

自然言語による画像データベース検索

杉浦 博 原田 将治 中谷 広正 伊東 幸宏

静岡大学工学部情報知識工学科
〒432 静岡県浜松市城北3-5-1

あらまし 本稿では、言語情報と画像情報とのマッチング手法を検討するために試作したシステムについて述べる。従来の画像データベース検索システムは、一般のユーザにとって扱いやすいものではなかった。そこで我々は、一般のユーザでも扱いやすいように、比較的自由的な語句や文体の自然言語を用いて、検索できるようにした。また、検索時に画像解析し画像特徴を抽出することによって検索要求とのマッチングを行っている。このような対話的に検索できる画像データベース検索システムの構築方法について検討する。

和文キーワード 画像データベース検索、自然言語、画像解析、言語解析

Image Retrieval by Natural Language

Hiroshi Sugiura Shouji Harada Hiromasa Nakatani Yukihiro Itoh

Department of Computer Science, Faculty of Engineering, Shizuoka University
3-5-1, Jyohoku, Hamamatsu, Shizuoka, 432, Japan

Abstract We present an image data retrieval system that accepts inquiries in Japanese and analyses image features online. We discuss how to interpret subjective expressions and how to deal with comparative sentences in inquiries. The system consists of two units, the linguistic data analysis unit and the image data analysis unit. The retrieval is performed as follows: The system analyses inquiries entered in Japanese sentences, and provides the retrieval conditions in intermediate representation format. Receiving the retrieval conditions, the image data analysis unit extracts the image features concerning the retrieval conditions and shows the possible images. Given the images, users can continue the interaction with the system until they are satisfied.

英文 key words image data retrieval, natural language, image data analysis, linguistic data analysis

1 はじめに

本報告では、一般のユーザにとって使いやすい画像データベースについて考察する。一般のユーザには、データベースの取り扱いに関する専門的知識、データベースに格納されているデータに関する専門的知識を期待することはできない。

そのようなユーザに対して、画像データベース検索システムを構築する際に、次のことを考慮に入れなければならない。

1. システムの使用上、できるだけ語彙や文体の面での制約がないほうが良い。
2. ユーザに、検索したい対象物の正確で客観的な表現を要求するのは無理がある。

したがって、実用的な画像データベース検索システムというのは、次のことを満たしているシステムと言えよう。

1. 検索要求はできるだけ制限のない語彙および文体で行えること。
2. 検索したい対象物の漠然としたイメージを表現するのに、主観的な印象を用いることができること。
3. 検索がシステムと対話的に進めていけること。

従来の画像データベース検索システムでは、め画像一枚ごとに内容を表すキー（記号）を覚えておき、そのキーを用いて検索する方法が主流であった[1]。このようなシステムでは、画像に与えられるキーはデータベースの管理者によって与えられるものがほとんどであった。したがって、検索する際に、画像の内容あるいはキーそのものがある程度知っていなければならない。これでは、一般のユーザにとって使いやすいとは言えない。また、全画像に対して、事前に人手でキーを与えておかなければならないという問題点もある。

最近では、画像処理研究と自然言語処理研究の成果を利用して、画像と言語との総合的な取り扱いが試みられている[2-5]。これによって、ユーザはいくつかの単語を用いて画像特徴を表現することがある程度可能となった。

例えば、加藤らのARTMUSEUM[3,4]では、絵画などの自然画像を対象とし、色彩に関する形容

詞と形の輪郭を表すスケッチ画をキーとしている。このシステムでは、人間の主観や視覚的な印象を反映させるという点で、一般のユーザの立場を重視したシステムといえる。

しかし、単語を用いた場合には、単に画像解析結果を適当な語彙と対応づけ、その語彙をキーとして検索を行うものが多く、同じ意味の言語表現でも、キーとして定義されている語彙を用いた言い回ししか受理できないとか、全体的な特徴しか指定できないなどの問題点があった。

感性語句を取り扱う研究としては、前述のARTMUSEUM以外に長町らのデザイン合成システム[5]がある。しかし、画像検索への適用を考えると、原データの解析結果と主観的な表現の解析結果との照らし合わせ作業が新たに必要となる。

我々は、自由な文体・語彙を用いた日本語文入力を受け付け、画像解析結果と照らし合わせて検索を行うシステムを、椅子の電子化カタログを題材として開発を進めている。本稿では、この実験システムについて述べる。特に今回は、修飾語として複数の感性語句を含む文を受理可能にしたので、そのことについても述べる。

