

構図が感性に与える影響を利用した 風景描写文から画像を作り出す試み

大久保 達真 大木 直人 寺本 邦夫 岡田 謙一 松下 温
慶應義塾大学 理工学部

長年の美術専門家の知識の蓄積により、風景画像は、用いられる構図により我々に異なった印象を与える。我々は、この構図が感性に与える影響を考慮し、小説などに出てくる風景描写文から風景画像と音風景を作り出す事を試みた。システムは、「山」「鳥」といった風景画像を構成するために必要な様々な画像パーツをデータベースに蓄えており、ユーザの入力をもとに、本論文で述べるメディアの補完を用いて、最適の画像パーツ群を検索し、それらを一枚の風景画像に配置し、同時に音風景を出力する。この技術により、一般ユーザでも、描写文を入力するだけで、簡単かつすばやくマルチメディア情報の製作を行なう事ができる。

Generation of Landscape based on Human Imagination

Tatsuma OHKUBO, Naoto OKI, Kunio TERAMOTO, Ken-ichi OKADA and
Yutaka MATSUSHITA

Dep. of Science and Technology, Keio University

To lessen the time and trouble of Multimedia authoring, we propose the complement of media environment and to apply it to Multimedia Authoring. The complement of media environment supports the user to retrieve media elements matching the user's image easily and composite them. To reflect the user's image on the output picture, the technique using the knowledge about the composition of the painting and the fuzzy set is discussed. We also implemented the system to demonstrate complement of media environment.

1 序論

近年の情報機器の高性能化や低価格化, WWW(World Wide Web)の普及により, マルチメディア情報発信の機会が増加してきた. これに伴い, 一般ユーザでも簡単にマルチメディア情報を創作できる制作環境の実現が期待されている. 現在, 一般ユーザのマルチメディア情報の制作は, まず, スキャナーで取り込んだ画像や CD-ROM などのデータベースにより提供される画像や音の中から利用したいものを収集し, 次に, エディターを用いて書いた文章に貼り付ける, という二つの作業過程に分類される. 前者は, マルチメディア情報の検索であり, 後者はマルチメディアオーサリングである. 一方, 現在のマルチメディア情報制作が, 複雑で時間のかかる退屈な作業であると報告されている [1]. それに対し, マルチメディアオーサリングの分野では, さまざまなハードウェア, ソフトウェアの統合, 動画や音声の同期の問題やオーサリング方法などの議論がなされている. 我々は, 特に, 一般ユーザのマルチメディア制作環境における検索作業の問題について,

- 1) ユーザが収集できるデータが少ないこと 通常データベース (写真集, CD を含む) は有料であり, 一般ユーザが多量のデータにアクセスできるわけではない. よって, ユーザのイメージにあうデータを見つけることが非常に難しい.
- 2) 一般ユーザが簡単に扱える適切な検索手法 現在普及しているデータベースは, 画像や音声データにキーワードやインデックスを付加して, 検索するものである. この検索手法は, 適切な論理式の扱いが困難であることや, 検索結果に満足できない場合がある. また, データ内容の特徴量 (画像における色彩や, 領域など) を用いたり, これとキーワードを組み合わせて検索する手法も提案されている [2]. しかし, データベースの内容について熟知していない一般ユーザにとって, これらの画像特徴量を多用することは困難であるうえ, 音声情報については, 適切な特徴量は提案されていない. さらに, 音声データは, それぞれある一定時間聞く必要があり, 満足できる結果を得るまでに, 時間がかかる.

の二点に注目した. これらの問題点を解決し, 一般ユーザでも, 簡単にマルチメディア情報を制作できる環境として, 以下の点を満たさなければならないと考える.

- 扱いやすい検索キー
- すぐに検索結果が得られる
- ユーザの感性を考慮した検索が行える

本論文では, 以上の要求を満たす制作環境を実現するため, 画像や音声部品ごとに管理し, ユーザのイメージを入力することにより, イメージにあった部品を検索し合成する技術を提案する. これにより, 一般ユーザでも, すぐに, 簡単に, マルチメディア情報を創り出すことができる.

