2004－AVM－44　(12)
2004／3／5

# 画像検索のための３Ｄインターフェース

小池真由美[*1]　青木輝勝[*2]　池田佳代[*3]　伊藤学[*4]　日高宗一郎[*5]

[*1]（有）エスパリエ　[*2]東京大学先端科学技術研究センター
[*3]（有）エクセリードテクノロジー　[*4]山形県デジタルコンテンツ利用促進協議会
[*5]国立情報学研究所

近年、インターネット等のネットワークを用いたコンテンツ流通が盛んに行われるようになり、音楽配信に加え映像配信までもがブロードバンド環境下のもと急速に普及しはじめている。このような背景のもと画像検索に関するニーズが急速に高まってきており、これに歩調を合わせるかのようにMPEG-7に代表される画像検索のためのメタデータの国際標準もほぼ固まりつつある。一方、現在画像検索技術は必ずしも実用レベルに達しているとは言えず、今後さらなる改良に向けた研究開発が必要であることは言うまでもない。本稿では、現在の画像検索の最大の問題は入力インターフェースにあることを言及するとともに、新たに超直感インターフェースとして３Ｄインターフェースを提案する。

# 3D Interface for Image/Video Retrieval

Mayumi Koike[*1]　Terumasa Aoki[*2]　Kayo Ikeda[*3]
Manabu Ito[*4]　Soichiro Hidaka[*5]

[*1] Espalier Inc. [*2] University of Tokyo, RCAST
[*3] Excellead Technology　*
[4] Yamagata Digital Content Center for Research & Promotion
[*5] National Institute of Informatics

*In recent years, content distribution through networks such as the Internet has been actively carried out, and image/video distribution as well as music distribution has become widespread in rapid pace thanks to the broadband network environment. Under this situation, there are growing needs for image/video retrieval. The International Standards of content metadata, such as MPEG-7, are almost finalized as if they keep line with such growing needs. On the other hand, technologies for image/video retrieval have not reached to practical level. As such, more research will be required. This paper mentions that the biggest issue in current image/video retrieval is in input interface and proposes 3D interface as the newly emerging super-intuitional interface.*

# 1．　はじめに

動画像を個人で収集・蓄積し，またネットワーク経由で通信する機会は数年前と比較して格段に増えている．事実，総務省の全国消費実態調査によれば，ビデオカメラの世帯普及率は約 40%，ビデオテープレコーダーの世帯普及率は約 80%であり，ますます増加の傾向を示している．また、ハードディスクレコーダ等もその利便性から急速に普及しつつある。

このため所有する多くの動画像コンテンツあるいはネットワーク上の動画像コンテンツから所望のシーンを検索したいというユーザの要求は非常に強まっている．

従来画像検索の研究においては，動画像中のどのような特徴量を用いて画像検索を行うか，その特徴量を動画像中からどのようにして抽出するか，あるいはあらかじめ検索目標となる画像が用意されており，どのようにしてその目標画像が動画像中のいずれにあるかを特定するか，などの観点から研究が進められてきた．しかし残念ながら今日に至っても，依然として個人が使いやすい形での画像検索技術は実用化されていないのが現状である。

このように画像検索の研究が数多くなされているにも関わらず，十分実用に耐えうる検索技術が出現していないのは，画像が本質的に持つ意味の多義性のために，キーワード付けや内容把握，内容検索の実現が簡単ではない、ということが最大の要因であると考えられる．人間は画像に対してキーワードのような言語的な認識だけでなく，意味的な認識と感覚的な認識とを組み合わせて認識しており、また文字検索と異なり，検索の前に検索対象がはっきりしていないことも少なくない。つまり、画像検索の難しさは，検索目標であるにもかかわらずその目標画像に対する記憶があいまいであり，目標画像のイメージを正確に描けないことに起因しているとも言える．したがって、高精度な画像検索の実現のためには，ユーザの検索要求をどのように入力し，システムがその入力をどのように解釈・処理し，結果をユーザに返すかというユーザインターフェースの観点からの積極的な検討が必要不可欠であり、特にユーザのクエリー生成をどのように支援するかは画像検索技術における最大の課題であると言える。