2 基礎的考察

2. 1 感性的・主観的な表現の取り扱い

2. 1. 1 感性語句

椅子の設計者であれば、椅子の画像の検索の際、客観的基準が存在する語彙だけを用いて検索要求を書くこともできるが、素人には無理である。普通は、「かわいい椅子」、「地味な色の椅子」のような表現が用いられる。この場合、椅子の「かわいさ」、色の「地味さ」と言ったものを客観的に定義することが難しい。感性語句とは、このように客観的な基準がはっきりしておらず、個人の感性に依存するような、直感的印象を表す語句のことを指す。

自然言語により画像データベースを検索する場合、感性語句を取り扱うことが極めて多くなる。そのため、感性語句に対する解釈方法を検討することが必要である。

2. 1. 2 比較表現

感性語句の解釈の仕方は人によって異なる。そのため、感性語句への対応は、個人差を考慮しなければならない。

その対応の一つとして、ユーザが比較表現を用いて対話的に検索が行えるようにすることが考えられる。たとえ1回の要求で希望する画像が検索できなくても、「もっとかわいい椅子」といった比較表現を用いて、最終的には、ユーザの希望する画像を検索できるようにして、個人差を吸収することができる。

2. 2 視覚と言語

画像と言語とを対応付けようとするとき、次のことに注意しなければならない。

1. 両者の形態が大きく異なっていること。
2. 自然言語には数多くの同義な言い換え表現が許されること。

そのため、両者の媒介となる表現が必要である。この媒介表現について、画像との接続・言語との接続という2つの側面から検討する。

2. 2. 1 画像との接続

画像解析を行うことを考えた場合、媒介となる表現は、次のことが要求される。

1. 比較的容易に抽出し得る特徴概念を基本要素としていること。
2. 画像解析アルゴリズムが前提としている画像的特徴の階層構造と、媒介表現を構成する階層構造とが一致していること。

2は、画像と媒介表現とのデータ変換において、両者の階層構造が一致していないと不必要な手間がかかるからである。

さらに、比較表現にも対応できるようにするためには、画像との接続を重視した表現において、次のことが要求される。

3. 各特徴において、異なる値の間の距離が定められるものを基本要素としていること。
4. 各々の特徴の取り得る値が連続的なものを基本要素としていること。

2. 2. 2 言語との接続

自然言語には、異なる言い方ではあるが、同義の意味を有する同義文が多数存在する。もちろん、受理可能な入力文に、語彙や文体の面での制約をなるべく無くすため、多様な文体の文に対しても対応していかなければならない。しかし、同義文に対して、文体ごとに個別に処理を行っていたのでは、合理的でない。したがって、言語との接続を検討した場合、媒介となる表現は次のようなものが望ましい。

1. 同義文に対して、その意味の同等性が容易に判定できるような定式化された表現であること。
2. 文体が異なっても、同義文の媒介表現が一定のアルゴリズムで処理できること。

2. 2. 3 画像と言語とのマッチング

媒介となる表現を、「画像との接続」、「言語との接続」、といった側面から検討したが、両者から望まれる媒介表現の形態は異なっている。そのため、媒介表現を2層に分け、各々のレベルで画像および言語との接続条件を満たせるようにする方針をとった。

画像よりの媒介表現としては、画像から直接抽出しうる物理的特徴空間上での範囲指定という形で構成することにした。なお現在、指定できる特徴として、椅子の背に対する色、ならびに、その形の一部の指定が可能になっている。また、色の物理的特徴空間としては、HSV空間(3. 3. 1)を用い、領域の指定は、3次元空間上で直方体となる領域で指定している。(図1)

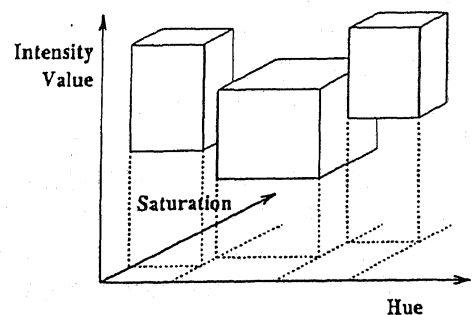
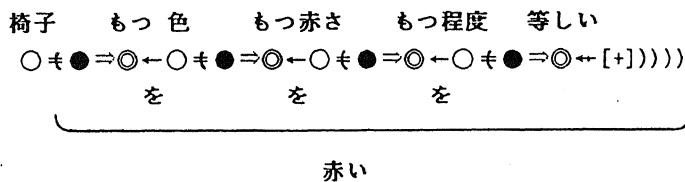


