

画像データベースの自然言語インタフェースについて On Natural Language Interface for Image Database

伊東幸宏, 中谷広正

Yukihiro ITOH, Hiromasa NAKATANI

静岡大学情報学部, 浜松市

Faculty of Information, Shizuoka University,
Hamamatsu, Shizuoka, 432

あらまし: 本稿では, 画像データベースの自然言語インタフェースの構築について述べ, その構築経験から人文科学分野の研究・思考を支援するためのユーザインタフェースの枠組について考察する. 画像データベースの自然言語インタフェースの構築では, 自然言語で表現される直観的印象概念を解釈して画像特徴に関する制約を生成するための特徴空間の構築が必要となる. この特徴空間構築プロセスを形状特徴空間の構築を例に説明し, このプロセスが一つの思考プロセスのモデルとなりうることを述べ, 人文科学研究支援システムの一つの可能な形態について述べる.

Summary: In this paper, we describe a method to construct a natural language interface for an image data retrieval system, and discuss a framework of an interface for a computer aided research system. In order to construct a natural language interface for an image retrieval system, we should construct a feature space on which subjective words representing user's intuitive impressions are arranged, and through which the system transforms user's inquiry sentences into representations of restriction on image features. We illustrate the process to construct a shape feature space to retrieve images of chairs. Then we discuss similarity between the process and thinking process of researchers of humanities, and show one possible framework of interface for a computer aided research system.

キーワード: 画像データベース, 自然言語インタフェース, 思考支援

Keywords: Image database, Natural language interface, Computer aided research

1 はじめに

多くの人文科学領域で画像データを含むマルチメディアデータベースが構築されつつある. しかし, それを使いこなす技術, すなわち, 広い意味でのユーザインタフェース技術は, 十分に検討されているとは言えない.

人文科学研究に積極的にデータベースシステムが活用されるようになるためには, まず, データベースの非専門家にとって扱いやすいインタフェースが提供されることが必要である. しかし, 単に表面的な取り扱いにくさを解消するだけでは十分ではない. データベースシステムを積極的に利用してゆくことによって, 人文科学の研究者の思索が触発されるような環境, 言い換えれば, ある種の思考支援機能を含んだ環境が提供されることが望ましい. このうち, 使いやすいインタフェース構築の問題に関しては, GUIの研究を含め多くの研究がなされてきている. しかし, 思考支援機能を含んだインタフェースについては, まだ研究が緒についたところであるというのが現状である.

我々は, これまでに, データベースの非専門家であるユーザにとって使いやすいインタフェースとして, 自然言語インタフェースシステム

を検討してきた. このシステムは, 椅子の電子化カタログから, 色や形の情報を用いて椅子画像を検索するという例題を想定して試作されている. このシステムは, 「柔らかい色」「陽気な形」といったような, ユーザの主観的印象を表す語彙を用いることができる, 「もっと淡い色」「もっとシャープな形」等, 比較表現を用いて対話的に検索を進められる, という特徴をもつ.

このシステムの試作の際に, 我々は椅子の形状特徴についての言語表現を解釈して, 椅子画像から画像解析によって抽出しうる特徴量に関する制約条件を生成するための形状特徴空間の構築を行ってきた. この特徴空間構築プロセスは, データベース中の画像データを直感的な印象概念に基づいてクラスタリングしうる空間と, その空間上での座標値と画像特徴の間の関係を定義する作業である. 我々は, このプロセスを基にして画像データベースを用いた研究・思考プロセスの一つのモデルを想定できるのではないかと考えている.

