルの圧縮・符号化

動的3次元モデルのうち、CG技術で作られたシーケンスは全フレームに渡って結線情報が一定に保たれているものが多い。その場合、幾何情報の時間的変化のみを符号化すればよく、文献45)のサーベイ論文でもいくつか手法が紹介されている。一方、多視点映像から生成され

る3次元映像も理想的には結線情報は時間的に変化しないことが望ましく、4-2-3節(1)で述べたように結線情報をなるべく保存して動的3次元モデルを生成する研究が行われている。しかし、各フレーム独立に3次元モデルが生成される場合や、位相が変化するなどして結線情報を変更せざるを得ない場合も存在する。

結線情報が時間的に変化する3次元モデルを圧縮・符号化する方法としてはGeometry Images³⁶⁾を時間方向にも拡張して2次元映像圧縮手法を適用しようとする方法^{46),47)}や、2次元映像で用いられているブロックマッチング法を3次元に拡張した方法^{48),49)}などが提案されている。

■参考文献

- 46) H. Habe, Y. Katsura, and T. Matsuyama : “Skin-off: representation and compression scheme for 3D video,” PCS, pp.301-306, 2004.
- 47) D.Thi P. Quynh, Y. He, X.M. Chen, J. Xia, Q. Sun, and S.C.H. Hoi : “Modeling 3D articulated motions with conformal geometry videos CGVs,” ACM Multimedia, pp.383-392, 2011.
- 48) S. Han, T. Yamasaki, and K. Aizawa : “Time-varying Mesh Compression Using Extended Block Matching Algorithm,” IEEE Transactions on Circuit and Systems for Video Technology, vol.17, no.11, pp.1506-1518, Nov. 2007.
- 49) T. Yamasaki and K. Aizawa : “Patch-based compression for time-varying meshes,” IEEE ICIP, pp.2945-2948, 2010.

(3) メッシュポリゴンの検索

3次元モデリングツールの普及に伴い、多種多様な3次元モデルが入手可能になった。また、SketchUp⁵⁰⁾などのようにユーザが作成したモデルを公開し、共有するサービスも存在する。大量の3次元モデルの生成・入手が可能となると、それらを如何に効率的に検索するかが課題となる。PrincetonのShape Benchmarkなど、検索用の標準データセットも幾つか公開されている。

3次元モデルの検索は、モデルの大きさ、回転、視点の違い、ノイズなどに対してロバストであることが望ましい。全体一致のみを扱うのか、部分一致も扱うのか、などの違いによっても必要とされる技術が異なる。3次元モデルのどのような特徴を扱って検索するかによって、例えば

1. 3次元モデルから頂点の空間分布や曲率の分布、3次元 SIFT 特徴量などを抽出し、特徴ベクトル空間で検索する方法
2. 3次元モデルをグラフ表現に変換し、グラフマッチングで検索する方法
3. 3次元オブジェクトのシルエットなど3次元モデルを画像に変換して検索する方法

などが提案されている。静止3次元モデルの検索については既に多数のサーベイ論文が発表されているので、それらを参照されたい^{52)~57)}。

一方、動的3次元モデルの検索は入手可能なデータセットが制限されているなどの理由でまだ一般的でない。これまで、別の動的3次元モデルをクエリとして、特徴ベクトル空間で検索する手法⁵⁸⁾、グラフベースで検索する方法⁵⁹⁾、モーションキャプチャデータ・2次元動画像・加速度センサなどをクエリとしてマルチモーダルに検索する方法⁶⁰⁾、などが提案されている。また、3次元形状、ならびにその時間変化が似ているシーケンスを検索することで、映像同士をつなぎ合わせて新たなシーケンスを生成する試みもなされている^{61),62)}。検索技術の応用例として、単に似た形状を検索するだけでなく、動的3次元モデルの時間的・空間的類似度を検索して似たシーケンス同士をつなげることで撮影時にはなかったシーケンスを生成する試み

どもなされている^{61),62)}.

