

## 解説

## コンピュータ・グラフィックスの基本的テクノロジーと最近の話題

## 3. 画像とCGの融合

A Fusion of Image and Graphics by Kazunori MIYATA, Masato KUROKAWA (Tokyo Research Laboratory, IBM Japan) and Gentarou HIROTA (Departments of Computer Science, University of North Carolina at Chapel Hill).

宮田 一 乗<sup>1</sup> 黒川 雅 人<sup>1</sup> 広田 源太郎<sup>2</sup>

<sup>1</sup> 日本アイ・ピー・エム(株)東京基礎研究所

<sup>2</sup> ノースカロライナ大学計算機科学科

## 1. はじめに

今、手元に、某アーティストのエンハンスド CD がある。この CD には、従来の音楽トラックのほかに、アーティストのプロモーションビデオや、過去のアルバム収録曲の一覧表などが入っており、PC で再生して楽しむことができる。

マルチメディアという言葉を目にするようになって久しくなるが、この CD のようなメディアの多様化にともない、実際のコンテンツ作成は複雑になるばかりである。この CD をぱっとみただけでも、簡単なオープニングのアニメーション、立体的なメニューやカラフルなグラフィックスを使ったボタン、デジタルムービーなどのコンテンツがある。一方、今までのパッケージ型のメディアとは対称的に、ネットワークによるマルチメディア情報の配信では、コンテンツの複雑さに加えて、情報提供のスピードも重要な要素である。したがって、手軽に見栄えのあるコンテンツが作成できるかどうかは鍵になってくる。

本稿では、マルチメディアのコンテンツ作成に重点を置いて、実画像に CG で作成された画像をモンタージュする手法や、画像認識を伴ったヘアスタイルのシミュレーション手法などの具体例を報告する。

## 2. デジタル画像加工の現状

筆者らが、PC の上でソフトウェアだけで再生される動画をみたのは、Macintosh 上で QuickTime を使って再生される、数秒ほどの猿の動画であった。今と比較したら、メモリ容量は低く CPU も非力であったため、小さいウィンドウの中でごちなく動くパラパラ漫画のようであった

が、非常に衝撃的であった。以後、ハードウェアの発展にともない、動画の再生だけでなく、編集作業まで PC でできるようになってきた。

本章では、デジタル編集の現状と、映像分野における CG の役割について解説する。

## 2.1 リニア編集からノンリニア編集へ

最近の記憶媒体の大容量かつ高速化、および JPEG や MPEG などの画像圧縮技術の一般化にともない、放送用の映像製作の現場では、従来のテープ主体のリニア編集システムから、テープ上の映像をデジタル化して一度ディスクに格納し、それらのデジタル画像をランダムアクセスして、時間軸上で編集するノンリニア編集システムへのパラダイムシフトがみられる。とくに、最近の特撮映画では、CG で作成した映像と実写映像との合成を行っていることもあり、ハリウッドにおけるデジタル映像編集への関心は、非常に高い。

8 年前に Macintosh 上の Avid 1 で始まったノンリニア編集システムは、その速さと汎用性において、シーケンシャルメディアであるテープを使った従来のオフライン編集に勝っており、その利用範囲を飛躍的に広めた。これらのノンリニア編集システムは、当初は、オフラインでの編集のみを目的としたシステムだったが、その役割も徐々に変化してきている。今やハイエンドの放送やポストプロダクション用のシステムでは、非圧縮（もしくは低圧縮率）のデジタル画像を高速なディスクアレイに格納し、画面のディゾルブ処理などの種々の視覚効果を行って映像を加工した後、その出力がそのまま最終結果となるような利用も可能となっている。

この傾向は、映像だけでなく、音楽の世界でも

そっくりあてはまる。従来テープとミキサで行っていた編集作業が、ハードディスクに蓄積したデジタルの音データを、PCのソフトで編集するようになってきている。しかも、一部の演奏パートを、コンピュータを使ってMIDI音源で再生させることも可能である。

一方、放送の分野とは別の形で発展してきたQuickTimeやVideo for WindowsのようなPC上でのデジタル映像は、当初は、質やフレームレートなどの点で問題があったが、圧縮技術の進歩とともに、CD-ROMなどのパッケージメディアやネットワーク上のコンテンツなど、ローエンドの映像の利用分野としての地位を固めた。

これらの映像を作成する際のノンリニア編集システムでは、エフェクトの種類や機能面では、放送の分野で使われているシステムには及ばないが、デスクトップで簡単な動画コンテンツを作成するには十分な機能をもったものが数多く出回っている。今や、ユーザは、自分の目的や仕上がりの映像の質に応じて、ハイエンドからローエンドまでの品揃えの中から好きなものを選択できる環境になってきている。

