

画像データベースに適したメタデータの分析と評価[†]

福本 徹*・赤堀侃司*

東京工業大学大学院社会理工学研究科人間行動システム専攻*

学校現場では、これまで写真やビデオといった静止画や動画が多く使われており、教育効果をあげてきた。近年、インターネット上には膨大な数のデジタル画像が蓄積されつつあり、画像データベースが注目を集めている。教室からインターネットに接続できるようになると、授業に必要な動画や静止画を、教師が的確に探し出せるようになることが求められる。本論ではメタデータ型の画像データベースを対象として、枠組みと記述内容について以下のように実験と考察を行った。まず、動画と静止画の両方を用いて画像に付与されたメタデータの枠組みについて、記述すべき内容とその記述順序を抽出した。次に、メタデータとしてキーワードを選び、画像に付けるための一貫性・妥当性のあるキーワード付けの基準について考察を行った。そして、以上の結果をもとに実際に検索システムを構築し被験者による検索実験を行い、その有効性を検証した。

キーワード：知的学習支援, WBT, 画像検索, データベース

1. はじめに

近年、デジタルカメラやスキャナといったデジタル入力機器や、大容量で低価格な記憶媒体が普及したことで、膨大な数の画像が蓄積されつつある。これらを効率的に管理するために、画像アルバムや画像ファイリングシステムといった画像データベースが注目を集めている。またインターネットの普及および高速化によってデータベースへのアクセスが容易となり、画像データに対する検索ニーズが高まっている。

一方これからの時代は、デジタル画像はコンピュータの世界だけではなくてきている。2000年末にはBSデジタル放送がスタートしたが、地上波においてもデジタル化が決定し、多チャンネル化時代を迎えることとなる。ビデオカメラにおいてもテープではなくディスクに記録するような製品が開発されている。このようなデジタル画像をハードディスク型VTRにそ

のまま記録し、デジタルカメラによる静止画とVTRによる動画を横断的・統一的に検索する場面が今後増すものと思われる。

学校現場では、これまで写真等の静止画やビデオやテレビ番組といった動画が多く使われており、教育効果をあげてきた。教室からインターネットに接続できるようになると、ビデオ・オン・デマンドの形で必要な時に必要な動画がインターネットを介して教室に配信されるようになるであろう。このときには、例えば授業に必要な動画や静止画を、教師が的確に探し出せるようになることが求められる。また、生徒が興味のある動画や静止画を、簡単に探せるようになることも必要であろう。その上、2001年度中には全ての学校がインターネットに接続されるようになった(文部省1998, 1999)。すると今後は児童・生徒がブロードバンド回線や画像機器を使いこなすことであろう。

画像に対する検索を行う場合のシステム構成は、キーとなる項目によって、特徴量型、感性型、メタデータ型の3種に分けることができると考えられる。特徴量型は、画像の色やオブジェクトの形状といった画像から直接抽出される特徴量によって検索を行う(串間ほか1999)。感性型は感性語などの感性に関わる情報を検索者が入力し、システムが色やテキストなどに変換して特徴量型と同じように検索する(木本1999)。

2002年7月24日受理

[†] Toru FUKUMOTO* and Kanji AKAHORI*: The Analysis and Evaluation of Metadata with Image Database

* Graduate School of Decision Science and Technology, Tokyo Institute of Technology, 2-12-1, O-okayama, Meguro-ku, Tokyo, 152-8552 Japan

メタデータ型はデータベースシステム提供者が画像に説明として様々なメタデータを予め付けておき、検索者はそれをもとに検索を行う(清木ほか 1996)。

特徴量型や感性型では、画像から機械的に特徴量を抽出する。そのため中間ほか(1999)にもあるようにオブジェクト等の認識精度が高くなく、そのため検索時にノイズが多く含まれて検索精度、特に適合率が低下する。

一方、メタデータ型では付与者が画像にメタデータの内容を記述するため、人手でのチェックが入っているという点で物体を対象とする検索に強く結果として検索精度が高くなる。しかしどのようなメタデータを付与するかが問題となるし、その上にメタデータを付与する手間が発生する。例えば国立歴史民俗博物館の歴史民俗画像データベース(照井 1998)や商用の画像データベースシステム(例えば CanDINet など)ではコンテンツ提供者が人手に頼ってメタデータを付与しているのが現状である。

我々は、一度人間の目による確認がなされていることによる検索精度の高さと、文書検索や Web 等の検索エンジンとの親和性を考え、メタデータ型の画像検索に注目している。

メタデータ型ではまず、メタデータそのものの枠組みを設計する。次に決定された枠組みの下で、付与者が画像にメタデータの内容を記述する。そして検索者はテキストなどを検索のためのキーとして入力し、システムが入力されたテキストと画像に付けられているメタデータの枠組みと内容とを照合することで検索を行う。

しかしメタデータの設計がうまく出来ていない場合には、画像データベースの用途や収録されている画像の内容に相応しくないメタデータの枠組みとなる場合もある。また、決められた枠組みの中であってもメタデータを付ける人によってはその付け方が不規則である場合もあり、不十分な場合もある。この場合は所望する画像が検索できないといった事態が発生してしまう。特に商用システムの場合にはデータベースの品質に影響する。また児童・生徒にとっては正しく検索できないということは、学習の妨げともなりかねない。そのため一貫性と妥当性のあるメタデータを設計し内容を記述する必要がある。つまりメタデータは、枠組みと記述内容という2つの面から成るのである。

