



解説

実写ベースの コンピュータ グラフィクス技術

新谷 幹夫 杉山 和弘
NTTサイバースペース研究所

コンピュータグラフィクスによる映像が身近になるにつれ、リアルなCGモデルを多数、効率よく作成することが求められるようになった。しかし、モデル作成は熟練者の手作業に頼っている状況で、効率化が立ち遅れている。このモデリングボトルネックを解消するため、実写からCGモデルやCG映像を作成する技術が盛んに研究されている。本稿では、これら実写ベースのコンピュータグラフィクス技術を解説し、今後の展望について述べる。

なぜ実写ベースか

近年のコンピュータグラフィクス(CG)技術の発展により、映画、コマーシャルはもとよりホームページなどでもCG映像は重要な役割を果たすようになってきた。CG映像の作成には、物体の形状や色、動きなどを定義するモデリング過程、CGモデルから画像を生成するレンダリング過程、の2つのプロセスが必要である。このうち、レンダリングは比較的自動化が容易であり、計算機の処理能力や記憶容量の増大の恩恵を受け、複雑なモデルから写実的な画像を容易に生成することが可能となった。

しかし一方、モデリングは「創造、造形」のプロセスで、自動化は容易でない。このため、実際のモデリング作業は造形能力を持つ熟練者の手作業に頼らざるを得ない状況である。しかも、レンダリングの進展により、より複雑なモデルを多数、迅速に作成することが求められるようになり、モデリングがCG映像作成のボトルネックとなっている。すなわち、モデリング作業の効率化、大衆化がCGの最重要課題であるといえる。

モデリング環境のヒューマンインタフェース向上はモデリング効率化に重要である。しかしこれが、大衆化につながるとは限らない。筆者自身の経験では、スケッチを書く程度の空間把握能力や表現能力がないと、どのような方法にしろ、良い「形」を作り出すことは困難であるような気がしてならない。

さて、実際に存在している物体をあるがままに「記録」するのであれば、必ずしも「高度な」造形は必要ない。むしろ、より機械的な「複製」作業と見なすこともできる。このような作業であれば、自動化が可能である。精度的な意味では、人手を介すより有利にすらなり得る。このような観点から、モデリングボトルネック解消のエースとして登場したのが、実写ベースのCG技法である。この分野はCG、CV、映像処理などの境界領域であるが、CGサイドからのレビューはあまり見かけない。そこで以下、実写ベースCG手法開発の経緯と技術の概要をCGの立場から解説する。この分野の研究開発を進めようとお考えの方々の一助になれば幸いである。

実写とCG研究

実写利用の推移

1970年代には隠れ面処理、曲面理論などCGの基本的な方法論が確立し、1980年代以降、写実的な画

像をいかに作るかという点に研究の焦点が集められた。このような中で、Blinnは実写を物体の表面に張り付けることで写実性を飛躍的に向上できることを示した。これが、テクスチャマッピング法であり、恐らく実写を用いる最初のCG手法だと思われる。図-1を見れば、テクスチャの威力を理解いただけよう。

次に注目を集めた実写CGはモーフィング手法である。モーフィング手法は、画像を変形させたり、画素値を変化させたりすることで、たとえば男性歌手の顔が女性の顔に滑らかに変化したり、犬やブタがしゃべるように口を動かすというような映像を合成する。むしろ、映像作成現場が先行して作品を作り始め、1990年くらいから研究が盛んに行われるようになった。動画作成では必要不可欠の技術として定着した。

情景の「記述」という観点で実写の威力を強烈に印象づけたのは、1995年のQuickTime VR¹⁾ 1) であろう。写真なのに、まるで3次元CGのように視線方向をくるくる対話的に変化できる。実写ならではの迫力と対話性の結合にインパクトがあった。これに触発され、実写ベース手法の研究は一躍脚光を浴びるようになった。翌年、光線空間法がStanford大²⁾、Microsoft³⁾、および東大³⁾から発表され、Image-based renderingが登場した。

一方、画像列から情景の3次元構造を復元する問題は、コンピュータビジョンにおける歴史的な研究課題であり、ステレオ法や動画像解析など、いわゆるShape from Xなる研究がきわめて多数発表されている。この流れの中で、CGへの応用は有望な新分野と見なされるようになり、精力的に研究が行われるようになった。特に、画像列からCGモデルや照明モデルを作成する研究が大変活発である。