本稿では，このような観点から、既存検索インターフェースの問題点を整理するとともに、画像検索に用いるユーザにとって直感的で使いやすい入力方式として，3D 入力インターフェースを提案する。本方式は人間の記憶が３D 的であることに着目した超直感的インターフェースであり、人間の記憶の仕方に合わせて検索クエリーを生成しようとする点が従来の検索インターフェースと最も異なる点である。

# 2. 画像検索に関する従来研究

## 2.1　画像検索技術の分類

画像検索に用いる入力インターフェースを入力内容の観点から大別すると、

(A)テキスト入力型
(B)略画入力型
(C)オブジェクト選択型
(D)画像探索型
(E)その他

に分けることができる。

上記分類(A)は、[1]～[4]などのようにテキスト語句によるもので，任意名詞句や制限された名詞句，また印象語や感性語，動作語などをユーザが入力し検索を行うものである．「富士山」や「走っている人」など具体的な場合に加え，「明るい空」や「丸い果物」などの主観を含む場合もある．テキスト語句による入力では，既存の語句検索と同様にシソーラスを用いて，ユーザの意図をより反映することも可能である．

一方、上記分類(B)は[5]～[11]などのようにラフスケッチを描くことでユーザのイメージを構成し，クエリーとするものである．これらのスケッチによる検索は，ユーザの略画の描き方にも精度が依存するが，一般的に厳密な検索に用いることは困難である．しかし，対象物に曖昧さがある場合には有用であると考えられる．

上記分類(C)オブジェクト選択型とは，例えば簡単な図形を矩形領域中に配置し，検索目標のイメージを構成したり，また検索目標のアイコンを配

置，変形しユーザの目標画像を示したりするものである．目標画像中のオブジェクトを略画により描く必要はないため容易な入力方式である一方，画像を構成するオブジェクトが限られてしまう，変形の自由度が限られてしまうなど，ユーザの直観的な入力を阻害する可能性もある．

上記分類(D)画像探索型とは，あらかじめ検索したい画像（キー画像）を所有している場合に限って利用できる技術であり、例えば放送映像中に所望のコマーシャル映像がどこにあるか，また編集前の映像が放送後の映像中のどのあたりに含まれているかなどの用途に利用される。これらの方式は検索目標の画像が厳密に用意できる点では，文章中からの単語の検索と同様高い検索精度が期待できるクエリー入力方式である．しかし特定用途向けのいくつかのわずかな例では実用性は高いものの，一般的な画像検索としてユーザが検索したいと考えている画像を参照画像として用意することは困難である。

本稿では画像検索のインターフェースとして(A)～(D)の４種類にして概観したが、実際に試作されているシステムもしくは商用化されているシステムは(A)テキスト入力型がほとんどで(B)略画入力型がわずかに存在する程度である。次の 2.2、2.3 ではこの(A)(B)の既存研究についてその概要をまとめる。

## 2.2　テキスト入力型インターフェース

テキスト入力型インターフェースは最も常識的なインターフェースのためほとんどの画像検索システムに採用されている。最新の研究開発事例としては、例えば[12]にて感性語による画像検索を報告している．これによれば，入力された感性語を対応テーブルによって配色パターンに変換し，この配色パターンとの類似度により検索を実現している．同研究において，感性語ごとに検索における感性語と色相・彩度・明度の関係を明らかにしている．

また、[13]では栗田らが「ロマンチックで暖かい」というような視覚的印象により検索を実現するため，利用者に対して学習用の絵画に対して印象語を付けてもらい，その結果から印象語と画像特徴との相関関係を学習し，検索に利用することを報告している．

[14]では，蓄積されているキーワードと検索キーワードとの不一致を避けるため，システムで規定するキーワードを用いる統制キーワード方式を用いている．

## 2.3　略画入力型インターフェース

[15]では西山らが人間の思考にあった画像表現をモデル化し絵画検索に応用し有効性を示している．これによれば，ユーザの記憶に残っているあいまいな部分を効果的に使うために，画像がだいたいどのように塗り分けされていたかという「領域情報」，人物・机などの「オブジェクト情報」，それぞれのオブジェクトの詳しい様子である「特徴情報」の３つを，略画として検索要求に表現している．

