

自然言語を利用した CG 創作環境の一手法

菅原 衣織[†] 伊藤 貴之[†]

[†]お茶の水女子大学 理学部 情報科学科 〒112-8610 東京都文京区大塚 2-1-1

E-mail: [†]{iori, itot}@itolab.is.ocha.ac.jp

あらまし 我々は 3 次元 CG 制作の入力手段として音声を使うことを考えている。その前段階として本研究では、音声認識結果である自然言語からの CG 制作ツールを実現する一手法を提案する。本ツールでは入力文からキーワードを抽出し、それに対応する 3 次元モデルとテクスチャ画像を検索して組み合わせる。3 次元モデルとテクスチャ画像には本ツール特有のタグが付与されるものとする。物体名を表す名詞が付与されると、その類似語や上位概念語も自動的に付与される。またユーザによって主観的に付与される形容詞や、色・周波数特性などの特徴算出値もタグに含む。これらのタグの検索結果として表示される物体に対して、入力文から指定される座標変換を施す。以上を反復することで感覚的な 3 次元 CG 制作環境を実現する。

キーワード 3 次元 CG, 自然言語。

A computer graphics modeling tool featuring a natural language based input interface

Iori SUGAHARA[†] and Takayuki ITOH[†]

[†]Department of Information Sciences, Faculty of Science, Ochanomizu University

2-1-1 Otsuka, Bunkyo-ku, Tokyo, 112-8610 Japan

E-mail: [†]{iori, itot}@itolab.is.ocha.ac.jp

Abstract We aim to apply a speech recognition technique as an input interface of computer graphics modeling environment. As a preliminary step of this goal, this paper proposes a computer graphics modeling technique featuring a natural language based input interface. This tool combines 3D shape models and texture images corresponding to the keywords extracted from input sentences. Here, we suppose labels are manually assigned to the 3D shape models and texture images. Synonym and the words of a superordinate concept are then automatically added as labels. Also, labels related feature values (e.g. color and frequency) are assigned to the texture images. Coordinate transformation specified from input sentences is also applied to the 3D shape models. We think this tool realizes intuitive computer graphics creation by iterating those processes.

Keyword 3D computer graphics, Natural language.

1. 概要

3 次元コンピュータグラフィックス(CG)の制作環境には、商用では Maya や 3D Studio Max など、またウェブ上のツールとしては Google Sketch Up など、既に多数のものが知られている。しかし、いずれも習熟するためには相応の知識や技術が必要であり、CG に関する専門知識の無いユーザや小さな子供にとって親しみやすいツールとは言いがたい。そこで近年ではスケッチインタフェースなどを導入した初心者ユーザ向けの

3 次元 CG 制作環境の研究も進んでいる。我々も同様に、専門的な知識のあるユーザではなく初心者ユーザを対象とした手軽な CG 制作環境の構築を目標にしている。

近年のスマートフォンの普及に伴って、iPhone4S から対応になった Siri や、Android の音声認識 API に代表されるように、音声認識技術の発達が著しい。我々はこの音声認識技術を用いて 3 次元 CG の感覚的な制作環境を構築することを目標にしている。音声入力を

導入することで例えば、マウスやタッチパネルなどの 2 次元操作デバイスがもつ 3 次元操作の難しさが緩和されることが期待される。また、スマートフォンやポータブルゲーム機などの小型端末ゆえの操作の困難さからも解消されると期待される。さらに、一同に介した多数のユーザによる協調的な制作が容易になると期待される。

本報告はその前段階として、自然言語を入力とした 3 次元 CG 制作環境の構想と、その実装について論じる。本手法では入力された自然言語からキーワードマッチングにより、あらかじめデータベースに登録されている 3 次元形状モデルとテクスチャ画像を抽出し、これらを組み合わせて表示する。

2. 関連研究

2.1 自然言語による 3 次元 CG 制作

自然言語による幾何形状モデリング手法は以前にも発表されている[1]。この研究では英語の SVO の文法に則って球や立方体などのプリミティブな 3 次元モデルを操作、配置し、組み合わせて物体を制作する 3 次元グラフィックシステム”Do”を提案している。CG の専門的な知識のあるユーザと知識のないユーザを対象とした実験を行っており、共通の物体を作成する際に掛かった時間や使った単語数を比較している。図 1 は”Do”を使って制作された雪だるまの CG である。



図 1. [1]による雪だるま作成例

(左)Do 開発者が作成した雪だるま
(中央)CG 非専門家が作成した雪だるま
(右)CG 非専門家が作成した雪だるま

実験の結果より、事前知識をほとんど要しない、オブジェクトの選択が容易であるという利点とともに、微調整がし難い、座標系の感覚がつかみにくいという問題点が挙げられている。以上の研究を踏まえ我々の提案する手法では、座標系を意識する必要のない手軽で感覚的な指示による 3 次元 CG 制作を目指す。