2 メディアの補完を用いたメディア部品の検索と合成

本研究では, マルチメディア情報の制作として, 様々なメディア部品を, ファジィ集合を応用したメディアの補完を用いて検索, 合成し, とくに, 風景描写文から風景画像と音風景を創り出すことを試みた. メディアの補完とは, 人間が想像を膨らます過程をモデル化し, それを検索に応用したものである. 例えば, ユーザがクリスマスカードを書くときについて考える. ユーザが伝えたい内容を思い浮かべると, ふとサンタクロースやクリスマスツリーを思い浮かべる. すると, その背景となる夜景や雪景色が付加され, さらにベルやクリスマスソングのひびきをなど思い出す. 我々は, このように, ある一つのこと (例では, クリスマスという言葉) から, 様々なそれに関する事象を次々に連想し, ユーザのクリスマスのイメージを作り上げていく過程をモデル化した. この過程を, 一つのメディアから他の欠落しているメディアを次々と補い完全なものにするととらえ, メディアの補完と呼ぶことにする.

メディアの補完では, 全ての画像や音声データをメディア部品として管理する. メディア部品とは, 牛の画像なら, 背景を含まない牛そのものの画像であり, 牛の音声なら, 他の音を含まない牛の鳴き声だけのデータである. 汎用の画像データは, 複数の物体や背景が写っているものが多い. しかし, 先に

述べたように、人間の頭の中に管理されている画像は、一つ一つの物体ごとであり、想像を膨らます過程で、それらから一つ一つ取捨選択し、イメージに合うものを合成している。よって、メディアの補完は、これらメディア部品から適切な部品を選択し、それらを適切な位置に構図を考慮して配置し、複数の部品を合成して出力する機能を持つ。この部品ごとの管理により、データの再利用性、検索において部品ごとの比較ができるなどの利点がある。

しかし、入力された言葉にあう画像や音データをメディア部品として検索するだけでは、ユーザのイメージに合うデータが得られたとは言えない。例えば、滝を検索する場合でも、ユーザがどのような場所のどのような雰囲気を想定しているかで、ほしいデータは全く異なる。また、入力されてはいないが、あった方がよいメディア部品もあるであろう。よって、補完されるメディアは、ユーザが入力する場所や季節や時間や、メディアにもたせたい雰囲気などのユーザの感性を考慮し検索したメディア部品に、ユーザが望むであろうと推定される部品を付加したものでなければならない。

場所や時間、感性などユーザの入力を、出力するマルチメディア情報に反映させるため、我々は、ファジィ集合を応用することにした。本研究において、あるメディア部品が、場所・時間・季節・雰囲気によって、イメージを構成するために、必要であるか無いのかを、定義する必要がある。例えば、セミというメディア部品は、夏の風景を表すには必要である場合が多いが、海の風景には必要ではない。そこで、要素が集合に属するか属さないかを明確に定義できない場合に用いられるファジィ集合 [3] を、選択した。

3 風景画像への感性の反映

システムの出力する風景画像と音風景は、感性を反映したものでなければならないが、音風景への感性の反映は、我々の以前の研究で行った [7]。風景画像に感性を反映させる要因としては、構図、色、物体のイメージが挙げられる。構図と感性の関係は、例えば、富士山のような三角形の山が真ん中にあると安定感があり雄大な印象を与えるというように、画像内での物体の輪郭や配置の仕方が、人に与える印

象に影響することである。色と感性に関する研究は多く行われていて、例えば白は清楚な印象を与えるなどで、風景画の着色に応用した例 [4] もある。物体そのものと感性の関係とは、例えば、“牛がいるとのどかな感じがする”などというもので、主に人間の知識や経験に基づくものである。本論文においては、主に構図と感性の関係に注目した。

