

文章表現からの画像生成手法

安倍 裕貴¹ 杉田 薫² 横田 将生²

1 福岡工業大学大学院

2 福岡工業大学

イメージ(心像)を共有する手段として、文字、音声(voice, sound)、画像(photo, graphics)、動画像(movie)が存在する。しかし、このような手段では遠隔地に存在する第3者に自分のイメージを的確に伝えるのは困難である。本研究では第3者とのイメージの共有を促進する手段の確立を目的として自然言語文章が意味する画像を自動生成する手法について検討している。本手法はリレーショナルデータベースとイメージ記述言語 L_{md} を利用して文章表現の意味を解析し、3次元空間上に配置された画像を生成することでイメージの共有を促進する。本稿では、文章表現から意味を解析する手順と現在の実装について報告する。

A Picture Creation Method Reflecting A Meaning of Sentence

Yuki Abe¹ Kaoru Sugita² Masao Yokota²

1 Graduate School of Fukuoka Institute of Technology

2 Fukuoka Institute of Technology

People usually try to share mental images by expressing them in character, audio (voice and sound), picture (photo and graphics), movie or so. However, remote people cannot so exactly share their mental images by these expressions. We have been developing a method for pictorial interpretation of text in order to help remote people in sharing their exact mental images. Our method has employed the mental image description language L_{md} for text meaning representation, which is associated with semantic information of the photos in a relational database to be arranged in a 3D space. This paper presents the flow of text understanding and the current stage of our implementation.

1. はじめに

イメージを共有する手段として、文字、音声(voice, sound)、画像(photo, graphics)、動画像が存在する。特に、インターネット上でイメージを共有する手段として、文字、音声、手話等で表現する手段、動画像を利用した手段、アバター及びエージェントを介する手段が一般的に利用されている。しかし、このような手段では遠隔地に存在する第3者に自分のイメージを的確に伝えるのは困難である。本研究では第3者とのイメージの共有を促進する手段の確立を目的として自然言語文章が意味する画像を自動生成する手法について検討している。本手法はリレーショナルデータベースとイメージ記述言語 L_{md} を利用して文章表現の意味を解析し、3次元空間上に写真を配置することでイメージを共有するための画像を生成する。

2. 関連研究

第3者にイメージを伝える手段に関する研究として、自然言語処理に関する研究が存在する[1][2][3][4]。このような研究では、人

間が日常的に用いる言葉の持つ意味をコンピュータ上で解釈する試みが実施されている。人間が日常的に用いる言葉は構文が厳密に定まっていだけでなく、曖昧な表現を多く含んでいる。このため、真偽のはっきりした命題として解釈できない文章や表現に含まれる意味を文章の前後から読み取る手段が必要となる。このような手段を実現可能な研究としてメンタルイメージモデルとこれを形式化するためのイメージ記述言語 L_{md} が提案[5]されており、その応用例としてクロスメディア変換(異種メディア相互変換)に関する研究[6]が行われている。しかし、これらの研究では知識の格納方法とコンピュータ上での実現方法については言及されていない。一方、データに関する関係表現の簡素化と高速な検索を実現するデータモデルとしてリレーショナルデータモデル[7]が提案されており、現在のソフトウェアにおいて普遍的かつ必要不可欠なデータモデルとなっている。

本研究では最も単純なイメージに対応する軌跡式(原子軌跡式)を利用したクロスメディア変換において、リレーショナルデータモデルに基づいた知識表現の導入を検討している。本稿では、文章表現から意味を解析する手順、

現在の実装および文章表現中に含まれる単語から画像への変換処理について述べる。

3. 文章表現からの画像生成手法

本手法では図1に示すように自然言語による文章から画像を生成するため、文章表現と表形式で表現された知識との照合が行われ、文章に対する軌跡式が生成される。軌跡式生成後、軌跡式で表現された空間的な位置関係に基づいてオブジェクトが画像上に配置される。図1は「猫が寝ころぶ」という文章から画像を生成する例を示している。まず、ユーザから「猫が寝ころぶ」と入力されると、「猫」と「寝ころぶ」という単語の組み合わせに対応した軌跡式が生成される。次に、この軌跡式から画像内に「猫」に対するオブジェクトが配置され、「寝ころぶ」に対する動作が「猫」オブジェクトに与えられる。

