

画像検索をサポートするカメラアングルを用いたコンテンツ分類

伊藤学¹⁾ 小池真由美²⁾ 池田佳代³⁾ 日高宗一郎⁴⁾ 青木輝勝¹⁾

¹⁾ 東京大学 先端科学技術研究センター 〒153-8904 東京都目黒区駒場 4-6-1

²⁾ (有) エスパリエ 〒175-0094 東京都板橋区成増 3-20-16

³⁾ (有) エクセリードテクノロジー 〒167-0054 東京都杉並区松庵 3-20-11

⁴⁾ 国立情報学研究所 〒101-8430 東京都千代田区一ツ橋 2-1-2

E-mail: ¹⁾{ito, aoki}@mpeg.rcast.u-tokyo.ac.jp ²⁾ koike@espalier.co.jp

³⁾kayo@excellead.jp ⁴⁾ hidaka@nii.ac.jp

あらまし 近年のデジタルコンテンツが爆発的な増加を受け、大量に蓄積されたコンテンツに対する効率検索が注目を集めている。しかしながら、現在主流の検索法はテキスト検索しかなく、効率的な手法は未だ十分とは言えない現状がある。本報告では、画像検索における効率の向上を大目標として、人間の経験値より共通に認識される、被写体の位置とカメラの高さや仰角など、それぞれのパラメータを組み合わせた分類法について述べる。さらにそれらを用いた画像検索テストベッド構築について報告する。

キーワード 画像検索 画像分類 カメラアングル メタデータ

Contents Classification Using Camera Angle to Support Image Retrieval

Manabu ITO¹⁾, Mayumi KOIKE²⁾, Kayo IKEDA³⁾, Soichiro HIDAKA⁴⁾ and Terumasa AOKI¹⁾

¹⁾ The University of Tokyo RECAST 4-6-1 Komaba, Meguro-ku, Tokyo, 153-8904 Japan

²⁾ Espalier 3-20-16 Narimasu, Itabashi-ku, Tokyo, 175-0094 Japan

³⁾ Excellead Technology 3-20-11 Syouan, Suginami-ku, Tokyo, 167-0054 Japan

⁴⁾ National Institute of Informatics 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo, 101-8430 Japan

E-mail: ¹⁾{ito, aoki}@mpeg.rcast.u-tokyo.ac.jp ²⁾ koike@espalier.co.jp

³⁾kayo@excellead.jp ⁴⁾ hidaka@nii.ac.jp

Abstract Recently, an increase explosive digital contents is continued. The efficiency retrieval to a large amount of accumulated contents attracts attention. However, the retrieval method used widely at present has only the text browsing. it cannot mean an efficient technique. This report describes the classification of narrowing down a candidate picture as a technique for aiming at improvement in retrieval efficiency. It is the classification, which combined each parameter, such as a photographic subject's position, height and elevation of a camera. There are recognized in common than human experience value. Furthermore, the picture retrieval test bed construction using them is described.

Keyword Image classification, Image retrieval, Camera angle, Meta-data

1. はじめに

近年、デジタル画像の爆発的な増加に伴い、効率の良い画像検索技術注目されてきている。従来画像検索の研究においては、画像中のどのような特徴量

を用いて検索を行うか、その特徴量を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、どのようにして画像群から特定するか、などの観点から研究が進め

られてきた。しかしながら、画像検索の研究が数多くなされていくにも関わらず、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多義性のため、キーワード付けや内容把握などの実現が簡単ではないということが、要因であると考えられる。画像検索の際に重要となるポイントは3つある。1つは自分のイメージする画像を探し出すための検索クエリ創出、2つ目はデータベースなどに蓄えられた画像に対し検索要求に対応させるための特徴量抽出、最後は両者のマッチングである。

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記憶しており、文書検索と異なり、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検索クエリとして生成でき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る。

一方、ユーザからの検索要求に対し該当する候補画像を絞り込むためには、画像の特徴量を抽出し検索対象とさせるメタデータ生成が重要な技術となる。一般的に文字をベースとした検索に対してメタデータ化を行う場合、撮影された情報や内容を書くことになるが、それらの情報を詳しく知っている者は、撮影者や編集者となる。この場合、詳しく知っているがゆえに、キーワードやタイトルなどの記述は被写体そのものの固有名詞を用いやすい。しかしながら、検索するユーザはそれらの固有名詞を知らない人がほとんどであろう。メタデータ化は、ユーザから送られる検索クエリに対し、万能性を追求すればするほど、候補映像として該当するが、膨大になってしまうという現状がある。

