

## 任意メディアにより問い合わせ可能な 分散マルチメディア情報検索方式

林孝志 鈴木源吾 飯塚裕一 小西一也 串間和彦

NTT サイバースペース研究所

本稿では、任意のメディア情報を検索キーに用いて、ネットワーク上に分散して存在するマルチメディア情報を一括検索できるメディア統合検索システムを提案する。本システムでは、異メディアのデータを関連付けた辞書により、入力された検索キーを適切にメディア変換できる。さらに、メディア変換辞書を与えられた学習用データから自動生成できることが特徴である。提案した方式に基づき、プロトタイプシステムを作成し、画像および文章の検索へ適用し、提案システムの有効性を確認した。

## Distributed Multimedia Information Retrieval by Free-Media Query

Takashi Hayashi, Gengo Suzuki, Yuichi Iizuka, Kazuya Konishi  
and Kazuhiko Kushima

NTT Cyber Space Laboratories

In this paper, we propose the integration retrieval system, which retrieves distributed multimedia information on the WWW by entering any multimedia key. Since a media translation dictionary connects different media data, the proposed system can translate a media of retrieval key into an appropriate media. In addition, the dictionary can automatically be made from the learning data. We show image and text retrieval results to examine the effectiveness of the proposed system.

### 1. はじめに

近年、図書館の書籍や美術館の絵画あるいは雑誌・新聞などの文章や写真を電子化したマルチメディア情報が増大している。マルチメディア情報は一般的に文書、音声・楽音、静止画、動画等のメディアを複合することにより表現される。そして、これらはデータベース管理システム、ファイルシステム等により、情報源として格納されており、ネットワーク上に分散している。

これらの既存情報源から、任意のメディアを用いて、所望の情報を検索できれば、多くのアプリケーションで有効である。例えば、ハミングでテーマソングを入力し、映画のワンシーンを検索したり、図や表に関する曖昧な記憶から過去の資料を検索することが可能となる。

そこで本稿では、任意のメディアを用いた問い合わせにより、ネットワーク上に分散している既存情報源からマルチメディア情報を検索するシステムを提案する。

## 2. 現状の検索技術と問題点

マルチメディア情報の増大に伴い、ユーザにとって直感的で分かりやすい検索技術の開発が期待されている。現在、実用化されているマルチメディア情報検索の多くは書誌情報に基づく検索である。書誌情報に基づく検索とは、マルチメディア情報に対して、テキストで表された書誌情報を付与しておき、検索キーとして入力されたキーワードと書誌情報とを照合し、マルチメディア情報を検索する方法である。また、静止画から静止画を検索するような同一メディアによる類似検索[1][2]も研究されている。類似検索とは、検索対象のマルチメディア情報から特徴量を抽出しておき、入力された検索キーから抽出された特徴量との間で距離計算（N次元特徴量空間から1次元距離空間への写像。この写像は特徴量を重視する値を示すので重みベクトルと呼ばれる）を行い、マルチメディア情報を検索する方法である。「距離が近いデータ同士は互いに類似している」という仮定（仮定1）に基づいている（図1）。検索結果が類似しているか否かの判断はユーザが行う。

このような現状の検索技術では、情報源の提供する検索キーのメディア種別とユーザが有する検索キーのメディア種別が一致しないと検索できないという問題点があった。

検索キーのメディア種別の不一致という問題点を解決するために、異メディアで表現された情報を関連付ける研究が行われている。例え

ば、文献[3]では、学習用の絵画（静止画）に対してユーザが印象語を付与して、得られた印象語ベクトルと抽出された静止画特徴量ベクトルとの間に正準相関分析を適用し、相関が最大となる線形写像を求めている。この線形写像が静止画とテキストとの間の関連付けを行っている。また、文献[4]では説明文付きの静止画について、説明文を単語の特徴量空間でクラスタリングし、近接している説明文に対応する静止画はその特徴量空間上で近接するとみなし、近づけている。静止画を検索キーとして入力し、類似画像の説明文が近接している部分を検索結果として出力する。

