

## 画像の描写に関する研究

井手 一郎, 田中 英彦

{ide, tanaka}@mtl.t.u-tokyo.ac.jp

東京大学大学院 工学系研究科\*

## 1 概要

昨今の計算機および周辺機器の飛躍的な進歩や情報通信基盤の整備などにより、マルチメディア・データベース(特に画像データベース)が注目を浴びている。様々な検索・生成手法が提案されているが、画像という対象を直接扱うことは困難であり、適切なキーを付加するのが現実的だと考えられる。

本研究では、人間による画像の描写の解析を通じ、適切なキーとはいかなるものかを模索し、それにより、効率的な検索を実現するデータベースの適切な構築法に関する知見を得ることを目的とする。

具体的には、描写の階層性という概念を提唱し、描写の多様性との関連を心理実験を通して調べる。

## 2 研究の動機

本研究の動機は、以下に示すような大枠の上に成り立っている。

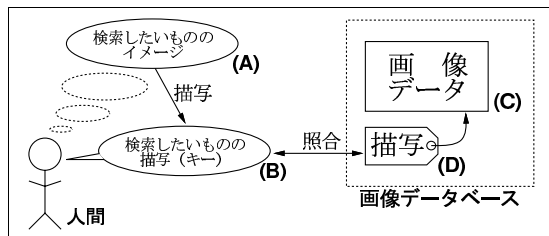


図1: 研究の動機

1. 検索の際には、求める画像のイメージ(A)が頭の中に存在する
  2. (A)を言葉で表現(=描写)し(B)、データベース検索のキーとする
  3. 検索結果(C)は、(A)に近いほどよい
- データベース中の画像データの描写(D)が(B)に近いほどよい
- 良い(D)を生成することが、良い画像データベースの作成につながる
  - 良い(D)に要求される条件は、(B)を研究することにより知見が得られる

∴ 人間による画像の描写の研究が必要

\*"Research on Image Descriptions Required for Multimedia Databases"

Ichirō Ide, Hidehiko Tanaka

University of Tokyo, Graduate School of Engineering  
7-3-1 Hongō, Bunkyo-ku, Tokyo 113, Japan

## 2.1 描写

本研究で「描写」とは、画像の内容を自然言語で表現することをいう。基本的には「[主語]が([目的語]を)[述語]」という形によるものを考える。描写対象は主語や目的語となる可視的事物(名詞)と、述語になる可視的動作・状態など(動詞)となる<sup>1</sup>。

## 2.1.1 描写の多様性

人間による画像の描写を考える際には、描写の多様性を考える必要がある。多様性が生じる理由として、次の2つが大きく関係すると思われる。

- 背景知識 ▪ 個人差が大きい
- 注視点 ▪ 同じ人でも状況によって変わる

ここでは多様性を取り上げたが、共通性も同様に存在し、3章で行なう心理実験を通して、両者の存在を明らかにする。共通性は、画像的に目立ったり、内容的に衝撃的であったりすることにより生じると思われる。

## 2.1.2 描写の階層性

理想的には、どのような画像でも、数多くの描写の集合で表し尽くせると仮定する。また、各々の描写はより細かい描写に分解できるので、階層性をもつ。つまり、画像は理想的には階層性をもつ木構造により描写できると考えられ、この構造を描写木と呼ぶ。このような構造を仮定すると、人間による描写は図2に示すように、描写木の一部を切り出したものとみなせる。

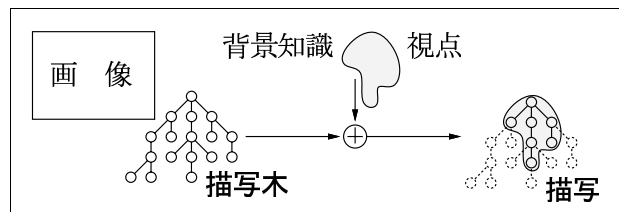


図2: 描写の階層性と人間による描写  
階層性に次の2種類<sup>2</sup>のものがある。

- 概念階層 ▪ ex. 自動車→乗用車→カローラ
- 部品階層 ▪ ex. 自動車→車体→バンパー

これらは必ずしも独立した階層構造ではなく、互いに密接に関連しているが、以下では主に後者の部品階層について取り上げる。

<sup>1</sup>ここでは研究の第1段階として、事物について論じる

<sup>2</sup>事物を例として取り上げるが、動作に関する階層性も同様にある

### 3 描写の多様性と階層性

描写の多様性は描写木のどの部分を切り出すかによって生じると考えられる。以下では、背景知識と注視点によって、切り出す部分がどのように異なるかを心理実験を通して示す。