ユーザの検索要求に対してどの程度の候補画像を返すかは、検索クエリとして送られてくる文字情報や数値情報と、画像の特徴量として抽出される情報とのマッチング精度が重要となってくる。文字情報の場合、前記したように画像のもつ多義性のため、

同じ被写体に対しても多くの表現（例えば、家、住宅、ホーム、建物、建造物など）があり、語彙階層を定義しなければならないなど、多くの課題がある。また、数値的な情報（色、明るさ、輪郭など）を用いた場合、人間の記憶の曖昧さを補うために、冗長性を広く設定する必要があるが、その分、候補画像が膨大になってしまうという弱点がある。

画像検索において、自分のイメージした画像を正確にクエリとして生成するインターフェースと、画像が持つ特徴を的確に抽出するメタデータ化が最も重要であり、この両者が同じ思想の元に生成されれば、マッチングの負荷が軽減するとともに、検索効率向上に大きく寄与できると考える。

本報告では現在存在する画像の検索効率向上に貢献することを目的として、これまで著者らが提案している[1]-[3]画像の見え方に着目した、人間の経験値より共通的に認識される情報として、被写体の位置や、被写体に対する相対的なカメラ位置及び仰角といった3つのパラメータを用いた画像分類法の説明と、それらを用いた検索テストベッド構築について述べる。

2. 画像検索における課題と解決策

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)画像探索型に分けることができる。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。(A)は、機械的な検索ツールをほとんど使わないことも、広く用いられている理由の1つと言える。

一般的に言えることではあるが、検索効率の向上を図るためには、分類を用いる方法がある。膨大な検索対象となる画像を、ある定義に従い分類することで、検索対象を絞り込むことができる。その後、テキストマッチングなどを行うことで、イメージした画像を効率よくユーザに返すことが可能となる。しかしながら、分類方法に人間の主観が入ってしまうと、前記したように画像の持つ多義性のため、解釈は無限大となり、結果としてテキスト検索と同じ問題を抱えてしまう。

著者らは、画像分類において、万人が画像を見て共通的に認識できる客観的な分類方法として、被写体の位置、被写体との相対的なカメラの高さ、及びカメラの仰角を用いた新しい分類法として“AVR Attribute 方式”を提案する。画像検索において、1つの手法のみで満足な性能を実現することは困難であり、複数の手法を組み合わせる使用が今後重要になる。このような観点に立てば、既存の様々な手法とは従属ではない独立手法を提案することは極めて重要である。また、検索において万人が共通に認識できる分類を用いることは、クエリーを創出するユーザと、画像からその特徴をメタデータとして生成するプロバイダが、同じ特徴量を生み出すことになり、マッチングの精度も向上すると考えられる。

3. AVR Attribute 方式とは

人間が1度見たことのある画像を探し出す際、頭の中でその画像をイメージする。「遠くに山があつて、その手前に川が流れていて〜。」など考えるであろう。また、今まで1度も見たことのない画像を検索する際も、これまで見たことのある画像と類似させてイメージするであろう。本研究のポイントは、これら万人が画像に対してイメージする共通項を探し出し、それを用いた特徴表現を行い分類することにある。



図1 風景画像

図1に、一般的な風景画像1を示す。これらからイメージされる特徴として、テキスト検索に用いるキーワードであれば「風景」「海」「街並み」など思いつくであろう。しかしながら、これらのキーワードは前記したように、あまりにも高い語彙階層にあり、このままクエリーとしてしまうと、候補画像は膨大になる。では、他に画像を見て分かる情報として考えると、

- ・ 遠くに海がある。
- ・ その手前に街並みある。
- ・ 高いところから撮影している
- ・ カメラは下を向いている。

などは容易に判断できるであろう。

これらの画像情報を整理して特長表現として用いることはできないか検討を行った。特徴表現するためには、万人が画像を見て共通の認識を持つ必要がある。つまり「被写体の位置や距離感」、「被写体に対する相対的なカメラの高さ」、「被写体に対する相対的なカメラの仰角」を証明する実験を行った。