構図と感性の関係に関する知識は、長年にわたり画家など美術家により培われ、現代では体系化されている。部品画像を中国花鳥画の構図規則に従い配置する研究では、構図知識をモデル化し、同じ部品を複数の異なる構図を用いて配置している [5]。しかし、この研究では、感性と構図の関係については議論されていない。本研究では、線や地平線が人間の感性に与える印象に関する知識を利用した [6]。画像の中の物体の輪郭は、線により構成されている。線には、(直線、曲線)(垂直線、水平線、斜線)(長さ)(太さ)といった幾つもの要素があり、これらが異なった印象を与える。例えば、水平線(広々とした、穏やかな)、垂直線(緊張感)、斜線(動的な)、直線(固い)、曲線(柔らかな)などがあり、太さや長さはその線の性質を強調する。また、直線の組合せで、三角形(正方向)となるものは、安定した印象を与える。この知識を利用するために、あらかじめ画像データベースを構築する際に、各画像パーツの形を解析し、それぞれの画像パーツの直線度、曲線度、水平線度、垂直線度、平均斜度、三角形度を求め、それらの値をもとに、各画像パーツのメンバシップ値を決定しデータベースに記述しておく。このメンバシップ値は、画像パーツ検索時に参照される。

また、風景画像において、空が広いと「広々とした」、空が狭いと「圧迫感」といった印象を与える。これは、画像における地平線の位置の高さと関係している。よって、ユーザの入力した感性語によって、地平線の高さを調節することにする。

ここで、「固い」や「圧迫感」などの構図と直接結び付いている言葉を、「構図感性語」と呼ぶ事にする。

4 システムの構築

システムの概要

我々の構築したシステムは、風景描写文を入力すると、風景画像とそれにふさわしい音風景を出力す

るシステムである。図 1 にシステムの概要を示す。

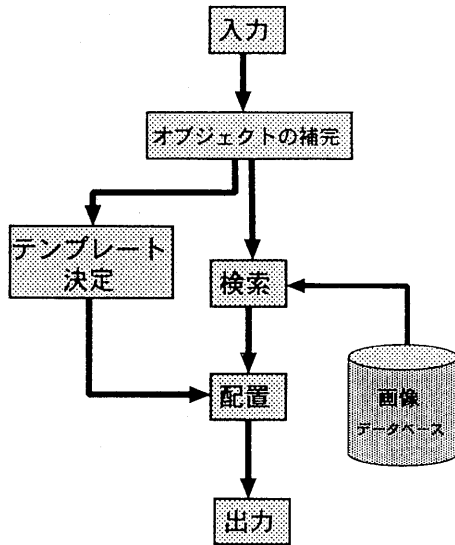


図 1: システムの概要

風景描写文の入力

まず、システムは、風景描写文を理解する必要があるが、自然言語理解の分野は、まだ研究途上である。情景描写文の解釈に関する研究も行われているが、これは我々の主題ではないので、本システムでは、風景描写文から、風景の合成に関係する単語を抽出する方法を採用した。抽出された単語は、山や鳥など部品の名前に関するもの（オブジェクト）、季節、時間、場所などの条件（客観条件）、ユーザーの出したイメージや雰囲気に関するもの（感性語）に分類される。

オブジェクトの補完

次に、ユーザに入力されていないが、その風景にあってしかるべきオブジェクトを補完する。例えば、港の風景には、カモメの鳴き声や灯台、漁船などは、必ず存在する。このようなオブジェクトは、ユーザの入力がなくても、システムにより自動的に付加される。我々は、この補完というものを4つの種類に分類した。

● 基本補完

風景画像を構成するのに必要最低限と思われるオブジェクトを補完する。おもに「空」がある。

● 上位補完

「港」だけが与えられた時に「海」、「セミ」だけ与えられた時に「森」といったように上位（親）オブジェクトにあたるものを補完する。

● 下位補完

「森」があった時に「セミ」、「海」に「船」など、下位（子）オブジェクトにあたるものを補完する。これは必ずしも必要でない。

● 関連補完

「滝」があるなら「川」があるというように、あるオブジェクトがあるなら、これがあるといった場合。

入力された単語群から補完を行なうわけだが、本システムでは補完テーブルというものをを用いている。補完テーブルには、あるオブジェクトと他のオブジェクトとの補完度が格納されている。補完度は0から1までの値を持ち、どの程度補完する必要があるかを意味する。0ならば全く補完する必要がない、1ならば必ず補完する必要がある、0.5ならば場合によっては補完する必要がある事になる。