そこで本稿では, このインタフェースシステムの概要と特徴空間構築プロセスについて述べ, その枠組みに基づいて, 思考支援機能を含めた

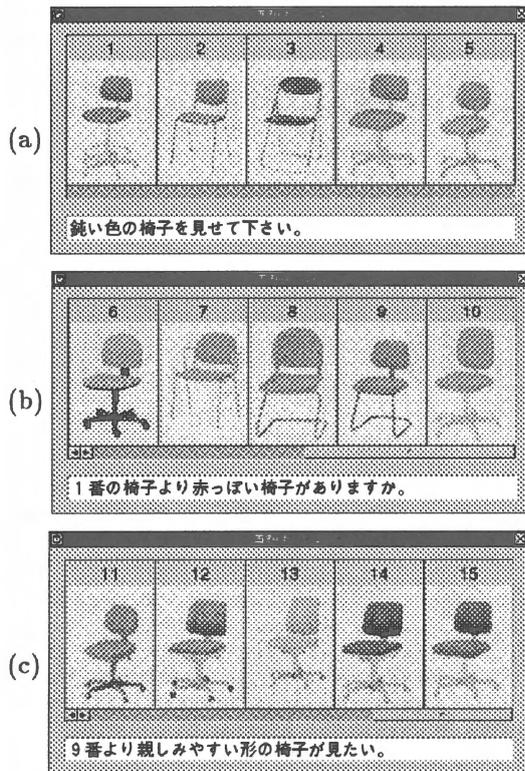


図 1: 検索例

人文科学研究にとって有用なインタフェースシステムへの拡張の方向性について述べる。

2 システムの概要

2.1 動作例

システムの検索例を図1に示す。まずユーザには日本語で欲しい椅子を入力してもらう。例えば、「鈍い色の椅子を見せて下さい。」と入力したとする。すると、システムは5枚の候補画像を提示する(図1(a))。ユーザは続けて、「1番の椅子より赤っぽい椅子がありますか。」と提示された画像との比較表現を用いて検索が行うことができる(図1(b))。提示された画像に満足がいけない場合はさらに続けて、「9番より親しみやすい形の椅子が見たい。」のように、対話的に要求を入力することができる(図1(c))。このように、ユーザは対話的に検索を行うことにより、欲しい画像の検索が行える。

2.2 処理の流れ

次に、処理の流れを簡単に示す。なお、ここでは議論を簡単にするため、色特徴を言及する表現を例にとって説明する。

ユーザから入力された自然言語文は形態素解析、構文解析が施され、意味表現[1]が生成される。この意味表現において我々は、「暖かい」、「かわいい」といった感性語句をそれぞれ、「暖かさ」、「かわいさ」といった感性的属性概念を用いて定義した。

次にこれらの感性的属性概念は色特徴空間上の検索領域に変換される。我々はこの過程を意味表現の解釈と呼ぶ。この解釈は予め感性的属性概念ごとに用意された領域、ならびに、ピーク点に基づいて行われる。この領域とは多くの人がその色に対して感性的属性概念で表されるような印象を持つ色特徴空間上での色の範囲である。また、ピーク点とは最もその印象を表すと思われる色特徴空間上での点である。

最後にシステムは指定された検索範囲内の色をもった椅子画像を探索し、候補画像として提示する。なお、画像中の椅子の色の値は、画像解析を行って自動的に抽出される。

以上で説明した「意味表現の解釈」処理は、ユーザの直感的印象を表現した概念表現を、以下のような特徴空間上での検索条件表現へと変換するプロセス、と捉えることができる。

- (1) 人間の印象を反映する。すなわち、同様な印象をもつ概念同士は空間上においても近い領域を占める。
- (2) 画像解析により特徴空間上の座標が求められる。

このような条件を満たす特徴空間として、色特徴に対しては色相、明度、彩度軸で構成するHSV空間がよく知られている。しかし、形状特徴に対しては、これまでのところそのような空間は提案されていない。そこで、我々は、以下のような方法を用いて形状特徴空間の構築を行った。詳細は第3章で述べるが、我々は、この特徴空間構築のプロセスが、データベースシステムによる人文科学の思考支援の一形態を示唆していると考えている。そこで、次章では、その形状特徴空間の構築プロセスについて述べる。

3 属性空間の構築 - 形状空間の場合 -

我々は、以下のようにして形状特徴空間を構築した。まず、イメージ空間の測定にはSD法[2]を用いることにした。SD法とは、心理実験を基に因子分析をするといった手法であり、抽象的概念から具体的な物品まで、さまざまな刺激が及ぼすイメージ・印象・雰囲気・感情といった側面での、心理効果を測定する方法として知られている。ここで簡単な流れを記しておく。



図 2: 33 種類の椅子

Chair Number 1			
直線的な	—:—o:—:—:—	曲線的な	
落ち着いた	—:—:—:—:—o	落ち着かない	
陰気な	—:—:—o:—:—	陽気な	
:			
まとまりのある	—:—:—:—o:—	まとまりのない	

図 3: アンケート用紙 解答例

Step1 なるべく形の異なる対象 (椅子) の画像を収集する。我々は実験では 33 種類の椅子の 2 値画像を用いた (図 2)。

Step2 <陰気な-陽気な>, <暖かい-涼しい>といった形容詞対からなるアンケート用紙を作成する (図 3)。我々は 20 種類の形容詞対を用いた。また、それぞれの対は 5 段階で評価してもらうことにした。