3.1 意味解析と軌跡式

原子軌跡式[6]は事物の時間的または空間的状态変化を属性ごとに表現可能なメタファンクションであり、下記の式で与えられる。

$$L(x, y, p, q, a, g, k) \quad (式1)$$

各変数は下記を示すパラメータである

(P1) x : 原因事物

(P2) y : 変化事物

(P3) a : 属性

(P4) p, q : 属性値

‘p=q’ : 状態無変化(維持)

‘p≠q’ : 状態変化

(P5) g : 変化の種類

‘g=Gt’ : 時間的変化

‘g=Gs’ : 空間的変化

(P6) k : 基準となる属性値(基準値)

上記のパラメータによって原子軌跡式には下記の意味が与えられる。

「事物‘x’が事物‘y’の属性‘a’の値をp’から‘q’まで変化させる。この時、‘p’および‘q’基準値‘k’に関して相対的である。また、‘g=Gt’であれば時間的な変化、‘g=Gs’であれば空間的な変化を示す。」

例えば、「バスが東京から大阪まで走る」と「道路が東京から大阪まで走る」の意味はそれぞれ(式2)および(式3)のようになる。

$$L(x, \text{バス}, \text{東京}, \text{大阪}, \text{位置}, \text{Gt}, k) \quad (式2)$$

$$L(x, \text{道路}, \text{東京}, \text{大阪}, \text{位置}, \text{Gs}, k) \quad (式3)$$

文章から軌跡式への変換の流れと軌跡式を利用して意味解析を行う例は図2に示す通りであり、各項目は下記の処理を示している。

(F1)形態素解析：文章中に含まれる単語を抽出する。

(F2)構文解析：単語の品詞に基づいて文法構成を抽出し、文法上の矛盾をチェックする。

(F3)意味構生成：意味表現を生成するため、文法構造に基づいて文章に含まれる各単語を軌跡式に当てはめる。

(F4)意味解析：文章から句を生成し、常識と公理によって意味の妥当性のチェックをする。

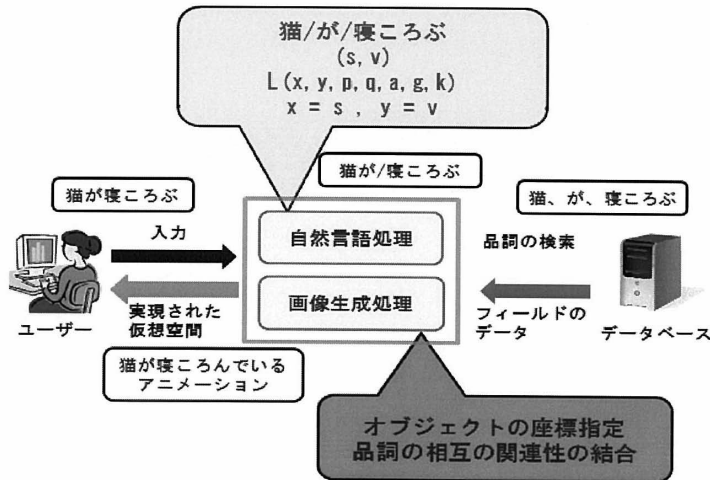


図1 文章表現から画像を生成する流れ