上述した、基本補完、上位補完、関連補完の場合は、対象となるオブジェクト（例えば入力を「船」とした時の「海」）の補完度は1となる。補完は、入力されたオブジェクト一つ一つについて行ない、閾値を決定してその閾値以上の補完度を持っているオブジェクトを補完する。閾値は、ユーザ側で変更する事ができる。閾値を高くすれば、文中の語句に忠実な風景画像が作成され、閾値を低くすれば想像力豊かな風景画像が作り出されるのである。

テンプレートの決定

抽出されたオブジェクト群にふさわしいテンプレートを選ぶ。テンプレートには、おもに、画像の背景とその背景の中で、画像パーツを配置するにふさわしい位置が、記述されている。例えば、単語群が海に関係の深いものであると判断すると、「海テンプレート」が選ばれ、そこには、海の上に空の広がる背

景、水平線の位置や船やカモメなどの画像パーツを配置するのにふさわしい位置が記述されている。補完された語句群よりテンプレートが決定される。テンプレートは検索キーワードを持っており、単語群とキーワードとのマッチングを行ない、適切だと思われるテンプレートを選択する。テンプレートは木構造になっており、適切だと思われるテンプレートが複数ある場合は、階層が最も下のテンプレートを用いる。

例えば、単語群に「海」「砂浜」があった場合、「海テンプレート」と「砂浜テンプレート」が候補として選択されるが、下位テンプレートにあたる「砂浜テンプレート」を用いる。

画像パーツの選択

複数のオブジェクトをテンプレートにもとづき、一枚の画像と音風景に合成するにあたって実際に風景画像と音風景に出力される画像パーツと音パーツは、特定のオブジェクトの子として管理されている。例えば、いくつかの山の画像パーツが、オブジェクトの山に含まれている。前過程で、選択されたオブジェクトに含まれる画像パーツのうち、最もユーザの入力した感性語と客観条件にふさわしいものが選ばれる。各画像パーツは、すべての感性語と客観条件に対して、メンバシップ値を持っている。感性語と画像パーツのメンバシップ値は、先に述べた構図を反映させたものである。オブジェクトに含まれる画像パーツのうち、もっともメンバシップ値の平均が高いパーツが、実際に合成されるパーツとして選択される。

画像パーツと音パーツの配置

まず、地平線の高さを決める。「広々とした」と似たような感性語が入力されると地平線を下げ、「圧迫感」に似た言葉が入力されると地平線を上げる。次に、それぞれのパーツをテンプレートの情報をもとに配置する。ただし、オブジェクトは、どのオブジェクトの上に置けるか、と言う情報を持っており、(例えば牛なら地面の上、船なら海の上)もしその位置に置けないならその近辺を探す。

音パーツの情報は、画像パーツの配置された位置とともに、我々が以前開発したシステム OKeS [7] に送られ、配置され音風景が構築される。

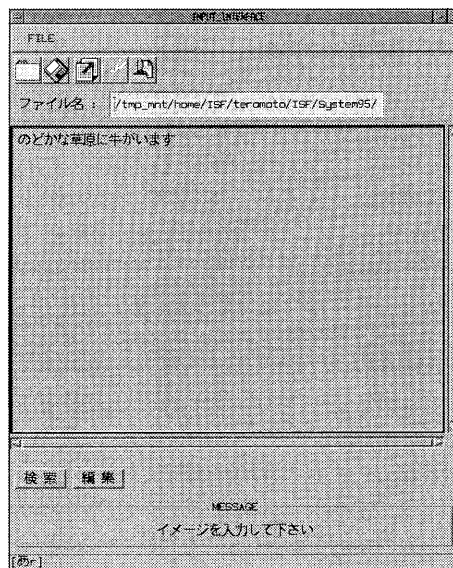


図 2: 入力インタフェース

5 システムの実装と実行結果

本システムは、SUN SPARC STATION の X ウィンドウシステム上で、Motif を用いて開発した。サウンドインタフェースは、OKeS を利用した。

実装したシステムの入力インタフェースを図 2 に示す。ユーザは、テキスト入力スペースに自由に風景描写文を入力できる。

出力インタフェースの例として、「のどかな草原に牛がいます」という入力に対する実行結果を図 3 に、「荒々しい山があります」という入力に対する実行結果を図 4 に示す。