絵画の分野では画像に含まれる内容や書誌情報を記述するために、画像記述言語という研究がなされて

きた。4W 法(守田ほか 1995)では、Who・What・When・Where の4つの視点に基づいて記述する。Who は絵画に描かれている対象や存在であり、What は動きや出来事を表し、Where は描かれている場所、When は描かれている時間を表している。この方法を用いた検索実験では、記述の難しさと対象を1つに限定することが問題とされている。PDL 法(LEUNG *et al.* 1992)は実体-属性-関係モデルに基づく記述法である。実体は対象物そのものを指し、属性はその対象物の特性や特徴、関係はその絵画に含まれる対象間に存在する関係を表すものである。PANOFSKY (1987)は美術史におけるイコノグラフィーを利用した3段階の分類法を提案している。第1段階を自然的意味といい、絵画中に描かれた人物などの対象および対象間の相互関係と、対象が表現する動作を示すものである。第2段階は伝習の意味といい、絵画に対するイメージを通してテーマや概念を把握するものである。第3段階を内的意味といい、時代や階級・宗教や哲学的心情などといった、作者や描かれた対象に固有の環境を調べることで把握するものである。絵画における象徴的価値を示すものである。しかしこれらの記述法は画家の意図や書誌情報に重きを置かれているものであり、近年増加しているデジタル画像に付与するメタデータに直ちに応用できるものではない。

一方、学校におけるインターネットの検索や情報過多に関する分析として、越桐(2001)は初等中等教育学校のwebページの管理者を対象とした調査を行っている。アンケート中では、不足している教育・学習情報として画像などの素材データを挙げた回答が34.2%で調査項目中第2位であった。今後インターネットの教育利用がますます進んで活用する方法論がある程度浸透すれば、教育・学習素材情報に対する要求が高まっていくことも予想される。また、情報受信時の問題点として過剰な情報からの取捨選択が困難であることを挙げた回答が61.5%で、調査項目中トップであった。このことは、教育・学習の場で利用可能な情報を探し出すことが容易でないことを示している。

教育用コンテンツを流通させることを目的として、IEEE1484 の Learning Object Model (LOM) が提案されている(先進学習基盤協議会 2000)。LOM では教材を再利用するために分類項目の標準化を提案している。これを用いると、各学習リソースのタイトルやジャンル、データタイプ、説明文、キーワードなどを教育用コンテンツに対して付与できる。このように教育分

野においても、学習素材に対してメタデータを付与し、検索や再利用を容易にする試みがなされている。しかし、画像等の学習コンテンツに対してキーワードを付与する場合の基準については明確に述べられていない。

また手間とコストの面を考えると、学習コンテンツとなる大量の画像に対して常に専門家がメタデータを付与できるとは限らない。コンテンツの検索や再利用といった流通性を確保するためには、非専門家であっても専門家と同じような品質となるメタデータを付与できるようにする必要がある。例えば写真家が、自分が撮ったすべての画像に自らがメタデータを付与してゆくことは現実的ではない。付与する場合にはおそらくアシスタントやコンテンツ流通業者等の、写真家とは別の人間が行うであろう。またアマチュアカメラマンが撮った画像の場合には、アマチュアという非専門家がメタデータを付与することになる。

我々の研究の目標は、非専門家であったとしてもコンテンツの検索や再利用といった流通性を確保するために、様々な画像に対して的確なメタデータが付与できるようにすることにある。

2. 研究の目的

本論文では、画像検索において画像に付与するメタデータの枠組みと記述内容について考える。画像のどのような部分に注目してメタデータを付与するかは、検索精度を向上させ、ひいては検索者が検索しやすくなるためには重要な鍵である。そこで、人間が画像のうちどのような部分に注目しているかを実験によって明らかにすることを第1の目的とする。このことが明らかになれば、実際に画像にメタデータを付与する際のガイドともなる。

また、先に述べたように画像データベース中に静止画と動画とが混在し、これらを統一的に検索することが今後多くなると思われる。それでは静止画と動画とでは注目する部分は異なるのか、もし異なるとすればどのような点についてなのかを見出すことが第2の目的である。もしそうであるならば、画像に付与するメタデータの枠組みを動画と静止画とで分け、それぞれに相応しいメタデータを付与することで、検索者がより検索しやすくなると考えられる。

以下、3章ではメタデータの枠組みに関して、4章ではメタデータの枠組みの中に記述される語句の内容に関して、5章では3章と4章で得た知見をもとに構築した検索システムとその評価について述べる。

3. メタデータの枠組み

本章では、メタデータの枠組みに関して行った実験とその結果について考察する。先に述べたように、きちんとした枠組みを設計することが、データベースの品質を左右する。従ってメタデータの枠組みはメタデータ型データベースの基礎となる。

3.1. 実験概要

実験は28名の短大生・大学生を被験者として、これを14名ずつ2つのグループA、Bに分けて行った。対象とした画像は動画、静止画各13種類であり、同一コンテンツを使用して、コンテンツ間の影響を可能な限り排除した。コンテンツの種類はニュース、スポーツ、ドラマ、風景映像、アニメなどである。

