

映像の知識を利用した感性による自動映像創作システム

齊藤 伸介 山岡 一夫 岡田 謙一 松下 温

慶應義塾大学理工学部

〒 223-8522 横浜市港北区日吉 3-14-1

{saito,yamaoka,okada,on}@mos.ics.keio.ac.jp

概要

3次元CGの発展にともない、3次元モデルやそれらを組合せた3次元仮想空間を構築することが可能となり、仮想都市などさまざまな応用分野が検討されている。将来、一般ユーザでも、これらの3次元モデルを用いたコンテンツを制作し、情報発信に用いたいというニーズが増加すると考えられる。3次元CGを用いて映像を制作する時の問題点として、仮想カメラの多くのパラメータを扱わなければならないことと、制作者の意図を映像に反映させることが困難であることがあげられる。本研究では、これらの問題を解決するために、「堂々とした」や「軽快な」といった抽象的な要求を理解し、その感性を再現できるように仮想カメラを制御するシステムを提案する。このシステムは、感性と映像の関係を知識として持っており、簡単に仮想空間内において撮影をすることができる。また、感性語を入力できる機能の有用性について評価したところ、3次元CGの仮想カメラの操作は一般ユーザにとって依然として困難なことであり、このような機能が映像創作の大きな助けになることがわかった。

The automatic image creative system which uses the knowledge of the image

Shinsuke SAITO, Kazuo YAMAOKA, Kenichi OKADA, Yutaka MATSUSHITA

Faculty of Science and Technology, Keio University

3-14-1, Hiyoshi, Kouhoku-ku, Yokohama, 223-8522 Japan

{saito,yamaoka,okada,on}@mos.ics.keio.ac.jp

Abstract

This paper describes the system for producing an image in 3-D virtual space. The system allows the user to simply tell the feeling they want to express, and controls the virtual camera automatically to produce an image that reflects the feeling. After the user inputs the impressionistic word, the system decides the composition of the image. This system holds the image knowledge which is a collection of the technique for composing the image. The image knowledge in this system was extracted from the knowledge of artists.

1. まえがき

近年の情報処理機器の高性能化にともない、3次元コンピュータグラフィックス(3次元CG)を用いたアプリケーションが注目されている。この3次元CGの発展は、3次元モデルやそれらを組合せた3次元仮想空間を構築することを可能にし、仮想都市や仮想オフィスなどさまざまな応用分野が検討されている。[1] 将来、一般ユーザでもこれらの3次元モデルを用いたマルチメディアコンテンツを制作し、情報発信したいというニーズが増加してくると考えられる。たとえば、ある商品の企画を上司にプレゼンテーションするときに、その商品の設

計データなどから作成した3次元モデルがあれば、自分で映像を作成し、資料として提示できる。また教育現場では、古代遺跡の3次元モデルを用いて教師が、独自の視点から映像を作成することによって、生徒により分かりやすく、強く印象づけることができるだろう。

このような3次元仮想空間に限らず、家庭用のビデオカメラを用いて撮影することを考えても、制作者の意図を反映した映像を作成できないことが多い。また、登場人物の微妙な心の移り変わりなどの演出を映像に反映させることは難しい。それは、映像に関する専門の知識を持たない一般の人々は、自分の感性をどのように映像に反映させてよいのかわからないからである。

本研究では、3次元仮想空間を用いて映像を制作するにあたって必要となるであろう「威厳を感じさせたい」「ふわっと浮かせたい」などの、ユーザが映像で表現したい感性を理解し、その感性が再現できるような映像撮影をおこなうように仮想カメラを制御するシステムを提案する。我々は、ある映画の1シーンを見たときに、その被写体が大きく映し出されていれば「迫力、偉大さ」を感じ、また逆に小さければ「不安感、卑小感」といった印象を受ける[2, 10]。本研究では、さまざまな映像の特徴のうち、映像撮影におけるカメラワークと関係が深いものを利用する。これらのカメラと被写体の相対的な位置、画面上での被写体の方向などのパラメータと心理的効果の関係をモデル化し、知識としてシステムに組み込む。