2.2 音声を利用したユーザインタフェース

音声を利用したユーザインタフェースの一例として、オノマトペを用いたマルチモーダルインタラクション[2]がある。これはオノマトペンというツールとして発表されており、マイクで拾った音声を音声認識で

判別し、タッチペンでの描画に反映させるという研究である。この研究では主にオノマトペ、つまり擬音語を採用しており、「てんてん」や「きらきら」といった言葉に対応して描画スタイルを変化させている。図 2 はオノマトペンの概念図と実物例である。



図 2. オノマトペン[2]による実行例

2.3 テキストラベルとスケッチによる画像生成

文字情報によるデータの検索とキーワードマッチングの手法として Sketch2photo[3]を紹介する。Sketch2photo はテキストラベルの注釈とフリーハンドで描いたスケッチから現実的な画像を生成する手法である。テキストラベルとスケッチが一致した複数の写真をインターネットから検索し、新しい画像を生成する。Sketch2photoによる画像生成の流れを図3に示す。最適な組み合わせを見つけるために、検索された画像にはそれぞれ数値が割り当てられており、それによって画像が選出されている。



図 3. Sketch2photo[3]による画像生成例

3. 提案手法

本章では提案手法について述べる。まず本手法における 3 次元 CG のプログラミングモデルを述べ、それに対応する自然言語について論じる。続いて本手法で自然言語を入力した際の処理手順について述べる。

3.1 3 次元 CG のプログラミングモデル

我々の実装では OpenGL というプログラミング環境をベースにして 3 次元 CG を表示する。このプログラミング環境では以下のような処理手順を要する。

初期化: ウィンドウの設定、視点や光源の設定など。

物体単位の処理: 物体形状（頂点情報および位相情報）の定義、物体に貼りつけるテクスチャ画像の設定、物体の座標変換（拡大縮小、平行移動、回転など）、など。

現時点での我々の実装は、この2種類の処理手順のうち物体単位の処理について、以下の2種類の自然言語入力を想定している。

- 1) 「A である B」(A は修飾語, B は物体名) という構文により, B に相当する物体形状と, A にふさわしいテクスチャ画像を検索して表示する。
- 2) 座標変換を表す予約語の入力により, 物体の拡大縮小, 平行移動, 回転などを指定する。

本手法における自然言語入力の例と, それに対応して構築されるデータ構造の例を, 図4に示す。本手法ではまず物体形状とそのテクスチャ画像を自然言語から検索し, 続いて必要に応じてその座標変換を自然言語から設定する。

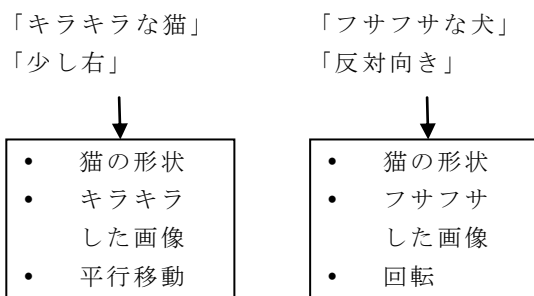


図4. 自然言語入力とデータ構造

以下本報告では, 物体形状とテクスチャ画像の検索方法と, 座標変換の操作について論じる。

3.2 3次元形状モデルとテクスチャ画像

本手法は自然言語入力のため既存のCG制作ツールと同様に精度の高い形状モデリングは困難である。既存研究[1]においてもプリミティブな3次元形状の組み合わせだったために創作の範囲が限定されていた。そこで我々は, 本ツール自体で3次元形状モデルを設計するのではなく, インターネット上にある3次元形状モデルデータベース(例えば3D Model Search Engine[4])からダウンロードした3次元形状を用いる。これらの3次元形状モデルはWRL形式やSTL形式といった標準的な書式で記述されている。

我々の実装では, 後述するタグを付与された3次元形状モデル群とテクスチャ画像群がデータベースに登録されていることを前提とする。そして自然言語を入力すると, その入力文を構成する単語とのマッチングにより, 3次元形状モデルとテクスチャ画像を検索する。以上の処理によって定義されたテクスチャ画像付きの3次元形状モデルを, 以下「キャラクタ」と称する。

3.3 タグとのマッチング

本手法で扱う3次元モデルやテクスチャ画像には, その内容を表す物体名や形容詞をユーザが手動でタグ付けすることを想定する。そしてシステムは, 手動でタグ付けされた単語の類似語や上位概念語を割り当て, またテクスチャ画像においては色などの特徴量に対応する単語を自動で割り当てる。以下にタグの付与方法を記す。

物体名: ユーザが手動で付与する。1つの3次元形状モデルやテクスチャ画像に対して複数の物体名を付与することも想定する。また, 同じ物体名が複数の3次元形状モデルやテクスチャ画像に対して付与されることも想定する。