任意のメディアを用いた問い合わせにより、既存情報源からマルチメディア情報を検索する場合、これらの従来法には以下の問題点がある。

### (1) 特徴量空間上で多峰性の分布をしているデータを関連付けるのが困難

従来法では、「互いに類似しているデータ同士は、距離が近い」という仮定（仮定1の逆）に基づいている。この仮定は関連づけるべきデータが特徴量空間上で多峰性の分布をしており、単純な重みベクトルでは距離が近くなならない場合、成り立たない。例えばリンゴについての文書に対して、赤いリンゴの静止画と青リンゴの静止画を関連づける場合を考える。このような場合、従来法ではデータの重心に関連付けられるか、あるいは、データ群が近接化される（図2）。つまり、特徴量が大きく異なり、近接していないものを関連付けることは難しい。

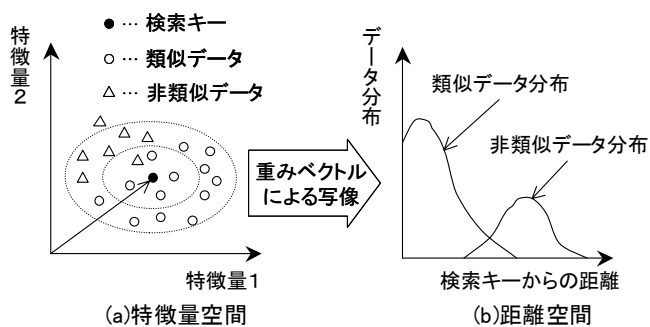


図1. 特徴量空間・距離空間における類似検索のモデル

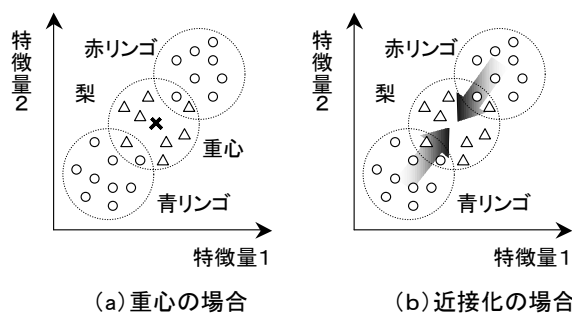


図2. 関連付けるデータの位置

## (2) 既存情報源の利用が困難

既存情報源はそれぞれ特徴量の種類や問い合わせの形式が異なる。利用者は情報源毎に、その形式にあわせて、異なる問い合わせを生成し、必要な情報を検索しなければならない。したがって、ネットワーク上に分散している様々な既存情報源からマルチメディア情報を一括検索することは難しい。

## 3. 任意のメディアを用いた問い合わせによる検索機能の提案

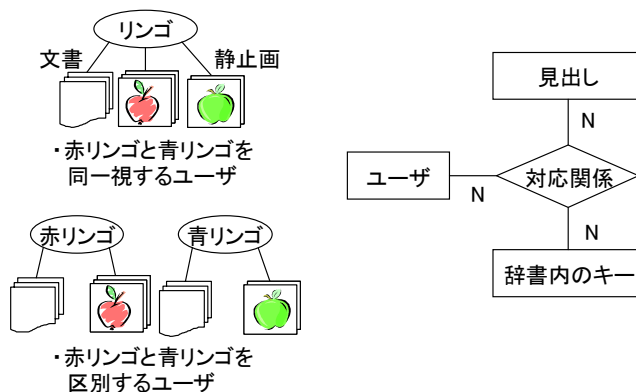
2章に述べた問題点を解決し、任意のメディアを用いた問い合わせにより、ネットワーク上に分散している既存情報源からマルチメディア情報を一括検索できるメディア統合検索システムを提案する。まず、任意のメディアで表現された検索キーを情報源毎に入力可能なメディアで表現された検索キーに変換するために、異メディアで表現されたデータを関連付けるメディア変換方式について検討する。さらに、複合メディアで表現されたデータを学習用データとして利用して、適切な関連付けを自動生成する方法について検討する。