我々のシステムは感性と映像の関係を知識として持ち、映像に詳しくない一般のユーザでも感性を反映した映像を得ることができる。

2章では3次元仮想空間における仮想カメラの基礎的な知識について整理し、3章では実際の映画におけるカメラワークとその心理的な効果を説明する。4章では我々が行った調査に基づき構図の知識を獲得する方法を述べ、5章では実装したシステムについて説明する。本システムはこれらの知識を利用して仮想空間内において3次元モデルを動かし、アニメーションを出力する。

2. 感性情報処理の研究

様々なメディアから感性情報を自動的に抽出しようという試みは多くなされている。

八村らは絵画から主要な色を抽出する画像処理手法について検討している[3, 4]。絵画の背景となっている部分の色と、絵の中で主にモチーフ（主題）を表現するのに使われていると考えられる主要色を自動的に抽出し、対象の絵画をいくつかの主要な色と、その空間的な分布で表現した。それぞれの色にはある種のイメージや印象が伴うことが知られており[5]、使われている色の印象語からその絵画の印象を近似的に表現できるとしている。

田中らは人間が画像のどの部分に注目しているのかを考慮して、画像から感性的な要因を的確に抽出することを試みている[6]。画像に適応して複数の画像処理手法を選択的に使用し、画像に使われている色のヒストグラムの特徴から、局所的あるいは大局的に代表色を抽出し、印象語と結び付けている。

鄭らは、画像から感性表現を抽出するには画像に含まれる「色彩」「モチーフ（画像の主題）」「構図」の3種類が重要であるとし、特に構図に関しての情報を自動的に抽出する研究を行っている[7]。

画像そのものから得られる情報だけでなく、画像に付加されているテキスト情報から感性的な情報を抽出する試みもある。渡辺らは絵画の解説文で対象情報・感性情報がどのように現れるのかを調査し、それを抽出するために必要な語彙の知識と構文的知識を整理した[8]。画像の内容を説明している自然言語テキストから対象の色の情報をとりだし、その情報を利用して画像から対象領域を抽出している[9]。

現在、マルチメディアコンテンツを制作するための各種ツールが市販されており、これらの機能を駆使すれば誰でも専門家並に優れた作品を作成することができる。

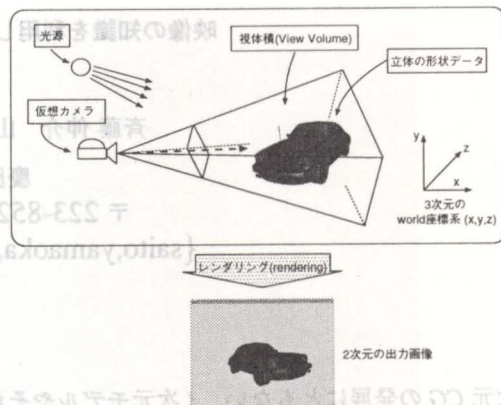


Figure 1. 3次元コンピュータグラフィックス

将来、一般のユーザでも、これらの3次元モデルを用いたコンテンツを制作し、情報発信に用いたいというニーズが増加すると考えられる。しかし、自分のイメージを的確に表現するためには、自己表現のための専門知識とその能力が必要である。人間と画像のコミュニケーションのうち、画像から人間へのインパクトは心理学の研究を応用したものが数多く報告されているが、人間が頭に思い描く心象を画像に伝えて表現する研究はほとんど行われていない。

このように、感性情報処理に関する研究は、メディアに含まれる感性を認識する技術が主であり、本研究のように感性をメディアに反映させる研究は、人間の表情合成に関することに見られるだけで、まだ緒に就いたばかりである。