Figure 1 Flow of Pictorial Interpretation of Text

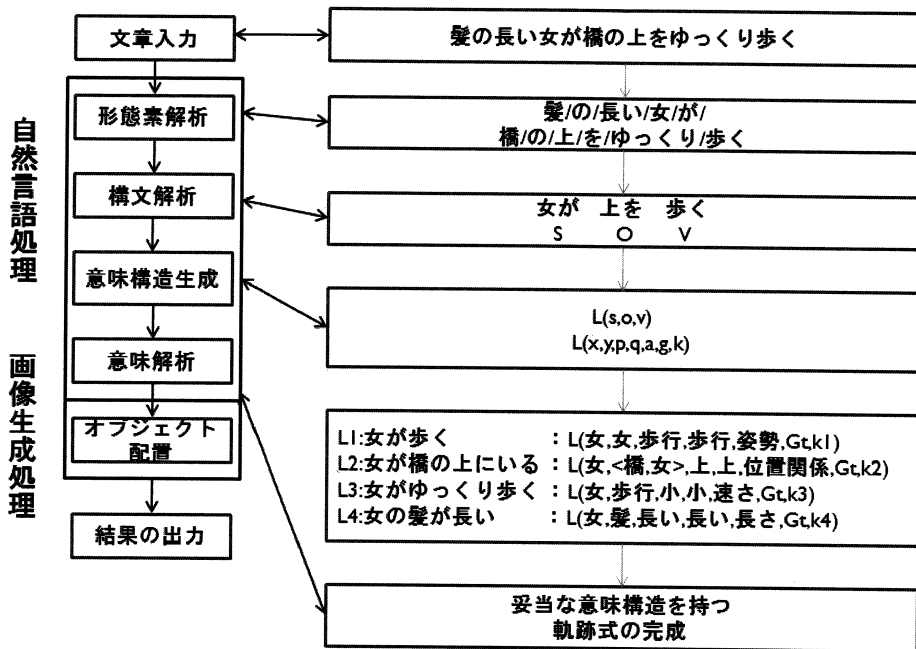


図2 意味解析の例と流れ
Figure 2 Flow of Text Understanding

図2は軌跡式を利用して「髪の毛の長い女がゆっくり橋の上を歩く」という文章の意味を解析する例である。まず、形態素解析によって、この文章から「髪」「の」「長い」「女」「が」「ゆっくり」「橋」「の」「上」「を」「歩く」といった単語が抽出される。次に構文解析によって「S: 女」「O: (橋の)上」「V: 歩く(歩行)」と解釈され、SVO型の文型に当てはめられる。そして、次のように軌跡式に単語を当てはめる。

L1=L(女, 女, 歩行, 歩行, 姿勢, Gt, k1)
(=女が歩く)
L2=L(女, <橋, 女>, 上, 上, 位置関係, Gt, k2)
(=女が橋の上にいる)
L3=L(女, 歩行, 小, 小, 速さ, Gt, k3)
(=女がゆっくり歩く)
L4=L(女, 髪, 長い, 長い, 長さ, Gt, k4)
(=女の髪が長い)

最後に意味解析によって常識や公理に基づいて各句の意味の妥当性のチェックをする。

3.2 画像生成とオブジェクトの配置

文章から軌跡式に変換後、その文章中に含

まれる事物とその状態変化を画像上に表現するためには、各事物の表現に必要なオブジェクトを画像に配置し、その空間的・時間的変化を画像中に反映させる必要がある。また、文章から複数の軌跡式が生成される場合には各軌跡式が表現する事物の時間経過による変化を表現する必要がある。この各軌跡式と軌跡式に当てはめられた句は木構造を用いてそれらの時間的・空間的関係を表現することができる。

なお、名詞や前置詞で表現されるオブジェクトの空間的・時間的関係を描画するために必要なパラメータは事物に関する知識に記述しておくことで、これらを画像内に反映させる。

4. 表構造による知識表現

表構造による知識表現とは自然言語処理で利用される単語や文法とそれらを解析するためのパラメータや画像との関係を表形式で表現したものであり、自然言語処理の各処理で下記の表が利用される。