技術の分類

入力画像から何を作り出すかという観点からみると、図-2に示すように研究を分類することができる。まず、画像列から新たな画像を直接生成する技術があげられる。モーフィング、QuickTime VR、光線空間法などがこのカテゴリーに入り、Image-based renderingと呼ばれる。広義には(正確な距離画像を伴った)CG画像を対象とする技術も含まれる。これ

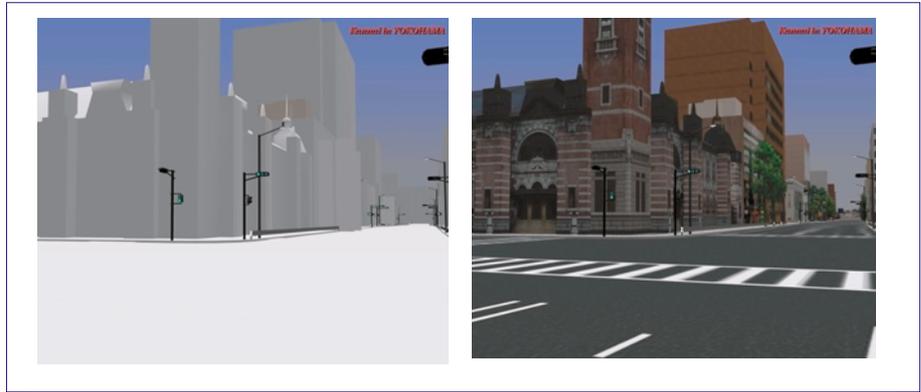


図-1 テクスチャマップの一例。左図がテクスチャなし、右図がテクスチャあり。
(NTT堀口賞一氏提供)

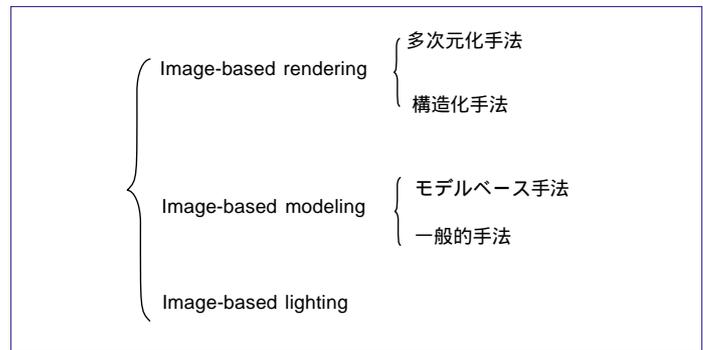


図-2 実写ベースのCG技術

に対し、画像列から3次元CGモデルを作成する手法はImage-based modelingと呼ばれている。ステレオ法や動画像解析の応用はこの範疇に入る。また、もう1つの分野として、画像から反射分布や照明モデルを抽出するものがある。テクスチャマップはこの分野のはしりと見なすこともできよう。あまり、名前を付けて呼ばれることはないようだが、本稿では他の2つにならい、Image-based lightingと呼ぶことにしよう。

画像から画像へ

ほとんどのCGシステムでは、3次元CGモデルに対して仮想カメラの位置や照明条件を指定し、画像を算出する。Image-based renderingでは、複数の撮影条件(主に視点)で撮影された画像から、他の撮影条件で得られるであろう画像を「補間」により作り出す。与えられる画像が密な場合は、画像系列を視点、視線などにより多次元化し、補間フィルタを適用することで合成が実現できる。一方、少数の画像しか与えられていない場合には、より組織的に画像を変形させることが必要となる。このため、画像間の対応関係を指定する必要がある。前者のアプローチ