### 3.1. 辞書を用いたメディア変換方式

異メディアによって表現されたデータが類似しているか否かはユーザによって判断される。しかし、2章で述べたように、類似しているデータ群が、特徴量空間上で近接しているとは限らない。したがって、各々の特徴量空間上で多峰性の分布をしていても、異メディアで表現されたデータを関連付け、メディア変換できる必要がある。

また、既存情報源を利用するためには、様々な検索エンジンへの入力を想定し、特徴量に限らずファイル等データそのものにメディア変換できる方が望ましい。

そこで、対応する異種メディアのデータを、同一のカテゴリを示す識別子を媒介にし直接結びつける辞書を用いたメディア変換方式(辞書方式)を提案する(図3)。このとき、対応す



(a) 見出しと辞書内のキー (b) メディア変換辞書のE-R図

図3. メディア変換辞書

る異種メディアのデータを「辞書内のキー」、識別子を「見出し」と呼ぶ。

一般の辞書でも見出しに対して説明があるように、メディア変換辞書では辞書内の複数のキー群によって見出しが説明される。また、同じデータであっても、ユーザの解釈が異なれば、複数の見出しに対応する。したがって、見出しと辞書内のキーはユーザを介して多対多の関係として表現する必要がある(図3(b))。

新規メディアの追加に関しても、辞書方式では統一的なデータ構造で異メディア間の関連付けを管理できる。つまり、新規メディアにより表現された辞書内のキーを見出しに追加すれば良く、追加が容易である。

### 3.2. 検索の流れ

メディア変換機能を用いた検索の流れを図4に示す。

#### 3.2.1. メディア変換機能

前節で説明したように、メディア変換辞書は、「見出し」に対応する辞書内のキーが、複数結びつくという構造となっている。メディア変換は、

- (1) ユーザ問い合わせ解釈機能
- (2) 見出しを引く機能
- (3) 辞書内のキーを引く機能

の3段階の機能から構成される。以下これらの機能について述べる。

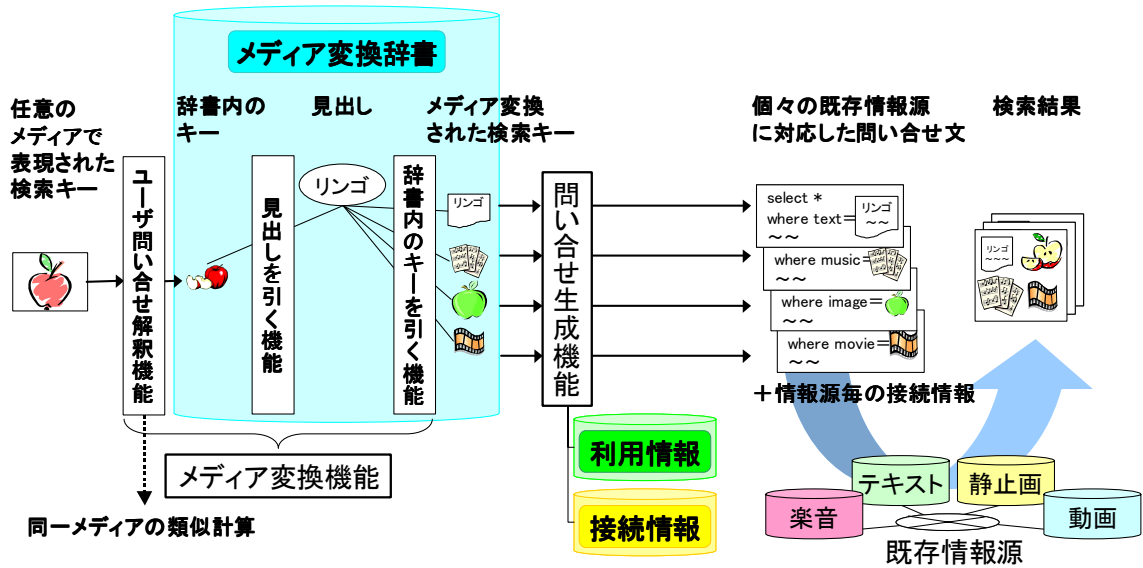


図4 . 検索の流れ

### (1) ユーザ問い合わせ解釈機能

ユーザが入力する検索キーが、辞書内のキーとして存在すれば、直接そのキーから見出しを引くことができるが、現実には、すべての情報を辞書内のキーとして管理することはできない。そこで、ユーザの入力した検索キーから同一メディア内の類似検索技術などを用いて、辞書内のキーを決定する。この機能をユーザ問い合わせ解釈機能と呼ぶ。

