

日本語文からの自然画合成
— 位置関係情報からの画像部品の配置 —

岡本 修作 前田 利之 中川 雅通 前原 文雄
松下電器産業株式会社 中央研究所

Abstract

コンピュータにより多数の画像部品を組み合わせて1枚の画像を合成するには、部品の検索と加工、配置という3つの操作が必要である。本報告では、これらの操作を自動化するために、情景を記述した日本語文を用い、その意味内容を自然な画像に自動的に合成する方式を提案する。本方式は、情景を記述した日本語文を解析して得られた、対象物(画像の合成に用いる画像部品)とそれらの位置関係から、まず対象物名によって画像部品を検索し、さらにあらかじめ与えられている手続きと属性値によって、画像部品の拡大・縮小倍率の計算と配置を行なうものである。本方式では、上下関係の木構造で階層化されたクラスで画像部品を管理し、クラスで手続きを、画像部品で属性値を記述しており、効率的なデータの蓄積が可能である。上記方式に基づき実験システムを作成し、その有効性を確認した。本方式はワープロにおける自然な画像の作画、スタジオにおける背景画の自動合成など幅広く利用可能な技術である。

Natural Image Synthesis from Japanese Sentence Input
— Placing Image Parts from Positioning Relation —

Shusaku OKAMOTO, Toshiyuki MAEDA, Masamichi NAKAGAWA and Fumio MAEHARA
Central Research Laboratories, Matsushita Electric Industrial Co., Ltd.
15, Yagumo-nakamachi 3-choume, Moriguchi, Osaka, 570 JAPAN

Abstract

In order to make a synthesized image from image parts using computer, 3 processes are needed, such as retrieving of image parts from database, processing of image parts, and placing them. We propose a method for synthesizing a natural image based on understanding a Japanese sentence, which describes a scene. This method is described below: At first a Japanese sentence is analyzed and results in information about objects and their positioning relations described in the sentence. Next, image parts corresponding to the objects are retrieved, and these parts are resized and placed, using pre-obtained attributes data and procedures. On this method, image parts, which are described with attribute data, are managed as hierarchical classes, described with procedures, thus it is possible to store data effectively. Based on this method, we have made a prototype and confirmed the effectiveness. This method is widely applicable, for example, drawing a natural image on a word-processor, or automatic image synthesis of background on TV recording.

1 はじめに

近年、ワードプロセッサや簡易印刷器などの普及にともない、さまざまな印刷物でイラストや写真などが頻りに用いられるようになってきた。これにともない、容易に画像を合成できるツールの必要性が高まってきている。

画像を合成するには、画像部品の検索・加工・配置の3つの操作が必要である。特に大規模な画像データベースから高速に目的の画像を検索できる機能は重要であり、例えば、階層構造で管理された画像データベースから、対話的に索引画像をたどって目的の画像を容易に検索できる手法や [3]、それを用いた画像合成システム [1] が報告されている。しかし、従来の画像合成システムは、検索・加工・配置の操作をすべて手作業で行なわねばならないため、画像を素早く合成できないことが問題であった。

これに対し、我々は情景を記述した日本語文を入力とし、その情景を表す画像を自動で合成する方法を提案する。同様のアプローチとしては、情景を描写した言語文と実際の画像との対応関係を分析し、入力文から得られた複数の対象物の空間的位置関係の制約条件を最も満足するような各対象物の位置を、ポテンシャルエネルギー極小化法によって決定しようという試みが行なわれているが [2]、この方法では、自動合成するために必要なデータが複雑なため、膨大な画像に対してその記述や修正が容易に行なえないことが問題となっている。

本報告では、画像部品を上下関係の木構造で階層化されたクラスで管理し、クラスに記述された画像の合成を行なう手続きと、画像部品に記述された手続きを必要な属性値とを用いることにより、入力文を解析して得られる対象物とそれらの位置関係から、対象物名によって検索された画像部品の拡大・縮小倍率の計算と配置の位置の決定を行なう方式について述べる。以下、最初に2章でシステム全体の処理の流れについて簡単に述べた後、3章でデータの管理方式とそれに基づいた画像の合成方法について詳しく述べる。4章で実験的に構築したシステムの概要を述べ実験結果について考察する。