[16]では望月らが, 画像を広く捉えたテクスチャ性を反映しかつ人間が画像の初見から受ける感覚に関連すると考えられるフラクタルベクトルを適用し，また複数点からの特徴量を抽出するためのブロックを，単純な分割ではなく位置可変に設定することで柔軟な検索が可能となることを示している．

一方、[17]では加藤がユーザの意図がシステム側に伝わる「柔らかいシステム」を実現するために，ユーザの表現していない意図をシステムが吸収するよう遺伝的アルゴリズムに対話機構を持ち込んだ方法を提案している．

# 3. 画像検索のための 3D インターフェースの提案

## 3.1　検索インターフェースの要求条件

種々の画像検索のインターフェースの性質をまとめた概念図を図 1 に示す．図 1 の横軸は，人間が画像検索を行う際に入力するクエリーの直観性・入力の平易さを表わすものとし、また縦軸は，検索に用いる特徴量をコンピュータが理解する際の難易度を表わすものとする．この図１ではコンピュータが理解しやすいほど，低レベルな特徴量，すなわちインデキシングコストが安くなることを示し，逆にコンピュータが理解し難い特徴量ほど

高レベルな特徴量であり，すなわちインデキシングに人手を要するなどコストが高くなることも同時に意味している．

すなわち，参照画像を用意し検索を行う方式は，画像の画素値という最も低レベルな特徴量を用いるためインデキシングコストが最も安いが，入力の平易さという点では最も難しい．一方，検索目標の画像をテキストで入力する方式としては，自然言語で入力するもの，抽象的テキストで入力するもの，具体的なテキストで入力するもの，制限付きテキスト（選択肢）で入力するものがあり，この順で直感的でユーザにとって入力しやすいと考えられる。しかし，用いる特徴量としてあらかじめキーワードを付与しておく必要があるか，あるいはキーワードと画素値との変換テーブルを用意しておく必要があるなどインデキシングにコストがかかってしまう．仮に付与するキーワードに制限を加えた場合でも制限の程度によるがインデキシングコストがかかってしまう。

また、キーワードを画素値に変換し検索を実行する手法もあるが、この場合にもあらかじめキーワードを付与する必要がないためインデキシングコストは低いものの、変換テーブルを何らかの方法で用意する必要があり，ユーザにいくつかの画像とキーワードを見せて対応を学習しておくなどの手間がかかる.、また、このような方式では一般的にキーワードを増やすためには，学習を繰り返す必要があり容易にキーワードを増やすことが困難であるという問題点がある．

さらに，このようなテキスト入力型の共通の問題点として、一意のキーワードで表現しにくい場合の対応が困難であることが挙げられる．また，特定の集団，地域のみでしか通用しないキーワードを付与してしまうことも予想され，そのような場合きちんとキーワードを付けてあるにもかかわらず，第三者には検索できないという状況が発生する．

ところが，このようなあいまいさのないキーワードの一例として固有名詞がある．特に人物名に関する固有名詞をキーワードとして画像認識技術を応用することで，自動でぶれのない確かなキーワードを付与することが可能になる．

一方、以上のようなテキストによる入力に対し，略画やオブジェクトをクエリーとして用いることで，ユーザのイメージを適切に示すことが可能となる．画素値や画素値から抽出される情報を特徴量として用いるため，インデキシングコストは低い．ただし，クエリーとしてユーザの目標画像のグラフ構造を入力する方式は，ユーザが検索目標とする画像の構造をあらかじめ推測する必要があり，また必ずしも直感的で分かりやすいとは言えないグラフ構造として入力する必要がある点において，あまり誰にでも可能な入力方式とは言えない．あらかじめ定められたオブジェクトを配置したり変形したりする入力方式は，この点でより直感的であるが，ユーザの入力自由度が制限されてしまう．直観性，自由度の点からこれらの両方式を凌駕するのが略画による入力方式である．ところが，この場合高精度な検索のためには検索目標の特徴を丁寧に描く必要がある．図１中で”実画”という入力方式は，略画の中でも特に丁寧に細かく描かれた略画で，参照画像に近いものを指している．現実の環境において，ユーザが所望のシーンを検索したいと考えるときに，丁寧な略画を描くことはあまり期待できないであろう．