なお、画像中に現れる色を対象とした検索には特徴量型が強いことが明らか（近藤ほか 2000）である。そのため模様やテクスチャパターンのような画像や、色に関する語句に対する評価は本論文では行わないこととした。

実験の手順は、1シーンの動画またはその動画の中から代表画像と考えられる1枚の静止画を提示し、その間に被験者が対象となる画像から思いつく単語や短い文をフリータムで用紙に記述した。これを付与語と呼ぶことにする。動画は同じシーンを3度繰り返し、静止画は同じ時間だけ提示した。本研究では画像に対する記憶ではなくどのように見ているかについて着目しているため、記述語数や時間の制限はしていない。グループAには先に動画を13種類続けて提示し、次に静止画を13種類続けて提示した。グループBはグループAとは逆に、先に静止画を、次に動画を提示した。本実験で用いた画像の種類を表1に、画像の例を図1～3に示す。

3.2. 実験結果と考察

・記述内容と順序

実際に付与された付与語の例を表2に示す。付与される順序は以下のような傾向が見られた。まず画像中の人物や場所が特定できる場合には固有名詞が記述された。例えばクリントン大統領、曙（力士のしこ名）などである。2番目にはその画像の状況が記述された。状況とはその場やその時のありさまのことと定義し、例えばサッカーやドラマなどである。3番目に場面である。場面とはその場の様子と定義し、ゴール前、話し合いなどである。1つの状況の中で様々な場面が展開され、サッカーという状況の中でパスやシュートと

表 1 画像の種類

番号	画像の内容
1	大統領の会談
2	記者会見
3	相撲：取組み
4	サッカー：ゴール前
5	サッカー：ゴール前
6	野球：タイムリーヒット
7	野球：ホームラン
8	ドラマ：話し合い
9	ドラマ：お茶の間
10	風景：海
11	風景：夕日
12	アニメ漫画
13	アニメ漫画



図 1 サッカー：ゴール前

いう場面が現れるのである。その後中心の動作、シュート、バッターが打つ、次に画像中の主となるオブジェクト以外の物体や背景など周辺の状況が記述される。その後周辺動作である。最後に美しい、綺麗などの画像から受ける印象や感じ方などの感性語が記述された。このような傾向で全体の 47.1% を占めた。

しかし画像中の場面が特徴的な場合は、中心の動作が最初に記述される傾向が見られた。例えば表 3 に示すように、図 1 の画像では、「シュート」という動作が特徴的なため、状況である「サッカー」よりも先に記述された。感性語についても同様であり、図 1 の画

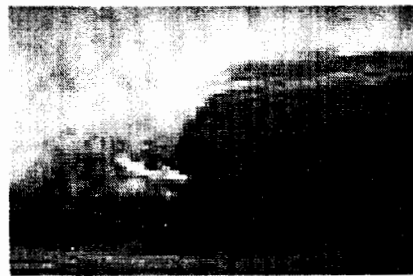


図 2 サーフィン

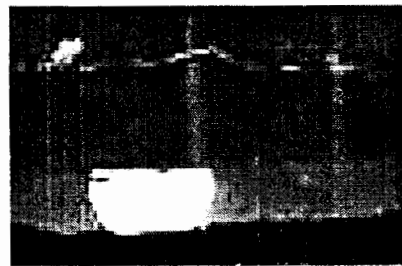


図 3 夕日

像ではゴールによる喜びを表す「入ったー！」という感性語が最初に記述された。また、固有名詞はそれ自体が特徴的であると考えられ、通常の画像の場合でも最初に記述されると言える。転置して記述された割合は全体の 20.5% であった。特に図 1 のコンテンツの動画では被験者 28 人中 20 人、野球のホームランのコンテンツの動画では 21 人がこの傾向で記述していた。

つまり、付与者にとってインパクトのある記述内容が先頭に転置されて記述される傾向が見られた。表 3 にあるように一般的には「サッカー→ゴール→シュート」という記述順序になるが、転置すると「シュート→ゴール→サッカー」という記述順序などとなる。

ある画像に対して付与者がメタデータを付与する際の提示順序として上述した順序に従うことで、付与者にとって効率の良い付与が行えると考えられる。このことを利用して画像にメタデータを付与する際にコンピュータから支援することが可能となる。

表 2 被験者による付与語の例：通常の場合

画像 No.	固有名詞	状 況	場 面	中心の動作	周辺の状況	周辺の動作	感性語
4		サッカー	ゴール	シュート	広告の看板		やったー！
1	クリントン		会談	笑う	花		
8		ドラマ	会議	主張する	男	歩く	
7		野球		打つ	観客	手をたたく	

表 3 被験者による付与語の例：転置する場合

状況	場面	中心の動作	感性語
サッカー	ゴール	シュート	入ったー！

中心の動作	状況	場面
シュート	サッカー	ゴール

感性語	状況	場面
入ったー！	サッカー	ゴール

表 4 付与された語数 (n=28)

番号	動 画	静止画	有意差
1	133	99	**
2	130	106	*
3	94	81	*
4	93	85	
5	95	76	*
6	82	100	
7	71	78	
8	115	84	**
9	112	121	
10	96	94	
11	138	92	**
12	122	88	**
13	103	104	