表1 画像認識シート

項目	パラメータ値		
被写体の位置	近距離 (約0~10m)	中距離 (約10m~1km)	遠距離 (約1km~∞)
カメラの高さ	地上	展望台などの高所	空撮
カメラの仰角	上向き	ほぼ水平	下向き

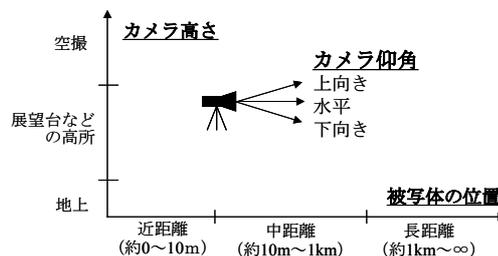


図2 パラメータ値の概念

被験者による認識実験では、表1に示す画像認識シートを用い、図2に示すパラメータ値の概念を参考に行った。表1において、“被写体の位置”とは画像に写っている被写体が指定された距離（近・中・遠距離）に存在するか否か示すもので、全ての値にマークが入ることもある。“カメラの高さ”とは、被写体の見え方に対する相対的なカメラ高さであり、地上、展望台などの高所、さらには飛行機など空からの空撮の3段階に区切った。さらに、“カメラの仰角”とは、地面に対して仰角が水平、上向き、下向きの3段階に別けた。両者とも3段階の選択肢より1つを選ぶことになる。実験にあたり16人の被験者に10枚の画像を提示した。その実験結果として、被験者の共通認識率を表2に示す。

被写体の位置では平均 87.5%, カメラの高さでは 99.3%, カメラの仰角では 92.5%, 全パラメータの総合では 81.3%という結果を得た。本実験では、被験者が 16 人、参照画像 10 枚という少数ではあったが、かなり高い確立で共通に認識していることがわかる。つまり、これらのパラメータによる認識が有効であることを示している。

表 2 被験者による共通認識結果

実験画像	パラメータ種			全パラメータ 総合
	被写体位置	カメラ高さ	カメラ仰角	
1 	93.8 ^(%)	100 ^(%)	100 ^(%)	93.8 ^(%)
2 	100	100	100	100
3 	87.5	93.3	100	87.5
4 	87.5	100	75.0	75.0
5 	75.0	100	100	75.0
6 	87.5	100	100	87.5
7 	100	100	50.0	50.0
8 	100	100	100	100
9 	68.8	100	100	68.8
10 	75.0	100	100	75.0
平均	87.5	99.3	92.5	81.3

本稿のパラメータの組合せは、被写体の位置では 8 通り、カメラの高さと仰角では 9 通り、両方で 72 通りの分類が可能となる。100%の精度は無いものの、一般画像に対して主観に左右されない分類ができたことは検索効率に大きく貢献できると考える。実際、筆者が個人で所有する一般的な風景画像 300 枚に対して、本方式を用いた分類を行ったところ、22 パターンに分類できた。

4. 画像検索テストベッド構築

これまで検討してきた分類を活用するためには、画像に対して検索対象となるメタデータに AVR Attribute 方式を取り入れる必要がある。採用するフォーマットは現在注目を集めている MPEG-7 を用いることを前提とする。MPEG-7 を用いることは、国際標準として一般性のある記述フォーマットであることと、このようなコンテンツ検索エンジンが分散設置された場合、機器・データ間で記述フォーマットの共通化・互換性確保をする事で、ユーザにとって検索しやすい環境を提供できると考えたためである。また、近年、これまで多く用いられてきた RDB (Relational Data-base) の他、XML をそのまま検

索対象可能なネイティブ XML データベースがリリースされていることも理由である。

次に、著者らが提案する AVR Attribute の記述法の検討を行う。方法はいくつかあるが、現在構築しているテストベッドは文字検索をベースとしている事から、文字情報としてコード化することとした。図 3 にコード化の概念を示す。図に示すように、被写体の位置、カメラの高さおよび仰角それぞれに 3 個ずつ、合計 9 個の文字を用い、該当する項目に“1”を、そうでない項目に“0”を記述した。