¹ QuickTimeはApple Computer, Inc.の登録商標。

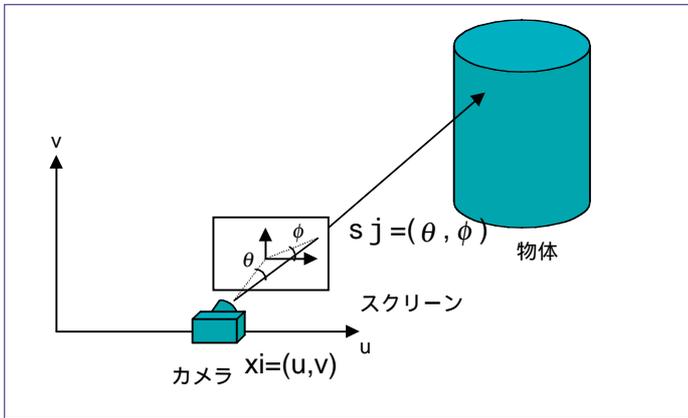


図-3 カメラモデルと光線

子を高次元化アプローチ,後者を構造化アプローチと呼ぶことにしよう。高次元化アプローチでは,処理が容易かつ安定であるのが利点であるが,撮影に手間や工夫が必要な点,データ量が大きくなる点などが問題となる。他方,構造化アプローチでは,画像間の対応決める処理に決定打がなく,「弁慶の泣き所」となっている。

を?

多次元化アプローチ

ピンホールカメラモデルで考えると,画像中の各

画素は視点を通る,ある方向の光線の強度(色)を記録している。図-3に示すように,視点の位置を x_i ,光線の方向を s_j とすれば,画像列は $\{(x_i, s_j)\}$ と表せる。これは,連続関数 $f(x, s)$ の標本と見なせる。すなわち,画像生成処理は,標本 $\{f(x_i, s_i)\}$ から任意視点・任意方向の画像 $f(x, s)$ を再構成する処理であると定式化される。ここで, $f(x, s)$ は x を通過し,方向が s である光線の強度を意味するので,これを光線空間⁴⁾(light field²⁾,Lumigraph³⁾)と呼ぶ。視点の位置 x の自由度に応じて,以下のような機能が実現できる。

(1)2D - 全周囲写真

視点の位置を固定し,さまざまな方向の写真撮ると,方向 (θ, ϕ) 2自由度の光線空間が構築できる。光学的アナロジーとしては全周囲写真である。この「全周囲写真」を視軸方向に投影することで,任意方向の画像が生成される。QuickTime VR¹⁾がこの実現例である。視点に関する自由度がないので,視点を動かすことはできない。

(2)3D - レンチキュラー画像

カメラを直線上あるいは円周上に移動して撮影すると,視点の位置 u で1自由度,方向 (θ, ϕ) で2自由

度,合計3自由度の光線空間ができあがる。視点の軌跡上では任意方向の画像を生成できる^{5),6)}。画像の例を図-4に示す。光学的アナロジーはレンチキュラー画像で,横方向の視差はあるが,縦方向の視差は記録されていない。したがって,視点が軌跡から大きく外れる場合には物体の距離情報が必要となる。



入力画像の一例



再現画像

図-4 光線空間法による処理例

(NTT植本尚子氏提供)

(3)4D - ホログラフィ

カメラを面上に移動して撮影すれば,視点位置 (u, v) で2自由度,方向 (θ, ϕ) で2自由度,合わせて4自由度の光線空間が得られる。面と交差する任意の光線が表現されるので,この面から覗き込んで見られるすべての画像を生成できる。視点の移動も,面の内側に入らない範囲で自由にできる。いわば,ホログラフィである。しかし,生の4次元デ