* $p < .05$, ** $p < .01$ (表 4~6 に共通)。

・付与された語の数

次に動画と静止画とで注目する点の数が異なるか否かをみるために、語の数に変化があるかどうかをカウントした。その結果を表 4 に示す。同一被験者が付与した語の数に対して t 検定を行ったが、動画の方が静止画より多いか、動画と静止画とで同じ程度であった。また、「走る」「座る」などの動作を表す語のみを抽出して表 5 に示すように同じく t 検定を行ったが、動画の方が静止画よりも数が多い傾向が見られた。静止画よりも動画の方が動作を捉えやすいことが推察される。そのため動きを示す画像では動きに注目することで、動画の方が静止画よりも語の数が増えたと考えられる。なお被験者のグループ A とグループ B とで両方の語数に対して t 検定を行ったが差はなかった。

被験者が画像から受ける情報量としては明らかに動画が静止画に勝っていると言える。しかし静止画はじ

表 5 付与語のうち動作を表す語数 (n=28)

番号	動 画	静止画	有意差
1	14	3	**
2	2	4	
3	6	4	
4	13	4	**
5	19	9	**
6	17	14	**
7	12	14	
8	18	11	**
9	11	9	
10	9	0	**
11	16	5	**
12	15	3	**
13	18	10	**

表 6 付与語の種類数 (n=28)

番号	動 画	静止画	有意差
1	55	41	**
2	60	39	**
3	52	35	**
4	39	31	
5	44	47	
6	51	30	**
7	31	33	
8	65	38	**
9	61	52	*
10	54	39	**
11	42	34	
12	51	46	
13	39	30	*

っくりと見る事が可能であるために、被験者は周辺状況に対してまで多くの語を付与することができる。そのため動きを示す画像以外の画像では、語の数は動画と静止画とではほぼ同じとなったと考えられる。

・付与された語の種類数

そして画像中のどのような点に注目するのかをみるために、被験者が付与した語のうち被験者の間で重複する語を省いて語の種類数をカウントした。その結果を表 6 に示す。このうち 1% 水準で有意差が見られた画像の内容を見てみると、画像 2 は記者会見、画像 3 は相撲の取組み、画像 6 は野球のタイムリーヒット、

画像8はドラマ中の話し合いのシーン、画像10は海の風景である。これらの画像に共通する特徴として、動画といってもあまり動きがなく静止画に似たものであった。つまり動画と静止画で情報量にあまり変化のない画像では、動画が静止画より語の種類の多い傾向が見られた。一方、動画では中心となるオブジェクトがよく動いているような画像では、動画と静止画とではほぼ同じであった。

なお被験者のグループAとグループBとで動画と静止画両方の種類数に対してt検定を行ったが差はなかった。

動画では画像中の情報に変化が少ないと、被験者には静止画のように感じられると思われる。また、静止画で記述された内容にプラスしてオブジェクトの動きに関する記述が見られた。つまり、まず動画を見ている状態として動いているオブジェクトに着目して記述を行う。同時に、静止画を見ている状態として周辺の状態や動作に関する記述を行う。このように、動画と静止画との両方を対象として見ていると考えられる。一方動画でも変化が多いコンテンツの場合には動いているオブジェクトにのみ集中して着目していると考えられることができる。

このように、画像に対する見方としては、静止画、動きの少ない動画、動画と3段階の認知形式が想定できる。特に動きの少ない動画に対するメタデータとしては、静止画と動画の両方の要素を記述する方法が考えられる。

3.3. メタデータの枠組みの基準

以上の結果から、以下のように基準を設定することができる。

最初にその画像中の人物や場所が特定できる場合にはその固有名詞、2番目にはその画像の状況、3番目に場面、4番目に特に動画においては画像中の主となるオブジェクトの動作、5番目にそのオブジェクト以外の物体や背景など周辺の状況、6番目に静止画あるいは動きの少ない動画では周辺の動作、最後に画像から受ける印象や感じ方などの感性語を記述する。また4番目にあげた画像中の主となるオブジェクトの動作については、静止画では欠落することを許容する。そして、インパクトのある記述内容が画像中に存在する場合には先頭に転置して記述する。

4. メタデータの記述内容

3章ではメタデータの枠組みについて述べたが、本

章ではメタデータとして実際にどのような語句が記述されるか、その記述内容について述べる。

本論文では特にプリミティブなメタデータとしてキーワードを扱い、専門家と比べて非専門家が記述する語句そのものの傾向について分析を行い、キーワードを付与する際の基準を明確にすることを目的とする。

4.1. 実験概要

被験者は大人7人（男性5人女性2人）である。まず、被験者には10種類の画像を1つずつ提示した。コンテンツの種類は風景、動物、静物、模様などである。3章の実験と同じく動画と静止画とで同一コンテンツを利用し、動画と静止画との提示順序でのグループ分けは行っていない。3章の実験結果よりグループ間の差が見出せなかったためである。そしてこれらの画像に対しキーワードを付与してもらった。被験者にはキーワード基準書や画像を見ながら思いつくままに付与してもらい、特に数の制限は加えていない。