このように、専用機をもった特別なスタジオのみで可能であった映像編集の作業が、映像データをデジタル化することにより、自宅のPC上でも編集が可能になってきている。近い将来には、ビデオカメラで撮影した自分の子供たちの映像をPCで編集してナレーションを入れ、できあがった成長記録をDVDに焼いて、子供の結婚祝いにプレゼントする親が出てくることだろう。

## 2.2 映像分野におけるCGの役割

映像分野におけるCGの役割は、2つに大別される。

1つは、従来のエフェクト装置がもっているような、映像の特殊効果としてのCG技術である。これは、2枚の画像を融合させるモーフィングなどの2次元の画像変換や、シーンの切替え手法であるディゾルブ処理、文字を3次元空間内で移動・表示させるようなグラフィック処理などを含む機能であり、リアルタイムでの処理が要求される分野である。

最近の天気予報や選挙速報番組などで多用されているバーチャルセットなどは、広い意味でこの最先端の技術といえる。天気予報の例では、青い

背景のセットを前にアナウンサーが天気予報を読み上げるシーンをビデオカメラで撮影しながら、ハイエンドのグラフィクス・ワークステーションで生成された日本地図などの3次元グラフィクスの映像をリアルタイムでクロマキー合成して、その撮影現場から直接放送している。

もう一方は、コンテンツ作成に用いられる素材としてのCGの技術である。

CG画像の一番の特色として、現実には存在し得ないものを映像化することがあげられる。たとえば、「ジュラシックパーク」では、標本でしかみることのできない恐竜をスクリーン上に甦らせ、「アポロ13」では、今となってはカメラでとらえることのできないロケットの打ち上げシーンを再現している。一昔前の特撮映画では、ロケットの模型や恐竜のぬいぐるみを作成して撮影していたものを、すべてコンピュータの中に構築された仮想空間の中で仮想カメラを通して撮影できるようになっている。これらの映画で使用されたCGの画像そのものの現実感はいうまでもなく、特筆すべきことは、俳優の動きや映画の舞台の背景などの実世界（実写画像）と、まったく違和感なく合成されている点である。

このように、CG画像と実写画像とを違和感なく合成するためには、実写画像を撮影したときのカメラの位置や向き、照明条件などを、CG画像を作成するときのパラメータとして与える必要がある。

これらの実世界とCGの仮想世界を結びつけるために必要なパラメータ取得の方法として、以下の章では画像解析を用いたいくつかの例を紹介する。

## 3. 自然画像とグラフィクスの合成

本章では、カメラから入力された実写画像（静止画像および動画）とCGで作られたオブジェクトを、自然にみえるような形で容易に合成することを目的とした例を紹介する。

この目的を解決するために考慮しなければならない条件として、少なくとも以下の項目があげられる。

1. CGと実写画像のカメラパラメータ（視点およびビューボリュームなど）の統合
2. 両者の照明条件の統合

3. 実写中のオブジェクトとCGによるオブジェクトの前後関係から生じるオクルージョンの適切な表現

4. 合成部分境界における不自然さの解消

これらの条件は、どれも容易に解決可能な問題ではない。

ここでは、そのための第一歩として、上記の条件の1.に焦点を絞り、簡単なマンマシンインタラクションと画像解析により、実写画像中の主要な平面の3次元構造を推定して、実写画像とCGのオブジェクトとを組み合わせた例を紹介する。

### 3.1 静止画像の場合

最初に1枚の静止画像から3次元構造を求める方法を述べる。ここでは対象となる動画像内のオブジェクトの形状に仮定を設けている。

実際には、ユーザが実写画像中で、実世界では矩形と思われる部分（以降、リファレンスレクタングルと呼ぶ。これは剛体かつ動かないという仮定をおく）を指定し、これを手がかりにカメラパラメータを算出する。

透視変換を仮定したカメラ座標系において、画像面上で与えられたリファレンスレクタングルが3次元上で矩形である条件、およびスケージングファクタとして、レクタングルのある一辺の実測値を与え、解析的に各点の3次元座標を求め、カメラの位置および方向の情報を得る。

カメラの内部パラメータとしては、実測によって得られた焦点距離と画角を使用し、撮影時にズームは使用していない。図-1に、静止画像上にCGで生成した戦闘機のオブジェクトを重ね合わせた例を示す。戦闘機の下に表示された四角が、リファレンスレクタングルである。