Step3 被験者にアンケート用紙を用いて OHP により提示された画像の形についてその印象を評価してもらう。我々は 26 名の被験者 (男子学生) に対して実験を行った。

Step4 アンケートの結果から因子分析を行い、イメージ空間の軸を決定する。

表 1 は因子分析の結果である。この結果から椅子の形の心理的意味は、互いに独立した 3 つの因子に依存しているといえる。また、この因子負荷行列を用いて因子スコアを推定し、それをプロットしたものを図 4 に示す。

図 4 は、横軸にそって第 1 因子を、縦軸にそって第 2 因子を表している。この図では左にいくほど「まとまりのある」、「親しみやすい」、「使いやすい」、「落ち着いた」、「単純な」、「上品な」、「地味な」、「安定した」、とい

SD スケール	因子負荷量			共通性
	第 1 因子	第 2 因子	第 3 因子	
まとまりのある-まとまりのない	0.93	-0.04	0.14	0.89
親しみやすい - 親しみにくい	0.92	0.10	-0.08	0.86
使いやすい - 使い難い	0.76	0.13	0.41	0.76
落ち着いた - 落ち着かない	0.76	0.41	0.39	0.90
単純な - 複雑な	0.72	-0.60	-0.00	0.88
上品な - 下品な	0.68	0.39	-0.18	0.65
派手な - 地味な	-0.66	0.19	-0.66	0.91
安定した - 不安定な	0.63	0.50	0.47	0.87
軽やかな - 重厚な	0.02	-0.92	-0.36	0.98
安っぽい - 豪華な	0.23	-0.91	0.01	0.87
暖かみのある - 冷やかな	0.33	0.87	-0.24	0.93
硬そうな - 柔らかそうな	-0.23	-0.85	0.06	0.78
すっきりした - ごてごてした	0.60	-0.73	-0.24	0.94
直線的な - 曲線的な	-0.15	-0.51	0.27	0.35
かわいくない - かわいい	-0.16	0.09	0.93	0.90
陰気な - 陽気な	0.20	0.05	0.89	0.83
かっこいい - かっこ悪い	0.09	-0.04	-0.84	0.71
古臭い - 新しい	0.54	-0.21	0.77	0.93
趣味的な - 実用的な	-0.59	0.19	-0.72	0.90
一般的な - 個性的な	0.66	-0.22	0.68	0.95
Commonality	8.29	5.22	3.73	17.2
共通性 (%)	41.4	26.1	18.7	86.1

表 1: 因子分析結果

た第 1 因子に含まれる印象が強くなることを示している。逆に右方向にいくほど反対の印象が強くなることを示している。また同様に、上方方向にいくほど第 2 因子に含まれる「軽やかな」、「安っぽい」、「冷やかな」、「硬そうな」、「すっきりした」、「直線的な」、といった印象が強くなっており、下方方向にいくほどその反対の印象が強くなっていくことを示している。

このように、因子スコアは因子にどの程度依存しているかを表しているもので、例えばこの図では 27 番の椅子は第 2 因子のプラス方向に対する依存が高い。すなわち、27 番の椅子は「まとまりのある」、「落ち着いた」形をしているという印象を与えると考えられる。逆に、8 番の椅子は反対方向に依存が高いため、「かわいくない」、「陰気な」形の椅子といった印象を与えると考えられる。

次に、我々は構成したイメージ空間と相関の高い物理的特徴空間を構成することにした。まず、我々は意味空間の因子と対応すると思われる物理的特徴概念を抽出することから始めた。

構成したイメージ空間の軸上に椅子を並べ、観察を行い、そこからその軸に影響すると思われる物理特徴を調査した。その結果、我々は椅子の形の印象は主にその椅子の脚の形に依存していると推定した。そこで、椅子の形に基づき 4 つのタイプに分類した (図 5)。タイプ 1 は脚が一本脚の回転椅子、タイプ 2 は 2 本脚の椅子、タイプ 3 は折りたたみ椅子、タイプ 4 は 4 本脚

