

自動演出生成方式によるシナリオ入力型 CG 動画制作システム

宮崎 誠也, 申 金紅, 青木 輝勝, 安田 浩

東京大学大学院 工学系研究科 〒153-8904 東京都目黒区駒場 4-6-1

E-mail: {seiya, j-shen, aoki, yasuda}@mpeg.rcast.u-tokyo.ac.jp

あらまし 本稿では、ブロードバンドを活用する映像コンテンツを大量に供給するためのツールとして、誰でも容易にかつ短時間で映像コンテンツを制作できることを目的にした CG 映像制作システム DMP のアーキテクチャを提案する。DMP は、一般的な CG 映像制作ソフトウェアとは大きく異なる発想で設計されており、映像のシナリオ文章をユーザからの入力とし、映像に必要な背景映像や CG、音声などのコンテンツをインターネット上から取り込み、シナリオ文章に従い自動的に演技・演出を付与することで映像を自動生成する。この手法により、制作者はこまかな演技や演出などの設定にわずらわされることなく、ストーリー制作や会話内容だけに集中できるため、映像制作のノウハウをまったく持たないユーザでもアイディアを即座に映像化できるのが特徴である。

キーワード 映像制作、演出、動画生成、演出知識ベース

Scenario Based CG Motion Picture Production System by Automatic Directorial Planning Method

Seiya MIYAZAKI, Jinhong SHEN, Terumasa AOKI, Hiroshi YASUDA

Faculty of Engineering, University of Tokyo 4-6-1 Komaba, Meguro-ku, Tokyo, 153-8904 Japan

E-mail: {seiya, j-shen, aoki, yasuda}@mpeg.rcast.u-tokyo.ac.jp

Abstract We propose DMP (Digital Movie Producer) system that provides non-professionals an easy way to create 3D CG movies. The users input is only text script and the elemental CG objects are searched automatically on the Internet or Content DB (Data Base). The goal of DMP is to make an instant movie automatically from the script using knowledge base of movies and content retrieval modules. Here, sophisticated knowledge base makes it possible to generate positioning tasks, detailed character actions and allow amateurs to concentrate on dialogue and story. For the function, even an amateur can create his/her own movie easily. In this paper, we suggest utilizing DMP system with directorial planning engine.

Keyword Filming, Direction, Motion Picture Generation, Directorial Knowledge Base

1. はじめに

近年映像コンテンツの制作や配信のコストが下がっており、素人でも思い通りに映像を制作したいというニーズが増している。しかし、従来の一般的な映像編集ソフトウェアや CG 制作ツールは細かい部分までマニュアルで設定をする必要があり、制作者は映像制作のノウハウや専門知識が多少なりとも必要である。そこで本稿では、ユーザからの入力に応じて演技や演出などの細かい設定を自動で設定可能な CG (Computer Graphics) 映像制作システム DMP (Digital Movie Producer) を提案する。DMP では、映像のシナリオ文章をユーザからの入力とし、映像に必要な背景映像や CG、音声などのコンテンツをインターネット上から取り込み、シナリオ文章に従い自動的に CG キャラクタを演技させ、カメラワーク等適切な演出を付与することで映像を自動生成する。本稿では、これを実現するため

の解決すべき問題点を明らかにし、解決手法を示す。

2. 関連研究

CG による映像を簡単に制作するための従来ツールとして、Alice[1], 3D Movie Maker[2]などがある。これらはマウスによる簡単な設定のみでインタラクティブに映像を制作できるのが特徴であるが、CG のクオリティが低く思い通りの映像を制作できなかったり、素材となるコンテンツがクロードな環境に依存したりしているなどの欠点が挙げられる。

テキスト文章やスクリプトから映像を自動生成する研究例として、Virtual Director[3], TVML[4], サイバーシアター[5]などがあるが、動作や演出の詳細な設定をしなくてはならず、制作者にある程度高度なノウハウや専門知識を要求している。

一方で、カメラワーク等の演出を映画撮影技法など

を用いて生成する研究例[6, 7]、人物の動作を効率的に生成する研究例[8]などがあるが、これらは映像のシナリオ文章を基に生成するものではない。

3. DMP の提案

3.1. 要求条件と解決すべき課題

映像制作システム DMP の要求条件を次のように設定した。

1. 映像制作のノウハウや知識を持たなくても制作できること
2. 誰でも簡単に扱える制作方法を提供すること
3. 多少クオリティを下げて良いから、手早く制作できること
4. コンテンツを再利用しやすい仕組みを提供し、低コストで制作できること