### 3.2 動画像の場合

上記の構造推定法を動画像に適用しようとした場合、動画像の各フレームにおけるリファレンスレクタングルの座標を求めることが必要になる。ここでは、ユーザが最初の1フレームに関して、リファレンスレクタングルを指定し、以降のフレームでは動き推定法を用いてリファレンスレクタングルの軌跡を追跡する。

ここで用いる動き推定法は2段階よりなる。最初に実写画像中の連続する2フレームにおける、各画素単位の動きを階層的な勾配法<sup>1)</sup>によって求める。

次に得られた画素単位の動きを、affine motion model<sup>2)</sup>で統合し、前フレームのリファレンスレクタングルに適用することによって、次フレームにおけるリファレンスレクタングルの位置を得ている。

図-2に動画像に適用した例を示す。図-2の(a)-(c)が実写画像であり、(d)-(f)が合成画像である。この実験では、45フレームからなる実写画像を用いている。各フレームは、三脚に固定されたカメラを移動させて撮影したものであり、カメラが右から左へむかって歩道上を横切っている。戦闘機が移動するCGをキーフレームアニメーションで作成し、リファレンスレクタングル上の1点を原点として、推定したカメラパラメータを用いて画像上に重ね、アニメーションを作成した。また、地面以外の実写中のオブジェクトの3次元位置はわからないため、オクルージョンは考慮されていない。

## 4. ヘアスタイルのシミュレーション

年頃の女性にかぎらず、ほとんどの女性は、おしゃれに対する関心が強い。とくにヘアスタイルは、本人の印象を大きく変える要因であるため、ファッション雑誌の記事の中でも、一番の注目を集めている。本稿で述べる手法は、このような女性（ばかりではないが）の変身願望を叶えるための、ヘアスタイルの自動合成に関するものである。

### 4.1 システム構成

システムのハードウェアの構成は、PCMCIAスロットを備えたPCおよび、デジタル・スチル・カメラの2つである。カメラに取り込まれた被験者の画像は、メモリカードに蓄えられ、PCMCIAスロットを介して、PCに転送される。

システムのソフトウェアの構成は、以下の4つからなり、取り込まれた被験者の顔を自動認識させて、データベースに蓄えられているヘアスタイル画像を被験者の顔に合成する。

- 顔画像の取り込みモジュール
- 顔画像の認識モジュール
- ヘアスタイル画像の合成モジュール
- ヘアスタイルのデータベース

### 4.2 顔画像の自動認識法

処理の対象となる人物画像（以降、顔画像と呼

ぶ)は、2次元のカラー画像であり、その撮影条件は以下のとおりである。

- 正面から見た1人の上半身が写っている。
- 髪が眉や目にかからないように頭の後ろで束ね、額を露出した状態で撮影されている。

この条件で撮影された顔画像から、顔領域(首を含む)と顔部品(目、口、鼻、顎、頭頂点)を自動的に抽出する手法について、以下に述べる。抽出されたデータは、次段の合成処理において、髪画像を補正する際に用いられる。

#### 4.2.1 顔領域の抽出処理

最初に、画像の領域分割処理<sup>3)</sup>に基づいて、顔をひとかたまりの肌色の領域として抽出する。た

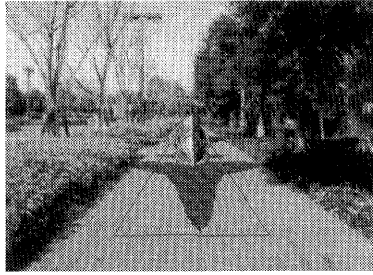


図-1 静止画像の合成例

だし、領域分割処理の結果は、必ずしも求めたい顔領域とは一致しないので、領域分割の結果得られた領域を組み合わせ、大局的に判断して顔と思われる領域を作成する必要がある。

本手法では、顔を構成する領域のもつ形状や色に条件を設定し、その条件を満たすまで領域分割を繰り返す方法を用いている。

#### 4.2.2 色差を用いた顔部品候補の抽出

求められた顔領域の中から、目および口の候補となる領域を、HSV表色系における色差とエッジ密度を参照して、抽出する。ここでエッジ密度とは、原画像にLoGフィルタ<sup>4)</sup>をかけ、エッジとみなされた画素の候補領域内における面積比である。

#### 4.2.3 ルール集合を用いた顔部品の決定

求められた顔部品候補領域に対し、それらの相互的な配置関係から、まず両目と口の組の候補を選択する。次にその候補から顔の中心線を決定し、画像のエッジ強度を参照して、鼻および顎の点を決定する。

