

画像検索のための3Dインターフェースの有効性の検証

小池真由美*1 青木輝勝*2 池田佳代*3 伊藤学*2 日高宗一郎*4

*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 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も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いるが、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した上で、筆者らが提案している3D入力インターフェースについて2D入力インターフェースに比して有効であることを検証する。

Verification of Effectiveness on 3D Interface for Image/Video Retrieval

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

*1 Espalier Inc. *2 University of Tokyo, RCAST
*3 Excellead Technology
*4 National Institute of Informatics

In recent years, content distribution through the Internet has been actively carried out, and image/video distribution as well as music distribution has become widespread in rapid pace. Under this situation, there are growing needs for image/video retrieval, however, 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 shows our proposed interface (3D interface) improved the precision of image/video retrieval compared to 2D input interface.

1. はじめに

近年の個人の画像の収集量は、かつてないスピードで増加し続けている。ビデオカメラの世帯普及率が約 40%、世帯普及率は約 80%という事実に加え、ビデオテープが DVD にとって代わり、ハードディスクレコーダが台頭し、映像のデジタル化が進んでいる。このことにより、大容量の画像が省スペースで劣化なく収集・保存できることになり、今後ますますの個人収集画像の増加を促すことが予測される。

このことから、自身の所有する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。個人保存の画像がデジタル形式に移行しているという事実は、この要求に応えやすい環境が整いつつあることをも示している。

さらに、ユーザの意向も自身の撮影した画像を個人保存や友人に送付することにとどまらず、不特定多数の誰かに見せたいという方向へ進んでいる。インターネットをはじめとするネットワークの世界では、BLOG 技術により、画像を含む個人の情報発信サイトが容易に日々更新可能となり、P2P 技術も発達しつつある。

では、このように増加の一途にあるユーザ所有の画像、インターネット上の画像から、所望の映像を検索する場合、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ることで認識しており、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。

2. 3D インターフェースの提案

2.1 従来の検索インターフェース

図 1 に、既存の検索手法のコンピュータにとっての難易度と人間にとっての難易度を示す。

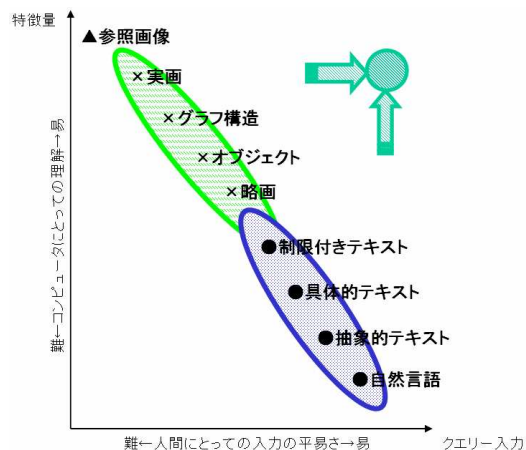


図 1 :種々の入力クエリーと用いる特徴量の関係 (図中右上の円形領域が、インデキシングコストを抑えつつ直感的な入力となる領域)

この図 1 より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。

しかしながら、「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る。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。そこで筆者らは特に(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することを目的に3D 検索インターフェースを提案している[1]。

2.2 3D 検索インターフェースの可能性

3D 検索インターフェースとは一般の静止画像 / 動画を検索する際に3D 情報を入力クエリとして用いる手法である。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。

2.3 3D インターフェースの設計

2.2 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、図3に示す3D入力インターフェースを設計している。

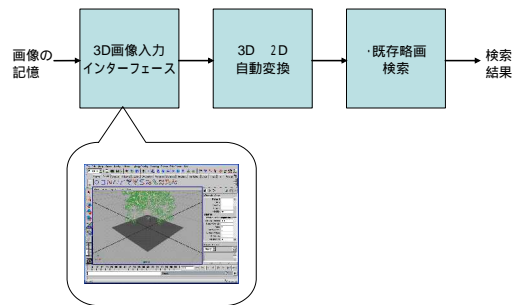


図 3 3D 検索インターフェース

図3において、「3D 検索インターフェース」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）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[2]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や映像、特に被写体として人物を含むものに対して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。

3. 画像検索のための3D インターフ

エースの有効性の検証実験

3.1 実験概要

本章では、2章で述べたように、本当に3Dでの

入力が直感的なものとなりえるか、を実験を通じて実証する。本実験の目的は、2個以上の物体の距離・遠近等の相関関係を表すのに2D描画（配置）に比して3D描画（配置）が有効かを検証することである。

実験を行うにあたっては被験者に検索すべき静止画像を記憶させ、2D略画入力（被験者にA4の用紙に略画を描かせる）と3D略画入力（被験者には2つの物体を配置させる）を行わせ、その精度を比較することにより行う。