2.1. 3次元仮想空間

3次元仮想空間内においては、そのユーザは視点、すなわち仮想カメラに投影される2次元画像を見ることによって仮想空間を認知する。3次元CGでは、物体の立体的な形状データ（座標で表された点・線・面で構成される）を光源によって照射された光を計算した上で描画(rendering)することにより、2次元の画像データが生成される[11] (Figure1)。

現在、広く普及している3次元CGライブラリでは視体積を操作するモデルとして一般的に

- 視点の位置 (position (x, y, z))
- 視点の方向 (orientation (x, y, z, angle))
- 画面の縦横比 (aspect ratio)
- クリップ面と視点の距離 (near distance, far distance)
- 画角 (height angle)

といった数多くのパラメータがある。仮想カメラを操作するということは、視体積のパラメータを操作することと等価である。仮想空間における映像を創作するということは、この仮想カメラを実世界のカメラのようにユー

ザの思い通りに動かし、実世界と同様な撮影ができることが望まれる。

しかし、これらのパラメータを直接操作する必要が
あるシステムでは、ユーザへの操作負担が大きく、仮想空間内のカメラの制御が容易ではないことが問題となっている。そこで、本研究では、抽象的な制御要求の中でも、ユーザが3次元仮想空間を用いて映像を制作する際に必要となるであろう「スピード感を出したい」や「威厳を感じさせたい」といった、感性語によって仮想カメラを制御するシステムを提案する。このシステムにより、一般ユーザでも感性語などの簡単な入力を行なうだけで、ユーザの意図を効果的に反映させた映像を得ることができる。

3. 映像と感性

本研究では、さまざまな画像特徴のうち、映像撮影におけるカメラワークと関係が深いものを利用する。本章では、どのように撮影したら目的の印象を反映させることができるかというパラメータを、映像作家によって培われてきた知識を分析することにより得ることを考える。

3.1. 映像撮影における感性

写真家や映画監督などの芸術家の作品は、本来彼らの想像力・感性や靈感による部分が多い。しかし、彼らは構図や色彩、照明などの基本的な知識を何も持たず創作しているわけではなく、これらの知識を踏まえた上で効果的に利用しているのである。

画家・写真家・映画監督などの芸術家は、色彩だけでなく、創作者の意図を反映した画面(映像)を構成することに頭を悩ませてきた[2, 10]。映像作家の感性がもっとも問われるのが、主役を画面上のどの位置に、どの方向に向け、どれくらいの大きさで配置するか、すなわち、映像の構図を決定することである。そこで、第3者に自分の表現したいものを伝えるには、全体的な視覚的效果を考えた映像構成が必要となる。映像構成の目的とは、見せたいものを的確に、バランス良く画面に配置し、効果的に動かすことである。

本研究では、芸術家の知識の一部をシステムに組み込むことによって、あたかも映像作家が撮影しているかのような仮想カメラの制御を実現し、一般ユーザでも手軽に芸術家なみの映像の制作が可能となる。

3.2. 本研究で用いる映像と心理効果

本研究では、映像の要素のうち、人間の印象に与える影響が大きい

- カメラのポジションおよびアングル
 - 主役の画面上での移動方向
 - 主役の大きさ
 - 主役のスピード
- をシステムに組み込んだ。

3.2.1 カメラのポジションおよびアングル

カメラのアングルは、低い位置から見上げたローアングル(仰角)、目の高さから水平に見たアイレベル、上から見下ろすハイアングル(俯瞰)、真上から見る垂直俯瞰の4つに分類できる。それぞれのアングルから撮影された画像は、以下の心理効果がある。