形容詞: ユーザが手動で付与する。ユーザの主観や趣向を反映するために有効である。1つの3次元形状モデルやテクスチャ画像に対して複数の形容詞を付与することも想定する。また, 同じ形容詞が複数の3次元形状モデルやテクスチャ画像に対して付与されることも想定する。

同義語や上位概念語: 手動でタグ付けされた物体名をもとに, 概念辞書を参照して自動的に付与する。

特徴量に対応する単語: テクスチャ画像に対して色や周波数などの特徴量を算出し, それに対応する単語を自動的に付与する。

なお我々の実装では, 概念辞書として日本語 WordNet[5]を利用している。WordNetでは各単語の概念が"synset"という単位にまとめられており, さらに関連のあるsynsetとリンクされている。我々の実装ではsynsetに属する同義語(synonym)と, synsetのリンクをたどって得られる上位概念語を用いる。

以上に記したタグ情報をもとにして, 本手法は以下の手順で最適な3次元モデルとテクスチャ画像のマッチングを図る。図5にタグ情報の格納例を示す。

1. 入力文中の物体名(またはその同義語)をタグに持つ3次元形状モデルを検索する。
2. 1.で該当する3次元形状モデルが複数存在する場合には, 手動でつけた3次元形状モデルにつけられた形容詞を比較し, 1個でも多く共通の形容詞を入力文に含む3次元形状モデルを選ぶ。
3. 入力文中の修飾語(またはその同義語)をタグに持つテクスチャ画像を絞り込む。図5では黄色い欄「ふわふわ」が該当する。
4. 3.で該当するテクスチャ画像が複数存在する場合には, 3次元形状モデルと同一の上位概念語を共通に有するテクスチャ画像を優先的に選ぶ。図5では緑色の欄「動物」が該当する。

5. 4.で上位概念語を共通に有するテクスチャ画像が複数存在する場合には、手動で3次元形状モデルとテクスチャ画像につけられた形容詞を比較し、1個でも多く共通に有するテクスチャ画像を優先的に選ぶ。図5ではピンク色の欄「かわいい」が該当する。

3次元形状モデルのタグデータ			
URL	keyword	synset	adjective
猫のポリゴンA	ねこ	動物	かわいい
猫のポリゴンB	ねこ	動物	かっこいい
犬のポリゴン	いぬ	動物	

テクスチャ画像のタグデータ			
URL	keyword	synset	adjective
雲のテクスチャ	ふわふわ	自然	あかるい
毛皮のテクスチャA	ふわふわ	動物	かわいい
毛皮のテクスチャB	ふわふわ	動物	かっこいい
綿のテクスチャ	ふわふわ	繊維	

図 5. タグ情報の格納例

3.4 データベースの設計

我々の実装ではデータベース SQLite を用いて3次元モデルおよびテクスチャのタグを管理している。我々の実装におけるデータベースの構造を図6に示す。

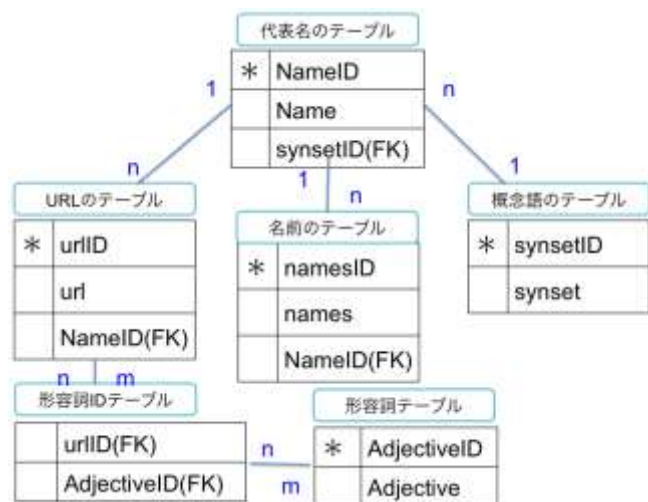


図 6. データベースの構造

3.5 座標変換

キャラクターに平行移動や回転などの座標変換の操作を加えるにあたり、本手法では数値表現ではなく口語的な文章表現と座標変換を対応づける。口語的な表現を採用することにより、3次元操作の簡易化を図るとともに、関連研究[1]のユーザテストで挙げられていた座標系の感覚がつかみにくいという問題点を緩和できると期待できる。図7に座標変換文と操作内容の例を示す。現時点で実装している操作内容は我々の主観に基いて調節されているため、今後はユーザアンケート

トを実施するなどしてユーザビリティを向上したい。

操作	入力文	内容
平行移動	みぎ	X軸正方向に移動
	まえ	Z軸正方向に移動
回転	まわれみぎ	Y軸を基準に90°回転
	のけぞれ	X軸を基準に90°回転
拡大縮小	おおきく	1.2倍に拡大
	とつてもちいさく	0.5倍に縮小