#### 3.1 心理実験

##### 3.1.1 実験手順

1. 白黒の静止画像を 15 秒間見せる
2. 画像を隠し、印象に残っている事物を 3 つ挙げてもらう → **注視点**
3. 再び画像を見せ、2. で挙げた 3 つの事物に関して、1 行程度で簡単に描写してもらう → **背景知識の度合**
4. 画像中の描写に値すると思われる事物を思いつまままに記してもらう<sup>3</sup> → **被験者の描写し得る事物**

##### 3.1.2 実験結果

以上の手順により、現時点では 2 枚の異なる画像<sup>4</sup>に関して 11 名の被験者の協力を得て、心理実験を行なった。まず被験者が 4. で挙げた、描写に値する事物を元に、描写木を作成した<sup>5</sup>。その上で、注視点、さらには描写木と多様性との関係について解析した。

##### ● 描写木について

描写される範囲は一定の大きさに収まるが、個人差はかなりある。特に、詳細度がまちまちである。表 1 に 4. で挙げられた事物の総数を示す<sup>6</sup>。

画	回答事物数 (全部)	回答事物数 (主要)
1	27	18
2	20	12

表 1: 描写木の大きさ

##### ● 注視点について

2. の回答を元に、注視点を解析する。共通性が非常に強く、画像 2 ではほとんどの人が同様の回答をした。表 2 に注視点の頻度順の累積割合を示す<sup>7</sup>。

画	1 位	2 位	3 位	4 位	6 位	8 位
1	24% 8*1	49% 4*2	—	67% 3*2	79% 2*2	100% 1*7
2	33% 11*1	64% 10*1	88% 8*1	100% 1*4	—	—

表 2: 注視点の共通性

画像中の事物の大きさや位置、コントラストなどの色彩的な問題、幾何学的特徴などが関連して、ある程度共通の反応を引き起こすのだろう。

<sup>3</sup>ここで挙げてもらう事物には 1. で挙げた 3 つを含む

<sup>4</sup>紙面の都合などによりここには掲載しない

<sup>5</sup>階層構造は画像と回答を参考にして筆者が構築した

<sup>6</sup>(主要)とは、1 件しか挙げられなかった事物を除いた数

<sup>7</sup>表の下段は (注視点数)\*(事物数)を示す

##### ● 描写の多様性と共通性

被験者全体の描写によって作成した描写木に対して、各被験者の描写を重ねると、図 2 で予想したような多様性が現れた。実際に 4. で回答があった事物の頻度分布を図 3 に示す。

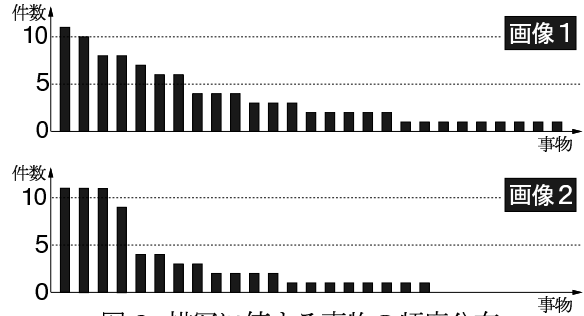


図 3: 描写に値する事物の頻度分布

### 4 結論

実験により、被験者による描写が一定の範囲に収まることが示され、画像中のあらゆる事物を記述した理想的な描写木を用意しなくとも十分であることが分かった。しかし、概念的には同じ事物や動作についても、様々な描写が可能であり、それらを同一とみなすために、概念辞書のようなものを用いる必要があるようだ。

注視点に関しては、実世界では動く事物に注視する傾向が場合によっては強いという面白い現象が見られたが、動く事物でも、人間は比較的注視されにくいという結果も出ており、今後原因を探りたい。

### 5 おわりに

本研究は、効率的な画像データベースの構築を行なう際に必要となる基礎的な技術として、画像をどのように描写するかという問題を扱った。将来的には、画像データに対して、描写を (半) 自動的に付加するような技術が必要になると思われるが、本研究はその際に描写に要求される条件・範囲を明かにすることに結び付く。

以上に示したのは、初歩的な実験結果であり、実験方法やデータ量など、依然不備な点や解析が不足している点が多々ある。また、画像や背景知識の定量的な評価がなされていないために、3. の結果と背景知識の関係についての解析ができていないので、今後定量化の手法を考えなければならない。

### 主要参考文献

1. 「画像の内容検索 – 電子美術館への応用–」  
加藤 俊一, 栗田 多喜夫;  
情報処理 Vol.33, No.5, pp.466-477; May 1992.
2. 「映像の内容記述モデルとその映像構造化への応用」  
柴田 正啓; 電子情報通信学会論文誌 D-II  
Vol.J78-D-II, No.5, pp.754-764; March 1995.