ローアングル 威厳、重厚感、尊大さや被写体の高さを強調する。

アイレベル 日常的な視点で、安心感や見る人との一体感が得られる。

ハイアングル やさしさ、おとなしさ、寂しさ、虚弱感などを表す。

垂直俯瞰 航空写真のように位置関係がデザイン化され、線運動が明確化される。

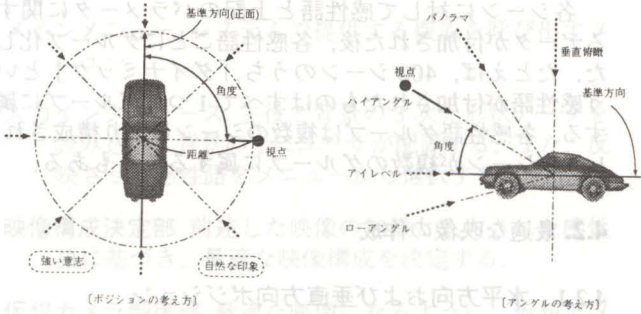


Figure 2. アングルとポジションの考え方

3.2.2 主役の画面上での方向

主役が画面上をどの方向に動くかによって、感性に与える影響が異なることが知られている。[10]

主役の画面上での方向とそこから受ける印象の関係をまとめると、Figure5のようになる。主役が画面上で右方向に動いているときには自然な動きのイメージとなり、左方向に動いているときには意志を持った動きを見ている人に感じさせる。

3.2.3 主役の画面上での面積

画面に占める主役の割合により異なる印象を見る人に与えることができる。主役の面積の違いによって、以下の心理的效果がある。

- 主役の面積が大きい 賑やかで活気がある。
- 主役の面積が小さい 落ち着いて上品な印象を与える。

4. 映像の知識の構築

4.1. 映画を用いた調査

我々は映像を見たときに得られる印象と仮想カメラのパラメータを関連づけるために、以下のような実験を行った。まず、映画20本とコマーシャルフィルムより、主役の動きによって感性を反映していると思われる400シーンを抽出し資料映像とした。これらの資料映像は音響が人間に与える影響を除くために音を伴わないものを用いた。これらの資料映像から受ける印象を被験者が複数の感性語で記述した。被験者は映像についての体系的な知識を持っていない学生20人であった。このとき、映像撮影についてまとめられた文献の中で使用されていた言葉80語と被験者がそれ以外に付け加えた感性語を利用した。実験後、実際に使用された感性語は60語であった。本論文で構築するシステムではユーザにこの60語を提示する。ユーザはこの感性語リストから自分の入力する語を選ぶことになる。次に、各シーン毎の始点・中間点・終点の3ヶ所でパラメータを計測した。計測したパラメータはカメラのポジション・アングル、主役の画面上での方向・面積・スピードであった。

各シーンに対して感性語と上記のパラメータに関するデータが付加された後、各感性語ごとにグループ化した。たとえば、400シーンのうち、「ダイナミック」という感性語が付加されたものはすべて1つのグループに属する。各感性語グループは複数のシーンにより構成され、1つのシーンが複数のグループに属することもある。

4.2. 最適な映像の作成

4.2.1 水平方向および垂直方向ポジション

各感性語グループに含まれるシーンは、それぞれ異なったパラメータを持っている。たとえば、「スピード感」という感性語が付けられたシーンにおける主役とカメラの相対的な位置関係はFigure3の様に分布している。この図は、各シーンにおいて、撮影を行ったカメラと主役との相対的な位置関係をまとめたものであり、各クラスターにいくつのシーンが含まれたかを示している。この図から、「スピード感」という印象を与えるようなシーンは主役の正面より水平やや右あたりから撮影したものが多かった、ということがわかる。

その他の感性語グループにおいても、ポジションとアングルに関してこの例のようにシーンが集中したクラスターが存在する。すなわち、その印象を反映するために芸術家がよく利用しているポジションとアングルである。

我々が実装するシステムでは「スピード感」を与えたいという要求があったとき、この最大値を持つクラスターに含まれるシーンのデータを用いることにした。

4.2.2 画面上における主役の方向

主役の画面上での方向は14クラスターに分類した。しかし、画面上での主役の移動方向はカメラと主役の位置に関係している。たとえば、被写体の正面に向かって左方向から撮影した場合、被写体は画面上で右方向を向くことになる。

