

自然言語による形状・シーンモデラにおける動的指示・表示手法の検討

柳井 康徳, 岡田 稔

早稲田大学大学院情報生産システム研究科

あらまし: 自然言語による三次元コンピュータグラフィックス (3D-CG) モデラの改良方式を提案する。本システムは入力音声であるため肢体不自由者でも容易に CG を作成することができるため、彼らの雇用機会拡大に繋がることを期待する。また、CG の非専門家にも CG 作成の有効な手段となるであろう。本稿ではこのモデラの概要説明に加え、従来手法の問題改善に繋がる提案手法について述べる。提案手法の実装方法として三つの幾何変換手法を検討した。さらに従来手法との対照実験を行い、提案手法の有効性を示した。

A Study on the Dynamic Instructions and Display Method by a Verbal Interface

Yasunori YANAI and Minoru OKADA

Graduate School of Information, Production and Systems, Waseda University

Abstract: We have developed a modeling system *Do* by a verbal interface. The system can be used by handicapped people because of its verbal interface and we expect that they will get employment opportunities. Additionally, it will be easier for non-experts to create computer graphics. In this paper we explain about the outline of the system and the proposed method in order to improve the problem of the former method. Three geometry transformation methods are considered for the implementation of the proposed method. Finally, contrast experiments of the former and proposed methods have been achieved to indicate the availability of the proposed one. As the result, we received the high evaluation on the proposed one.

1 はじめに

既存の多くのグラフィックスソフトウェアでは、キーボードやマウスのようなハードウェアデバイスの使用が不可欠である。そのため肢体不自由者がそれらのソフトウェアを使用することは困難である。そこで、筆者らは自然言語を入力とした三次元コンピュータグラフィックス (CG) による形状とシーンのモデリングを行うシステム *Do*[1] の開発を進めている。自然言語による言語的指示が可能になれば、肢体不自由者にも CG の作成が容易となることが期待される。また、*Do* は [2, 3] と同様、デザイナー等に対する初期発想支援を目的としている。

関連研究として、西山らによる Picnyck[4] は「広

大な」「清らかな」といった感性語を含んだ風景描写文を入力とし、それらの言葉にふさわしい風景画像を創造する。現在の *Do* では感性語は扱っていないが、感性語によってオブジェクトの質感を変化させる操作の導入は有効であると考えられる。Clayらによる *Put*[5] は単純に既存のオブジェクトの位置を参照し、その位置情報に基づいて配置を行う。それに対し、我々の *Do* は *Put* のオブジェクト配置の考えを基本として導入し、さらに移動、拡大・縮小、回転、CSG 演算といったモデリングシステムとしての多くの機能を備えている。さて、会話には any, some, more, little などの曖昧表現が存在し、オブジェクトに対して言語的指示を行う際にしば

しば用いられる。そのため、これらの曖昧表現の解釈方式をいかに実装するかということは *Do* にとって重要な課題である。辻らによる先行研究 [6] では any, some, other, another の解釈方法について検討している。さらに筆者らは more, little の二つの曖昧表現によるオブジェクト制御の基礎検討を行った [7]。more, little の曖昧表現を用いた言語的指示の際には直前の指示に対応した状態変化が必要であるが、従来法では直前の指示に対応していなかった。[7] によりこの問題は改善され、直前の指示に基づくより自然な会話に近い指示が可能となった。

本稿では動的指示・表示手法を導入した提案手法について検討する [8]。提案手法では、*Do* の基本方式においてオブジェクトの移動、拡大・縮小、回転等の変換にアニメーションを用い、従来手法では不可能であった視覚的判断による形状操作が可能となる。この操作により、作成時間および入力テキスト数の削減、操作の利便性および視覚的効果の向上等の効果が期待される。動的指示・表示手法の関連研究として江村らのキャラクタアニメーション技術 [9] が挙げられる。これはシナリオ文から複数の基本動作を抜き出し、その合成動作をキャラクタに適用する。*Do* では動的指示・表示手法により、移動、拡大・縮小、回転の幾何変換の合成が可能となる。