■参考文献

- 50) <http://www.sketchup.com/intl/en/index.html>
- 51) <http://shape.cs.princeton.edu/benchmark/>
- 52) P. Shilane, P. Min, M. Kazhdan, and T. Funkhouser : “The Princeton Shape Benchmark,” Shape Modeling International, 2004
- 53) J.W.H. Tangelder and R.C. Veltkamp : “A survey of content based 3D shape retrieval methods,” Shape Modeling Applications, pp.145-156, 2004.
- 54) B. Bustos, D. Keim, D. Saupe, and T. Schreck : “Content-Based 3D Object Retrieval,” IEEE Computer Graphics and Applications, vol.27, no.4, pp.22-27, Jul./Aug. 2007.
- 55) Z. Qin, J. Jia, and J. Qin : “Content based 3D model retrieval: A survey,” International Workshop on Content-Based Multimedia Indexing, pp.249-256, 2008.
- 56) J.W.H. Tangelder and R.C. Veltkamp : “A survey of content based 3D shape retrieval methods,” Multimedia Tools and Applications, 39, pp.441-471, 2008.
- 57) 大淵竜太郎 : “3次元形状の検索,” 講座：第7回 マルチメディア検索の最先端, 映像情報メディア学会誌, 2010年7月号, pp.93-98, 2010.
- 58) T. Yamasaki and K. Aizawa : “Motion Segmentation and Retrieval for 3D Video Based on Modified Shape Distribution,” EURASIP Journal on Applied Signal Processing, vol.2007, Article ID 59535, 11 pages, 2007.
- 59) 唯野隆一, 山崎俊彦, 相澤清晴 : “スケルトンを用いた Time-Varying Mesh からの動き抽出と類似動作検索,” 映像情報メディア学会誌, vol.62, no.5, pp.729-742, May, 2008.
- 60) D. Kasai, T. Yamasaki, and K. Aizawa : “Retrieval of Time-Varying Mesh and Motion Capture Data Using 2D Video Queries Based on Silhouette Shape Descriptors,” IEEE ICME, pp.854-857, 2009.
- 61) J. Xu, T. Yamasaki, and K. Aizawa : “Motion Editing for Time-Varying Mesh,” EURASIP Journal on Advances in Signal Processing Volume 2009, Article ID 592812, 9 pages, doi:10.1155/2009/592812, 2009.
- 62) P. Huang, C. Budd, and A. Hilton : “Global temporal registration of multiple non-rigid surface sequences,” CVPR, pp.3473-3480, 2011.
- 63) G.G. Slabaugh, W.B. Culbertson, T. Malzbender, R.W. Schafer : “A survey of methods for volumetric scene reconstruction from photographs, Volume Graphics, 2001.

■2 群-4 編-4 章

4-3 圧縮・符号化

■2 群-4 編-4 章

4-4 2D-3D 変換

(執筆著者：山田邦男) [2011年7月 受領]

2D-3D 変換とは、既存の 2D の画像（以下、動画映像を包含する）を立体視が可能な画像に変換する手段の総称である。その意味する範囲は非常に広汎であり、ハリウッド映画の制作において 1 フレームずつ人手で奥行きをつけていく作業が“2D-3D 変換にて制作”と言われる一方で、民生用の 3D Ready のテレビにおいて 2D の番組をリアルタイムで変換して擬似ステレオで観視可能にする付加機能も“2D-3D 変換機能”と呼ばれる。これらの例からも分かるように、2D-3D 変換に関しては、そのキーワードとしての分かりやすさが災いして、本来同列に行うべきでないような比較や議論がなされたり、処理そのものに過剰な期待や忌避が存在することも否めない状況にある。

本節では、3D 画像処理技術としての 2D-3D 変換技術を核に、分類や使用例などについて整理して説明する。また、この分野においては、具体的手法についての報告が内容的に充分でない、もしくは過去の文献や特許で提示された手法がその後どの程度実施されたか確認が困難であるようなケースが多いので、実用例の紹介として筆者らによる変換手法のアルゴリズムやアプリケーションについて記す。

4-4-1 2D-3D 変換の目指すもの

2D-3D 変換の究極の目標は 2D 画像をもとにシーンの立体的構造を正確に再現することである。手作業で人間が判断した奥行きをつけていくような 2D-3D 変換については 3D 画像処理技術と言うよりも編集作業のための環境の話になってしまうため本節では触れない。以下はリアルタイム・非リアルタイム、事後修正の有無を問わず、2D 画像を入力すると自動的に 3D 画像（ステレオ画像、視差データ、奥行きデータなど）が生成されるような処理システムについて述べるものとする。

このような処理において、動画の場合はオブジェクトの速度などによって前後関係の推測を行ったり、静止画の場合は何らかの意味でのシーン内容の推定を行うことにより奥行きを決めたりといった処理を行うが、これらは不良設定問題であるため単純な処理での解決は難しい。コンピュータビジョンの知見を有効利用することにより一定の奥行き推定精度の向上は可能であるが、究極的には画像のシーン内容の意味理解を行う必要があり、処理工数が膨大になる上にどのようなシーンに対しても有効性を担保するための完成度の向上へのハードルが極めて高い。