### (2) 見出しを引く機能

検索キーが持つ意味から見出しを特定するために、ユーザ問い合わせ解釈機能により特定された辞書内のキーから、見出しを引く機能である。

### (3) 辞書内のキーを引く機能

検索キーを異なったメディアに変換するために、特定された見出しから、その見出しに結びつく任意のメディアの辞書内のキーを検索する機能である。

### 3.2.2. 問い合わせ生成機能

メディア変換された検索キーを受け取り、ネットワーク上に分散している個々の既存情報源に対応した問い合わせ文を生成する。

但し、類似検索エンジン毎に特徴量や重みベ

クトルの種類、数が異なるのが一般的である。さらに、問い合わせ形式も SQL などを独自に拡張したもので類似検索エンジン毎に異なる。図5に静止画の類似検索エンジン ImageCompass[1]と QBIC[2]の問い合わせ例を示す。

そこで、問い合わせ生成機能には、以下の機能要件が求められる。

### (1) 個々の情報源に対応した適切な問い合わせ生成

類似検索では、重視する特徴量が異なると、得られる検索結果が異なる。そこでユーザが情

#### ImageCompassの問い合わせ例

```
select x.filename
from Image x, x.subimages y
where y.features similar(40,'0hue',0.8,'L1',...)
SubImageFeatures.MakeValue('0hue,17.0,1.0,2,...');
```

(説明) 指定した特徴量ベクトル(MakeValueで特徴量名、次元数、抽出された特徴量ベクトルを指定)に類似の(similarで要求数、特徴量名、重み、評価関数を指定)部分画をもつ画像のファイル名を取得する。

#### QBICの問い合わせ例

```
select ObScoreFromStr
('ObColorFeatureClass file=<server, "filename">' weight=50.0 and
ObTextureFeatureClass weight=30.0, db2image),
Content(db2image, 'jpg'), FILENAME(db2image)
from db2image_table order by 1;
```

(説明) 指定したサーバにある画像ファイル<server, "filename">を画像キーとして全体色を重み付け( weight=50.0 )により、テクスチャを重み付け( weight=30.0 )により類似計算を行う。

図5 . ImageCompass と QBIC の問い合わせ例

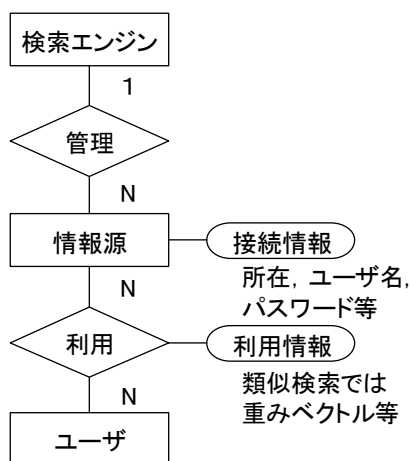


図6 . 利用情報と接続情報

報源を利用する際、適切な検索結果を得るための問い合わせの生成が望まれる。

## (2)情報源追加への柔軟な対応

情報源毎に物理的・論理的所在や接続のための情報(ユーザ名やパスワードなど)が異なる。また、類似検索エンジン毎に問い合わせの形式も異なる。検索対象として情報源が追加された時に、これらの異種性を解消し、柔軟に対応することが望まれる。

提案システムでは、機能要件(1)を満たすために、利用情報としてユーザ毎に重みベクトルを設定・管理する。また、機能要件(2)を満たすために、情報源毎の所在や接続のための情報を接続情報として管理する(図6)。

メディア変換されて得られた辞書内のキーは、利用情報が参照され、特徴量毎の重みベクトルが付加されて個々の情報源にあった問い合わせが生成される。さらに、接続情報を参照して、問合せが各情報源に送られ、検索が実行される。