被験者にはキーワード付け基準として、FUKUMOTO *et al.* (2000) や、画像検索に関する専門家3人による議論を通しての意見を参考に、

- ・妥当性のあるキーワードを付ける
- ・多くのキーワードを付ける
- ・主観的なキーワードと客観的なキーワードとを区別する

の3つを示した。

その後に画像検索分野を研究している専門家3人で、別個にキーワード付与を行った。そして先の被験者による付与結果と専門家による付与結果をつき合わせて、双方のキーワードの内容を専門家3人で評価した。以下では実験の結果を分析し考察を加える。

4.2. 実験結果

- ・キーワードの数とその妥当性について

表7にあるように10種類で動画静止画合わせて20通りの画像に対して総計で416個のキーワードが付与された。1通りの画像あたり約20個である。画像中に見られる物体の数とおおよそ同じ数のキーワードが得られた。動画と静止画とで個数の平均値についてt検定を行ったが、被験者・専門家とも差は見られなかった。専門家による評価で新たに加えられたキーワードは動画で48個、静止画で55個である。専門家による付与では、画像中にわずかに写る物体に対してもキーワードを付与する傾向が見られた。特に多くの物体が写りこむ画像で客観的なキーワードについては、多く付与された。例えば図4の画像に対しては、「猫」

表 7 付与されたキーワード数

	総 計		主 観		客 観	
	動画	静止画	動画	静止画	動画	静止画
被験者 7人	157	156	35	39	122	117
専門家 3人	48	55	2	5	46	50
小 計	205	211	37	44	168	167
合 計	416		81		335	



図 4 被写体が単一の画像

「アップ」「かわいい」と動画・静止画とも被験者と専門家との間でキーワードのずれは全く見られなかった。しかし図5の画像に対しては、被験者が付与したキーワードが動画で9個、静止画で14個、専門家が新たに付与したキーワードが動画で1個、静止画で7個であり、キーワードを付与する際における単語の違いが静止画で見られた。

次に、図6の画像に対して被験者は動画・静止画とも「赤」「水滴」や「気持ち悪い」というキーワードを付与した。一方、専門家は「模様」「テクスチャ」という的確なキーワードを付与した。こうした模様には画像中の物体が特定しにくいいため、被験者がキーワードを付与できなかったものと考えられる。画像の特性という意味では、模様やテクスチャといった、一見ただけでは物体が判別しにくい画像に対しては、キーワードの付与が困難であろうと推察できる。つまりデータが付与できないということはメタデータ型検索の限界であり、模様やテクスチャには特徴量型検索が適していると考えられる。

検索のためのキーワードという観点では、適合率という精度を向上させるために、画像中のわずかに写る物体にまでキーワードを付与する必要がある。しかし



図 5 多くの物体が写る画像

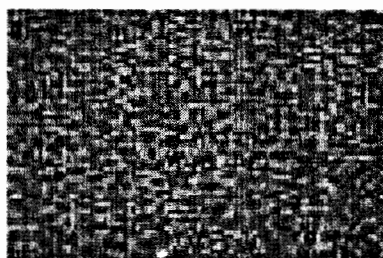


図 6 テクスチャ

妥当性という観点では、どのような物体がどのようなイメージでどれくらいの大きさで写っているかという情報が、実際に検索者が目的に合った画像を入手するために画像データベースを検索する際には必要と考えられる。このような情報は先に被験者に示したキーワード付け基準では画像に付与されない。画像中の物体の特徴に応じた、キーワードに対する重要度やキーワードが示す物体の画像中での占有面積といった何らかの指標付けが必要であると考えられる。

・主観的/客観的キーワードについて

表7にあるように主観的なキーワードは合計で81個であり、全体に占める割合は19.5%であった。動画と静止画とで個数の平均値についてt検定を行ったが、被験者・専門家とも差は見られなかった。また、専門家による評価では新たに動画で2個、静止画で5個が付与されたが、被験者による主観的なキーワードの数と比べても1割程度である。一方、客観的なキーワードは専門家による評価によって新たに動画で48個、静止画で55個が付与された。

キーワード数の差を考えると、主観的なキーワードは被験者と専門家との間にそれほど差が見られないと言える。一方客観的なキーワードは被験者と専門家との間にばらつきが見られ、一定していないことが言える。

主観的なキーワードがあまり多くなくしかも被験者

と専門家ではらつきがないということは、画像の特徴を表現する言葉が限られていると考えられる。この限られた言葉からキーワードとして付与するために、偏差が少ないのである。一方、客観的なキーワードは、被験者が主として画像の中心となっている物体に対してキーワードを付与したと思われる。これに対し、専門家は画像の特徴を丹念に拾い上げることで新たに付与するキーワードが増加したと推測できる。

前に述べたようにキーワードに対する指標付けを行い、より正確なメタデータの付与が求められる。

4.3. 考察

以上の実験結果を踏まえて、4.1.に述べたキーワード付与の基準は、以下のようにまとめられる。

・妥当性

対象となる画像データベースの用途を分析して、どのレベルでの区別が必要かによって妥当性を判断する。先に挙げた例のように「自動車」と「ボルシェ」は同じでよいのかどうか、といった点に注目する。汎用的なデータベースの場合は、「自動車」という一般名と「ボルシェ」という固有名詞の両方を付与する。