現状の 2D-3D 変換においては、処理能力やメモリといったリソースに応じて実際の奥行き精度向上を目指しつつもエラーの回避は不可避であり、そのような状況にあっても破綻や違和感が少なく、できるだけ立体感や奥行き感に富む、見て心地よい (Visually Pleasing) 立体画像の生成が模索されている。これはコンピュータビジョン的であると同時に画像表現的アプローチでもある。

4-4-2 2D-3D 変換の代用品としてではない積極的利用

2D-3D 変換の技術的なターゲット以前の話として、2D-3D 変換が行われる主な動機は 3D コ

コンテンツの供給であり、これは過去の立体ブームと言われたものがコンテンツの不足を主な原因として終焉したとされることに対する反省に基づくものとされる。これはネイティブな 3D の代用品としての 2D-3D 変換という考えであるが、これとは別により積極的な目的に基づく 2D-3D 変換の利用も行われている。

それは幾つかの撮影条件に関するものである。まずは長距離のズームングに関するものである。2 眼カメラで遠方の被写体を撮影する場合、両チャンネルの視差がほとんどつかずに立体感に欠けることが多いが、2D-3D 変換により好結果を得られることがある。同様な理由で 2D-3D 変換は空撮などにも向いている。また、逆に 2 眼カメラで至近の被写体を撮影する場合、視差がつき過ぎるもしくは被写体が一方のカメラでしかカバーできないという状況がある。このような被写体に対して 2D-3D 変換を適用することにより好結果を得られることがある。これに類するものとして、格闘技のリングのロープが 2 眼撮影では視差がつき過ぎることがあるという問題も 2D-3D 変換によって解決可能と言われる。

更に、2D-3D 変換がコントローラブルな 3D 画像生成技術であることを活かして、安全性や疲労に応じて立体感の強度や輻輳を調整するという考え方も非常に有効である。

4-4-3 2D-3D 変換の分類

ここでは 2D-3D 変換技術について、用途、仕様、原理などといった複数の見地から分類を行いながら同技術をめぐる諸事情を紹介する。

(1) 業務用か民生用か

多くの場合、業務用途における 2D-3D 変換はコンテンツ制作のためのツールであり、民生用途における 2D-3D 変換はコンテンツの楽しみ方の幅を広げる付加機能である。よって、業務用途においては出力画像の品質の責任を制作者が負うので、あるシーンでは非常に効果的であるがあるシーンでは強烈に破綻してしまうような変換機でも構わない。一方、民生用途では、ある程度の調整が可能であるにしろ、出力画像の品質は基本的に機器が継続的に担保する必要がある。

また、近年のデジタル家電やスマートフォンのような民生用途における付加機能として用いられる 2D-3D 変換の場合、SoC、システム IC への実装が前提になるので、アルゴリズムの計算コスト、メモリなどの必要なリソースについても配慮が必要である。

(2) リアルタイムか非リアルタイムか

もともと非リアルタイム変換は主に静止画処理と制作に使用されることが多い。事実上非リアルタイムでも十分な用途であってもリアルタイム変換機が使用されることも多いが、リアルタイム変換においては変換エラーの修正ができないことが前提になる。このため、エラーの発生量の縮減や目に付きやすいエラーの回避のための努力が必要である。また、継続的にコンテンツ観視を行う場合の急激な変換性能の変化は強い違和感を与えることがあり、ベストケースの品質向上よりワーストケースの品質向上のほうが、全体の印象の改善のためには効果的なことがある。

(3) 高奥行き精度か非高奥行き精度か

明示的に精度向上の努力を放棄することは通常ありえないとして、困難ななかでも出来る限りそれぞれのシーンオブジェクトの奥行きを推定を努力するアプローチが存在する一方、シーン全体にほぼ適正と考えられる奥行きを与えるようなアプローチも存在する。これらの明快な

分類は困難であるが、上記のように正確さを追求することは処理工数が膨大な画像理解に通じるといふこともあり、リソースに応じてそれぞれ工夫している状態にある。

なお、対象によってはかなり精度の低い奥行き精度から作成したステレオ画像でも視覚的に十分な品質を有するケースが見られる。このあたりは呈示される空間知覚の手がかりが必ずしも幾何学的に正しい必要はないとの見解¹⁾を裏付けるものかもしれない。