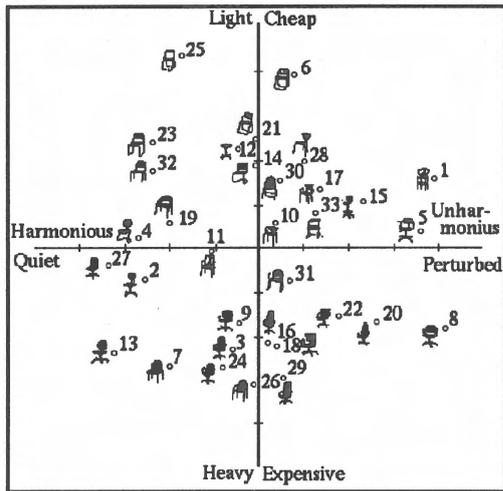


図 4: 構成したイメージ空間 (断面図)
(横軸: 第 1 因子, 縦軸: 第 2 因子)

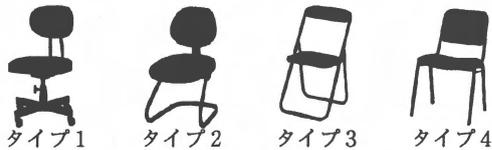


図 5: 椅子のタイプ分け

の椅子, といった4つのタイプに大別した。

しかし, 脚のタイプだけではイメージ空間に対応付けることは難しい。そのため, さらに分類した椅子のタイプごとに調査を行った。例えば, 図 6 は, タイプ 4 の椅子を第 2 因子の軸上に並べたものである。この図から左側の椅子ほど黒い部分の面積が大きいことが分かる。そこで, 我々はタイプ 4 の椅子が第 2 因子に影響を与える特徴として式 1 で表される特徴概念を定義することにした。ここで, 「椅子の面積」は図 7 のように画像を 2 値化し, その画像から抽出される黒い部分の面積を指し, 「椅子の凸包の面積」とは図 7 のように 2 値化した画像の凸包の面積を指す。

$$\frac{\text{椅子の面積}}{\text{椅子の凸包の面積}} \quad (1)$$

他のタイプもそれぞれ同様にして定義した式を表 2, に示す。我々はこれらの式の値をそれぞれ変形させ, イメージ空間軸上の値と対応させるようにした。

図 8, 9 にイメージ空間と形状特徴空間を示す。図 8 に我々の心理実験により構成されたイメージ空間の 3 つの軸を示し, 図 9 に我々が定義し

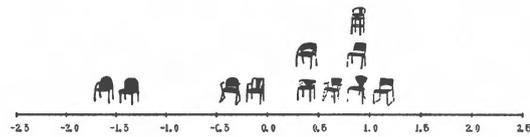


図 6: 第 2 因子軸上のタイプ 4 の椅子



図 7: 椅子の面積とその凸包の面積

た式を用いて算出した形状特徴空間の 3 つの軸を示す。心理的イメージ空間と形状特徴空間のそれぞれ軸の相関係数は, 第 1 因子が 0.99, 第 2 因子が 0.96, 第 3 因子が 0.95 と高い相関が得られた。

4 思考支援に向けて

研究者が画像データベースを用いて行う思考には様々なタイプが考えられるが, そのうちの一つのタイプとして, 次のようなもの考えることができよう。それは, 研究者は雑然と集められたデータ画像を俯瞰すべき着目点を定め, その視点に基づいてデータを整理して, ある種の一般性, 普遍性を見いだす, というプロセスを踏む思考である。このタイプの思考を行う場合, データを常にデータベース作成者が定めた座標軸に拘束されて取り扱うのでは柔軟な発想が妨げられる恐れがある。

データベースを用いた研究者の思索を支援するためには, データを整理すべき独自の座標軸の設定を支援すること, その座標軸を用いて直接データを取り扱えることが必要であろう。

すなわち, 画像データベースを利用した人文科学研究のプロセスモデルを以下のように考える。

- (1) 大量のデータを俯瞰すべき視点についての着想を得る。
- (2) その着想に基づいて, 緒データを整理すべき空間を想定する。その空間上では, 緒データが, 上述の視点から見て特徴的な様々な概念に対応したクラスターを構成し, それらの概念間の関係が明瞭に判断できなければならない。
- (3) そのような空間上に整理されたデータから, 一般的, 普遍的な概念間の関係を導き出す。