2 処理の流れ

入力文の内容を合成画像として再現するために、全体の処理を、入力文から対象物及びその位置関係の情報を獲得する言語処理と、その情報を用いて換

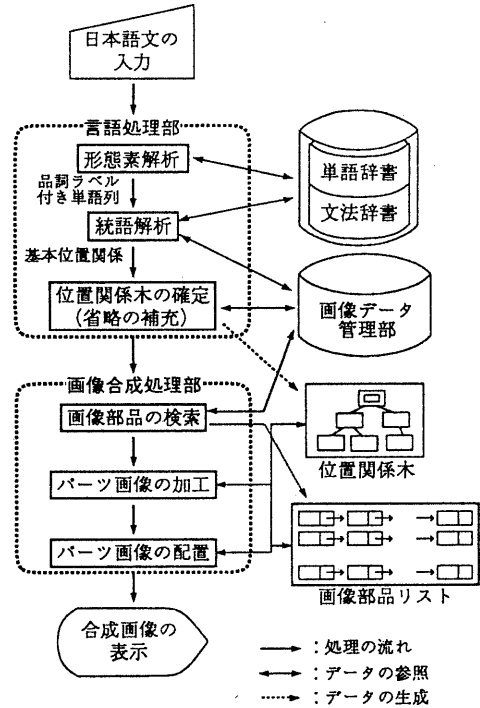


図 1: 処理の流れ

索・加工・合成の操作を自動で行なう画像合成処理の2つに分け、システムを構築している。本章では、まず本方式の全体像をつかむため、日本語文を入力して画像が合成されるまでの手順を、図1に従って説明する。

まず入力された日本語文を形態素解析によって単語列に変換し、各単語の品詞を確定する。続いて統語解析を行ない、入力文に現れるすべての基本位置関係を抽出する。基本位置関係とは位置関係およびその関係にある2個の対象物の3つ組のデータで、第1項の対象物に対して第2項の位置に第3項の対象物があることを示しており、次の形をとる。

$$(\text{対象物} \quad \text{位置関係} \quad \text{対象物}) \quad (1)$$

最後に、得られた基本位置関係から位置関係木を作成する。位置関係木とは入力文から抽出された基本位置関係を統合し木構造で表したもので、上位ノードの対象物に対する下位ノードの対象物の位置を、両ノードを連結するアークに付与された位置関係で示している。以上で言語解析部での処理が終了し、作成された位置関係木が画像合成部に渡される。

なお、入力文において、何らかの理由で対象物や位置関係が省略された場合には、対象物に関する知識を用いてそれらの補充を行なう必要があり、これについては3.2で述べる。また、本システムでは静止画の合成を目標としており、テンス、アスペクトの抽出までを考慮した統語解析は行なっていない。

画像処理部では、まず合成に用いる画像部品が画像データベースから検索される。検索された画像部品は、各対象物ごとにリスト形式で登録される（これを画像部品リストと呼ぶ）。続いて、各画像部品を加工し位置関係木に基づいてそれらを配置していく。合成画像は、背景画像上へ複数のパーツ画像を重ね合わせたものとして表現するため、画像部品はあらかじめ、背景とパーツに分けられて管理されている。加工は検索で得られた画像部品のうちパーツに対して行なわれる。すべての画像部品の加工が終了すると、最後にこれらを位置関係木に基づいて配置し処理を終了する。

3 画像データの管理と画像の合成

2章で説明した処理においては、合成のための手続き、属性値、画像部品（以下これを画像データと呼ぶ）をどのように管理するかが問題となる。

この問題を解決するために、本章では、3.1でオブジェクト指向に基づいた画像データの管理方法について、3.2で画像データを用いて画像を合成する手法について述べる。

3.1 画像データの管理方式

3.1.1 画像データ管理部

現在、我々の扱っている画像部品は、自然の風景と部屋の中の様子を表すために必要なものに限定しているが、将来的には特定の分野に限定されず、多岐にわたったものをめざしており、