更に、技術的な精度と全く違う次元において、現実と異なる奥行き構造を呈示せざるを得ないケースもまま存在する。それはいわゆる飛び出し画像である。通常の2Dコンテンツはいわば窓の向こうの世界を捉えたものである。それを3D化する場合は一部のものを除いて多くのオブジェクトを奥側に配置するのが正解になる。しかし“3D=飛び出し画像”という認識は根強く、事実上そのような変換が求められる場合もある。

(4) 動き利用かフレーム内処理かハイブリッドか

ここでは技術的な手段による分類を行う。

水平方向への単純な動きをしているオブジェクトが存在する動画シーケンスからシーンの前後関係をうまく見出して3D化を行うという考えは順当のものであり、三洋電機から多く動き利用の手法の特許提案がなされている。この手法では動きが停止したときに奥行き情報が得られない、複雑な動きに適さない、動きによっては前後が入れ替わるという問題がある。

フレーム内処理では画像のテクスチャのみから奥行き推定を行うため静止画の変換が可能である。マーキュリーシステムの手法²⁾が知られるほか、後記のJVC方式もこの手法である。

動き利用の方式の静止画で奥行き情報が失われたときにフレーム内処理に切り替えるのがハイブリッド方式で、三洋電機がLSI開発も行っている³⁾。更に、最近の東芝の方式⁴⁾では動き抽出の精度向上やコンピュータビジョン系の知見を組み合わせた高度な処理へと発展させている。

各タイプについて、ほかにも膨大な数の手法が提案されているが、具体的な記載は、ここでは上記の広く知られた国内のものにとどめる。なお、符号化における動きベクトルを奥行き推定に利用するアプローチもありえるが、符号化における動きベクトルは残差処理をミニマムにするものであり、必ずしも実際のオブジェクトの動きの反映を補償するものでないで注意が必要である。

4-4-4 2D-3D変換アルゴリズムの実例 (JVC方式)

本稿では筆者らが手がけた2D-3D変換技術について簡単に紹介する^{5),6)}。本手法では下記の条件をクリアすることを目標とした。

- (a) 本当のステレオ撮影 (ネイティブ3D) に迫る自然で迫力のある立体感を出来るだけ多くのタイプの画像において実現する。
- (b) ワorstケースについても“いまひとつ立体感に欠ける”もしくは“少しだけ不自然な部分があった”程度のレベルでとどめることとし、極端な破綻にはならないようにする。
- (c) フルハイビジョン以上の解像度のリアルタイム変換を一般的なプロセッサにおいて可能にする軽い演算量。
- (d) 低遅延と低コストを実現するフィールドフレームメモリの不使用。

図4・1にJVC方式のアルゴリズムの概要を示す。全体の処理は極めて単純で、まず入力画像の絵柄に対応するシーンが単純な曲面構造で近似され、曲面に画像のテクスチャの色の赤・青

に基づく凹凸をエンボス状に加味される。ここまでの奥行き構造の推定であり、このエンボスつき曲面をもとにして入力画像をシフトすることによりステレオペアが構成される。ここでは入力画像（原画）を右チャンネル、シフト画像を左チャンネルとしている。

曲面近似においては、入力画像の上下部分のテクスチャの高域成分を評価して、3タイプの基本奥行きモデルと呼ぶ標準曲面のブレンドを行う。

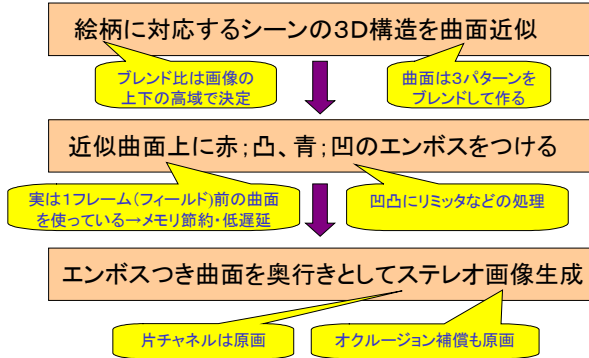


図 4・1 JVC 方式の概要

図 4・2、図 4・3 に奥行きモデルとそれが適用されるシーンの例を示す。図 4・2 のような球面構造の基本奥行きモデル Type 1 は典型的な遠近法的シーンまたはシーン構造判定不能の場合を想定し、図 4・3 のような上半分が円筒構造、下半分が球面構造の基本奥行きモデル Type 2 は遠近法的シーンでより空が開けたシーンを想定し、図 4・4 のような上半分が平面、下半分が曲線スロープの基本奥行きモデル Type 3 は遠景と地水面からなるようなシーンを想定している。