以上の手順で求められた自動認識の結果の例を、図-3にあげる。

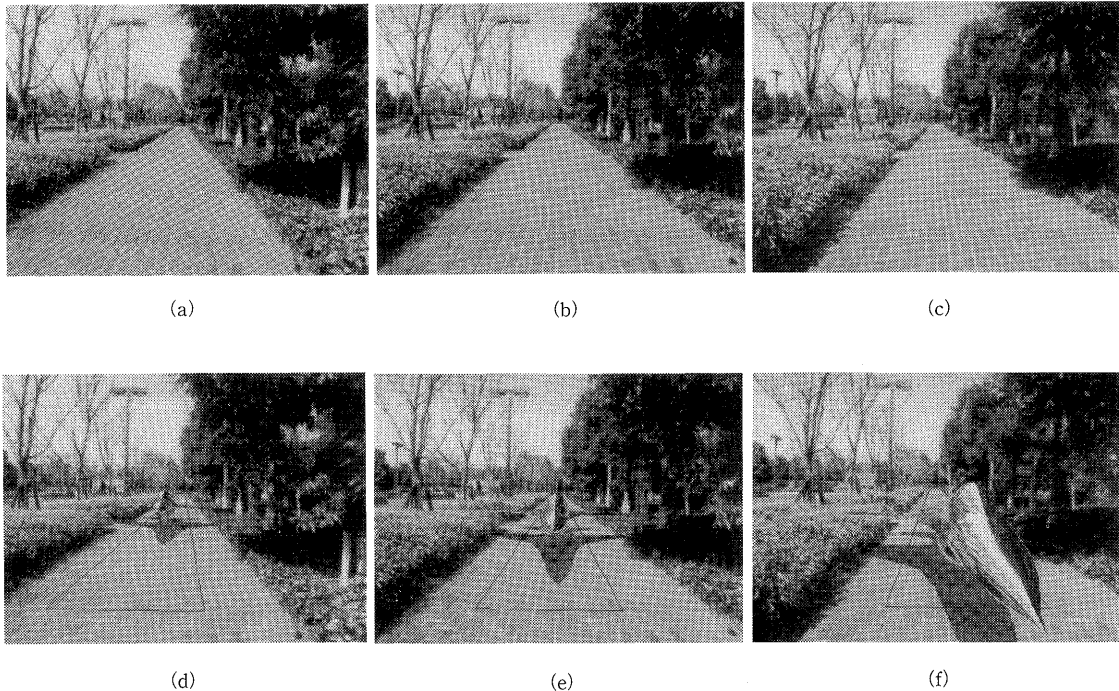


図-2

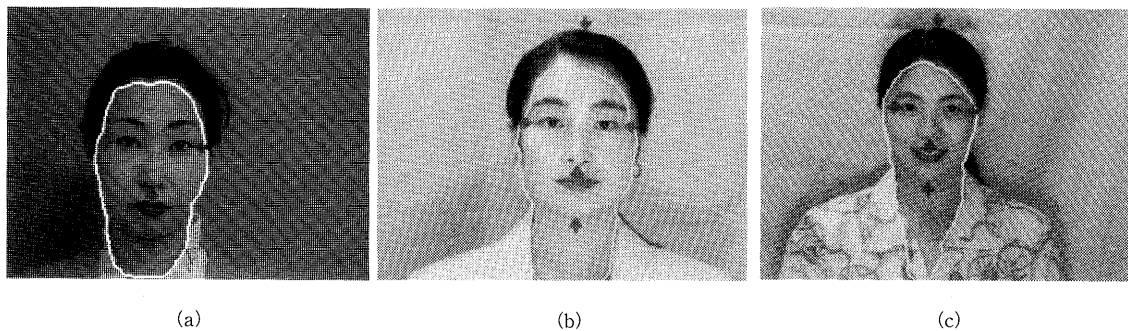


図-3 認識結果の例



図-4 ヘアスタイルの合成例

### 4.3 ヘアスタイルの合成法

自動認識された顔画像の特徴データを用いて、以下に述べるような処理手順に従い、入力された顔画像にヘアスタイルの画像を合成する。

#### 4.3.1 ヘアスタイルのデータベース

ヘアスタイルのデータベースの素材となる画像は、ヘアスタイルのモデルの上半身を、正面から撮影したものを用意する。この素材の画像を、髪の領域を残して背景を青の単色で塗りつぶした画像に加工して、ヘアスタイルの画像として蓄えておく。また、素材の画像の顔の特徴データも、合わせてデータベースに蓄えておく。