- 画像を自動で合成するための手続きが容易に記述でき、かつ拡張性にも優れていること
- 画像の検索や画像のデータの新規登録、更新などが迅速かつ容易に行なえること

の2つの要求を同時に満たすようなデータの管理方式が必要となる。

まず1つめの項目は、システムの実用性を左右する最も重要なポイントである。自然に見えるような

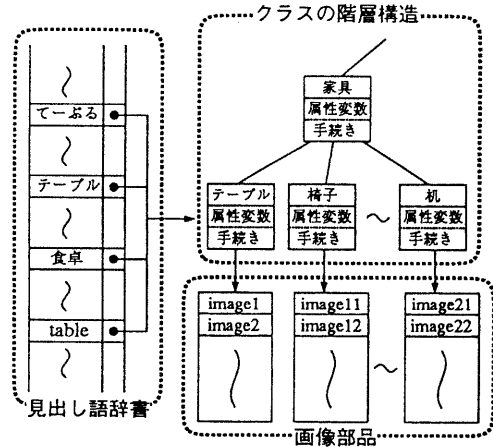


図2: 画像データ管理部

合成画像を作成するには、すべての画像部品に対して、どんな画像部品をどう加工してどこに配置するかをすべて記述できることが望ましい。この方法では微妙な加工の方法や配置の位置など詳細なデータの記述は容易に行なえるが、画像部品の追加とともにその記述量が爆発的に増加するため、拡張性に乏しい。そこで我々は、1つめの要求を満たすための方法として、オブジェクト指向に基づいたデータの管理方式を導入する。すなわち、すべての画像部品は概念の上位下位関係で階層化されたクラスの具体的実体として管理され、クラスには、画像部品を加工、配置するための操作手続きとそれに必要な変数（属性変数）を、画像部品には、その属するクラスで宣言された変数の具体的な値（属性値）を、それぞれ記述しておく（場合によってはクラスで属性変数の初期化が同時に行なわれる）。画像合成の際には、処理に応じて必要な手続きが呼び出され、それらは画像部品の属性値を引数として実行される。この方法では、類似した対象物に対する処理をまとめ、クラスに手続きとして記述するという一般化を行うことができるので、合成に関する知識の記述量が大幅に改善される。また、ほとんどの手続きは各画像部品が固有に持つ属性値を参照して実行されるため、合成画像の自然さもそれほど失われることはない。

次に2つめの項目に関しては、見出し語辞書を導入することで解決した。見出し語辞書とは、キーワードとその概念を表すクラスへのポイントとを組にし見出し語データを集めたもので、これによって、任意の対象物を表す名詞を入力文で用いることが可

能となった。

図2に画像データ管理部の基本構成を示す。

3.1.2 手続きと属性値の記述

次に、属性値とそれに対する手続きが実際にクラスと画像部品にどのように記述されるかを、図3を用いて具体的に説明する。図3は、クラス「テーブル」について記述した例である。

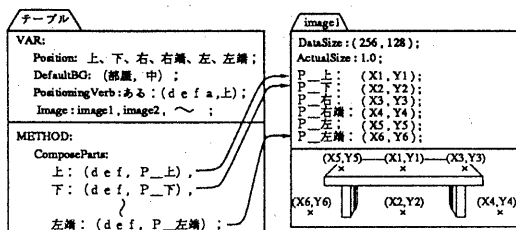


図3: クラスと画像部品におけるデータの記述

クラスは、それに所属する画像部品与える属性値の変数宣言部 (VAR) と、それらの画像部品に対して配置を行なうための手続きの宣言部 (METHOD) からなる。VAR では4つの属性値 Position、DefaultBG、PositioningVerb、Image が宣言されている。Position は、クラスの対象物に対して指すことのできる位置関係を列挙する。ここでは、「テーブル」に対して {上、下、右、左、右端、左端} の位置を指せることを表しており、「テーブル」に属する画像部品の1つ image1 ではそれらの具体的な位置が記述されている。DefaultBG は、クラスの対象物が通常存在する状況、背景を指定する。ここでは、「テーブル」が通常置かれている背景と場所が「部屋」の「中」であると宣言されている。DefaultBG を宣言する時は、該当クラスがパーツであること、参照するクラスが背景であることを確認する必要がある。PositioningVerb は動詞による位置が省略された場合に、補充される位置関係のデフォルト値を指定する。Image は、クラスの対象物を表す画像部品のファイル名を列挙する。次に METHOD では手続き ComposeParts が宣言されている。ComposeParts は位置関係別にそのクラスの画像部品に対して、他のクラスに属する画像部品を配置する操作が記述されている。

以上で説明した属性値は、従属属性 (クラスに宣言された変数に対応した属性値) であり、この他に、絶対属性 (クラスに依存せずすべての画像部品に与

えられる属性値) DataSize、ActualSize が記述されている。これらは、スケーリングの倍率を計算する手続き CalcZoomRatio で用いるもので、3.2.4で詳しく説明する。