このうち, (2) の空間の想定プロセスは, 前述の特徴空間の構築プロセスとほぼ同一のア

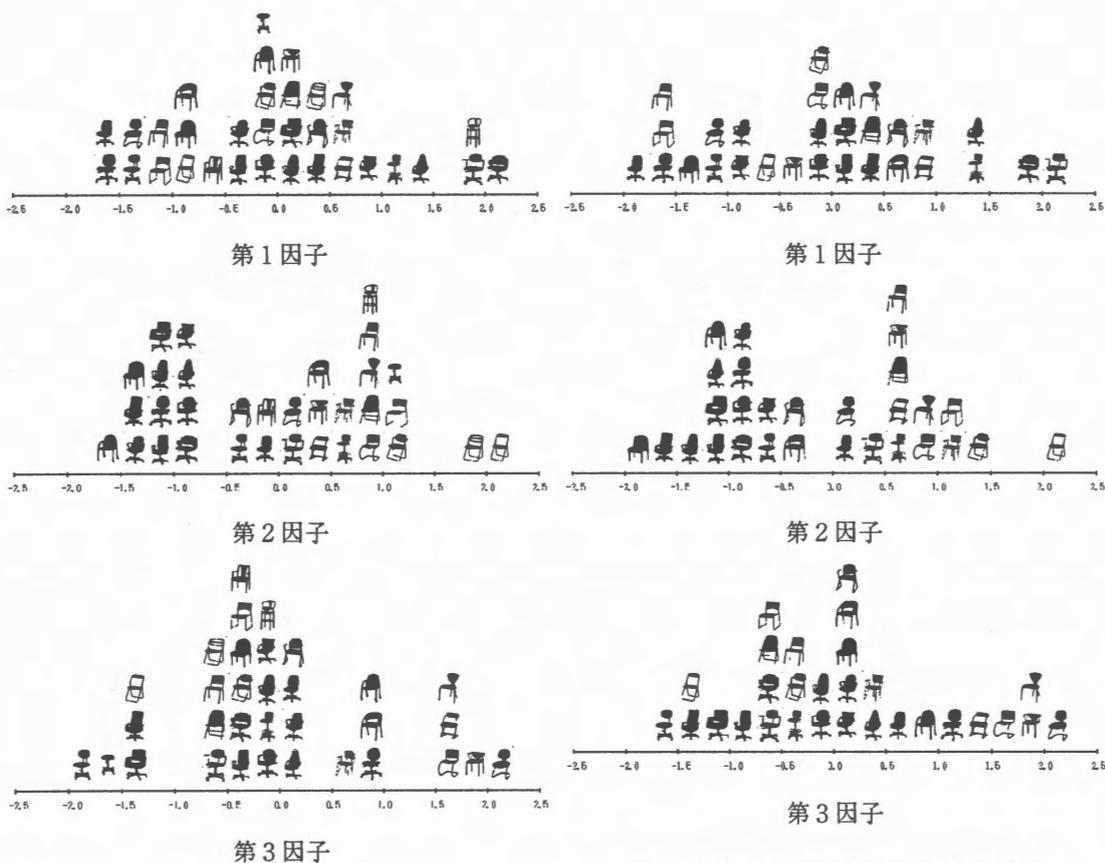


図 8: 心理実験で得られたイメージ空間

プローチで考えることができると思われる。すなわち、研究者は、まず、緒データを俯瞰すべき着目点を想定し、その着目点から見て特徴的な概念を拾い出す。データベース中のテストデータに対してそれらの概念を基準とした評価を行い緒データを整理すべき座標空間に関する仮説を生成する。次いで、その空間と画像データがもつ緒特徴量とを関係づけた上で、仮説を検証するために、より広い範囲のデータに対し概念記述を用いた検索を試行して評価する。その結果に応じて仮説の修正を繰り返すことによって、(3)のステップに耐えうる特徴空間を想定できるのではないだろうか。

このように考えると、前章で述べたような特徴空間の構築支援機能を持ち、概念表現を直接入力できるインタフェースシステムが、人文科学支援の一つの形態をなすと考えることができるのではないだろうか。

我々は、ユーザの直感的印象概念を整理するための座標空間の構築にSD法を用いた。この方法は、複数の被験者による直感的評価に基づいている。場合によっては、このような方法は適用できないことも考えられるので、この他の

図 9: 画像特徴より算出した特徴空間

方法論を用いてこのフェーズを支援するためのツールも必要となろう。また、特徴空間と画像特徴との関係を定義する際には、座標軸に沿って実画像を配置してそれを人間が観察して抽出すべき特徴量と、座標値と画像特徴量との間の関係式を求めるといった方法をとった。このフェーズは、このままでは画像処理や数式処理に不慣れたユーザでは実行できない。従って、このフェーズを支援するためのツールが必要である。そのためには、例えば、ニューラルネットや機械学習の理論を援用することなどが考えられる。