図 7 座標変換文とその操作内容の一例

3.6 GUI 機能

我々の実装における GUI を図8に示す。画面中央は実行結果を表示する描画領域である。画面左側には、キーワードを入力するためのテキスト記入欄や、表示されたキャラクターの拡大縮小、回転などの操作のためのチェックボックスを備えている。画面右側には、タグによる優先度が次点以下だったテクスチャを表示する領域を備えている。このとき、優先度が次点以下だったテクスチャをクリック操作で選択することで、3次元モデルに割り当てるテクスチャ画像を切り替えられる。



図 8. GUI 上での表示例

4. 実行結果

本手法を起動して自然言語を入力した例を図9に示す。この例では入力文として「しましま」「ねこ」というキーワードを入力している。このとき描画領域には、縞模様のテクスチャがマッピングされたねこのポリゴンが表示され、右側の領域には、キーワードと一致した複数のテクスチャ画像が表示される。このとき右側の領域では、「しましま」に加えて「ねこ」の上位概念語である「動物」とのタグを有するものを上半分に、「しましま」のタグのみを有するものを下半分に表示する。現在の実装では出現したキャラクターの初期位置は全て一定となっているが、座標変換の入力文によっ

てキャラクターの位置を操作することで、図 9 のように複数のキャラクターを画面上に表示することが可能である。



図 9. GUI 上での表示例

5. まとめと今後の課題

本報告では、自然言語入力によって 3 次元形状モデルとテクスチャ画像を検索することで、感覚的に 3 次元 CG を制作する一手法を示した。

今後の課題として、以下の点に取り組みたい。

形容詞のタグの自動化. 現時点での我々の実装では、3 次元形状モデルやテクスチャ画像についている形容詞のタグは手動での付与になっている。テクスチャ画層に関して、形容詞のタグつけの一部を自動化するために画像認識の適用を考えている。具体的には、各テクスチャの周波数分布や色分布などの特徴量を算出し、データベース上の他のテクスチャ画像と比較し、類似するテクスチャ画像と同じタグを付与する。これにより、主観性と客観性の両方を考慮したテクスチャ画像の推薦ができるのではないかと考えている。

座標変換の操作内容の改良. 先にも述べたように現時点での我々の実装では、座標変換の操作内容は我々の主観に基いて調節されている。このユーザビリティを向上させるために、ユーザアンケートを実施し、その結果に沿って操作内容を改良したい。また、キャラクターの初期位置も同一にするのではなく、3 次元モデルの上位概念を引用し、空に浮かんでいる物体は初期位置を空中に設定したり、筆記用具やパソコンなど地面に初期位置を設定しては不自然なものを別の物体の上に配置するなど、遊び心を交えたものにしようと検討している。

小さい子供の興味を引くような GUI の改善. 本手法は CG の専門知識のないユーザを対象にしている。GUI をより親しみやすいデザインにすることで、いろいろな人にとって使いやすいシステムを開発したい。

色調や現実性を考慮したモードの設定. 形容詞や特徴量に関するタグを利用して多様な作風を目指す機能

も追加しようと検討している。例えばいままでに選ばれたテクスチャ画像の特徴量や形容詞のタグと同一のタグを有するテクスチャ画像を優先的に選ぶことで、似たような色合いや雰囲気 of テクスチャが優先的に選ばれ、統一感のある空間を創ることが期待できる。あるいは上位概念語のタグを利用して、あえて現実には起こりえない組み合わせのテクスチャ画像と 3 次元形状モデルを選ぶことで、現代アート風の空間を創ることが期待できる。

ユーザテストの実施. 本手法のシステムの改良を進めていくにあたってユーザビリティの向上は必須の課題であると考ええる。一例として、ある同一の風景を制作させ、制作時間とユーザの満足度を測る、というような形でユーザテストを検討している。

参 考 文 献

- [1] 河合善之, 岡田稔, 自然言語による幾何形状モデリングと画像合成の一手法, 情報処理学会論文誌, 42(5), 1161-1169, 2001.
- [2] 神原啓介, 塚田浩二, オノマトペを用いたマルチモーダルインタラクション, 人工知能学会全国大会論文集, 1C2-OS4b-12, 2011.
- [3] Tao Chen, Ming-Ming Chen, Ping Tan, Ariel Shamir, Shi-Min Hu, Sketch2Photo: Internet Image Montage, SIGGRAPH Asia '09, Article No. 124, 2009.
- [4] Princeton Shape retrieval and Analysis Group, 3D Model Search Engine, <http://shape.cs.princeton.edu/search.html>
- [5] 独立行政法人情報通信研究機構(NICT), 日本語 Word Net, <http://nlpwww.nict.go.jp/wn-ja/>