以上、本方式でとりあげた手続きと属性値を、表1にまとめた。

表1: 本方式で扱う手続きと属性値

属性の種類	手続き	属性変数
絶対属性	省略補充	なし
	検索	なし
	加工 (CalcZoomRatio)	・ X 方向と Y 方向の画素数 (DataSize) ・ 対象物の物理的な大きさ (ActualSize)
	配置	なし
従属属性	省略補充	◎デフォルトの背景 (DefaultBG) ◎動詞の省略による補充位置 (PositioningVerb)
	検索	◎クラスに属する画像部品 (Image)
	加工	なし
	配置 (ComposeParts)	・ クラスで宣言された位置の座標 (Position)

3.2 位置関係に基づいた画像の合成

文の意味を表す画像を自動で合成するためには、合成に必要な対象物およびそれらの位置関係は欠かせない情報である。従って、入力文でこれらの情報が省略されている場合は、言語解析の処理が終了後画像合成に移る前に、クラスや画像部品に蓄積されている知識を用いて省略を補う必要がある。ここでは、いくつかの情景を描写した文例を対象として抽出された2つの場合の省略、すなわち、動詞による位置関係の省略、背景となる対象物の省略、に関して簡単な補充を行なっている。

さて、我々の扱っている画像部品はすべて2次元データで蓄積されているため、奥行き、視点、光源などを考慮し、各部品の立体的な加工や合成された画像の陰影処理などを行なうことは容易ではない。ここでは、合成画像は背景画像上への加工したパーツ画像の重ね合わせ (配置) で表現する。さらに、

画像の合成における各操作も、対象物名をキーワードとした検索、拡大・縮小のみを扱う加工、画像部品に記述された属性値との照合による配置、と単純化し、これらの高速な処理を実現している。

ここでは、3.1.2で説明した手続きと属性値を用いて、上記の省略の補充の方法と、画像を合成する手法について詳細に説明する。なお、以下の処理については、統語解析を行ない入力文の表層に現れる基本位置関係が得られたものとしており、統語解析の手法については文献[4]を参照されたい。

3.2.1 動詞による位置関係の省略の補充

入力文において動作主体の位置が動詞によって共起される場合、動作主体の位置を示す単語はしばしば省略される。例えば、「部屋の中に机がある」「部屋に机がある」の2つの文において、ほとんどの人は同じイメージを抱くであろう。それは、前の文で、「机」の「ある」位置が、「部屋」の「中」と明示されているのに対し、後の文では、「ある」という動詞が「部屋」「机」という対象物によって「中」という省略された位置関係を共起したからである。このように動詞によって省略された位置関係を補充するために、我々は各クラスで PositioningVerb という属性値を用意した。以下で図4と例文2を用いてその方法を説明する。

「テーブルにりんごとみかんがある」 (2)

まず、例文2から次の基本位置関係が抽出される。(この場合、位置関係がないときに動詞がその役割を担うことを文法規則に記述しておく)

(テーブル v ある りんご)
(テーブル v ある みかん) (3)

これは「テーブル」に対する「りんご」と「みかん」の位置が動詞「ある」で表されていることを意味するもので、実際にはこれがどのような位置関係になるかを、それぞれのクラスの PositioningVerb で調べる。PositioningVerb では、動詞「ある」の場合は動作の対象が何であっても (defa) その位置関係は「上」であると記述されているので、基本位置関係3は

(テーブル 上 りんご)
(テーブル 上 みかん) (4)

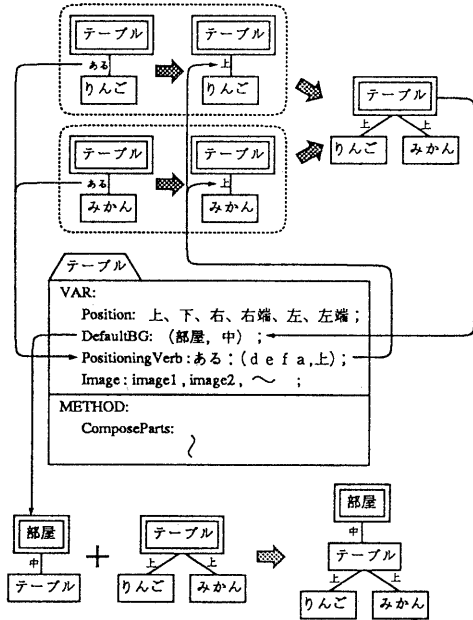


図4: 省略の補充

に変更される。さらに、他にもそのような基本位置関係あるかを調べ、なければ位置関係の補充処理を終了する。なお、このような知識表現を用いた動詞「ある」による位置関係の補充は、他の動詞による位置関係の省略にも適応することが可能である。例えば「道路を人が走る」という文を画像に表すには、クラス「道路」において