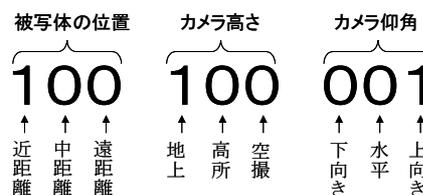


図 3 AVR Attribute のコード化

メタデータ構成としては、コンテンツ内容として、タイトル、撮影日、撮影場所やフリーテキストやキーワード、さらに ID やコンテンツ実態の参照先など記述している。また、AVR Attribute に関しては、図 4 参照に示すように<Semantic>を用い、<Label>に AVR Attribute を明記し、<Definition>にコードの記述を行った。

```

<Semantic>
- <Label>
  <Name>AVR Attribute of Image</Name>
</Label>
- <Definition>
  <FreeTextAnnotation>100100001</FreeTextAnnotation>
</Definition>
</Semantic>

```

図 4 AVR Attribute の記述例

作成したプロファイルをもとに、Web検索可能な実験用検索テストベッドを構築した。図5にその構成を示す。データベースソフトとしては、XQuery対応のオープンソースでネイティブXML対応のeXist[4]を採用し、MPEG-7をそのまま検索対象とすることを可能とした。また、一般的な文章などを用いた検索に対応できるよう、日本語形態素解析アプリケーション Sen[5]も搭載している。データベースには、コンテンツ実態、サムネイル画像、AVRプロファイルが一対に格納されている。主な機能としては、コンテンツ検索

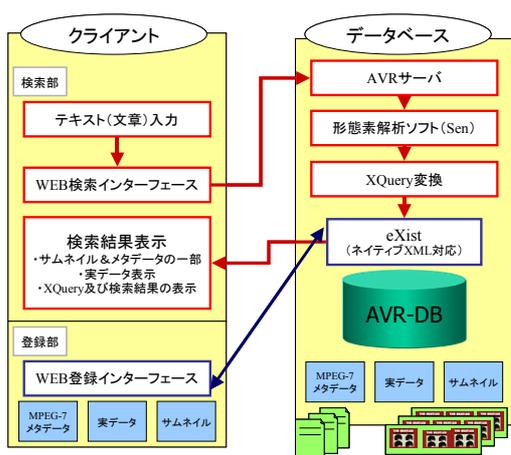


図5 検索テストベッドの構成

機能、検索結果表示機能、コンテンツ登録機能、形態素解析機能、XQuery表示機能、検索時間表示機能などがある。図8に検索結果表示の例を示す。



図6 テストベッド検索結果表示の例

6. まとめと今後の課題

本報告では現在存在する画像の検索効率向上に貢献することを目的として、画像の見え方の共通性に着目し、人間の経験値より共通的に認識される情報として、被写体の位置や、カメラの高さ及び仰角といった3つのパラメータを用いた画像分類法について述べた。さらに、それらを用いた検索テストベッドを構築し運用を開始した。本方式を用いることで、主観に左右されない見え方を基にした、類似画像をまとめることが可能となり、従来からある検索技術と併用し用いることで時間的、作業的な検索効率向上に大きく貢献できると考える。

今後の課題としては、被験者を増やしより正確な

データの収集や、パラメータの選択肢の増減と、特に重要なのはメタデータ化の自動抽出である。現在は、AVR Attribute の生成は全て人間の目視によって判断し手動で記述している。しかし、膨大なコンテンツに対して全てそれを行うことは現実的ではない。そこで今後は、コストを抑えたメタデータ生成を研究する予定である。

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)」の一環として行われたものである。

文献

- [1] 伊藤学、小池真由美、池田佳代、日高宗一郎、青木輝勝、“映像検索のためのクエリー生成とインターフェース構築”、情報処理 FIT2004, J-070, pp.361-362, Sep,2004
- [2] 伊藤学、小池真由美、池田佳代、日高宗一郎、青木輝勝、“奥行き情報を用いた画像検索インターフェースの検討”、画像電子第213回研究会 pp.57-61, Oct, 2004
- [3] 伊藤学、小池真由美、池田佳代、日高宗一郎、青木輝勝、“被写体の位置情報を用いたコンテンツ分類”、情報処理AVM第47回研究会 Vol.2004, No.126, pp.39-44, Dec,2004
- [4]eXist <http://exist.sourceforge.net/>
- [5]Sen <http://yamaguch.sytes.net/~tora/opensource/sen/>