・キーワードの数

その画像中に在るものの数プラス、画像から受けるイメージの分として最低1つ必要である。

・重要度

その画像から大きく受けるイメージまたは大きな面積を占めるもの、つまりその画像のテーマや題名となりえるものを、重要度「大」とする。東京都内から撮影する富士山の写真のように、被写体が小さくてもインパクトが大きくてテーマとなりえる画像が考えられるからである。重要度「小」として、画像中に少しだけ写り込んでいるものや、その画像から受けるイメージが小さい場合に付けられる。原画像のサイズにしたときに肉眼で確認できる程度の大きさを想定している。重要度「中」として、「大」でも「小」でもないものとする。重要度による検索結果の順序付けを考えると、重要度「小」と付与したことによる検索者の見落としをできるだけ回避するためである。

なお「大」「中」「小」は、システム内ではそれぞれ「3」「2」「1」等の数字で記述する方が処理しやすいのは言うまでもないが、人間が見る際の感じ方としては「3」より「大」が望ましいと考えられる。

・主観的/客観的区別

客観的なキーワードとしてその画像中に在る具体的な物を特定する単語すべてと、主観的なキーワードと

してその画像から受けるイメージを最低1つ記述する。そしてそれぞれのキーワードに「主観」「客観」や、これらをコード化した数字などを付加しておく。

5. 検索システムの構築と評価

3章および4章で述べた知見に基づき、実際に検索システムを構築し、被験者による評価を行った。

5.1. 評価概要

被験者は成人男女合わせて10人であり、Web等の検索エンジンの使い方には習熟している。画像データとして、テキストチャパターン（幾何学模様）、会社の社内風景、山や海といった自然の風景、といった画像100枚を検索対象としてデータベースに登録した。また簡易シソーラスとして、画像に付与したキーワードの類義語を複数の国語辞典よりピックアップし、システムのキーワード検索部に組み込んだ。

実験には2つのシステム、AとBとを用意した。キーワードはシステムA、Bとも被験者とは別に3名の成人によって付与した。システムAは、3章および4章で述べた画像に対するキーワード付与基準を盛り込んだシステムである。また検索結果である画像を表示する際には、検索結果である画像のスコアが高い順に表示する。具体的には以下ようになる。

被験者が入力した検索キーワードが n 個であった場合、それらを $k1, k2, \dots, kn$ とし、ある画像 X に付与されたキーワード kp, \dots, kq ($1 \leq p, q \leq n$)と一致したとする。当該検索キーワードに対する画像 X のスコアは、 kp, \dots, kq の重要度をそれぞれ重み付けとして $w1, \dots, wq$ とすれば、

$$S(X) = \sum_{i=1, \dots, q} wi$$

となる。本実験では重要度大、中、小の場合に重み付けをそれぞれ5、3、1とした。なお、重要度の最適な数値化については今後の課題である。

一方、システムBは、付与基準を考慮せずにキーワードを付与したシステムである。先にシステムB、次にシステムAの順で付与を行い、システムBにおいてキーワード付与基準による影響を取り除いている。被験者にはこの両方のシステムで検索を行ってもらった。

システムAにおいて画像に付与されたキーワードは平均20個、システムBでは平均7個である。t検定によってシステムAがシステムBよりも1%水準で有意にキーワード数が多いことを確認した。またz検定において、システムA・システムBとも画像間

で画像に付与されたキーワード数にばらつきがないことを確認している。なお主観的なキーワードはシステム A にのみ付与しており、システム B では付与されていないことを確認した。

画像に付与されたキーワード部分を格納するデータベースのスキーマは、システム A では画像 ID と 3.3 節で述べた記述順序通りにキーワードを格納している。

表 2、表 3 の例では画像 ID とキーワードの組として、

- ・「画像 4」に対して「サッカー、ゴール、シュート、広告の看板、やったー！」

- ・「画像 5」は「シュート、サッカー、ゴール、やったー！」

である。システム B についても同様であるが、順序はランダムである。そして被験者が入力したキーワードを、システムではこのデータベース中のキーワードと各々左側から照合してゆく。

実験の手順は、被験者を 5 人ずつ 2 群に分け、最初にシステム A で次にシステム B で検索を行う群とし、最初にシステム B で次にシステム A で検索を行う群とした。前者を被験者 A 群、後者を被験者 B 群とする。検索手順としては西山ほか (1996)、前田ほか (1999) などの手法にならない、まず被験者にデータベース中に含まれる画像を 1 枚ずつ 30 秒間提示し、その後画像を隠して記憶を頼りにする状況の下で検索を行ってもらった。そして検索をスタートしてから提示した画像を発見するまでの秒数を計測した。これを 10 枚の画像にわたって行った。次にクライアント側の Web ブラウザのキャッシュをクリアし、画像を読み込んで表示するという条件を同じにした。そして使用するシステム A、B を変えて再び検索を行ってもらった。

最後に被験者にアンケートに答えてもらった。アンケートの項目としてはシステム B と比較した場合のシステム A での検索のしやすさ、検索結果の見やすさ、入力したキーワードに対しての検索結果表示の妥当性と、自由記述である。

5.2. 実験結果と考察

実験結果を表 8 に示す。被験者が 10 人と数が少ないため、ウィルコクソンの順位和検定を用いて、実験結果に対する検定を行った。

検索に要する時間については、図 6 のようなテキスト画像を除いてはシステム A のほうがシステム B よりも 5% 水準で有意に短い。システム A は該当のキーワードの重要度による重み付けを行ったスコア順

表 8 検索時間 (秒)

画 像	システム A	システム B	有意差
ビジネス	17.0	21.3	*
ビジネス	16.7	20.4	*
テキストチャ	23.6	23.6	
テキストチャ	22.5	22.9	
青 空	19.5	20.4	
青 空	17.7	18.7	*
夕 日	17.8	21.4	*
夕 日	17.6	20.7	*
子 供	18.5	20.7	*
子 供	18.7	20.5	*

* $p < .05$.