PositioningVerb: 走る:(defa, 上);

と書いておけばよい。

3.2.2 背景の補充と位置関係木の確定

省略された位置が補充されると、位置関係木を次のようにして作成することができる。まず、すべての基本位置関係に対して、最初の項の対象物を上位ノード、最後の項の対象物を下位ノード、真ん中の項の位置関係をアークに置き換える(これを基本位置関係木と呼ぶ)。基本位置関係木が1つしかない場合はそれを位置関係木として処理を終了する。そうでない場合は、任意の2つの異なる木に対して同じ対象物を表すノードを一つにまとめる操作を、木が1つになるまで続け、出来上がった木を位置関係木とする。

さて、作成された位置関係木の根ノードとなる対象物は、他のすべての対象物の位置を支配しており、画像合成における基準となる。従って、位置関係木の根ノードとなる対象物のクラスは背景でなければならぬ。位置関係木の根ノードが背景かパーツかは、そのノードの対象物のクラスに DefaultBG が宣言されているかどうかで容易に判断できる。もし、背景でない場合は次のようにして位置関係木を修正する。まず、DefaultBG で背景と位置を参照し、その背景を上位ノード、位置をアーク、現在ある位置関係木の根ノードを下位ノードとする基本位置関係木を作成する。この基本位置関係木と元の位置関係木を位置関係木の作成の時と同様の操作でまとめればよい。

3.2.3 画像部品の検索

画像部品の検索は、対象物の形状、色などの属性は用いず、次の手順で簡単に行なう。まず、対象物を表す名詞を見出し語辞書から探しだし、見出し語データのポイントをたどって該当クラスを検索する。そして、このクラスに属する画像部品をすべて画像部品リストにロードする。これを、入力文の解析で得られた位置関係木のすべてのノードに対して行なう。画像部品リストは、再度画像を検索する手間を省くためにデータを一時的に記憶しておくもので、各対象物に1個づつリストが対応している。実際の画像の加工、配置においては、各対象物のリストに登録されている画像部品のどれか一つを用い、他の部品との交換は、その部品が登録されているリストの他の部品と順次入れ換えればよい。

3.2.4 パーツ画像のズームング

パーツ画像を背景画像上に配置したときに、パーツ画像の大きさは合成画像の自然さを出すための大きな要因の一つであり、パーツ画像のズーム率を計算する式は、任意に選ばれた背景画像とパーツ画像の組み合わせに対して比較的妥当な値が得られる一般性の高いものでなければならない。ここでは、位置関係木の根ノードの画像部品 (IM_S) に対するその他のノードの画像部品 (IM_D) のズーム率 r を式5で定義する。

$$r_{cal} = \sqrt{\frac{X_S Y_S}{X_D Y_D} \frac{D}{S}}$$

として

$$r = \begin{cases} r_{max} & (r_{cal} > r_{max}) \\ r_{cal} & (r_{min} \leq r_{cal} \leq r_{max}) \\ r_{min} & (r_{cal} < r_{min}) \end{cases} \quad (5)$$

ただし、 X_S 、 Y_S は IM_S の X 方向 Y 方向の画素数 (IM_S の DataSize)、 X_D 、 Y_D は IM_D の X 方向 Y 方向の画素数 (IM_D の DataSize)、 S 、 D はそれぞれ IM_S 、 IM_D の表す対象の物理的な大きさ (ActualSize) を、従ってここでは合成のもとになる背景画像のズーム率は行なわずパーツ画像のみとする。まず式5の1番目の式は、対象の物理的な大きさを用いてズーム率を計算するもので、大きさの比較が正確にできるように背景画像に合わせたパーツ画像のデータ領域の正規化を行なっている。また2番目と3番目の式は、ズーム率にその上限と下限を設定したものである。これによって、例えば非常に小さなズーム率のため対象物が消滅するなど、対象物の画面表示に関する問題を防いでいる。なお式5で用いる絶対属性値 ActualSize に関しては適度な測度が見つかっておらず、現在は数値で表すことのできる簡単な基準を設け、データベースの作成者がその基準に基づいて決定している。