2 自然言語による形状・シーンモデラ *Do*

Do の処理の流れを図 1 に示す。現在は音声の誤認識による副作用を防ぐため音声認識部は未実装であり、バーバルインタフェースとしての検討のために英文テキストを入力として基礎検討を行っている。テキストは構文解析によって各単語に対する品詞が判別され、構文解析木が生成される。次いで中間表現としてのニーモニックを生成する。ニーモニックにより 3D シーン情報は更新され、再描画が行われる。ユーザはその描画を見て次の指示を行う。

Do の構成を図 2 に示す。*Do* は入力部、命令解析部、3D シーン情報更新部、API インタフェースの四つのプログラムで構成されており、各プログラム間はソケット通信によりデータの送受信が行われる。現在、構文解析には Sekine らによる Apple Pie Parser [10] を用いているが、解析モジュールの変更により他の言語に対応させることも可能である。構文解析には単語・品詞辞書、ニーモニック作成にはプリミティブカラー辞書、プリミティブオブジェクト辞書がそれぞれ参照される。

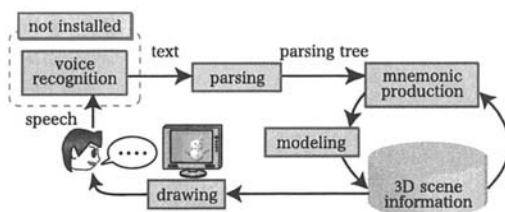


図 1: *Do* の処理の流れ

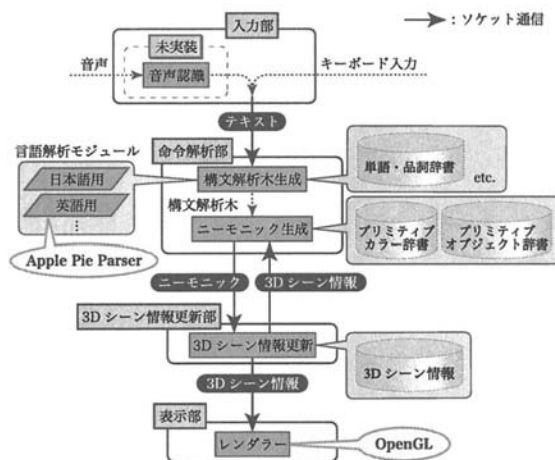


図 2: *Do* のシステム構成図

3 静的手法・動的手法による指示と表示

3.1 静的指示方式と動的指示方式

静的指示方式はオブジェクトの目標状態を指定した指示方式を指す。また、動的指示方式は目標状態を定めず状態を継続的に変化させる指示方式を指す。

3.2 静的表示方式と動的表示方式

静的表示方式とは現在の状態と次の目標状態の二つのフレームを瞬間的に切り替えて表示する方式である。それに対し、動的表示方式とは現在の状態から次の目標状態までの状態変化をアニメーションにより表示する方式である。

3.3 各指示と表示の組み合わせ

上記の二つの指示方式と二つの表示方式により、それらの組み合わせは四通り存在する。しかし、動的指示方式と静的表示方式の組み合わせに関しては、3D シーン情報が更新されているにも関わらず、表示に反映されないという実用性のないシステムになる。したがって、*Do* ではその組み合わせを除いた三つの組み合わせを考える。以下、各組み合わせを用いた従来手法、提案手法について述べる。

3.4 静的指示・表示手法による従来手法

従来手法は静的指示方式と静的表示方式のみを用いていた。従来手法の問題点は、ユーザの意図とスクリーン上の結果が異なることが多く、位置などの調整に何度も指示が必要となることである。

3.5 動的指示・表示手法を導入した提案手法

従来手法の問題を改善するため、筆者らは動的指示・表示手法を導入した提案手法を考案した。提案手法では指示方式に静的指示方式と動的指示方式の両方式を用い、表示方式は従来の静的指示方式を動的表示方式に変更する。そのため、提案手法ではオブジェクトの状態を継続的に変化させ、ユーザの意図した状態に達した時に停止命令等を与え、状態変化を止める操作が可能となる。スクリーンを見ながら即時的な指示が行えるため、従来手法に比べて指示回数および入力テキスト数の削減、操作の利便性および視覚的効果等の向上が期待される。