に表示するため、検索結果の全体から課題として提示した画像を判別しやすいことが推察できる。

一方、図 6 のようなテキストチャといった、一見しただけでは物体が判別しにくい画像に対しては、検索時間に有意差は見られなかった。例えば図 4 の画像に対して、アンケートの自由記述部分によると、被験者は「模様」や「赤」「気持ち悪い」というキーワードを入力して検索を行っていた。こうした画像は画像中の物体が特定しにくいため、被験者は画像の特徴をうまく捉えたキーワードを連想できず、一方、キーワードを付与する場合にも的確な付与ができなかったものと考えられる。つまりデータが付与できないということはメタデータ型検索の限界であり、木本 (1999) にもあるように模様やテキストチャには特徴量型検索が適していると考えられる。

アンケートからは、「システム A は的確なキーワードで検索がしやすい」というキーワードに関する意見が 7 人から得られた。また、「分類したまま保管したい」「自分が持っている画像を分類したい」という保存方法に関する意見が 4 人から得られた。その 4 人に追加してインタビューすると、デジタルカメラを積極的に利用していて、貯まっている画像の分類に困っているとのことであった。しかし「キーワードを付与する手間がかかりそう」「簡単にキーワードを付与できると良い」という、キーワードの付与に関する意見も 3 人から得られた。また、「模様の画像の検索はやりにくかった」という、テキストチャ画像に対するメタデータ検索の欠点を指摘する意見が 8 人から得られた。

検索結果の見やすさおよび検索結果表示の妥当性については、「物体が大きく写っている画像から順に表

表 9 アンケートの結果

項 目	システム A	システム B	有意差
検索しやすさ	4.4	3.6	*
検索結果の見やすさ	4.6	3.6	**
入力したキーワードに 対する結果の妥当性	4.4	3.4	**
総合評価	4.4	3.6	*

* $p < .05$, ** $p < .01$.

示されるので、課題の画像を探しやすい」との意見が6人から得られた。また「大きく写っているものを見たまに入力すればよいようで使いやすい」と妥当感を持っているとの意見が7人から得られた。しかし「模様の画像は、検索結果の順序がばらばらで見にくい」という意見が3人から得られた。例えば図6のような赤の模様、緑の模様である別の画像、また銀色の模様である別の画像に対して同じ「テクスチャ」というキーワードが付与されている。そのため、被験者からみると規則性がなく表示されているように感じられ、結果として検索結果の中から目的の画像を探しにくいことになってしまったと考えられる。先にも述べたが何らかの形で特徴量型を組み合わせた必要性が示唆された。

主観的キーワードの使われ方について検索ログを分析すると、青空の画像を検索する際に「夏」というキーワードを6人の被験者が入力していた。夕日の画像においては「秋」というキーワードを4人の被験者が入力していた。子供の画像においては「かわいい」というキーワードを7人の被験者が入力していた。このような主観的なキーワードはシステムAにのみ付与しており、システムBでは付与されていない。一方でアンケートの自由記述には「印象でも画像が検索できて驚いた」という記述が2人から得られた。画像から感じるままをそのままキーワードとして入力することで、検索しやすく結果としてシステムAでは検索時間が有意に短くなったと考えられる。

これらアンケートと検索時間の測定および検索ログにより、本システムにおけるメタデータは、模様等のある種の画像を除いては総じて有効であることが示された。

しかし、メタデータの簡単な付与方法については今後の課題である。先にも述べたように商用のシステムであれば、メタデータを付与する手間をかけることで、画像が検索されて、その結果売り上げの向上につなが

るというサイクルがある。そのため、商用のシステムの運営者をはじめとして手持ちの画像を流通させたい状況であれば、メタデータを付与するコストは厭わないと考えられる。一方で、個人や小グループで使用する場合には、付与する手間を敬遠することも十分に予想される。また、テクスチャ画像のような画像中の物体を判別しにくい画像に対しては、メタデータ型検索と特徴量型検索との組み合わせを取り入れる必要があることが明らかとなった。この点についても今後の課題である。

6. まとめと今後の課題

本論ではメタデータ型の画像データベースを対象として、枠組みと記述内容について以下のように実験と考察を行った。

まず、動画と静止画の両方を用いて画像に付与されたメタデータの枠組みについて被験者による実験を行った。その結果メタデータとして記述すべき内容とその記述順序を抽出した。最初にその画像中の人物や場所が特定できる場合にはその固有名詞、2番目にはその画像の状況、3番目に場面、4番目に画像中の主となるオブジェクトの動作、次にそのオブジェクト以外の物体や背景など周辺の状況、そして周辺の動作、最後に画像から受ける印象や感じ方などの感性語が記述された。またインパクトのある記述内容が先頭に転置されて記述される傾向が見られた。そして、語数と語の種類数からは、静止画と動きの少ない動画は細部にわたって、動きの激しい動画は動いているオブジェクトに注目して、付与語が付与される傾向が見られた。