### 3.3. 辞書の自動生成機能

メディア変換辞書とは見出しと辞書内のキーの対応関係であり、全てを人手で作成しては膨大な労力がかかり、現実的ではない。そこで、本論文ではWebページ、電子百科事典などの複合メディア情報を学習用データとして

利用してメディア変換辞書を自動生成する方法を提案する。

1章の従来法の問題点で述べたように、どのデータとどのデータを関連付けるべきかの指標が無いまま、Webページや電子百科事典から無差別に、辞書内のキーを生成するのでは、ユーザが望む検索結果を得ることはできない。

有効な辞書内のキーとは、ある見出しについて、それをできるだけうまく代表しているデータである。例えば、文献[3]では、特徴量空間上で収集されたデータの重心を代表点としている。しかし、前述のように、この方法はデータの分布が多峰性の場合、必ずしも有効ではない。

そこで、辞書内のキーとして、どれを採用すれば類似検索がうまく行くかを評価基準として、直接、学習データ中から選別する方法が考えられる。

この時、あるカテゴリ識別子に関連付けられた複数のデータの中から、それを検索キーとして用いた時に、高い精度・再現率の検索結果を返却しうるデータを、なるべく少数選別する必要がある。

これは組み合わせ最適化問題に帰着し、計算の複雑さの意味で、本質的に困難な問題に属する。そこで、厳密な最適解ではなく、近似解を短時間で求めるアルゴリズムとして、遺伝的アルゴリズム[5]、焼きなまし法[6]などが適用できる。精度と再現率を要約した指標であるF尺度[7]と選別個数の最小化からなる評価関数Fitnessを以下に提案する。

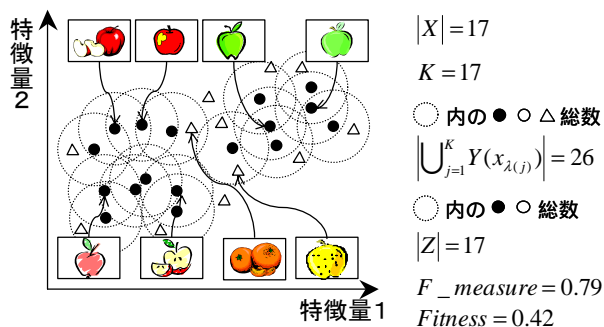
特定のユーザにとって、特定の見出しに関連付けられているデータ集合を $X$ とし、 $X$ に属している各データを $x_i$ とする。

$x_i$ それぞれを検索キーとして、対象情報源を検索した時、上位T件の検索結果をそれぞれ $Y(x_i)$ とする。

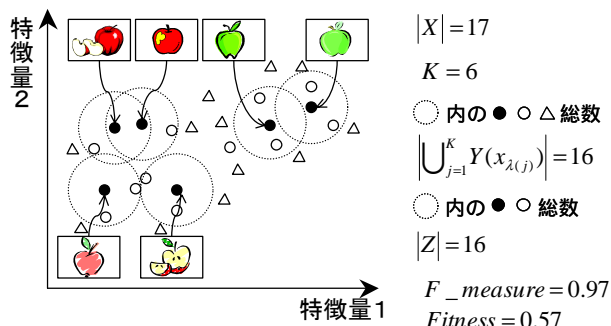
$X$ から任意個数 $K(K \neq 0)$ 個のデータ $x_{\lambda(1)}, x_{\lambda(2)}, \dots, x_{\lambda(K)}$ ( $\lambda(j)$ は1からNのいずれかで、重複しない)を選別し、これらを検索キーとして用いた時、検索結果の和集合