4 アニメーションのための幾何変換手法

提案手法における動的指示・表示方式に必要なアニメーションの実装方法、特に幾何変換の方式として、筆者らはキーフレームアニメーション、公差行列、公比行列について検討した。

4.1 キーフレームアニメーション

スプライン曲線のようなキーフレームアニメーションを用いるには、アニメーションの始点と終点に対応する初期状態と目標状態があらかじめ必要となる。しかし、動的指示ではアニメーションによる状態変化の途中で停止命令が与えられることにより目標状態が決定されるため、キーフレームアニメーションを用いることは不可能である。

4.2 公差行列

アフィン行列による変換は式 (1) で表される。

$$P_{n+1} = H_f P_n \quad (1)$$

ここで P_n は第 n ステップの点、 $H_i (0 \leq i \leq f)$ は P_n から P_{n+1} の状態へ変化するアニメーションの第 i フレームのためのアフィン行列を示す。 f は一つの動作に対応するフレームの総数である。

公差数列 d は公差数列の考えを変換行列に適用したものであり、静的指示方式において d , H_i は式 (2), (3) によりそれぞれ求める。

$$d = \frac{1}{f}(H_f - E) \quad (2)$$

$$H_i = E + id \quad (3)$$

ここで E は単位行列を表す。公差行列による手法は要素の線形補間であるため一般に $|H_i| \neq 1$ となり、回転に対する変換は非剛体運動になってしまう。回転の例を図 3(a) に示す。動的指示方式では H_f が与えられないため、システム側で d を決定する。

4.3 公比行列

公差行列と同様に、公比行列はスカラーにおける公比数列の考え方を変換行列に適用したものである。まず静的指示における各操作の変換方法を述べる。移動に関する指示 “Move the object $dir(u,v,w)$.” が与えられた場合、 H_f は式 (4) であり、 H_i は (5) により求められる。

$$H_f = \begin{pmatrix} 1 & 0 & 0 & u \\ 0 & 1 & 0 & v \\ 0 & 0 & 1 & w \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (4)$$

$$H_i = \begin{pmatrix} 1 & 0 & 0 & u/f \\ 0 & 1 & 0 & v/f \\ 0 & 0 & 1 & w/f \\ 0 & 0 & 0 & 1 \end{pmatrix}^i \quad (5)$$

拡大・縮小に関する指示 “Make the object $k(k_x, k_y, k_z)$ times.” が与えられた場合、 H_f は式 (6) であり、 H_i は式 (7) により求められる。

$$H_f = \begin{pmatrix} k_x & 0 & 0 & 0 \\ 0 & k_y & 0 & 0 \\ 0 & 0 & k_z & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (6)$$

$$H_i = \begin{pmatrix} k_x^{\frac{1}{f}} & 0 & 0 & 0 \\ 0 & k_y^{\frac{1}{f}} & 0 & 0 \\ 0 & 0 & k_z^{\frac{1}{f}} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}^i \quad (7)$$

x 軸回りの回転に関する指示 “Rotate the object θ degrees around x axis.” が与えられた場合、 H_f は式 (8) であり、 H_i は (9) により求められる。

$$H_f = \begin{pmatrix} \cos \theta & -\sin \theta & 0 & 0 \\ \sin \theta & \cos \theta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (8)$$