3.2.5 背景画像上へのパーツ画像の配置

各パーツ画像のズーム率の計算が終ると、パーツ画像を背景上に配置する。パーツ画像の配置は位置関係木の根のノードから順に各ノードのクラスに記述された手続き ComposeParts に基づいて行なわれる。ComposeParts は、そのクラスに属する画像に他の画像部品を配置するための操作を記述したもので、配置可能な位置関係と1個以上の配置データ (配置が可能な対象物クラス名と配置の座標値の2項組) からなり、入力データと配置データとの照合によって実際の配置の位置が決定される。

さて、あるアークに着目した場合の上位ノードの画像部品に対する下位ノードの画像部品の配置方法を説明する。まず、上位ノードのクラスの手続き ComposeParts からアークに付与された位置関係と同じものを選択する。次に、選択された位置関係の配置データの第1項に登録されているクラスと下位のクラスとを比較する。下位ノードのクラスが、配置データの第1項に記述されたクラスもしくはその下位のクラスと一致した場合は、上位の画像部品に対して下位の画像部品が配置可能であるとみなし、

配置データの第2項に示されている座標に下位の画像部品を配置する。それ以外の場合は配置ができないものとみなす。ただし、配置データの第1項で指定されたクラスが defa である場合は、下位の画像部品がどのようなクラスに属するものでも配置可能とする。図5は、例文2に対して作成された位置関係木をもとに、画像を合成する手順を示したものである。

このように、入力された位置に対応する場所を属性値から探しだし、その位置に画像部品を配置する手法により、高速な画像合成が実現できている。

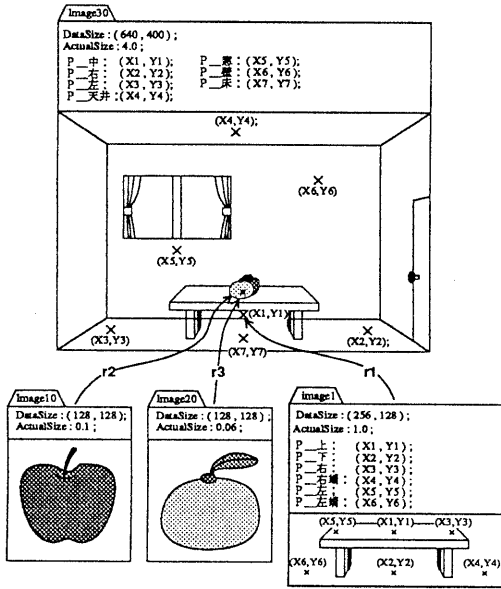
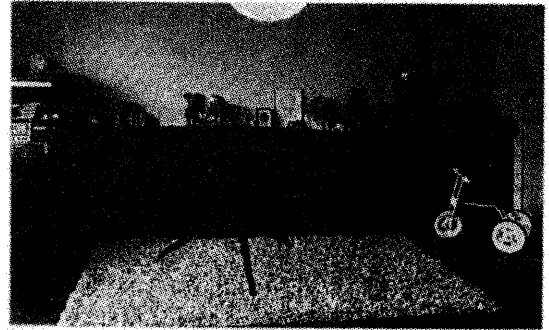


図5: 背景上へのパーツの配置

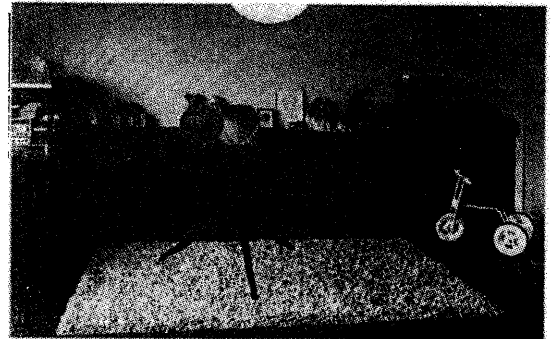
4 実験例と考察

本研究で試作したシステムは、情景を描写した日本語文を解析して、その内容を表す自然画像を合成するものである。実験装置はCPUに80286を搭載したパーソナルコンピュータを用い、言語処理部はProlog、画像処理部はC言語でインプリメントされている。ただし、高速性を要求される画像処理に関しては試作したメディア分散処理ユニットを用いた。実験に用いる画像部品は、背景画像を5種類、パーツ画像を30種類用意し、各画像部品のDataSizeは、合成画像の表示範囲となる背景画像はすべて固定の