第一文では文中に「空」「山」「地面」という単語は存在しないが、システムでそれらが必要であると判断され出力されている。第一文、第二文を比較してみると、同じ山でも入力される感性語に応じて変化している事が分かる。また、地平線の高さも変化しており、第一文の出力では第二文よりも空が広く現われていて「のどかな」感じを感じさせる。

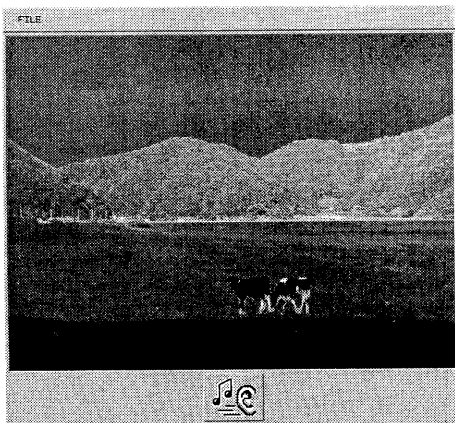


図 3: 入力 : のどかな草原に牛がいます

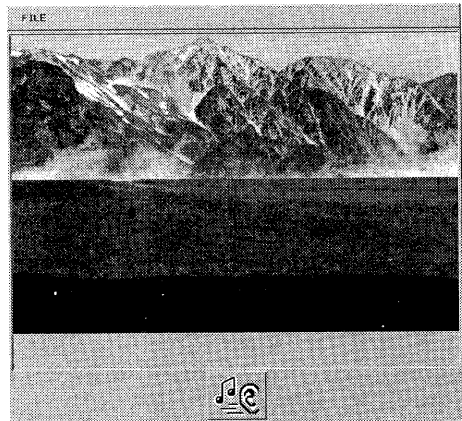


図 4: 入力 : 荒々しい山があります

6 まとめ

マルチメディア情報の製作環境の改善のため、われわれは、メディアの補完と構図知識を利用して、複数の画像や音パーツを検索し、合成する技術を提案した。この技術により、一般ユーザでも、簡単にマルチメディア情報を製作する事が可能になった。

また、ユーザの感性を反映することの重要性とその実現方法として、構図知識を以下に画像パーツ検索に利用するかについて述べたがその効果については、現在、評価実験を行なっている。

マルチメディア情報発信の環境は、急速に改善されてきているが、一般ユーザにとって、従来専門家が行なってきたように思い通りの情報を製作するのは、非常に困難である。この障壁を取り除き、誰でも簡単に思った時に望みの情報を発信できる環境を作り上げるために、本研究が貢献できれば幸いである。

参考文献

- [1] N.Layaida, J.Vion-Dury: *Multimedia Authoring: A 3D Interactive Visualization Interface based on a Structured Document Model*, Symbiosis of Human and Artifact, (1995)
- [2] 高畑, 坂本, 上野, 上森 : ユーザの検索意図を反映した画像提供方式, 信学技法, 1E93-

116,(1994)

- [3] 馬野 : ファジィ集合論とソフトウェア, 情報処理, Vol.30, No.8, pp.922-930(1989)
- [4] 寺野, 増井 他: ファジィ論理を用いた風景画の着色, ファジィ学会誌, Vol.5, No.2, pp.375-385(1993)
- [5] 劉, 青木: 中国花鳥画における構図規則に基づいた画面分割構図法, 信学論 (D-II), J73-D-II, No.10, pp.1696-1706(1990)
- [6] 視覚デザイン研究所 : 構図エッセンス, 視覚デザイン研究所 (1983)
- [7] 寺本, 大木, 阿部, 内田, 岡田, 松下: 人間の感情を考慮した音風景の構築, 情報処理学会全国大会, 第 50 会, 4-327,(1995)