図１　:種々の入力クエリーと用いる特徴量の関係
（図中右上の円形領域が，インデキシングコストを抑えつつ直感的な入力となる領域）

以上の状況をすべて考え合わせると、インデキシングコストを抑えつつも高精度な検索を行うためには，テキスト語句の厳密性と略画のあいまいさとを組み合わせたクエリー入力インターフェース，すなわち、定義が明確な名詞（固有名詞など）とあいまいな記憶ながら概略だけは描ける略画の２つの入力を組み合わせたインターフェースこそが最も優れていることが容易に推測される。

## 3.2　直感的な入力インターフェースとは？

3.1 では、今後あるべき画像検索インターフェースの姿について試論し、テキスト入力型とあいまい略画の入力の組み合わせこそが最適なインターフェースであることを論じた。本節では略画入力に関してさらに深く検討してみることとする。具体的にはどのようなインターフェースにすれば最も直感的な略画がクエリーとして記述できるかについて検討する。

さて、「絵を描く」という行為はたとえ簡易略画とは言え一般的には一部の人間を除くと非常に苦手としているのが現状であろう。この要因としては、

(1)正確な形状・色の再現が困難であること。
(2)この世の物は元来すべて3Dであるのに2Dで表現しなければならないこと。

の２点が挙げられる。本節ではこれらのうち特に(2)の負担を軽減することによってより直感的なインターフェースが実現できることを示す。

一例として図 2 を取り上げ、また、あらかじめこの画像を見たことがある検索者が再度この写真を探す場面を想定することにする。

図 2　検索したい画像

この場合、略画を描くためにはこの被験者の頭の中では２D 的な情報、すなわち、「左右にりんごがあり、左のりんごは右のりんごと比べて高さ、幅がおおよそ半分ずつである。またこの２つはほとんどくっつきそうなくらい近づいている。」という記憶がなされていなければ正確な略画を描くことはできないことになる。しかし人間は本来このような記憶法をとっていないことははほとんど自明であろう。

一方、この写真を見たときに一般的には人間は２D 写真でありながら３D 的な情報、例えば「手前のりんごと奥のりんごは同じくらいの大きさで奥行き方向に 50cm くらい離れている」などの見方で脳に記憶させている。

このため、より直感的な入力インターフェースを設計するにあたっては 3D モデリング的手法を取り入れることが望ましいという結論に辿りつく。

## 3.3　3D インターフェースの設計

3.2 ではこれまで以上に直感的な略画入力インターフェースを設計するために 3D モデリング的な手法を取り入れることが有効である根拠を示した。この考えをもとに、本稿では図３に示す３D 入力インターフェースを提案する。

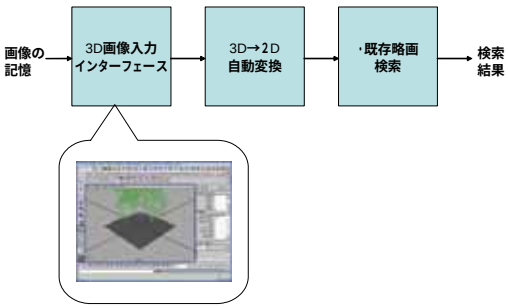


図 3　3D 入力インターフェース

図 3 において、「3D 画像入力インターフェース」とは、一般的な市販 3DCG 制作ソフトウェアのようなインターフェースを有し、直方体、立方体、球、回転体などが極めて容易に作成できるモジュールである。続く「3D→2D 自動変換モジュール」は本提案方式のキーとなる部分であるが、一般的に使われているカメラワークの各種技法（クローズアップショット、ウェストショット、ミディアムショット、ニーショットなど）を 15 種類程度を用いてそれぞれ並行的に作成した 3D モデルを”撮影”をする。これはまさに 3D モデリングを 2D 略画に変換する処理に他ならない。最後にこのように撮影された 15 枚程度の略画を従来同様の略画検索ツール[18]に入力させ、最終的な演算結果（検索結果画像）を得る。

一般的なカメラワークに関してはすでにある程度技法が確立しており、15 種類程度の技法に基づき撮影をしておけば通常の写真、映像にはほとんど対応可能であり、これこそが本方式において最も重要な技術根拠となっている。