図1： 領域の指定法

(1)「赤い椅子」



(2)「色が赤い椅子」

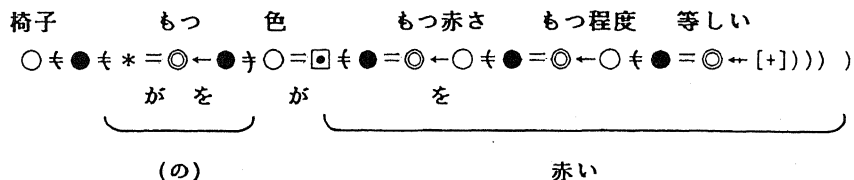


図2： 意味表現

以後、この画像よりの媒介表現を中間表現と呼ぶことにする。

また、言語との接続を重視した表現としては、同義文に対する柔軟性を特徴とする高木・伊東の意味表現[6]を採用することにした。

例えば、(1)「赤い椅子」、(2)「色が赤い椅子」の意味表現は図2ようになる。(1)・(2)の意味表現は、各々の文の構文的差異を的確に表現していると同時に、両者の意味の同等性も表現している。また、この意味表現を構文構造の差異にかかわらず、均一に読み取って意味処理を行う基本アルゴリズムも既に提案されている[6]。

2. 2. 4 解釈方法

次に、意味表現から中間表現への解釈方法(変換方法)を検討する。

検索に多く使用されると思われる感性語句は、客観的な解釈はできないが、ある程度、各人の評価尺度を表明することは可能である。例えば、物体の「かわいさ」であれば、

- ・ 全体的な大きさが小さい
- ・ 形状が丸みを帯びている
- ・ 色合いがピンクのようなもの

が「かわいい」というように表明できる。

そこで、感性語句に対する意味表現として、例えば、「地味な」の内部意味構造は、感性的属性概念「地味さ」を用いて定義する。その感性的属性概念ごとに、物理的特徴空間上で、その特徴を示すと思われる領域、ならびに、最も

その特徴を示していると思われる点を与える。それらを基に意味表現から中間表現に解釈することにした。詳細は3.4節で述べる。

3 システムの実装

本システムは、自然言語による画像データベース検索のための枠組みを検証するために試作したものであり、現在のところ必要最小限の部分が実装されているだけである。状況としては、椅子のカタログ写真集からユーザの希望する椅子を探し出すということ考えた。

3.1 システムの構成

システムは、言語処理部と画像解析部とから構成されている。

言語処理部では、主にユーザとのインタフェースを管理し、画像解析部では、画像データベースを管理している。

図3は、システムの構成を示したものである。ユーザの検索要求は、漢字仮名まじりの日本語文でキーボードから入力される。すると、言語処理部で、受け取った日本語文字列に対して、形態素解析、構文解析を施し、単語間の意味的接続関係を抽出する。次に、その単語の内部意味構造を参照しつつ、意味表現を生成する。その意味表現の解釈を行い、中間表現を生成し、画像解析部にその中間表現でのマッチングを依頼する。最後に、画像解析部が返してきた画像のデータを受け取り、その画像を提示する。

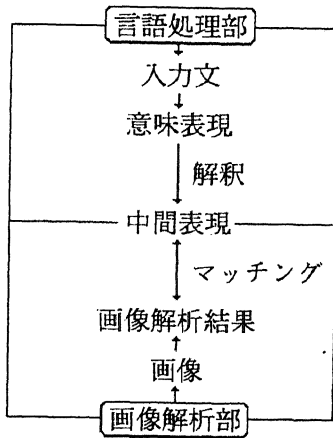


図 3 : 処理の流れ

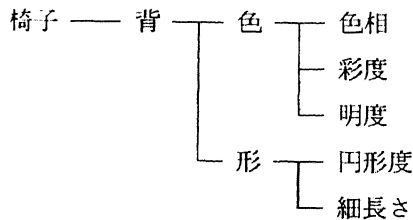


図 4 : 中間表現の階層構造

画像解析部では、蓄積されている画像を順に解析し、中間表現で指定されている条件とマッチングを行う。そこで、マッチした画像のデータを言語処理部へ受け渡す。次の入力文が「もっと赤い椅子を探して。」などの比較表現であるときに、その画像のデータが用いられる。ある画像に対して、一回解析を行うと、その解析結果を原画像と対応づけて蓄積し、次回からの解析の手間を省くようにしている。