次に、メタデータとしてキーワードを選び、画像に付けるための一貫性・妥当性のあるキーワード付けの基準について考察を行った。また画像に対してキーワード付けの実験を行った。その結果、妥当性のあるキーワードを付けること、多くのキーワードを付けること、主観的なキーワードと客観的なキーワードとを区別すること、重要度などのキーワードそのものに対する指標付けを行うこと、といった付与基準が見出された。複数人で画像にキーワードを付けるにあたっては、このような基準は必要不可欠であると考えられる。

そして、実際に検索システムを構築し被験者による検索実験を行い、以上で述べたメタデータの枠組みと記述内容の有効性を検証した。しかし、メタデータの簡単な付与方法については今後の課題である。また、模様のような画像に適した検索手法である特徴量型検

索を取り入れることも今後の課題である。

メタデータの枠組みに関してはISOでも標準化の動きがあり、MPEG-7という規格が提唱されている(柴田1999)。MPEG-7ではマルチメディア情報の内容を記述するための枠組みを規定することを目指しているが、画像中のカラーヒストグラムや音声の周波数など、Motion Picture Expert Groupであって画像を中心とした専門家集団ということもあり、静止画や動画から直接抽出できる内容に力点が置かれている(MPEG)。そのため本論文で述べたような言語的な情報に関する部分は検討作業が遅れているのが現状である。2002年2月に最初のバージョンが国際規格となったが、すでにISOでは次のバージョンを制定しようという方向で進んでおり、本研究で得た知見をMPEG-7の次のバージョンに対して提案してゆくことも今後の課題である。

参 考 文 献

- CanDINet 画像検索サービスシステム. <http://www.candinet.canon.co.jp/>
- FUKUMOTO, T. *et al.* (2000) The criteria and evaluation of metadata/Keyword in Image Retrieval, International Conference on Computers in Education/International Conference on Computer-Assisted Instruction 2000 : 542-546
- 木本春夫(1999) 感性語による画像検索とその精度評価. 情報処理学会論文誌, **40**(3) : 886-898
- 清水 康, ほか(1996) 意味の数学モデルによる画像データベース探索方式とその学習機構. 電子情報通信学会論文誌, **J79-D-II**(4) : 509-519
- 近藤邦夫, ほか(2000) 画像データベースのためのイメージカラー検索手法. 映像情報メディア学会誌, **54**(11) : 1615-1622
- 越桐國雄(2001) インターネットの教育利用の現状'00.1. <http://www.osaka-kyoiku.ac.jp/educ/enq00/enq00a.html>
- 串間和彦, ほか(1999) 色や形状等の表層的特徴量にもとづく画像内容検索技術. 情報処理学会論文誌, **40** SIG3(TOD1) : 171-184
- LEUNG, H.C. *et al.* (1992) Picture retrieval by content description. *Journal of Information Science*, **18** : 111-119
- 前田茂則, ほか(1999) 釈明情報の提示を行う対話型画像検索システム. 電子情報通信学会論文誌, **J82-D-II**(10) : 1617-1625
- 文部省(1998, 1999) 情報化の進展に対応した教育環境の実現に向けて. http://www.mext.go.jp/b_menu/shingi/chousa/shotou/002/toushin/980801p.htm http://www.mext.go.jp/b_menu/houdou/11/01/990101.htm
- 守田奈緒子, ほか(1995) 絵画の索引法: 段階的絵画解釈を応用した三つの索引法によるデータベースの作成. アート・ドキュメンテーション研究, **4** : 3-16
- MPEG Homepage : <http://www.cse.it/mpeg/>
- 西山晴彦, ほか(1996) 画像の構図を用いた絵画検索システム. 情報処理学会論文誌, **37**(1) : 101-109
- PANOFKY, E.(1987) イコノロジー研究: ルネサンス美術における人文主義の諸テーマ. 美術出版社, 東京
- 先進学習基盤協議会(2000) 教育及び学習における情報技術の標準化について. <http://www.alic.gr.jp/>
- 柴田正啓(1999) MPEG-7の標準化動向. 画像電子学会誌, **28**(7) : 298-303
- 照井武彦(1998) CDROMとネットワークによる民俗・歴史・考古画像データベースの提供方式の開発研究. 平成7年度~平成9年度科学研究補助金(基盤研究 A-1) 研究報告書

Summary

A large number of still images and the motion pictures have been used and raised educational effect in the school. By the digital camera and the digital large-scale storage having spread in recent years, a huge number of digital images are being accumulated and an image database attracts attention. If it can connect now with the Internet from the classroom, it will be searched for that the teacher can start looking for now the still image and motion picture required for lesson.

In this paper, the experiment and consideration were made as follows about the framework and the contents of description for the image database of meta-data type. First, the contents that should be described, and its description sequence were extracted about the framework of the meta-data that compose of both still image and motion picture.

Next, the keyword was chosen as meta-data and about the criteria of the keyword the consistency and the validity for attaching to the image was considered.

And, the retrieval system was actually built the above result, the retrieval experiment by the subject was made, and the validity was verified.

Key Words: IMAGE RETRIEVAL, DATABASE, WEB BASED LEARNING, INTELLECTUAL LEARNING SUPPORT

(Received July 24, 2002)