## ４．まとめ

本稿では、現在の画像検索の最大の問題は入力インターフェースにあることを言及し、理想的な画像検索インターフェースとしてテキスト入力と略画入力を組み合わせたインターフェースこそが最適であることを主張した。このとき、略画入力インターフェースをより直感的に使いやすくするために 3D 入力の概念を導入し、その具体的な設計について提案した。

今後は提案インターフェースについて実装を進め、より詳細な評価実験を行う予定である。

謝辞：本研究は総務省戦略的情報通信研究開発推進精度研究主体育成型研究開発平成 15 年度「簡単映像コンテンツ制作のための高度映像検索技術に関する研究（研究開発）」（研究代表者:青木輝勝（東京大学））の一環として行われたものである。

## 文献

[1]芝田滝也,加藤俊一,"街路の景観画像データベースのイメージ語による検索",信学論 D-I Vol.J82-D-I No.1, pp.174-183, Jan.,1999

[2]宮森恒,粕谷英司,富永英義,"動作語を用いた問い合わせによる映像検索方式",信学論 D-II Vol.J80-D-II No.6, pp.1590-1599, June,1997

[3]原田将治,伊藤幸宏,中谷広正,"感性語句を含む自然言語文による画像検索のための形状特徴空間の構築",情処学論 Vol.40, No.5, May,1999

[4]椋木雅之,田中大典,池田克夫,"対義語対からなる特徴空間を用いた感性語による画像検索システム",情処学論 Vol.42, No.7, July, 2001

[5]Shih-Fu Chang, William Chen, Horace J.Meng, Hari Sundaram, Di Zhong, VideoQ:An Automated Content Based Video Search System Using Visual Cues, Proc. of the 5th ACM international conference on Multimedia, 1997

[6]M.Flickner, H.Sawhney, W.Niblack, J.Ashley, Q.Huang, B.Dom, M.Gorkani, J.Hafner, D.Lee, D.Petkovic, D.Steele, and Peter Yanker, Query by Image and Video Content:The QBIC System, IEEE Computer Magazine, Vol.28,No.9,pp.23-32,Sep., 1995.

[7]金原史和,佐藤真一,濱田喬,"形状分解によるユーザの視点に基づいたシルエット画像検索,"情処学論 Vol.36, No.12, Dec.,1995

[8]黒川雅人,洪政国,"形状情報を用いた画像の類似検索システム",情処学論 Vol.32 No.6,June 1991

[9]金原史和,佐藤真一,浜田喬,"プリミティブ分解による多様な検索条件を扱いうるカラー画像検索",情処学論 Vol.37, No.11, Nov.,1996.

[10]椋木雅之,美濃導彦,池田克夫,"対象物スケッチによる風景画像検索とインデックスの自動生成",信学論 D-II Vol.J79-D-II No.6, pp.1025-1033, June,1996

[11]小早川倫広,星守,大森匡,照井武彦,"ウェーブレット変換を用いた対話的類似画像検索と民俗資料データベースへの適用",情処学論 Vol.40, No.3, Mar.1999

[12]木本晴夫, "感性語による画像検索とその精度評価",情処学論 Vol.40 No.3, pp.886-898,1999

[13]栗田多喜夫,加藤俊一,福田郁美,坂倉あゆみ, "印象語による絵画データベースの検索",情処学論 Vol.33 No.11, pp.1373-1383,1992

[14]浦谷則好,柴田正啓,野口英男,相澤輝昭, "静止画検索システム FORKS の試作" ,情処学論 Vol.28, No.7, Jul.,1987

[15]西山晴彦,松下温, "画像の構図を用いた絵画検索システム", 情処学論 Vol.37 No.1, pp.101-109,1996

[16]望月貴裕,伊藤崇之, "画像の全体構成による柔軟な画像検索の画像検索の一手法", 信学技報 IE2000-155,2001

[17]加藤宗子,"柔らかい画像検索における特徴選択", 信学論 D-II Vol.J-80-D-II No.2, pp.598-606,1997

[18]青木秀一,青木輝勝,安田浩,"動画像からのシーン検索のための略画処理手法の提案",情報処理学会 CVIM 研究会,2002.1