3.2 実験準備

実験準備として実験素材の撮影を、以下の手順で行なった。

- (1) 机に同じ大きさのオレンジ（直径約7cm，高さ約6cm）を隣り合わせに2個置く。この時、2個のオレンジの間の距離を10cmとする。
- (2) デジタルカメラを(1)の2個のオレンジ間の中心位置に水平に備え、カメラのレンズからオレンジの手前の端までの距離を20cmとする。
- (3) (1)(2)の状態を「基本位置」として撮影する。（図4参照）



図4 実験の「基本位置」

- (4) 右のオレンジを10cm奥に配置し撮影する。
- (5) (4)の手順で、右側のオレンジを更に10cm奥に配置、撮影を繰り返し、40cmの位置まで撮影。
- (6) 次に2個のオレンジの左右間を20cmに離す。
- (7) (4)(5)と同様の手順を繰り返す。

以上の手順で合計10枚の写真を撮影し、A4サイズにてカラー印刷した。A4サイズを選択した理由は、画像検索の際に、目的画像をコンピュータ上で、手書きする場合、その最大サイズはA4までになるためである。



図5 実験手順(4)の位置

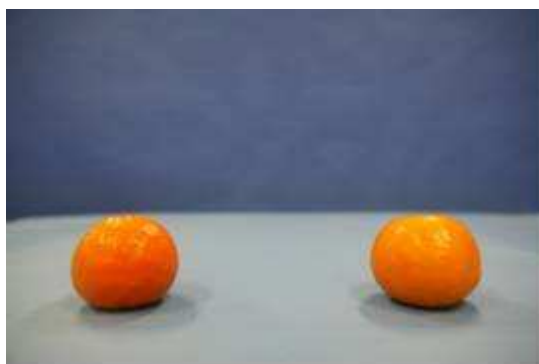


図6 実験手順(6)の位置

3.3 実験手順・内容

実際の実験は以下の手順で実施した。

- (1) 被験者の視線位置をカメラレンズと同位置にする。
- (2) 撮影時と全く同じ位置に2個のオレンジを配置し、その前面に図4の写真なたて置き、被験者は、この状態で撮影したものであることを確認する。
- (3) オレンジと写真を被験者の視界外におく。
- (4) 図5の写真を被験者に見せ、大きさや位置関係を記憶するよう伝える。この時の写真位置は、(2)で図4の写真を見せた場合と同位置とする。その後、写真を被験者の視界外におく。
- (5) 被験者にA4の白紙を渡し、(4)の写真と同様の位置関係と大きさで2個のオレンジを描画するよう指示する。この時、(1)の視線にいったん戻ってから描画してもらう。
- (6) 再度(4)を繰り返す。
- (7) 左側のオレンジを図4の基本位置に置いた上で、被験者に1個のオレンジを渡し、(4)の写真と同

様の位置関係と大きさとなるよう配置するよう指示する。この時、(1)の視線にいったん戻ってから配置をする。

以上が実験手順と内容である。(5)が2D描画、(7)が3D描画という位置づけである。

3.4 測定・比較方法

被験者の2D描画(配置)および3D描画(配置)結果の測定・比較方法を記載する。ここでは

- ・2個の物体の位置関係
- ・それぞれの物体の大きさ

の2点を重視した測定・比較方法を採用した。

まず、2D測定と3D測定の両方に共通に使用される座標を作成した。これは、「2.2 実験準備」で撮影した10枚の写真のオレンジの位置を黒丸で示した座標である。これを以下、「測定座標」と呼ぶ。

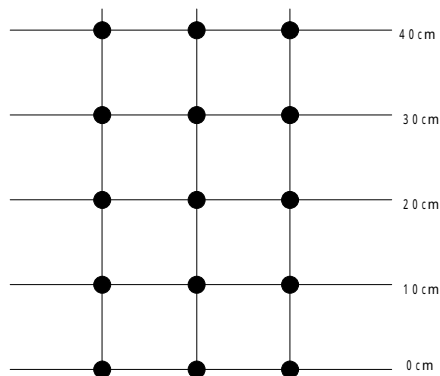


図7 測定座標

(1) 2D測定方法

- 「2.3 実験内容・手順」(4)の写真の左側のオレンジの中心を基点1、右側のオレンジの中心を基点2(基点1より右に10cm、奥に10cmの位置)として写真に印をつける。
- 2.3節(5)で被験者が描画した左側のオレンジの中心を上記(a)の基点1にあわせて重ねる。
- 被験者が描画した右側のオレンジの中心位置を確認し、「2.2 実験準備」で撮影した10枚の写真と最も近いものを選択する。このことで、右側のオレンジが測定座標のどこに位置しているかを確認、記録する。この時、写真の右側のオレンジの直径と、被験者の描いた右側のオレンジの直径の70%が重なる写真を選択することで被験者の示した位置を特定する。

(2) 3D測定方法

- 「2.3 実験内容・手順」(7)で被験者が配置した位置を記録しておく。
- (a)の結果を測定座標に記録する。

(3) 比較方法

- 測定座標に基点1および基点2を記録する。
- (1)-(c)の記録と(2)-(b)の記録を比較し、いずれが(a)に近似しているかを確認する。

3.5 実験結果

被験者6名に対して行った実験結果を図8、図9に示す。被験者に対しては「3.3 実験手順」以上の説明はせず、すなわち、実験意図は説明せずに実験を行った、また、一人の被験者が複数回実験を繰り返すと学習効果が生じる可能性があるため、被験者1名につき1回のみ実施した。