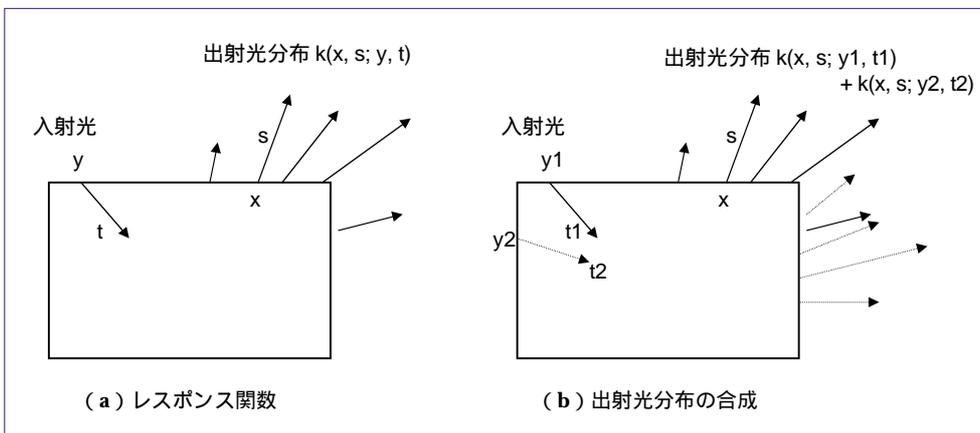


図-5 レスポンス関数と出射光分布の合成

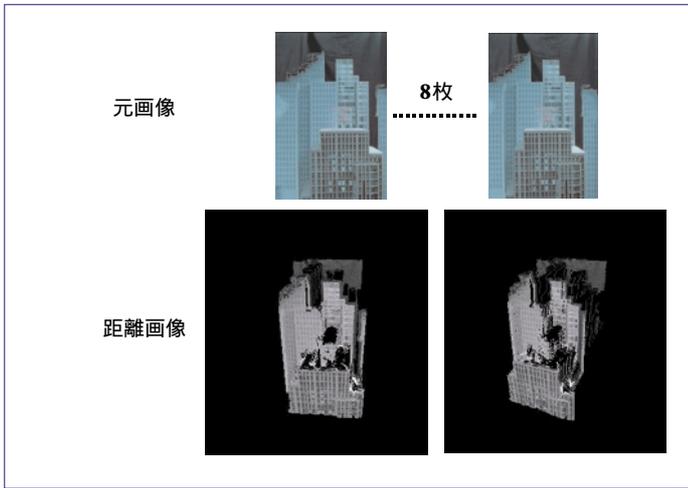


図-8 マルチベースライン法の処理例
(NTT山田正紀氏提供)

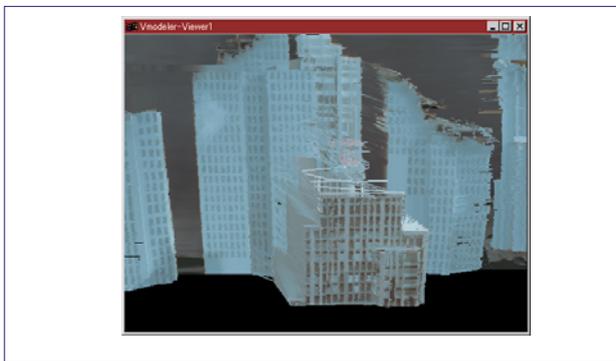


図-9 EPI解析処理例
(NTT南田幸紀氏提供)

視変換の不変量である cross-ratio を利用した手法¹⁰⁾ や消失点を利用した手法¹¹⁾ も知られている。

画像から形状へ

CGでは、アニメーション作成のために力学シミュレーションを行ったり、照明シミュレーションを行ったり、ほかのCGモデルや映像と合成を行ったりすることが多い。このような場合には、前述の Image-based rendering では不十分で、やはり形状とテクスチャを持ったCGモデルが必要となる。特定のモデルを仮定するものと、一般的な投影幾何的性質を利用する手法とがある。

モデルベースの手法

ビルディング、樹木、顔など、対象が決まっている場合は、対象形状の特性を考慮して形状復元を行うことが有効である。

(1) 建造物

建造物の概観は直方体や角錐などのプリミティブ

を用いて、比較的容易に作ることができる。そこで、これらプリミティブで作成したラフなCGモデルを、複数の写真と照合することでリファインしてゆく手法がパークレー大¹²⁾から提案されている。元となるCGモデルは手作業で作成し、写真のエッジとモデルのエッジの対応をいくつか与える。モデルの投影像と写真の差が最小となるようにプリミティブのパラメータ(直方体の辺の長さなど)を調整し、モデルを改良してゆく。比較的単純な形状に対しては強力な手法である。最小化プロセスの安定性などが明確化されることを期待する。

(2) 樹木

樹木の形状は、L-systemに代表されるような成長規則で記述することが有効である。坂口らは、複数の樹木の写真から成長規則を抽出することを試み、良好な結果を得ている¹³⁾。