(1) 形態素解析：単語

単語と事物、属性、品詞との関係が記述

- される
- (2) 構文解析：文型
品詞と語順との関係が記述される
 - (3) 意味構造解析：軌跡式
軌跡式と文型、属性との関係が記述される
 - (4) 意味解析：常識・公理
存在しえない単語間の関係が記述される

表 1 は形態素解析において知識として利用するデータベースの例であり、下記のパラメータが登録されている。

- (T1) ID：単語の重複のチェックやソートを行う為の番号
- (T2) word：単語
- (T3) Type of Media：メディアの種類
- (T4) filename：単語に対応するメディアのファイル名

このような表を利用して、単語とデータベースの照合と画像の選択が実行されることで、文章表現中に含まれる単語の意味に沿った画像の配置が実現可能となる。

5. 実装

本手法に基づいたソフトウェアを実現するために、PowerMacG5 および WindowsXP 上で実装中である。開発言語は C 言語、ライブラリは libjpegv6b[8], OpenGL, GLUT を使用している。libjpegv6b は JPEG 画像の入出力、OpenGL と GLUT はオブジェクトの描画用に使用している。本研究の主体となる表構造に関するモジュールは CSV ファイルをデータベースとして利用可能なモジュールを独自に実装している。

図 3 は現在実装中のソフトウェアであり、「月見には団子が必要だ」と入力した場合に表示される画像を示しており、月と団子が表示されていることが確認できる。このソフトウェアは MacOSX と WindowsXP のクロスプラットフォームでの動作し、下記の機能が実現されている。

- (F1) XYZ 方向への視点操作
- (F2) 単語に対応した JPEG 画像の読み込み
- (F3) 複数の画像の 3 次元空間上への配置
- (F4) 形態素解析

図 4 は実装中のソフトウェアのシステム構成図であり、現在、3DCG インターフェース、自然言語処理モジュール、DBMS モジュールが実装されている。図 5 から図 7 は現在実装されている主要なモジュールである。

3DCG インターフェースは (F1) から (F3) を実現するために下記のモジュールで構成されている。

- (3DCG1) 3D 表示の初期設定：initDisplay3D
- (3DCG2) マウス操作：mouse
- (3DCG3) キーボード操作：motion, key, special
- (3DCG4) JPEG 画像からのテクスチャ生成：createTexImageJPG
- (3DCG5) 描画：display, reshape
- (3DCG6) 画像貼付用パネル生成：makePhotoPanel
- (3DCG7) 形態素解析の実行用メニュー：leftMenu

自然言語処理モジュールは (F4) を実現するために下記のモジュールで構成される。

- (N1) 形態素解析：morphemeAnalysis
- (N2) 単語に対応する画像の描画：putObject

DBMS は CSV ファイルをデータベースとして利用するために下記のモジュールが実装されている。

- (DB1) データベースのオープン：DBOpen
- (DB2) キーワード検索：DBSelect
- (DB3) 検索結果の取得：DBDataFieldFetch
- (DB4) データベースのクローズ：DBCclose

5.1 形態素解析アルゴリズム

図 8 は今回実装した形態素解析のアルゴリズムを示している。このアルゴリズムは、まず、単語候補の最初の文字の位置を文章全体の一文字目に、単語候補の最後の文字の位置を文章の最後の文字に設定し、データベースと照合を行う。データベースに一致する単語がなかった場合には、データベースに一致する単語が見つかるか、単語候補の最後の文字が単語候補の最初の文字の位置に一致するまで、単語候補の最後の文字の位置を一文字分だけ一文字目に近づける。単語候補の最後の文字を走査した結果、データベースに一致する単語が見つからなかった場合には、単語候補の最初の文字を単語候補の最後の文字の位置に一文字だけ近づける。データベースに一致する単語が見つかった場合には、単語を抽出し、次の単語を探すため、単語候補の最初の文字の位置を単語候補の最後の文字の位置 + 1 に設定し、単語候補の最後の文字の位置を文章の最後の文字の位置に設定し、データベースと照合を行う。以上の走査を文章中のすべての文字に対して実行することで文章中に含まれる全ての単語を抽出することが可能となる。