#### 4.3.2 ヘアスタイル画像の補正

まず、自動認識された顔の特徴データから、顔の幅、傾き、中心の位置などを計算する。これらの結果を用いて、ヘアスタイル画像を、入力された顔画像の大きさ、位置、および顔の傾きに合うように補正する。

次に、顔の輪郭の個体差による髪のかかり具合の違いを補正するために、画像のワーピング処理<sup>9)</sup>を施して、ヘアスタイル画像の顔輪郭が顔画像の顔の輪郭にフィットするように、ヘアスタイル画像の変形操作を施す。

#### 4.3.3 ヘアスタイル画像の合成

最後に、変形されたヘアスタイル画像の背景部

分(青色の領域)を切り抜き、合成時のエッジ領域における不自然なぎざぎざを除くために、ヘアスタイル画像を顔画像にアルファ・ブレンディング<sup>9)</sup>して画像の合成処理を終了する。

#### 4.4 シミュレーション結果

実際に、本稿で述べた手法でヘアスタイルを合成した結果を、図-4に数例あげる。

以上で述べた手法を用いることにより、短時間で自然な仕上がりのヘアスタイルを合成することが可能になる。1つのヘアスタイルあたり、5秒から20秒程度の処理時間(CPUにIntel Pentium 75 MHzを使用時)で合成できるため、たとえば、美容院の数分の待ち時間の中に、客は好みのヘアスタイルを10パターン以上試すことが可能である。

#### 5. おわりに

今、女子高校生を中心に大流行している「プリント倶楽部」も、自分の手によるマルチメディアのコンテンツ作成の例だと考えられる。このように、インターネットを含めたマルチメディアの技術の発展にともない、メディアの権限は、徐々に消費者の側に移行してきている。

マルチメディアの基礎であるCGの技術は、日進月歩で進歩はしているが、実写画像のリアリティには、到底かなわないし、表現の幅も狭い。したがって、何もかもCGで作成するのではなく、CGでできるものはCGで、できないものは実写映像でと、両者の棲み分けを行い、両者を融合させてコンテンツを作成するのが自然であると考えられるし、そのほうが遙かに質の高い映像が作成できる。実際に、映画の世界では、実写とCGの映像をまったく違和感なく融合させる技術が盛んに取り入れられている。

近い将来には、CGで作成した俳優が活躍するアクション映画が、話題をさらうことだろう。この場合のアカデミー賞の受賞式には、果たして誰が参列するのだろうか？

(本稿については、以下のアドレスも御参照ください。

<http://w3.trl.ibm.com/w3/exthome/projects/>

s7340/cgimage/cgi.htm)

#### 参考文献

- 1) Horn, B. K. P. and Shunk, B. G.: Determining Optical Flow, Artificial Intelligence, Vol. 17, pp. 185-203 (1981).
- 2) Wang, Y. A. and Adelson, E. H.: Layered Representation for Motion Analysis, CVPR '93, pp. 361-366 (1993).
- 3) 長尾 真: 画像認識論, コロナ社 (1983).
- 4) Marr, D.: ビジョン—視覚の計算理論と脳内表現, 産業図書 (1987).
- 5) Beier, T. and Neely, S.: Feature-Based Image Metamorphosis, Computer Graphics, Vol. 26, No. 2, pp. 35-40 (1992).
- 6) Foley, J. D. et al.: Computer Graphics—principles and Practice, pp. 835-840, Addison Wesley (1992).

(平成8年10月31日受付)



宮田 一乗 (正会員)

1962年生。1984年東北大学工学部応用物理学卒業。1986年東京工業大学大学院総合理工学研究科修士課程修了。同年日本アイ・ビー・エム(株)東京基礎研究所入所。現在、同研究所主任研究員。フラクタル、自然形状のモデリング、および画像処理の研究に従事。



黒川 雅人 (正会員)

1960年生。1984年京都大学工学部情報工学科卒業。1986年同大工学研究科情報工学専攻修了。同年日本アイ・ビー・エム(株)東京基礎研究所入所。画像データベース、動画像処理の研究に従事。平成3年度研究会賞受賞。



広田源太郎

1963年生。1986年大阪大学基礎工学部生物工学科卒業。1988年同修士課程修了。同年日本アイ・ビー・エム(株)東京基礎研究所入所。1994年ノースカロライナ大学計算機科学科入学。現在、Argumented Realityの研究に従事。