(3) 顔

人の顔は千差万別とはいうが、基本的な構造はさほど変わらない。そこで、昔から「標準」モデルを顔写真にフィッティングさせる手法などが研究され、最近では、顔形状データベースから写真に似た顔を探してきて、写真に合わせて変形させることで顔モデルを生成する手法が提案されている¹⁴⁾。

一般的手法

CV分野で開発された形状復元の諸技法は、汎用的で、自律ロボットや物体認識などさまざまな応用が可能である。しかし、CGへの応用では、生成画像の出来栄え、すなわち「見てくれ」が最重要項目となる。特に、物体の見え隠れ(オクルージョン)は視覚的に重用である。このためには、情景を構成する面の復元、入力画像中のオクルージョンの対処、が特に重要であるといえる。この観点からみると、以下の手法が有利であると思われる。

(1) マルチベースラインステレオ法

位置と姿勢が既知の複数のカメラから得られた画像を用いる。基準画像の各画素において、他のすべての画像における投影画素との類似度の和が最大になるように、奥行き値を決定する。このため、一部の画像で他の物体に隠されていても、残りの画像で現れていれば、奥行きが正しく抽出され得る。また、密な奥行き地図が得られるので、面の再構成に有利である。モデルの作成例を図-8に示す。運動物体を対象とすることも可能で、3次元映像への応用が期待されている¹⁵⁾。

(2) EPI解析

カメラの視軸を固定して真横に動かすと、像点も真横に移動し、その移動速度は奥行きに反比例する。

この性質を利用して形状復元する手法をEPI (Epipolar Plane Image) 解析法と呼ぶ。すべての像点が真横に動くので、横方向の接続を求めやすいのが利点である。また、物体が他の物体に隠れていく過程を解析できるので、オクルージョンの処理が容易である。欠点としては、カメラの移動に関して強い制限があることだが、カメラ姿勢の推定などにより改善が可能であると思われる。図-9に処理例を示す¹⁶⁾。

このほか、因子分解法¹⁷⁾によるシステムや自己キャリブレーション法¹⁸⁾によるシステムが開発中であり、前述、がどのような形で解決されるのが注目している。

画像から光源・反射率を

写真の画素の値は、物体の反射率、物体に照射する光の強度、により決まる。すなわち、正確なCGモデルを作成するには、写真の画素値からとを分離・抽出することが望ましい。まず、このための基本ツールとして、多シャッター速度法が開発された。この手法は、写真やCCDのダイナミックレンジが発光体と反射体の輝度差に比べて著しく狭い点を補う。輝度に対する感度がシャッター速度で異なることを利用し、異なったシャッター速度で複数の写真を撮影し輝度に変換することで、高ダイナミックレンジの輝度分布を得ることができる¹⁹⁾。複数の既知の視点から得られた高ダイナミックレンジ画像と形状モデルが与えられれば、反復的手法で物体の反射特性、発光体の輝度が計算できることが知られている²⁰⁾。

画像から物体の形状、反射率、照明条件を抽出することは人間にとっても難しい²¹⁾。実際、我々はパースペクティブや照明条件の矛盾には、かなり寛大であるように思われる(たとえば、CG画像作成で見た目の「写実性」を高めるため、影を付けない光源や、特定の物体しか照射しない光源など、物理的に存在し得ないものを頻繁に使う)。照明・反射特性をどの程度復元すれば、どれだけの視覚的効果が高まるのか、この分野の重要なポイントであると思う。

今後の展望

light field rendering専用ハードウェアが開発されたり、image-based modelingのプロトタイプが展示会などに展覧されたり、実用化もそう遠くはないと

思われる。レーザー距離計測器も廉価になり、GPSなどの高精度化も見込まれ、現実世界と仮想世界との垣根はますます低くなる。3次元地図など、大規模なCG応用が盛んになるものと考えられる。今後、CGコンテンツの作成が容易になり、だれでもホームページなどで気楽に発信できるようになることを期待する。また、映像がCG化されるようになれば、映像と映像、映像とCGがシームレスで合成されるようになる。さらに、3次元化された映像がネットなどで送信できるようになれば、3次元ネットテレビが実現できる。たとえば相撲中継を土俵の中から観戦する、ということも不可能ではなくなる。見たいときに、見たい所が見たい方向から見れる、新たな映像時代が始まろうとしている。ホログラフィ技術など3Dディスプレイデバイスが進歩すれば、究極の3次元映像が楽しめるようになるであろう。