表 1 形態素解析用データベースの例
Table 1 An Example of Database for Morpheme Analysis

ID	word	type of Media	filename
int	string	string	string
0	tukimi	Photo	CU088.jpg
1	syuriboh	Photo	syuri.jpg
2	mega	Photo	TKY200706070333.jpg
3	kasa	Photo	CU051.jpg
4	dango	Photo	CU011.jpg
5	japan	Photo	CU013.jpg
6	koino-bori	Photo	CU026.jpg
7	kabuto	Photo	CU036.jpg
8	tukimi	Photo	CU087.jpg
9	kani	Photo	CU111.jpg
10	kaganimoti	Photo	CU134.jpg
11	hagita	Photo	CU143.jpg
12	mame	Photo	CU163.jpg
13	moti	Photo	CU149.jpg
14	魚	Photo	CU168.jpg
15	弓道	Photo	CU159.jpg

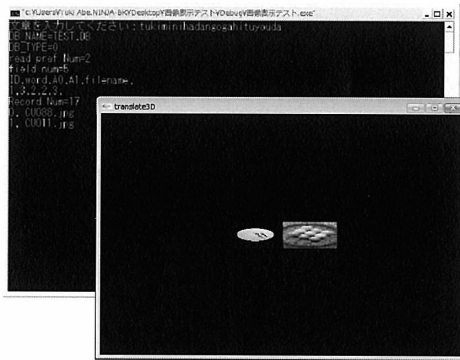


図 3 実装中のソフトウェア
Figure 3 Current Version of Implementation

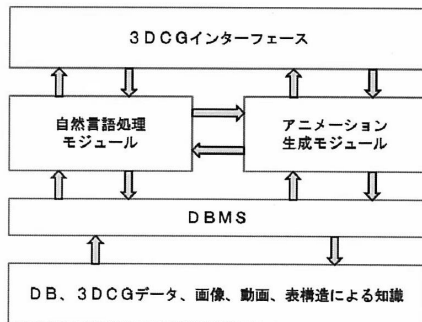


図 4 システム構成図
Figure 4 System Organization

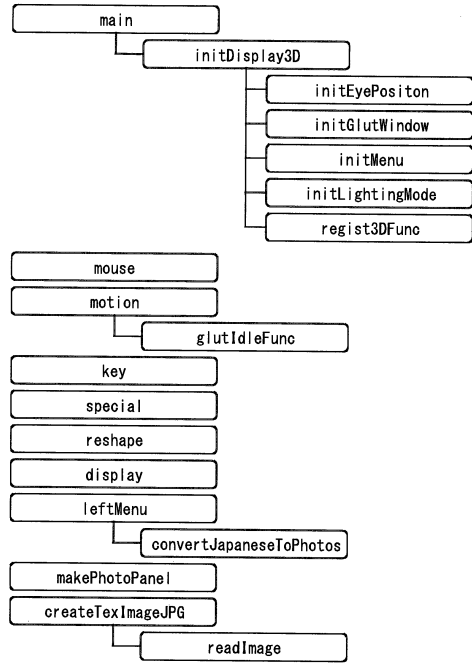


図 5 3DCG インターフェースのモジュール構成
Figure 5 Module Organization for 3DCG Interface

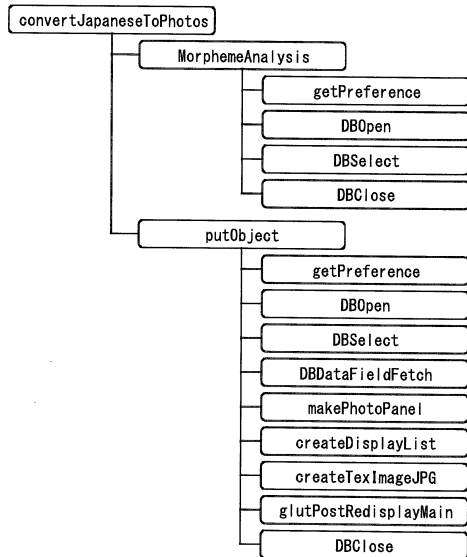


図 6 自然言語処理モジュール
Figure 6 Text Understanding Modules

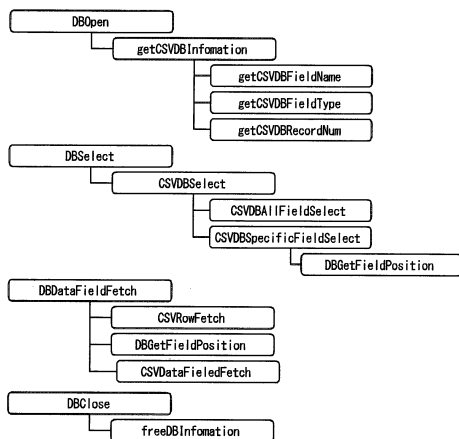


図7 DBMSのモジュール構成

Figure 7 Module Organization for DBMS

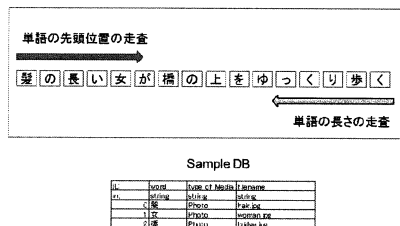


図8 形態素解析のアルゴリズム

Figure 8 Algorithm for Morpheme Analysis

6. まとめ