そこで、この要求条件を実現するために、次の手法を提案することにする。

1. シナリオ入力方式：映像のシナリオ文章をユーザからの入力とする
2. コンテンツ検索再利用方式：映像に必要な背景映像や CG、音声などのコンテンツをインターネット上のコンテンツデータベースから取り込む
3. 演技演出自動生成方式：シナリオ文章に従い自動的に CG キャラクタを演技させ、カメラワーク等適切な演出を映像制作技法に関する知識ベースを持つシステムが付与する

これらの一連の処理を概略図として図 1 に示す。

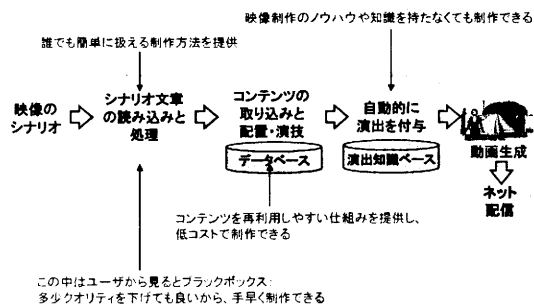


図 1 DMP 処理外略図

このアーキテクチャを実現するために、以下の技術的な重要検討課題が考えられる。

1. ユーザから入力されるシナリオ文章をシステムはどのように処理をしたら良いか？
2. 複数の素材コンテンツをインターネット上から取り込み、それらを適切に配置・演技させるにはどのようなアーキテクチャが必要だろうか？
3. ユーザに代わって映像演出をシステムが自動的に付与し、効果的な映像を生成するにはどうす

れば良いか？

以下の節ではこれらの課題を解決するための手法を示していく。

3.2. シナリオ処理方法と DMPML

ユーザにより記述されたシナリオ言語は、そのままではコンピュータが扱うのが困難であるため、XML の表現形式として DMPML(Digital Movie Producer Markup Language)を定義する。DMPML は、シナリオ文章を表現するのに必要十分な情報量を持ち、かつコンピュータが最も処理しやすい表現形式でなくてはならない。前者の要求条件を満たすために、我々はさまざまなジャンルのシナリオを調査し、さらに放送作家や映像プロデューサー・動画像処理の専門家の意見を参考にし項目をまとめた。また後者の要求条件を満たすために、DMPML に階層構造を持ち、各パートの階層で違う表現形式で持たせることにした。

具体的には、構造化されたシナリオ文章を記述するシナリオ文章パート、映像を構成する素材コンテンツに関する情報を記述するコンテンツ管理パート、動画を生成するのに必要なすべての詳細情報を記述する映像表現パートの計 3 つのパートを持たせ、同じ内容のことで、自然言語 (抽象的) → 5W1H (より具体的) → コンテンツ情報 (具体的) → 動作パラメータ情報 (具体的) と、各パートの階層で違う表現形式を持たせた。

例として、シナリオ文章パートの演技部分における表現を示す。

```
<Direction name="Daughter">
  <Text>Daughter takes the goods.</Text>
  <Structure>
    <mpeg7:Who>
      <mpeg7:Name>Daughter</mpeg7:Name>
    </mpeg7:Who>
    <mpeg7:WhatObject>
      <mpeg7:Name>the goods</mpeg7:Name>
    </mpeg7:WhatObject>
    <mpeg7:WhatAction>
      <mpeg7:Name>take</mpeg7:Name>
    </mpeg7:WhatAction>
  </Structure>
</Direction>
```

この例では、“Daughter takes the goods”という演技指示を<Text>タグではそのまま自然言語で記述し、<Structure>タグ以下で、MPEG7 の記法にしたがった 5W1H の形式で“Who”, “What”, “WhatObject”, “WhatAction”, “Why”, “How”の各項目に分けて記述する。この二通りの記述方法をすることで、自然言語処理はシナリオ入力時の 1 回だけで済まされる。

現在は高度な自然言語処理を避けるために、語彙や

文法が決められている制約のある英語を入力とし、あらかじめ準備した数十通りの予約語とシナリオ文章の品詞がマッチするかどうかで判断している。

またシナリオはある状態遷移に従うため、これを利用して映像を生成するために必要な情報がシナリオに正しく記述されているか否かを確認する、意味内容のバリデーション（判定）を行い、システムが映像生成する段階で致命的なエラーが生じないようにさせている。

3.3. 演技生成とメタデータ

DMP では映像を構成する素材のコンテンツ（登場人物、背景など）を制作する機能はなく、既存のモデリングソフトウェアで作られたコンテンツを検索・再利用する手法をとっている。