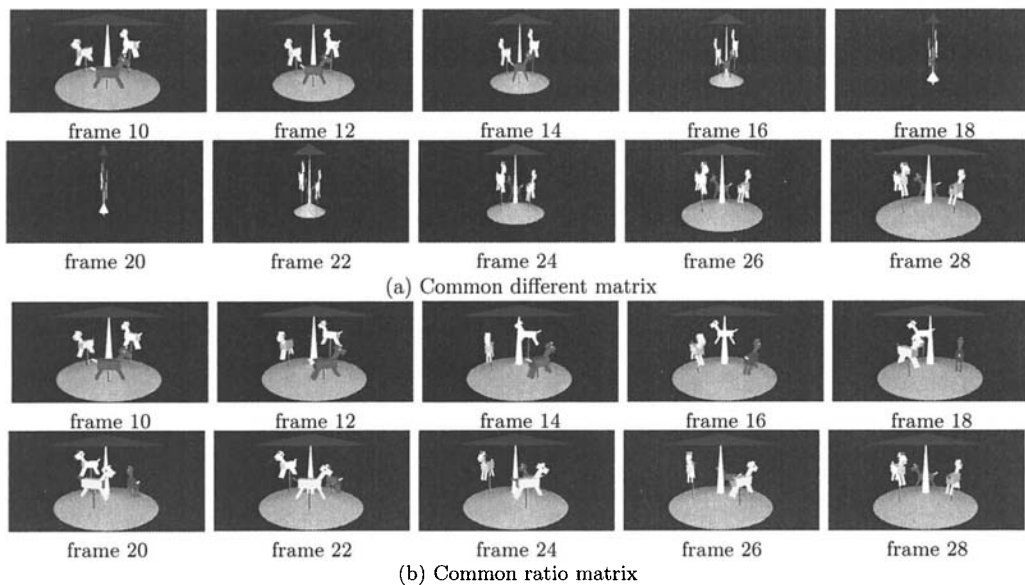


図 3: Examples of rotating a teapot 180 degrees.

$$\mathbf{H}_i = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \frac{\theta}{f} & -\sin \frac{\theta}{f} & 0 \\ 0 & \sin \frac{\theta}{f} & \cos \frac{\theta}{f} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}^i \quad (9)$$

同様に, y 軸回り, z 軸回りの回転に対する \mathbf{H}_i はそれぞれ式 (10), (11) により求められる.

$$\mathbf{H}_i = \begin{pmatrix} \cos \frac{\theta}{f} & 0 & \sin \frac{\theta}{f} & 0 \\ 0 & 1 & 0 & 0 \\ -\sin \frac{\theta}{f} & 0 & \cos \frac{\theta}{f} & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}^i \quad (10)$$

$$\mathbf{H}_i = \begin{pmatrix} \cos \frac{\theta}{f} & -\sin \frac{\theta}{f} & 0 & 0 \\ \sin \frac{\theta}{f} & \cos \frac{\theta}{f} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}^i \quad (11)$$

公比行列においても, 動的指示の場合は \mathbf{H}_f が与えられないため, システム側で \mathbf{H}_i を決定する.

公比行列による回転の例を図 3(b) に示す. 公差行列に対し, 公比行列による変換では $|\mathbf{H}_i| = 1, \forall \mathbf{H}_i$ となり, 剛体運動であるため歪みの問題も存在しない. したがって, アニメーションのための幾何変換手法として公比行列を採用した.

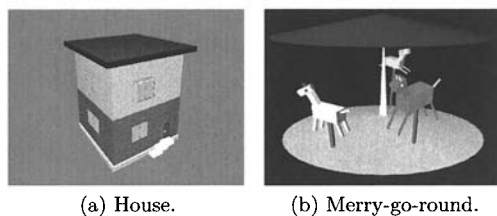


図 4: コストに対する対照実験で作成した図

4.4 任意軸に対する回転とスケーリング

任意の軸に対する回転とスケーリングのためのアフィン変換は式 (12) で実現できる.

$$\mathbf{U}\mathbf{H}_i\mathbf{U}^{-1} \quad (12)$$

ここで \mathbf{U} は座標変換行列である.

4.5 指示の合成

$\mathbf{T}_i, \mathbf{S}_i, \mathbf{R}_i$ をそれぞれ第 i フレームのための移動, 拡大・縮小, 回転のアフィン行列とする. これらの合成例として, $\mathbf{T}_1, \mathbf{S}_1, \mathbf{R}_1, \mathbf{R}_2, \mathbf{T}_2, \mathbf{S}_2, \dots$ がこの順で与えられた場合, これらアフィン行列の合成 \mathbf{H}_i は式 (13) で求められる. これらの積は左から右の順で求める.

$$\mathbf{H}_i = \mathbf{S}_1\mathbf{S}_2 \cdots \mathbf{R}_1\mathbf{R}_2 \cdots \mathbf{T}_1\mathbf{T}_2 \cdots \quad (13)$$