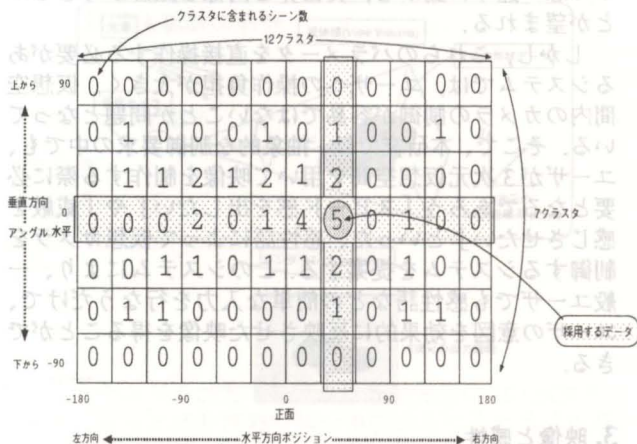


Figure 3. 「スピード感」グループにおけるシーンの分布

そこで、システムでは被写体の画面上での方向を決定する際に、まずポジション・アングルを決定したクラスターに含まれるシーンの中で、最も出現頻度の高い被写体の方向を選択する。

4.2.3 感性語の比較

本調査により、ある感性を与えるようなシーンは、特定のポジション・アングルから撮影したものが多くことが分かった。カメラのポジション・アングルと感性語の関係のうち代表的なものをFigure4に示す。

被写体の右前方からローアングルで撮影すると試聴者に力強さを感じさせ、逆に左上側からならば虚弱感を感じさせることなどがわかった。

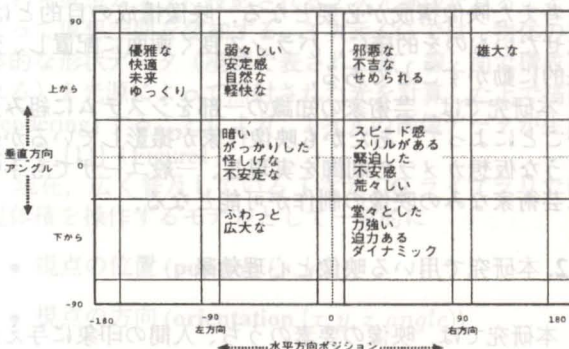


Figure 4. 水平方向ポジション・垂直方向アングルと感性語の関係

また、画面上における主役の方向と感性語の関係をまとめるとFigure5のようになった。Figure5の左の列は主役の方向を示している。真中の列は構図についての説明があった芸術書に記載されていた、主役の方向による印

主役の移動方向	心理的效果 (与える印象)	
	印象 (芸術書より)	アンケート結果
	こちよさ	快適な心地よい希望・未来優雅な
	最も自然な動き	自然なスムーズ軽快なふわっと
	不安感	不安怪しげな無力な
	スピード感	スピード感不吉なスリルがある
	意志の強さ	力強い荒々しい権威的な
	強い意志	堂々としたインパクト豊厚な
	より強い意志	向かってくる迫ってくるパワフル

Figure 5. 画面上の被写体の方向と感性語の関係

象である。右の列は本実験によって得られた結果である。芸術書をまとめたものと本実験の結果は似たような雰囲気を示す感性語の組合せになっている。

4.2.4 仮想カメラの制御

仮想カメラの初期位置は被写体の正面である。ここから、画面上での被写体の方向、面積、領域が入力された感性語により決定される各パラメータを満たすように仮想カメラを移動させていく。

4.2.5 キーフレームアニメーション

本研究ではアニメーションを作成するためにキーフレーム法を用いた。キーフレーム法とはアニメーションの流れの中で動きの主要な変化点を描き、その途中の他数の絵を自動的に補完することにより、一連のアニメーションとして出力する手法である。

本研究において、主要な変化点として 400 シーンの時間軸における始点・中間点・終点の 3 か所について上記パラメータの計測をした。そして、入力された感性語の中で最も多かったデータの組合せをその感性語のデータとし、キーフレームデータとして採用しアニメーションを作成した。