値とし、対象物の形状が様々であるパーツ画像は自由に設定できるようにした。以上の構成で、簡単な日本語文を入力して画像を合成する実験を行なった。



「部屋の中にテーブルがある」



「テーブルにりんごとみかんがある」

図6: 画像合成の例

言語文を入力してから合成画像が表示されるまでの時間は、言語の解析はほぼ一定で2.5秒程度、また、画像の処理はDataSizeによってまちまちであるが、背景画像のようにDataSizeの大きなもので約3秒程度、100×100画素程度の小さなものでは1秒以下で、処理時間のほとんどがディスクからの画像データの読み込みによるものであった。図6に画像の合成例を示す。

本システムをデータの蓄積に関与していない多数の被験者に操作してもらい、入力的方式や合成された画像について積極的に意見を述べてもらった。

まず入力方式、すなわち日本語文から画像を合成するアプローチに関しては、「まだまだ使いにくい」という意見が多く、その理由として主に「入力文の文法的な制限が強い」「対話的な合成が行なえない」

の2項目が挙げられた。最初の項目に関しては、文法規則を書き直すことによってかなりの改善が可能である。2番目の項目は、対話文の解析や対話データの管理など難しい課題も多く、現時点では対話システムへの発展は考えていない。このように入力に日本語文を用いたことに関しては、以上のように否定的な意見が多かったが、画像検索に応用すれば有効であるとの意見も得られた。今後、画像合成のための入力文の解析方法をさらに検討する必要がある。

次に合成した画像の内容については、「かなり不自然である」という意見が多く、その理由として主に「陰影が表現されていない」「いろいろなパーツ画像の向きが統一されていない」「位置や大きさが実際の世界に比べて適当でない」の3項目が挙げられた。このうち最初の2項目は視点や光源など空間的に見た場合の不自然さに関するもので、2次元の画像データを扱っている現在のシステムで解決することは難しい。3番目の項目に関しては、3.2.4で述べたズーム率の計算式5 (CalcZoomRatio) の上限値 r_{max} と下限値 r_{min} 、もしくは各画像の属性値 ActualSize が適当でなかったことが原因であると思われる。これに関してはデータのチューニングは難しいが、ある程度の精度でパーツ画像の配置は行なわれているため手作業でパーツ画像の位置や大きさの微調整を行うことは容易であり、本来の目的である画像合成の作業能率の向上という点で本方法の有効性を確認できた。

5 おわりに

情景を記述した自然言語文をその情景を表す画像に変換する方法を提案し、それに基づいた画像合成システムを試作した。特に本システムの特徴として、動詞によって省略された位置関係、背景を表す対象物の省略の補充について、また、オブジェクト指向に基づいて画像部品のデータを経済的に記述管理し、画像を自動で合成する方法について述べた。特に、位置に対応する場所を属性値と照合することによって、高速な画像合成を実現できた。

しかし、今回報告したシステムは4章でも述べたように、まだまだ解決すべき課題が多く残っている。また、以上の検討は画像のデータベースが非常に小規模な現在のシステムでの場合であり、これが大規模化された時のシステムの性能の劣化、特に階層構造で画像部品を管理する場合、クラスノードの付け換えや新たなクラスの導入が行なえるか、そ

の時に無矛盾性の検証が容易に行なえるか、画像部品の検索やパーツ画像の配置が短時間でなれるか、が問題となってくるだろう。

現在、処理の高速化や高機能化に向けてシステムをワークステーションに移行中であり、同時に、画像合成のための入力文の解析方法や、自動でパーツ画像を配置するための手続きについて見直しを進めている。

参考文献

- [1] 井上, 柴田, 中須: 画像合成のための部品画像ファイルシステム, 電子情報通信学会論文誌 J72-D-2, 11(1989), 1824-1832.
- [2] 山田, 西田, 堂下: 2次元平面におけるポテンシャルモデルを用いた位置関係の推定, 情報処理学会論文誌 (1988), 455.
- [3] 柴田, 井上: 画像データベースの連想検索方式, 電子情報通信学会論文誌 J73-D-2, 4(1990), 526-534.
- [4] 前田, 岡本, 中川, 前原: 自然言語入力による自然画合成(一言語解析方法の検討一), 平成3年情処秋全大 3-371(1991).