5 まとめ

画像データベースの自然言語インタフェース構築について述べ、そこで考えた特徴空間構築プロセスが、画像データベースを利用した人文科学研究支援の一つのプロセスモデルと考えられることを示した。

今後、まず、この考え方を、実際の人文科学研究に具体的に適用することが可能か否かを検証することが必要である。さらに、課題として、前章の最後でふれた2つのフェーズの支援ツールの設計・開発を進めてゆくことが挙げられる。

タイプ分け		第1因子の関数	第2因子の関数	第3因子の関数
タイプ1	肘なし	$\frac{ Lu-Ll }{Lu+Ll} - \frac{Se}{Sf}$	$-\frac{ Lr-Lb }{Lr}$	(b)
	肘あり	(a)	$-\frac{Sa}{Sh}$	
タイプ2		$\frac{2(W-H)}{MAX(W,H)} - \frac{4\pi Sb}{Lb^2} - \delta$	$\frac{ Lb-Le }{Le} - \frac{Sa}{Sh}$	δ
タイプ3		$\frac{H-W}{MAX(W,H)} - \frac{Sb}{Sa}$	$-\frac{Sa}{Sh}$	$\frac{4\pi Sb}{Lb^2} + \delta$
タイプ4		(c)	$-\frac{Sa}{Sh}$	$\frac{Su-Sl}{4Sb} + \frac{ Lu-Ll }{4Lb}$ $-\frac{ Le-Lb }{Le} + \frac{ Lr-Lb }{Lr}$

(a)			(b)	
タイプ分け	関数		タイプ分け	関数
$\frac{W-H}{MAX(W,H)} (=A) > 0.3$	$\frac{ Lu-Ll }{Lu+Ll} - \frac{Se}{Sf}$		$MIN(\frac{ Le-Lb }{Le}, \frac{ Lr-Lb }{Lr}, \frac{ Lt-Lb }{Lt}) (=B) < 0.03$	$\frac{W-H}{100MAX(W,H)}$
$A \leq$	$\frac{Se}{Sb} < \frac{Sr-Sb}{Sb}$	$\frac{Se}{Sb}$		+B
0.3	$\frac{Se}{Sb} > \frac{Sr-Sb}{Sb}$	$\frac{Sr-Sb}{Sb}$	$B \geq 0.03$	$\frac{W-H}{(100MAX(W,H)) - B}$

(c)	
タイプ分け	関数
$\frac{ Lb-Le }{Le} < MIN(\frac{ Lb-Lr }{Lr}, \frac{ Lb-Lt }{2Lt})$	$\frac{Sb}{Sa} + \frac{10Sd}{ScN} - \frac{ Lb-Le }{Le}$
$\frac{ Lb-Le }{Le} > MIN(\frac{ Lb-Lr }{Lr}, \frac{ Lb-Lt }{2Lt})$	$\frac{Sb}{2Sa} + \frac{5Sd}{ScN} - MIN(\frac{ Lb-Lr }{Lr}, \frac{ Lb-Lt }{2Lt})$

表 2: 定義した関数一覧

参考文献

[1] 高木, 伊東: “自然言語の処理,” 丸善, 1987.

[2] 岩下豊彦: “SD法によるイメージの測定”, 川島書店, 1983.

[3] 原田, 杉浦, 大庭, 中谷, 伊東: “自然言語による画像データベースの検索”, 人工知能学会全国大会 (第7回) 論文集, pp593-596, 1993.

[4] 杉浦, 原田, 中谷, 伊東: “自然言語による画像データベース検索”, 電子情報通信学会技術研究報告 Vol94 No.51, pp55-62, 1994.

[5] H.Nakatani and Y.Itoh: “An Image retrieval system that accepts natural language,” AAAI-94 Workshop Notes on Integration of Natural Language and Vision Processing, pp7-13, 1994.

[6] S.Harada, H.Sugiura, H.Nakatani and Y.Itoh: “On constructing pictorial feature space for image retrieval,” IJCAI-95 Workshop Notes on Representation and Processing of Spatial Expressions, pp103-117, 1995.

[7] 原田, 田中, 伊東, 中谷: “画像検索のための形状特徴空間の構築”, 電子情報通信学会技術研究報告 Vol95 No.320, pp7-12, 1995.