5. 実装

5.1. システムの概要

本システムは、ユーザが映像で表現したい感性語と 3 次元形状データを入力すると、映像の知識に基づいて感性を反映した映像になるように仮想カメラを制御し、アニメーションを出力する (Figure6)。

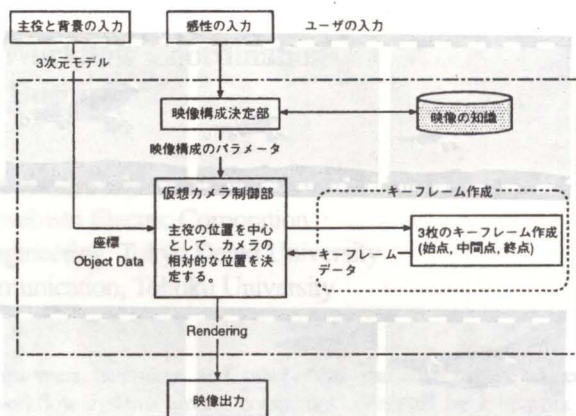


Figure 6. システムの概要

本システムは Silicon Graphics 社の Indy 上で Silicon Graphics 社の 3 次元グラフィック・ツールキット Open Inventor を用いて実装し、以下の機能により構成されている。

メインメニュー システムで提供する機能呼び出すメニュー。

入力インタフェース 主役と背景の 3 次元形状データを読み込み、背景の上にユーザが配置する。また、反映させる感性語をメニューから選択する。

映像構成決定部 前述した映像の知識と入力された感性語に基づき、最適な映像構成を決定する。

仮想カメラ制御部 最適な映像になるように、仮想カメラを制御する。

5.2. 実行例

実際に本システムに感性語を入力して得られた実行例を Figure7 に示す。

「ダイナミック」を入力した場合には、主役が画面右下から左上に動く「意志の強さ」を表すように、主役の前方左下から撮影するようにカメラを制御する。また、感性語を「軽快な」に変更すると、主役が「自然な動き」を表す画面左から右に動くようにカメラ制御を行なう。

5.3. 評価

本システムにより出力された映像がユーザの指定した感性語を反映しているかどうかを調べるための評価実験を行った。まず、被験者に提示する映像 24 シーンを本システムを用いて作成した。このとき作成した映像の主役は Figure7 の様な飛行機を用いた。被験者としては映像についての体系的な知識を持っていない 25 人であった。被験者が画像からどのような印象を受けるかを調べるために、イメージの有力な計測手段である SD (Semantic Differential) 法を用いた。被験者にシステムから出力した映像を見せ、アンケートに回答してもらった。

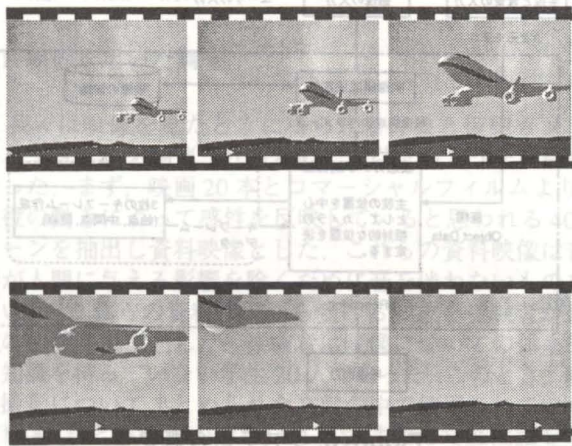


Figure 7. 実行例 入力：ダイナミック

評価の結果、代表的なものを Table1 に示す。この表は左の列が実験に用いる画像を得るためにシステムに入力した感性語であり、右側にこれらの映像に対し被験者が感じた印象を示している。たとえば「ダイナミック」と入力して得られた画像を提示したところ、被験者は「パワフル」「ダイナミック」といった印象を受けたということがわかる。この表から「パワフル」「弱々しい」といった力の強さや弱さを表す力量性、もしくは、「ダイナミック」「軽快な」といった動きの度合を表す活動性に関しては良く表現できている。しかし、「緊迫感」「寂しげ」といった感性語に対しては被験者の感じ方に、ばらつきが生じ評価できなかった。この原因としては、これらの感性はカメラワークが与えるの影響よりもテンポの速い効果音や暗い照明といった別の要因の影響が大きいことが考えられる。