3. 2 中間表現

今のところ、取り扱っている特徴は、椅子の背の色と形状に限られている。その他の特徴は、現在検討中である。したがって、中間表現の階層構造は、図 4 のように極めてシンプルに定義できる。

3. 3 画像解析

データベース中の画像は、RGB 表色系で量子化されているデータである。各々の画像から、

椅子の背の部分抽出する方法は、次のようにして自動的に行っている。

1. 領域分割をする。
2. 背景の明度を閾値として、椅子の領域を抽出する。
3. 膨張・収縮することによって、背・座・脚のような椅子の部分ごとに、領域を分け、その領域ごとに重心を求める。
4. 椅子の背として、最も高いところにある重心の領域を背とする。

しかし、この方法では、背と座が一緒になっているような椅子には対応していないので、まだ検討する必要がある。

3. 3. 1 色特徴[11]

色空間としては、人の感覚に合っていて、均等色空間として色差が定義されている修正マンセル色空間がある。また、その色空間を近似している C I E (1976) $L^*a^*b^*$ 色空間は、R C B 値から計算で求めることができる。今までのシステム [13, 14] では、この色空間を用いて、各感性語句に対応する色の領域を、直方体の組み合わせで近似する方針をとっていた。

しかし、この色空間では、等明度でも色相ごとに彩度の最大値が異なっている。すなわち、色空間上で凹となる領域が存在する。これでは、「明るくて暖かそうな色」のような修飾語に複数の感性語句が使われる場合、色が存在しない凹領域に検索範囲のピーク点を決定してしまうことがある。基本的に、複数の感性語句によって修飾されている場合の検索範囲のピーク点は、各々の感性語句の検索範囲のピーク点の中心をとるようにしているからである。(3. 4. 3)

一方、J o b l o v e らの色空間でも、色相・彩度・明度の三属性を用いており、凹領域は存在しない。しかし、この色空間は円柱形をしており、人間が見た色の違いの差と、この色空間での色の差が、明度等によって異なる。また、個々の感性語句に対応する色の領域を、直方体の組み合わせで近似することは難しい。

そこで、C I E (1976) $L^*a^*b^*$ 色空間と、J o b l o v e らの色空間を用いて、色空間に凹領域がなく、さらに C I E (1976) $L^*a^*b^*$ 色空間で考えていたように、直方体の組み合わせ

で感性語句に対応する領域が近似できるような新しい色空間を考えた。本稿では、この色空間をHSV空間と呼んでいる。

HSV空間の具体的な作り方は、次のようにした。

1. CIE(1976)L*a*b*色空間において、各明度での彩度の最大値の平均値を求める。
2. 明度軸方向の色の違いは無視して、Jobloveらの色空間の彩度の最大値が、ほぼ先に求めた平均値になるように調整する。

このHSV空間における、色相・彩度・明度の値を椅子の背の色の特徴量としている。

3. 3. 2 形状特徴

現在使用している椅子の原画像は、椅子のカタログ写真を基にしており、その写真は、やや斜めの方向の角度から撮影されたものとなっている。このような画像から椅子の背の部分抽出し、その背についての特徴量を求めても正しい値とは言えない。

そこで正面図にする必要がある。ここでは、透視変換法[12]を用いて正面図にしている。また肘掛けのある椅子にも対応できるように、椅子の背の左右対称性を利用して、正面図の左半分の形だけを用いて新たに正面図を作成している。

今のところ、椅子の背の形状特徴として、正面図より円形度(丸み度)・細長さを算出している。

3. 4 中間表現の生成法

「地味な色」、「もっと地味な色」、「暖かくて鮮やかな色」、「鈍い青色」という表現を例にとって説明する。なお、図5-図7では、分かりやすくするため、彩度-明度の2次元平面になっているが、実際には、色相も含めた3次元空間である。