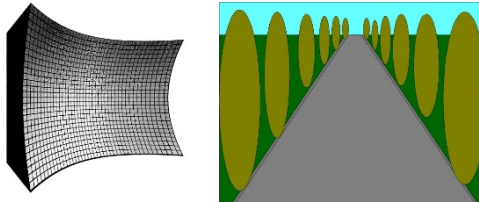


図 4・2 基本奥行きモデル Type 1 と適用シーン

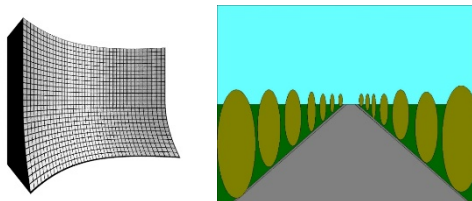


図 4・3 基本奥行きモデル Type 2 と適用シーン

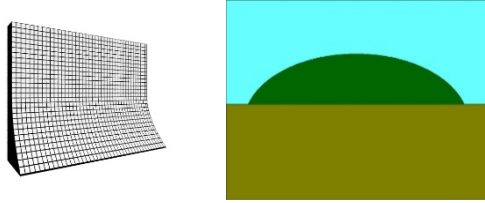


図 4・4 基本奥行きモデル Type 2 と適用シーン

テクスチャの高域による単純な判定であるため、想定通りのシーン構造が選択・構築されるわけではないが、シーンチェンジに応じた奥行き感の変化を実現している。

上記曲面に赤(R)信号が手前、青(B)が奥になるようなエンボスを施す。R信号が強い部分を手前にする理由の一つは、R信号の大きさが、順光に近い環境でかつテクスチャの明度が大きく異ならないような条件において、被写体の凹凸と一致する確率が高いという経験則によるものである。B信号が強い部分を奥にする理由の一つは、光線の散乱により遠くにあるものほど、青寄りに見えるという物理法則的一般論（空気遠近法）によるものである。更にもう一つの理由として、暖色系は色彩学における前進色であり、寒色系（後退色）よりも奥行きが手前に認識されるという特徴があり、この認識傾向を実際の奥行きに反映することにより立体感を強調している。また、この赤の飛び出しについては所定の値の上限を定めている。また、最近では赤青以外の種々の判定基準も導入している。なお、エンボスは実際にはフレーム相関を利用して1フレーム前の曲面に施している。これによってブレンド比決定結果を保持するフレームメモリが不要になり、処理による遅延を非常に小さいものにすることができる。

本アルゴリズムはハードウェア実装がなされ、IPのライセンスビジネスのほか、業務用変換機として商品化され⁷⁾、映画制作、スポーツ・イベント中継などの実績を重ねている。更に、民生向けカムコーダにも実装されている⁸⁾。

4-4-5 まとめ

2D-3D変換をとりまく諸事情と実用化の一例を紹介した。2D-3D変換はRecognitionの側面とGenerationの側面を具備するゆえの難しさと自由さが混在すると言うことができる。3D映像や3DTVの帰趨と絡み今後の展開が注目される。

■参考文献

- 1) 金子：“空間知覚の基礎,” 映情誌, vol.58, no.7, pp.909-914, 2004.
- 2) 特開 2000-228778
- 3) 森, 村田, 山下, 前中, 岡田：“新方式 2D/3D 変換マルチパーパス LSI,” SANYO Technical Review, vol.30, no.2, pp.27-35, 1998.
- 4) 三田：“2Dto3D コンテンツ変換技術,” 2010 年映情年次大会, S2-3.
- 5) 山田：“ハイビジョン対応リアルタイム 2D/3D 変換装置,” 信学技報, IDY2006-82, HI2006-122, 2006.
- 6) K. YAMADA and Y. SUZUKI：“Real-Time 2D-to-3D Conversion: Basic Concepts and Practical Use,” (Invited), International Display Workshops, 3D3/DES3-2, 2010.
- 7) <http://www.jvc-victor.co.jp/pro/monitor/imageprocessor/lf-2d3d1/index.html>
- 8) <http://www.jvc-victor.co.jp/dvmain/gz-hm990/index.html>