Table 1. 評価結果

入力した感性語	被験者が受けた印象	
スピード感	軽快な	スピード感
パワフル	パワフル	堂々とした
ダイナミック	パワフル	ダイナミック
堂々とした	堂々とした	パワフル
軽快な	軽快な	虚弱
弱々しい	虚弱	不安
緊迫感	評価不能	
寂しげ	評価不能	

6. むすび

本研究では、3次元CGに関する知識を持たない人でも感性を反映した映像創作を簡単に行うために、映像の知識を利用して仮想カメラの制御を行う自動映像創作システムを提案した。まず、3次元CGを用いて映像創作する場合に考えられる問題点について考察した。これら

の問題を解決するために「ダイナミック」などの感性語を入力するとそれを理解し、映像の知識に基づいて仮想カメラの自動制御を行う手法を提案した。我々のシステムは感性と映像の関係を知識として持ち、映像に詳しくない一般のユーザでも感性を反映した映像を作ることができる。システムが出力した映像を評価した結果、力量性や活動性についてはよく表現できていることがわかったが、「緊迫感」「寂しげ」といった照明や音の影響が大きいと思われるものについては表現できなかった。また、感性語を入力できる機能の有用性について評価したところ、3次元CGの仮想カメラの操作は一般ユーザにとって依然として困難なことであり、このような機能が映像創作の大きな助けになることがわかった。

今後の課題としては、作成した映像に効果音の付加、複数のオブジェクトへの対応、色彩や照明の知識の利用、などが挙げられる。

このような知識を組み込んだ映像創作システムを構築することによって、将来、映画の1シーンなども作成可能であると考えられる。

References

- [1] Hagsand, O., Interactive Multiuser VEs in the DIVE System, IEEE Multimedia, Spring, 1996.
- [2] 視覚デザイン研究所編, 構図エッセンス, 視覚デザイン研究所, 1983.
- [3] 八村広三郎, 英保茂: 絵画における感情情報の抽出, 情報処理学会人文科学とコンピュータ研究会, Vol. 24-2, pp. 9-16 (1994).
- [4] 八村広三郎, 英保茂: 色彩分布と印象語に基づく絵画データの検索, 情報処理学会人文科学とコンピュータ研究会, Vol. 27-6, pp. 37-44 (1995).
- [5] 小林重順: カラーイメージスケール, 講談社 (1990).
- [6] 田中昭二, 石若通利, 井上正之, 井上誠喜: 自然画像への印象キーワード自動付加に関する一考察, 電子情報通信学会技術研究報告, HIP96-20, pp. 19-24 (1996).
- [7] 鄭絳宇, 安部憲広: 構図の伝える感性情報の抽出と利用, 文部省重点領域研究報告書 04236107「感性情報処理の情報学・心理学的研究」, pp. 123-126 (1995).
- [8] 渡辺靖彦, 中村裕一, 長尾真: 絵画解説文の対象情報・感性的情報の抽出, 情報処理学会人文科学とコンピュータ研究会, Vol. 20-2, pp. 13-20 (1993).
- [9] 渡辺靖彦, 長尾真: 画像の内容を説明するテキストを利用した画像解析, 電子情報通信学会技術研究報告, PRU95-138, pp. 19-26 (1995).
- [10] PHOTO TECHNIC 編, 画面構成のテクニック, 玄光社, 1995.
- [11] Neider, J., Davis, T. and Woo, M., OpenGL Programming Guide(日本語版), アジソン・ウェスレイ, 1993.