ここで、仮に DMP システム用のクラウドな環境でしか動作しないコンテンツフォーマットにすると、実用化段階において素材コンテンツが多く集まらず、この枠組みを普及させることは難しいと予想される。そのため、コンテンツのデータ形式に依存しない、オープンな環境を実現するのに最適な枠組みが必要がある。そこで、コンテンツ毎に付与する検索と利用のための情報を保持するコンテンツ形式に依存しない汎用的なメタデータを定義する。表 1 にこのメタデータの属性とその内容の一部を示す。

表 1 コンテンツメタデータ

属性	属性の内容
Format	CG コンテンツのフォーマット形式
Type	キャラクタコンテンツを示す属性
Gender	"Male", "Female"
Age	"Middle", "20"など
Voice	さらに Style と Language の小属性に分けられる Style は "Deep", "Mellow" など Language は "English" など
Clothes	さらに Shirt, Skirt, Trouser, Dress などの小属性に分けられる
Action	演技可能な動作を示す "Backspin", "Moon walk" など

またキャラクタが複雑な演技を実現するために、配置制約情報を利用したアノテーションベースのキャラクタアニメーション手法を用いる。現在 DMP は、"sit on the bench" や "talk to someone", "appear in the screen" など、約 50 通りの動作をサポートしているが、例えば背景コンテンツのベンチの場所や向きなどの情報は、配置制約情報をメタデータに記述し、これを参照して

動作を実現させている。メタデータの例を次に示す。

```
<Object name="wood_bench" type="Chair">
  <Region>
    <Param name="BACKREST" type="Rectangle" x0="3.0"
      y0="0.5" z0="1.0" x1="5.0" y1="1.0" z1="1.0"/>
    <Param name="CHAIR" type="Rectangle" x0="3.0"
      y0="0.5" z0="1.0" x1="5.0" y1="0.5" z1="1.5"/>
    <Param name="LEGS" type="ComplexType"
      operator="LEG1 OR LEG2 OR LEG3 OR LEG4">
      <Param name="LEG1" type="Cylinder" x="5.0" y="0.0"
        z="1.5" r="0.1" l="0.5" vx="0.0" vy="1.0" vz="0.0" />
      <Param name="LEG2" type="Cylinder" x="5.0" y="0.0"
        z="1.0" r="0.1" l="0.5" vx="0.0" vy="1.0" vz="0.0" />
      <Param name="LEG3" type="Cylinder" x="3.0" y="0.0"
        z="1.5" r="0.1" l="0.5" vx="0.0" vy="1.0" vz="0.0" />
      <Param name="LEG4" type="Cylinder" x="3.0" y="0.0"
        z="1.0" r="0.1" l="0.5" vx="0.0" vy="1.0" vz="0.0" />
    </Param>
    <Param name="FRONT" type="Direction" vx="0.0"
      vy="0.0" vz="1.0"/>
  </Region>
  <ConstraintRegion subject="Character" action="sit">
    <Param name="on" type="Rectangle" x0="3.3" y0="0.5"
      z0="1.1" x1="4.7" y1="0.5" z1="1.4"/>
  </ConstraintRegion>
  <RecommendedRegion subject="Character" action="sit">
    <Param name="on" type="Point" x="4.0" y="0.5" z="1.2" />
    <Param name="d" type="Direction" vx="0.0" vy="0.0"
      vz="1.0"/>
  </RecommendedRegion>
</Object>
```

3.4. 演出生成と演出知識

演出をまったく付与せずにすべてのシーンをマスターショットのみで撮影すると、臨場感がなく味気のない仕上がりになってしまう。一般的には、シナリオの内容にあわせて適切なフレーミングやモンタージュ・ズーム等の演出をおこなった方がより効果的で臨場感のある映像が生まれると考えられている。

映像演出の自動付与を実現するために、演出に関するルールベースを用いシナリオ文章に応じて適切なカメラワークやライティング等の演出を施すための手法を提案する。もっともどのような演出が最適かというルールはなく、実際映画などでは監督の個性や感性、センスに従う部分が多い。しかし一方で視聴者が自然に感じる素直な演出方法には従来からある程度「定跡」となっている技法（イディオム）やノウハウが確立している。演出ルールは、これらの映像制作技法を実写映像の映像分析を通して蓄えることにする。またどの程度の演出をサポートすべきかという問題があるが、視聴者が不自然に感じないような、最低限レベルでの演出を施せることを最初のステップの目標とする。

DMP でサポートしている演出の一部を表 2 に示す。この演出知識は、シナリオ文章に記述された内容をも

とに、カメラのフレーミングやカット方法、ライティングなどの情報を提供する。ここで、WS はウエストショット、BS はバストショット、CU はクローズアップ、POV はポイントオブビュー、OTS は肩なめショットを示す。