$$\bigcup_{j=1}^K Y(x_{\lambda(j)})$$

- ○ ... 「リンゴ」のデータ集合 X
- ... 選別されたデータ(K個)
- △ ... その他のデータ
- ... 上位 T 件の検索結果の範囲



(a) 全てのデータを選択した場合



(b) データ選別の最適な組み合わせ

図7. データの選別

を求め、

適した検索結果である  $X$  と  $\bigcup_{j=1}^K Y(x_{\lambda(j)})$  の共通集合  $Z$  を求める。

$$Z = X \cap \left( \bigcup_{j=1}^K Y(x_{\lambda(j)}) \right)$$

結果の評価指標として、F 尺度 ( $F\_measure$ ) を用いる。F 尺度とは精度 ( $Precision$ ) と再現率 ( $Recall$ ) を重み付き平均した指標である。但し、 $\alpha$  ( $0 \leq \alpha \leq 1$ ) は精度と再現率のどちらを重視するかを表すパラメタであり、あらかじめシステム管理者が設定する。また、 $|A|$  は集合  $A$  の要素数を表す。

$$Precision = \frac{|Z|}{\left| \bigcup_{j=1}^K Y(x_{\lambda(j)}) \right|}$$

$$Recall = \frac{|Z|}{|X|}$$

$$F\_measure = \frac{1}{\frac{\alpha}{Precision} + \frac{1-\alpha}{Recall}}$$

したがって、「高い精度・再現率を示す検索結果を返却しうるデータを検索キーとして、なるべく少数選別する」ために、F 尺度を最大化し、選別個数  $K$  を最小化するデータの組み合わせ  $x_{\lambda(1)}, x_{\lambda(2)}, \dots, x_{\lambda(K)}$  を  $N$  個のデータから求める。

F 尺度を最大化し、選別個数を最小化するための評価関数  $Fitness$  として、次式を提案する。但し、 $\beta$  ( $0 \leq \beta \leq 1$ ) は F 尺度の最大化と選別個数の最小化のどちらを重視するかを表すパラメタで、あらかじめシステム管理者が設定する。評価関数があらかじめ設定した値以上になった時、得られたデータの組み合わせ  $x_{\lambda(1)}, x_{\lambda(2)}, \dots, x_{\lambda(K)}$  を選別されたデータとする。

$$Fitness = \beta \times F\_measure + (1 - \beta) \left( 1 - \frac{K}{|X|} \right)$$

上記処理を見出し数分、ユーザ数分そしてメディア種別数分繰り返すことでメディア変換辞書が自動生成される。

図7は、見出し「リンゴ」に関連付けられているデータ集合を用いて、データの選別を行った様子を示している。図7(a)は、関連付けられているデータ全てを選択した場合、図7(b)は、データの選別を最適に行った場合を示す ( $\alpha = \beta = 0.5$ )。

## 4. 提案方式の試作実装と評価実験

### 4.1. 検索実験

前章の提案に基づき、プロトタイプシステムを実装した。本プロトタイプシステムはユーザが入力する静止画またはキーワードから、辞書に基づくメディア変換を行い、ユーザにとって同じ意味を表す静止画またはテキストの情報の一括検索が可能である。静止画



図8. プロトタイプシステムを用いた検索例

(PhotoDisk<sup>1)</sup>の写真)の情報源は画像類似検索エンジン ImageCompass[1]によって管理されており、テキストの情報源は全文検索エンジン namazu[8]によって管理されている。これらの検索エンジンは、ユーザ問い合わせ解釈にも用いられる(検索キーからの辞書内のキー特定)。また、キーワードでのみ検索可能なWebページを検索対象の情報源として加えた。この時、必要な作業は接続情報の追加のみである。

プロトタイプシステムを用いた検索例を図8に示す。入力した検索キーは文字盤が黒い「時計」の静止画である。静止画を検索キーとしてアプリケーションに入力すると、メディア変換機能により辞書内のキー(静止画およびテ

<sup>1)</sup> <http://www.photodisc.com>, PhotoDisc, Inc.