本稿では知識表現の簡素化と処理の高速化を目指し、軌跡式と表構造による文章表現からの画像生成手法について検討を行い、文章表現から意味を解析する手順、現在の実装および文章表現中に含まれる単語から画像への変換処理について述べた。本手法はリレーショナルデータベースと軌跡式を利用して文章表現の意味を解析し、3次元空間上に写真を配置することでイメージを共有するための画像を生成する。現在、文章中に含まれる単語に関係付けられた画像を3次元空間上に表示可能なソフトウェアを実装済みである。今後の課題としては、知識表現の具体化、意味構造解析の実現方法の具体化、アニメーションの実現方法の具体化、3DCG ファイルや動画画像ファイルへの対応、これを反映したソフトウェアの実現と評価が挙げられる。

7. 参考文献

- [1] 東条 敏, 知の科学 言語・知識・信念の論理, オーム社, ISBN 4-274-20211-9 (2006).
- [2] 清水 浩平, 萩原 将文, 形容詞共起を用いた単語の印象推定法, 電子情報通信学会論文誌 D, Vol. J89-D, No. 11, pp. 2483-2490 (2006).
- [3] 横田 将生, 人間の自然言語理解に関する一つの心理実験, 電子情報通信学会論文誌 D, Vol. J71-D, No. 10, 2120-2127, (1988).
- [4] Yokota, M., A Theoretical Consideration on Robotic Imitation of Human Action According to Demonstration plus Suggestion, Proc. of AISB' 07 (Convention on Artificial and Ambient Intelligence), pp. 307-316, Newcastle University, Newcastle upon Tyne, UK (2007).
- [5] Yokota, M. et al., Mental-image directed semantic theory and its application to natural language understanding systems. Proc. of NLP RS' 91, 280-287 (1991).
- [6] Yokota, M., Capi, G., Cross-media Operations between Text and Picture Based on Mental Image Directed Semantic Theory. WSEAS Trans. on Information Science and Applications, 10-2, 1541-1550 (2005).
- [7] E. F. CODD, A Relational Model of Data for Large Shared Data Banks, Communications of the ACM, Vol. 13, No. 6, 377-387 (1970).
- [8] LibjpegV6b, Independent JPEG Group, <http://www.ijg.org/>