参考文献

- 1) Chen, S. E.: QuickTime VR - An Image-based Approach to Virtual Environment Navigation, Proc. Siggraph'95, pp.29-38 (1995).
- 2) Levoy, M. and Hanrahan, P.: Light Field Rendering, Proc. Siggraph'96, pp.31-42(1996).
- 3) Gortler, S., Grzeszczuk, R., Szeliski, R. and Cohen, M.: The Lumigraph, Proc. Siggraph'96, pp.43-54(1996).
- 4) 柳澤, 苗村, 金子, 原島: 光線空間を用いた3次元物体の操作, テレビ誌, Vol.50, No.9, pp.1345-1351(1996).
- 5) Shum, H-Y. and He, L-W.: Rendering with Concentric Mosaics, Proc. of Siggraph'99, pp.299-306(1999).
- 6) Uemoto, N., Horiguti, S., Nagai, S. and Sugiyama, K.: A Ray-Based Scenery Rendering Method Using Geometric Information of Objects, Proc of UM'99, pp.41-48(1999).
- 7) Seits, S. and Dyer, C.: View Morphing, Proc. Siggraph'96, pp.21-30(1996).
- 8) 向川, 中村, 太田: 2枚の顔写真を用いた任意方向の顔画像の生成, 情報処理学会論文誌, Vol.37, No.4, pp.635-644(Apr. 1996).
- 9) Avidan, S. and Shashua, A.: Novel View Synthesis by Cascading Trilinear Tensors, IEEE Trans. on Graphics, Vol.4, No.4, pp.293-306(1998).
- 10) Havaldar, P., Lee, M-S. and Medioni, G.: View Synthesis from Unregistered 2-D Images, Proc. of Graphics Interface'96, pp.61-69(1996).
- 11) Horry, Y., Anjyo, K. and Arai, K.: Tour Into the Picture: Using a Spidery Mesh Interface to Make Animation from a Single Image, Proc. of Siggraph'97, pp.225-232(1997).
- 12) Debevec, P., Taylor, C. and Malik, J.: Modeling and Rendering Architecture from Photographs: A Hybrid Geometry- and Image-based Approach, Proc. Siggraph'96, pp.11-20(1996).
- 13) 坂口, 大谷, 中津: 実写映像に基づいた三次元樹木モデルの生成, 電子情報通信学会J82-D-II, No.9, pp.1469-1477(1999).
- 14) Blanz, V. and Vetter, T.: A Morphable Model for the Synthesis of 3D Faces, Proc. of Siggraph'99, pp.187-194(1999).
- 15) Narayanan, P. and Kanade, T.: Constructing Virtual Worlds Using Dense Stereo, Proc. of ICCV'98, pp.3-10(1998).
- 16) Shinya, M., Saito, T., Mori T. and Osumi, N.: VR Models from Epipolar Images: An Approach to Minimize Errors in Synthesized Images, in LNCS 1352 Computer Vision-ACCV '98, Vol.II, Springer, pp.471-478(1998).
- 17) Tomasi, C. and Kanade, T.: Shape and Motion from Image Streams Under Orthography: A Factorization Method, Int'l J. Computer Vision, Vol.9, No.2, pp.137-154(1992).
- 18) Xu, G. and Zhang, Z.: Epipolar Geometry in Stereo, Motion and Object Recognition, Kluwer Academic Publishers, Dordrecht(1996).
- 19) Debevec, P. and Malik, J.: Recovering High Dynamic Range Radiance Maps from Photographs, Proc. of Siggraph'97, pp.369-378(1997).
- 20) Yu, Y., Debevec, P., Malik, J. and Hawkins, T.: Inverse Global Illumination: Recovering Reflectance Models of Real Scenes from Photographs, Proc. of Siggraph'99, pp.215-224(1999).
- 21) Nishida, S. and Shinya, M.: Use of Image-based Information in Judgements of Surface-reflectance Properties, J. Opt. Soc. Am. A, Vol.15, No.12(1998).

(平成11年11月25日受付)

- : ハイフン?
- : ダーシ?