3. 4. 1 感性語句の解釈

「地味な色」のように、程度が特に指定されていない場合は、「地味さ」に対応する物理的的特

徴空間上の領域をそのまま検索範囲とする。

「かなり」、「少し」といった程度を表す副詞は、その程度の強いと思われるものから順に1から5までのレベルに振り分け、程度副詞の修飾を受ける場合は、そのレベルにより中間表現は図5のように決定される。例えば、「かなり地味な色」などの一番程度の強いものの場合、ピーク点から対応する領域までの距離が1/4の領域を検索範囲としている。

3. 4. 2 比較表現の解釈

「もっと地味な色」のような比較表現の場合、比較対象の特徴点と感性的属性概念「地味さ」に対応する物理的特徴空間上のピーク点、なげに、副詞句のレベルをもとに図6のように決定される。例えば、「もう少し地味な色」といった副詞句のレベル4の比較表現の場合、現在点

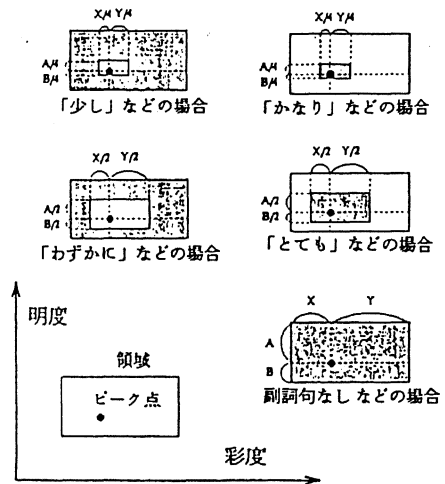


図5： 程度副詞の解釈法

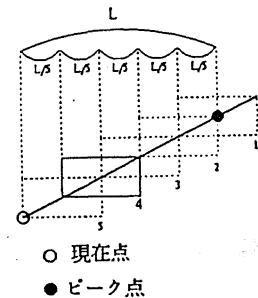


図6： 比較表現の解釈法

表 1 : 「鈍い青色」の解釈例

鈍い			青			鈍い青		
ピーク点			ピーク点			ピーク点		
色相	彩度	明度	色相	彩度	明度	色相	彩度	明度
なし	3.0	2.0	237.6	10.0	5.0	237.6	6.5	3.5
検索範囲			検索範囲			検索範囲		
色相	彩度	明度	色相	彩度	明度	色相	彩度	明度
0.0-360.0	0.0-4.0	1.0-4.0	216.0-252.0	5.0-10.0	1.0-6.0	216.0-252.0	2.5-7.0	1.0-5.0

とピーク点の距離Lの2/5だけピーク点に近づけた点を中心として、各成分が中心から±L/5以内の領域を検索範囲としている。

3. 4. 3 複数の感性語句を含むときの解釈

「暖かくて鮮やかな色」・「鈍い青色」のように複数の感性語句で色を修飾している場合、「青」などの主に色相に依存している語句を含むかどうかで場合分けをした。

(1) 主に色相に依存した語句を含まない場合

「暖かい」・「鮮やか」の個々の検索範囲の平均を検索範囲とする。またピーク点は、基本的に個々のピーク点の中心で、先の検索範囲から外れる場合は、検索範囲内で一番それに近い点とする(図7)。