図 5: 雪だるまの見本と提案手法により各被験者が作成した雪だるま

表 1: 筆者の入力テキスト数と作成時間

オブジェクト	コスト	
	静的方式	動的方式
家	作成時間(分)	45
	テキスト数(行)	290
メリーゴーランド	作成時間(分)	145
	テキスト数(行)	1205

表 3: 表 2 のアンケート結果

評価項目	評価				平均
	被験者				
	A	B	C	D	
利便性	4	4	4	3	3.75
視覚的效果	4	4	4	4	4.00

表 2: アンケート評価項目

	評価・理由		
	従来手法	1・2・3・4	提案手法
操作の利便性			
理由			
視覚的效果			
理由			
自由回答			

5 実験と考察

5.1 作成時間と入力テキスト数の比較

5.1.1 実験方法と結果

筆者は従来手法と提案手法の両手法により図 4 に示す家とメリーゴーランドを作成し、作成時間と入力テキスト数を比較した。実験結果を表 1 に示す。

5.1.2 考察

表 1 の実験結果より、家の作成において、提案手法は従来手法に比べ、作成時間と入力テキスト数はそれぞれ約 33%、38% 削減された。また、メリーゴーランドの作成では、作成時間と入力テキスト数はそれぞれ約 14%、32% 削減された。これらの結果から、提案手法は従来手法に比べ、より操作しやすい方式であることが示唆された。

5.2 利便性と視覚的效果の比較

5.2.1 実験方法と結果

操作の利便性と視覚的效果については、四名の被験者に従来手法と提案手法によってそれぞれ同じオブジェクトを作成させ、表 2 に示すアンケートを行った。被験者に作成させたオブジェクトは図 5(a) に示す雪だるまである。実験前に、雪だるまの作成条件として以下の四つの項目を満たすよう指示した。

- 目、帽子、鼻、腕、ボタンを三つ付ける。

- 帽子と腕には傾きを持たせる。
 - 雪だるまの色は各方式で同じ色にする。
 - プリミティブオブジェクトは立方体と球のみ。
- 実験結果を図 5(b)–(e)、表 3 に示す。利便性の評価に対する理由として以下の項目が挙げられた。
- 意図した通りにオブジェクトを変形できる。
 - オブジェクトの変形過程を見ることができ、特に目標状態に近づいたときは微調整しやすい。
- 視覚的效果として以下項目が挙げられた。
- アニメーションは面白く、楽しく作成できる。
- 自由回答で挙げられた意見を項目別に以下に示す。
1. 使用できると便利な操作
 - UNDO 操作
 - 左右対称にする操作
 - 複製する操作
 2. 動的指示・表示方式に対する意見
 - 動的指示による状態変化の速度が遅い。
 - 何倍速という指示を用いてアニメーション速度を変化させたい。
 - 速度を上げる際、“faster”のような短い指示を行いたい。
 - 数値指定したときに途中で止められない。
 3. 画面に対する意見
 - オブジェクトの接触がわかりづらい。
 - 視点位置が変えられない。
 - 画面から出た際の対処が必要。
 4. その他
 - 色の数が少ない。
 - 入力テキストの履歴を使用したい。
 - sphere や cube といったオブジェクトの形状を目的語として使用したい。

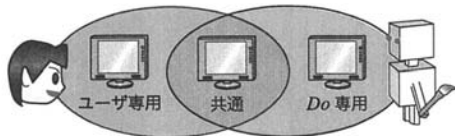


図 6: Do のメンタルモデル

5.2.2 考察

アンケートの結果、操作の利便性と視覚的効果ともに提案手法を高く評価する回答が得られた。ただ、被験者一人が利便性の評価を 3 としており、これは自由回答の意見からアニメーション速度が変更できない不自由さが影響したと思われる。自由回答で挙げられた問題の対策案を同項目別に以下に示す。

1. 使用できると便利な操作

これらの実装は比較的容易であり、入力テキスト数の削減や作成時間の短縮に繋がる重要な操作であるため早急に実装する必要がある。

2. 動的指示・表示方式に対する意見

倍率を用いた速度指示は拡大・縮小と同様の構文解析により実現可能であろう。この操作により速度は自由に変更でき、速度の問題に対処できるようになる。また指示の簡易化、静的指示の変形途中の訂正に関しては、構文中の最重要語を解析する手法、静的指示と動的指示の優先順位についてそれぞれ検討する必要がある。

