社団法人 情報処理学会　研究報告
IPSJ SIG Technical Report

2004－DBS－134　(I)　(22)
2004／7／13

# 楽曲メディアデータと画像メディアデータ間における連想検索方式

中 西 崇 文†　北 川 高 嗣††

本稿では，楽曲メディアデータと静止画像メディアデータ間の異種メディア間連想検索の実現方式について示す．本方式は，メディアデータを対象としたメタデータ自動抽出機構である Media-lexicon Transformation Operator を言葉の相関を計量できる意味の数学モデルのメタデータ空間により連結することで異種メディア間の検索を実現する．本方式により，これまで独立に実装された異種メディアデータを統一的に扱うことが可能になる．これにより，新しい情報生成が可能となり，既存のデータの利用機会を増大させ，データベース群の利用価値を飛躍的に増大させることが可能となると考えられる．

# A Method of Associative Heterogenous Mediadata Search for Music Data and Image Data

Takafumi NAKANISHI† and Takashi KITAGAWA ?

This paper presents an implementation method of associative heterogenous mediadata search for music data and image data. This method is realized by measuring correlation of the impression words extracted from media-lexicon transformation operator (automatic metadata extraction method) for each mediadata. The feature of this method is to bridge over heterogeneous mediadata which exist independently as different database resources.

## 1. はじめに

現在，コンピュータネットワーク上に多種多様なメディアデータ群が散在している．また，それらを検索対象とするシステムの実現が行われつつある．これらの多種多様なメディアデータ群を扱うシステムにおいて，人間の感性は重要な問題のうちの 1 つとなっている．

我々は，メディアデータに対応するメタデータを言葉によって表現し，検索者の与える文脈に応じた意味的解釈を伴う間接的な検索方式として，メディアデータを対象とした意味の数学モデルによる，意味的連想検索方式 [3),4),6)] を提案している．これにより，統計的に意味素を抽出して意味の解釈を実現する従来の研究 [1)] と比較して，言葉の意味を文脈に応じて解釈する機構より，言葉と言葉，あるいは，言葉とメディアデータ間の意味的な関係を与えられた文脈や状況に応じて動的に計算することが可能となる．現在の実現システムでは，文脈の様相の数は約 $2^{2000}$ であり，ほぼ無限の文脈を表すことが可能である．

さらに我々は，文献 [2),4),7)~9)] でメディアデータのメタデータを自動抽出するための実現方式について示している．特に文献 [2)] では，メディアデータの印象を表す語をメタデータとして自動抽出する枠組みとして，Media-lexicon Transformation Operator を示している．さらに，文献 [7),8)] ではメタデータを抽出する際に，人間の感性や感覚に基づいた関数を適用する方式を示している．メディアデータを対象としたメタデータ抽出において，人間の感性や感覚を解釈する機構が導入されれば，人間の感性や感覚に合致した検索が可能となると考えられる．これにより，メディアデータが人間に与える印象を抽出するメカニズムの解明の第 1 歩になると考えられる．

本稿では，楽曲メディアデータと静止画像メディアデータ間の異種メディア間連想検索の実現方式について示す．本方式は，文献 [2),7),8)] で示される Media-lexicon Transformation Operator を言葉の相関を計

† 筑波大学大学院システム情報工学研究科，つくば市
Graduate School of Systems and Information Engineering, University of Tsukuba, Tsukuba, Ibaraki 305-8573, Japan
e-mail：takafumi@nalab.is.tsukuba.ac.jp
†† 筑波大学大学院システム情報工学研究科，つくば市
Graduate School of Systems and Information Engineering, University of Tsukuba, Tsukuba, Ibaraki 305-8573, Japan
e-mail：takashi@is.tsukuba.ac.jp

量できる意味の数学モデルのメタデータ空間により連結することで異種メディア間の検索を実現する．異種メディア間の連想検索に関して，これまで文献 [10] では，静止画像メディアデータから自動的に顔の表情を生成する方式について実現してきた．

本方式は，様々な異種のそれぞれのメディアデータから特徴量としてメディアデータの印象を表す言葉 (印象語) を抽出することにより，意味の数学モデルによるメタデータ空間で異種のメディアデータ間においてもそれらの印象の相関を計量が可能である．これらにより，本方式は，これまで独立に実装された異種メディアデータをメタで連結させ，統一的に扱うことが可能になる．また，異種メディアの連結による新しい情報生成が可能となる．さらにこれらから，既存のデータの利用機会を増大させ, データベース群の利用価値を飛躍的に増大させることが可能となると考えられる．

## 2. 意味の数学モデルの基本構成

本節では，人間が様々な印象を表す際に用いられる様々な単語 (以下，印象語) によって表現した問い合わせに対応したメディアデータを検索することを目的とした意味の数学モデルによるメディアデータを対象とした意味的連想検索方式の概要を示す．詳細は，文献 [3),4),6)] に述べられている．

(1) **メタデータ空間 $\mathcal{MDS}$ の設定**

検索対象となるメディアデータをベクトルで表現したデータをマッピングするための正規直交空間 (以下，メタデータ空間 $\mathcal{MDS}$) を設定する．

(2) **メディアデータのメタデータをメタデータ空間 $\mathcal{MDS}$ へ写像**

設定されたメタデータ空間 $\mathcal{MDS}$ へメディアデータのメタデータをベクトル化し写像する．これにより，同じ空間に検索対象データのメタデータがメタデータ空間上に配置されることになり，検索対象データ間の意味的な関係を空間上でのノルムとして計算することが可能となる．