(2) 主に色相に依存した語句を含む場合

色相だけは、「青」などの主に色相に依存した語句の値を使い、彩度・明度については、(1)と同じように処理する(図7, 表1)。

4 システムの仕様・検索例

4. 1 仕様

現在のシステムの使用を表2に示す。なお、椅子の画像は、椅子のカタログ写真を基にしてスキャナー(EPSON GT8000)を用いて取り込んでいる。

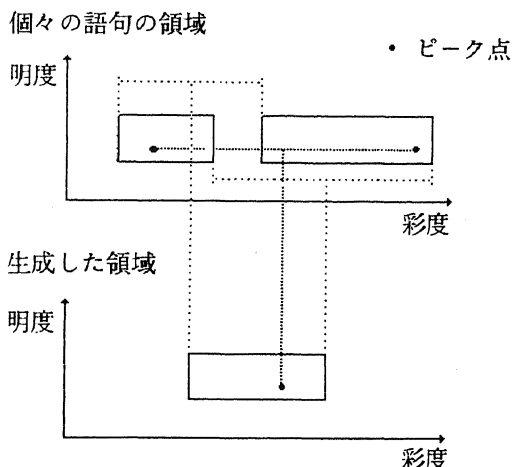


図 7 : 複数修飾の解釈法

表 2 : システムの仕様

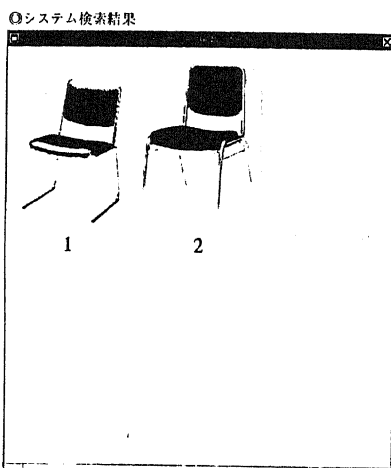
受理可能な語句		例
色を表す語句	453語	赤、涼しい
形を表す語句	6語	丸い、縦長だ
名詞	2語	椅子、背
程度副詞	8語	もっと、少し
接尾辞	3語	～ぼい、～的な
動詞	4語	見せる、探す
計	476語	
その他		助詞、助動詞、単位名詞等

検索可能な画像の数	167枚
-----------	------

4. 2 検索例

検索例を図8に示す。最初の入力文により、システムは候補の画像を最高5枚まで番号とともに提示する。そこで、希望する画像が得られ

入力文「陽気な色の椅子を見せて下さい。」



入力文「2番の椅子よりもっと淡くて
ピンクの椅子がありますか。」

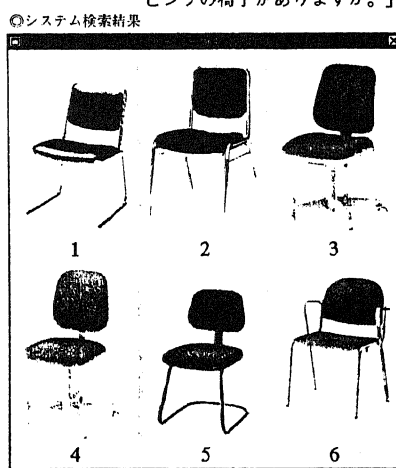


図 8 : 検索例

ない場合でも、2つ目の入力文のような比較表現を解釈して、ユーザの要望に柔軟に対応することができる。

5 おわりに

本稿では、一般のユーザを対象にした、画像データベース検索システムの一手法を提案した。まず、画像情報と言語情報という形態の異なっているものを対応づけるために、両者の媒介となる表現が必要であることを述べた。さらに、感性語句を受理可能にしたことによって、ユーザとのインタフェースが円滑に行えるようになった。

今後は、形に関する検討を行い、さらに広範な表現を取り扱えるようにする必要がある。

謝辞

日頃、ご指導頂く阿部圭一教授に感謝致します。本研究の一部は、文部省科学研究費補助金および電気通信普及財団の補助により行われた。

参考文献

- [1] 坂内：“画像検索技術”，電子情報通信学会論文誌 Vol. 71 No. 9, pp. 911-914(1988年).
- [2] 栗田, 下垣, 加藤：“主観的類似度に応じた

画像検索”，情報処理学会論文誌 Vol. 31 No. 2, pp. 227-237(1990年).

- [3] 加藤, 栗田, 坂倉：“フルカラー絵画データベース ART MUSEUM”，情報処理学会研究報告 IE88-118, pp. 31-38(1988年).
- [4] 加藤, 栗田：“画像の内容検索”，情報処理 Vol. 33 No. 5, pp911-914(1992年).
- [5] 長町：“感性工学”，海文堂(1992年).
- [6] 高木, 伊東：“自然言語の処理”，丸善(1987年).
- [7] 小林：“配色センスの開発”，ダヴィッド社(1990年).
- [8] 日本カラーデザイン研究所：“カラー・イメージ事典”，講談社(1992年).
- [9] “日本工業規格 JIS Z 8721 標準色票”，日本規格協会(1959年).
- [10] 相賀：“色の手帳”，小学館(1986年).
- [11] 高木, 下田：“画像解析ハンドブック”，東京大学出版会(1991年).
- [12] 山口：“コンピュータグラフィックス”，日刊工業(1979年).
- [13] 大庭, 伊東, 中谷：“自然言語による画像データベースの対話的検索”，人工知能学会研究会報告 SIG-HICG-9202(1992年).
- [14] 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検索”，人工知能学会全国大会(第7回)論文集 pp. 593-596(1993年).