3. 画面に対する意見

これらの問題は、視点位置を移動させる指示の考案、また、現在の唯一の画面をメイン画面とし、その他に異なる視点からの複数のサブ画面を用いる方式、HMD を用いた VR 方式による解決が考えられる。

4. その他

使用可能な色が少ないという指摘を受けたが、色の追加は容易であるため早急に対処する。入力テキストの履歴に関しては、自然言語による履歴の指示方法について検討が必要である。例として、何回前の指示かを数値で与えることや、前回の移動、拡大・縮小、回転といった特定の動作を再指示することも考えられる。形状を目的語として扱いたいという意見に対しては各オブジェクトに名前と形状を与える必要がある。さらにオブジェクト名と形状名の名前空間について検討する必要がある。

5.3 メンタルモデル

Do のメンタルモデルを図 6 に示す。本実験ではユーザと Do システムとで共通のスクリーンのみを用いたが、例として、複数のオブジェクトが存在し、

全体のバランスを考慮しつつ細部のモデリングを行う必要があるような場合には、ユーザ専用、Do 専用のスクリーンをそれぞれ用いた方が便利であろう。

6 まとめ

従来手法における過大な作成時間や入力テキスト数の問題に対し、動的指示・表示手法を導入した提案手法について述べた。動的方式により、オブジェクトの状態変化を継続的に行い、変化中の状態をリアルタイムで見ながら停止指示を与えるという従来の Do では不可能な操作が可能となった。また、静的方式との対照実験とアンケートにより、動的方式の利便性、視覚的効果の向上が示された。動的方式の考案により、CG による静止画作成ツールとして考えられていた Do は様々なシステムに拡張される可能性も見出された。今後の課題として、表 2 の自由記述で挙げられた問題点の改善、音声入力の実装が挙げられる。また、アニメーションを活かした Do の拡張性についての更なる検討も重要課題である。

参考文献

- [1] 河合善之, 岡田稔: “自然言語による形状モデリングと画像合成の一手法”, 情処論, Vol. 42, No. 5, pp. 1161–1167 (2001).
- [2] S. Mizuno, M. Okada and J. Toriwaki: “An Interactive Designing System with Virtual Sculpting and Virtual Woodcut Printing”, *Computer Graphics Forum*, Vol. 18, No. 3, pp. 184–193, 409 (1999).
- [3] 前野輝, 岡田稔, 鳥脇純一郎: “直観的・対話的自由曲面モデラの変形における制御点調整法”, 情処論, Vol. 45, No. 2, pp. 647–658 (2004).
- [4] 西山晴彦, 大久保達真, 松下温: “Picnyck: 風景描写文から風景画像の創造”, 情処論, Vol. 38, No. 5, pp. 997–1007 (1997).
- [5] S.R. Clay and J. Wilhelms: “Put: Language-Based Interactive Manipulation of Objects”, *IEEE Computer Graphics and Applications*, Vol. 16, No. 2, pp. 31–39 (1996).
- [6] 辻美恵子, 木崎正憲, 岡田稔: “自然言語による CG モデラにおける選択の任意性の基礎検討”, 平成 14 年度電気関係学会東海連大, No. 637 p. 319 (2002).
- [7] 柳井康徳, 岡田稔: “言語指示による 3D-CG モデラにおけるオブジェクト制御の基礎検討”, 平成 17 年度九連大, No. 13-2A-11, p. 545 (2005).
- [8] Y. Yanai and M. Okada: “A Study on a Geometry Transformation Method for a Geometry and Scene Modeler by a Verbal Interface”, *Proc. of GRAPHITE2006*, pp. 221–230 (2006).
- [9] 江村恒一, 青木輝勝, 青木浩: “キャラクタアニメーションにおけるシナリオ駆動型動作合成”, 情処研報, No. 2004-AVM-47-14, pp. 83–87 (2004).
- [10] S. Sekine, R. Grishman: “A Corpus-based Probabilistic Grammar with Only Two Non-terminals”, *Proc. of 4th IWPT*, pp. 216–223 (1995).