(1) 2D描画(配置)結果

2D描画(配置)の位置関係結果を図8に示す。2個のオレンジの基点を赤色で示し、被験者の描画位置を実験の順に数字で示す。

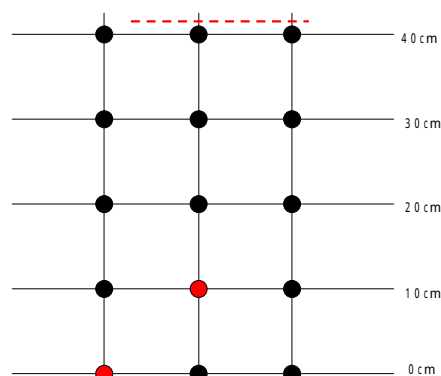


図8 2D描画(配置)位置関係結果

図8で示すように、右側のオレンジの横位置については被験者6名全てがほぼ正確に描画し、奥行き位置に関しては2名のみがほぼ正確に、他の4名に関しては、40cmより更に奥に離れた位置に描画しており、2D略画では奥行き感が高精度で表現できていない様子がわかる。

次に、描画の大きさについて表1に示す。

表1の結果で興味深いのは、被験者から見て手前に位置する左側のオレンジの大きさが比較的精度よく描けていることである。最も差異のある被験者2でも7.7cmと、写真の長径である6.2cmの

124%の長径の大きさで描画,全体平均すれば10%に満たない誤差である。それに対し,右側のオレンジは,写真の長径が4.4cmに対し,被験者1のケースである2.5cmはわずか57%という大きさであり,全体平均では24%の誤差が生じている。

表1 2個のオレンジの大きさの描画結果(長径)
(単位:cm)

	写真	1	2	3	4	5	6
左側	6.2	6.0	7.7	6.0	7.3	6.5	6.7
右側	4.4	2.5	4.8	3.0	5.5	3.6	5.3

(2) 3D描画(配置)結果

3D描画(配置)の位置関係結果を図9に示す。2個のオレンジの基点を赤色で示し,被験者の配置位置を実験の順に数字で示す。

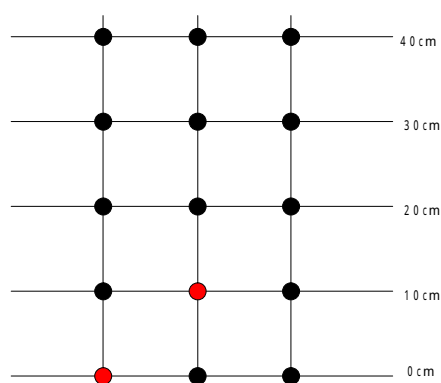


図9 3D描画(配置)位置関係結果

図9に示すように,右側のオレンジは,基点2の横方向に約5cm,奥行き方向に約13cmの範囲で配置されている。この結果は,当初筆者らが予想した通り,3D検索インターフェースは2D検索インターフェースと比較して奥行き感が高精度に入力できることを示している。

以上の図8、図9より,実験結果をまとめると下記の考察が導かれる。

- (1) 2D描画(配置)の場合,2個の物体の相関関係は,横より奥行きを把握することが難しい。
- (2) 2D描画(配置)の場合,2個の物体のうち手前にあるものの大きさが把握しやすい。
- (3) 2D描画(配置)の場合,デッサン力,記憶力による個人差が描画結果に現れやすい。
- (4) 3D描画(配置)の場合,2D描画(配置)

より2個の物体の奥行き相関関係が把握しやすい。

4.まとめ

本稿では,2個以上の物体の距離・遠近などの相関関係を表すのに,2D描画(配置)に比して3D描画(配置)が有効であることが検証できた。

今後の課題としては,

- (1)同実験で被験者を増やす
- (2)2個の物体の配置を変える
- (3)2個の物体を異なる形状とする
- (4)2個の物体を異なる色とする
- (5)2個の物体を異なる大きさとする
- (6)対象物体を増やす

等を踏まえてさらに評価実験を行い,3D描画(配置)が人間にとって直感的で表現しやすい手法であることを検証していく。

尚,最終目的は,この3D描画(配置)を画像検索適用していく,画像検索3Dインターフェースを実現することであるから,3Dモデリング,2Dモデリングのソフトウェアを利用した実験も並行してすすめ,その結果を実装に役立てる予定である。

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。尚,本実験に参加いただいた6人の被験者に心より感謝申し上げます。

文献

- [1] 小池真由美,青木輝勝,池田佳代,伊藤学,日高宗一郎「画像検索のための3Dインターフェース」情報処理学会 オーディオビジュアル複合情報処理研究会,2002.3
- [2]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