表 2 演出知識例

	映画 (スビルバーグ風)	ドキュメンタリー
全体	カメラの動き=あり	カメラの動き=ズーム少
状況: 会話シーン	フレーミング=BS/POV/OTS	フレーミング=WS
動作: 感情表現	フレーミング=CU 動作=ズーム	フレーミング=BS
動作: 見る	フレーミング=CU	フレーミング=BS

演出の種類は分析によっていくらかでも増やすことができるが、詳細に分類していくことで映画・ドキュメンタリーなどのジャンルによる違いだけでなく、監督による違いなども可能になる。

DMP は、このフレーミングサイズなど抽象的な結果を基に、カメラのポジションやアングルなど具体的な数値を求めたり、ライティングやアニメーションエフェクト・カット方法・効果音を生成し演出を実現する。

4. DMP の実装

以上シナリオ入力・処理方式、コンテンツ検索・利用方式、演出自動生成方式の3つの方式を統合したシナリオ入力型動画制作システムを実際に構築した。

DMP のシナリオ入力ユーザインタフェースを図2に示す。本稿で提案した DMP システムは、個人用途として気軽に日記や旅行記などの映像作成ツールとして用いられるだけでなく、初心者向けの映像制作教育ツールとしてや、今後さらに映像のクオリティが上がることで映像メディアを駆使した e-ラーニングやストーリーミング配信用商品広告など業務用途の映像を低コストで制作するツールとして用いることができるようになると思われる。また CG を利用しているため、短時間に動画の絵コンテや動画のプロトタイプ・新しいジャンルの実験開拓などで威力を発揮すると予想される。

一方で、テキストからの映像生成のため、抽象的な動画や動きの激しい映像などには不向きであると予想されるため、これらは従来の制作手法を用いるのが良いと考えられる。

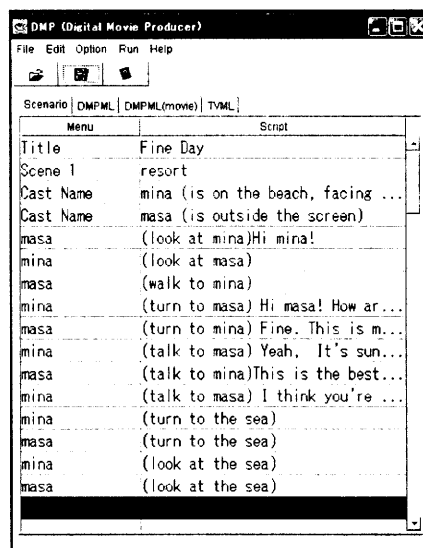


図 2 シナリオ入力インタフェース

5. 今後の課題

映像表現のバリエーションを増やすことでより多くのジャンルに対応できる映像生成を実現すること、リアルな表現や誇張した表現を実現する CG クオリティを達成することが今後の課題になる。

文献

- [1] M. Conway, et al., Alice: Lessons Learned from Building a 3D System for Novices, CHI, 2000.
- [2] 3D Movie Maker, <http://www.microsoft.com/catalog/display.asp?site=267&subid=12&pg=1>
- [3] K. Manos, T. Panayiotopoulos, G. Katsionis, "Virtual Director: Visualization of Simple Scenarios," 2nd Hellenic Conference on Artificial Intelligence, SETN, 2002.
- [4] 林正樹ほか, "テレビ番組記述言語 TVML の言語仕様と CG 記述方法", 第 3 回知能情報メディアシンポジウム, pp.75-80, 1997 年
- [5] Hiroshi Matsuda, Yoshiaki Shindo, "Prototype of Cyber Theater Scenario Language," Proceedings of IMSA 2003, Florida, Aug. 2003.
- [6] Li-wei He, Michael F. Cohen, David H. Salesin, "The Virtual Cinematographer: A Paradigm for Automatic Real-Time Camera Control and Directing," Proc. of SIGGRAPH, pp. 217-224, 1996.
- [7] Kevin Kennedy, Robert E. Mercer, "Planning Animation Cinematography and Shot Structure to Communicate Theme and Mood," Symposium on Smart Graphics, Hawthorne, NY, 2002.
- [8] Okan Arikan, David A. Forsyth, James F. O'Brien, "Motion Synthesis From Annotations," Proceedings of SIGGRAPH 2003, CA, Aug. 2003.
- [9] "Movie Analysis for Directorial Planning Engine", Seiya MIYAZAKI, Jinhong SHEN, Takafumi YUKI, Mitsuru KANEKO, Terumasa AOKI, Hiroshi YASUDA, IMSA2003, Hawaii, USA, T576MP, Aug.2003.