

高度画像検索のための直感的インターフェース

Intuitive Interface for Advanced Image/Video Retrieval

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}
 Mayumi Koike Terumasa Aoki Kayo Ikeda Manabu Ito Soichiro Hidaka

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、そのデジタル化も進んでいることから、自身の所有のコンテンツやインターネット上の画像から所望シーンを検索するユーザ要求は非常に強まっている。

しかし、残念ながら、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。これは、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。従って、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

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には、テキスト語句の厳密性と略画の曖昧さとを組み合わせたクエリー入力インターフェース、即ち、定義が明確な名詞(固有名詞等)と曖昧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[1]が、本稿では、このうちの略画入力の方に焦点を当てる。

「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で、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する[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/動画像を検索する際に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、あらかじめこの画像を見たことがある検索者が、この写真を探す場面を想定することにする。

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殆ど自明であろう。

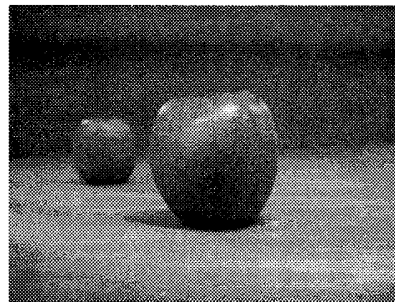


図2 検索したい画像

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提案

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提案している。

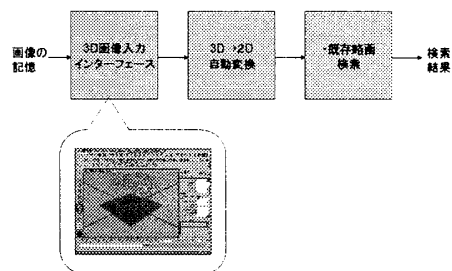


図3 3D検索インターフェース

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。

^{*1} (有) エスパリエ ^{*2} 東京大学

^{*3} (有) エクセリードテクノロジー

^{*4} 国立情報学研究所

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. 3Dインターフェースの有効性の検証実験

2章で述べた如く、この世の物は元来3Dであることから、従来の2D入力インターフェース入力での課題は、奥行きと位置関係の表現であろう。

そのため、『3Dでの入力が直感的なものとなりえるか』、つまり、奥行きと位置関係を2Dインターフェースに比してより正確に表現できるか、を筆者らは実験を通じて検証してきた。

過去の実験により、

- (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個の物体の奥行きの相関関係が把握しやすい。

という4点の結果が導かれている。[3]

つまり、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)が有効であることが検証できている。

上記実験は同形状、同色の2個の物体を用いて行なわれたため、課題として2個の物体の配置、形状、色、サイズを変更した実験や、対象物体を増やすこと等が残されたが、本稿では、このうち『2個の物体を異なる色とする』、これを選択した。

この理由は、『暖色系は浮き出て見え、寒色系は沈んで見える』という人間の目の立体知覚のメカニズム[4]や、さらに一般的には、白は前進色、黒は後退色であり、すなわち白は大きく見え、黒は小さく見える、という人間の目の特色が存在することから、そのメカニズムが2D描画、3D描画にどう影響するかを、特に2個の物体の距離・遠近等の相関関係において確認するためである。

上記に2点挙げたばかりでなく、色相に関する課題は多岐に亘る可能性が高いが、本稿では、同形状・同サイズの白と黒の球を用いた予備実験の結果を述べる。

3章冒頭で述べた過去の実験と同様の手法を利用したため、詳細はほぼ文献[3]に記載の通りであるが、準備として図3のように白と黒の球各一個の写真を数枚撮影後、以下概略により実験を進めた。図3を例にとって示す。

- (1) 図3の写真将被験者に記憶してもらう。
- (2) 奥側の球(図3では黒の球)を固定した状態から、被験者に白の球を図3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。
- (3) 図3の写真を再度被験者に記憶してもらう。
- (4) 被験者にA4の白紙を渡し、図3と同様の位置関係と大きさを手前側(図3の場合は白の球)を描画するよう指示する。この時写真撮影時の視線で描画してもらう。
- (5) 図3とは配置が逆、つまり白の球が奥で黒の球が手前というパターンを試験する。

以上において、(2)が2D描画、(4)が3D描画という位置づけとして分析を行なった。

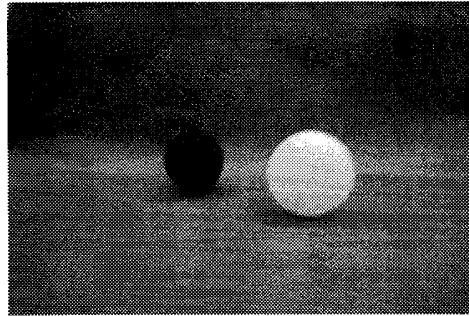


図3 検索対象画像の例

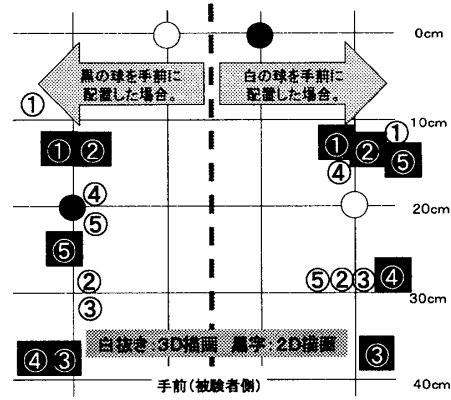


図4 実験結果(白と黒の球)

この結果、図4に示すように、奥行き・位置関係において、色による差異も2D描画と3D描画の差異も顕著に見られない結果となった。今後更に、手書き表現における色の影響をさらに研究し、検索インターフェースに有効に適用していく予定である。

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16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被験者に心より感謝申し上げる。

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 オードイ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
- [3] 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インターフェースの有効性の検証」情報処理学会 オードイオビジュアル複合情報処理研究会, 2002.6
- [4] 河合隆史, 田中見和 "次世代メディアクリエータ入門① 立体映像表現" カットシステム, 2003.11