(a)文字盤が白い  
ストップウォッチ (b)文字盤が黒い  
目覚まし時計

図9. 最適化により得られた「時計」のキー

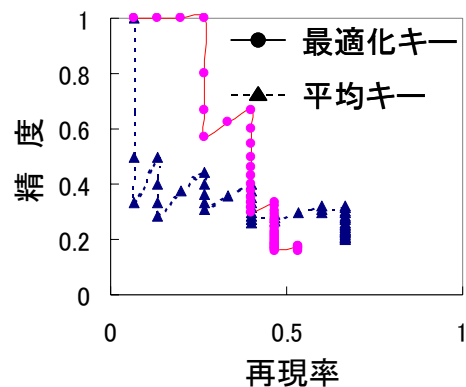


図10. 最適化キーと平均キーによる  
精度・再現率グラフ

キスト)が引かれる。この辞書内のキーにより類似検索・全文検索を行い、検索結果が得られる。ImageCompassの検索結果は、同じように文字盤が黒い「時計」やダーツ的などが検索されているが、メディア変換辞書を介することで文字盤の白い「時計」の静止画も検索されている。また、従来、キーワードでのみ検索可能であった情報源(Webページ)もメディア変換機能によって静止画の検索キーで一括検索できる。

このように、ユーザが入力した画像キーから静止画およびテキスト情報の一括検索が可能である。同様にテキスト情報も検索キーとすることができる。

#### 4.2. 辞書の自動生成の評価実験

提案した辞書内のキーの選別方法の有効性を評価するために、評価実験を行った。

画像データベース全640枚中、15枚の時計の画像を同じ「時計」という見出しに対応する

とみなす。このデータを用いて、辞書内のキーとして「時計」全体をうまく代表している辞書内のキーを生成する。3.3節で述べた評価関数（予備実験により  $\alpha = 0.5$ ,  $\beta = 0.9$  とした）を遺伝的アルゴリズムにより最適化する。評価関数があらかじめ設定した閾値を超え、最大値（Fitness=0.43）を示した検索キー（2枚）を図9に示す。また、上記、最適化により得られた検索キーと15枚の「時計」画像の特徴量ベクトルを平均した検索キーとを検索に用いたときの再現率・精度を図10に示す。

## 5. 考察

4.1節の検索実験で示したように、本方式では、異メディアで表現された辞書内のキーにより、メディア変換機能を実現している。また、問い合わせ生成機能により異なる検索エンジン（ImageCompass, namazu）に管理されている情報源を一括検索できることが示された。

静止画の類似検索のみでは特徴量空間上で近接しているデータを類似しているデータと見なし、検索結果として出力する。しかし、ユーザが類似していると判断するデータは、必ずしも特徴量空間上で近接しているとは限らない。4.2節の実験で選別された白い文字盤の時計と黒い文字盤の時計とを検索キーとして用いることで、特徴量ベクトルを平均した検索キーよりも高い精度・再現率を実現できることが示された。つまり、複数の辞書内のキーを持つことで、特徴量空間上で多峰性の分布をしているデータに対しても、有効な検索を実現できる。

## 6. まとめ

本稿では、マルチメディア情報群に対して、任意のメディア情報を検索キーに用いて、一括検索できるメディア統合検索システムを提案した。本システムでは、異メディアのデータを関連付けた辞書により、入力された検索キーを適切にメディア変換し、既存情報源を検索できる。また、評価関数を最適化することで、メデ

ィア変換辞書を与えられた学習用データから自動生成できる。提案に基づき、プロトタイプシステムを作成し、画像および文章の検索へ適用し、特徴量空間上で近接しているとは限らないデータの検索に対し、高い再現率・精度を実現できることを確認した。

## 参考文献

- [1]串間他：“オブジェクトに基づく高速画像検索システム：ExSight”，情報処理学会論文誌，Vol.40，No.2，pp.732-741，1999.
- [2]IBM Almaden Research Center：“Query by Image and Video Content:The QBIC System”，IEEE Computer，Vol.28，No.9，pp.23-32，Sept 1995.
- [3]栗田他：“印象語による絵画データベースの検索”，情報処理学会論文誌，Vol.33，No.11，1992.
- [4]岡他：“画像とテキストの自己組織化データに基づく画像理解方式の提案”，信学技報，PRMU-98-74.
- [5]安居院他：“ジェネティックアルゴリズム”，昭晃堂，1993.
- [6]長尾：“最適化アルゴリズム”，昭晃堂，2000.
- [7]徳永他：“言語と計算 5 情報検索と言語処理”，東京大学出版会，1999.
- [8]馬場：“日本語全文検索システムの構築と活用”，ソフトバンクパブリッシング株式会社，1998.