(3) **メタデータ空間 $\mathcal{MDS}$ の部分空間 (意味空間) の選択**

検索者は与える文脈を複数の単語を用いて表現する．検索者が与える単語の集合をコンテキストと呼ぶ．このコンテキストを用いてメタデータ空間 $\mathcal{MDS}$ に各コンテキストに対応するベクトルを写像する．これらのベクトルは，メタデータ空間 $\mathcal{MDS}$ において合成され, 意味重心を表すベクトルが生成される．意味重心から各軸への射影値を相関とし, 閾値を超えた相関値 (以下，重み) を持つ軸からなる部分空間 (以下，意味空間) が選択される．

(4) **メタデータ空間 $\mathcal{MDS}$ の部分空間 (意味空間) における相関の定量化**

選択されたメタデータ空間 $\mathcal{MDS}$ の部分空間 (意味空間) において，メディアデータベクトルのノルムを検索語列との相関として計量する．これにより，与えられたコンテキストと各メディアデータとの相関の強さを定量化している．この意味空間における検索結果は，各メディアデータを相関の強さについてソートしたリストとして与えられる．

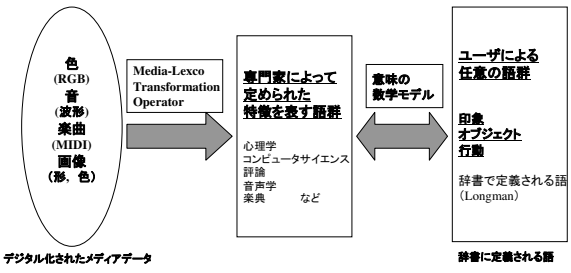


**図 1** Media-lexicon Transformation Operator の概要.
Fig. 1 Media-lexicon Transformation Operator.

## 3. Media-lexicon Transformation Operator の実現

本節では，Media-lexicon Transformation Operator の概要を示し，さらに，画像メディアデータ，音楽メディアデータを対象とした実現方式ついて示す．詳細は，画像メディアデータを対象としたものは文献 [4),8)]，音楽メディアデータを対象としたものは [2),7)] で示している．Media-lexicon Transformation Operator の概要図を図 1 に示す．Media-lexicon Transformation Operator$\mathcal{ML}$ は一般的に次のように表される．

$$\mathcal{ML}(Md) : Md \mapsto Ws. \quad (Md：メディアデータ, Ws：(重み付き) 単語群) \quad (1)$$

### 3.1 画像メディアデータを対象とした Media-lexicon Transformation Operator の実現

本節では，画像メディアデータを対象とした Media-lexicon Transformation Operator の概要 [4),8)] を示す．本機能は，画像メディアデータから，画像メディアデータに印象に合致した語を抽出する．

本機能は，3 つのステップからなる．

**Step 1** ：色印象行列 $C$ の作成.

色彩とその色彩から想起される印象の関係を示し

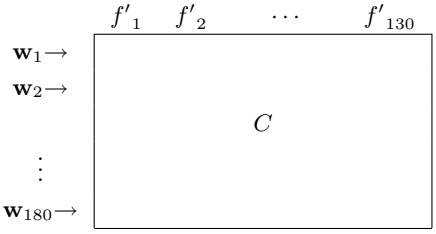


**図 2** 色印象行列 $C$
Fig. 2 Color-impression Transformation Matrix $C$.

た統計データとしてカラーイメージスケール [12] がある．カラーイメージスケールを用いて，有彩色 120 色，及び無彩色 10 色の基本色 130 色とその印象を表す 180 語により 180 行 130 列の色印象行列 C とする．色印象行列 C の各要素はカラーイメージスケールによる色と印象語の関連の強さを示す数値データである．色印象行列 C は図 2 に示す．

**Step 2** ：対象静止画像から色彩情報 **m** 抽出.

静止画像から色彩情報が抽出し，その色彩情報は静止画像全体における基本色 130 色の占める割合で構成される画像色彩ベクトル **m** によって表現される．画像色彩ベクトル **m** を次に示す.

$$\mathbf{m} = (m_1, m_2, \cdots, m_{130})^T. \quad (2)$$

**Step 3** :印象語抽出.

色印象行列 $C$ ，及び画像色彩ベクトル **m** を用いて，画像メタデータ **w** の抽出を行う．画像メタデータ **w** は，色印象行列 180 個の印象語と同一の印象語で特徴付けられるベクトルである．これにより画像メディアデータに合致した印象語が抽出される.

$$\mathbf{w} = C\mathbf{m} \quad (3)$$

### 3.2 楽曲メディアデータを対象とした Media-lexicon Transformation Operator の実現

本節では，楽曲メディアデータを対象とした Media-lexicon Transformation Operator の概要 [4),8)] を示す．本機能は，楽曲メディアデータから，楽曲メディアデータに印象に合致した語を抽出する.

本機能は，4 つのステップからなる.

**Step 1** ：変換行列 $T$ の作成.

印象と楽曲構造要素との相関関係を調べた基礎心理学研究に Hevner の研究 [13)~16)] がある．Hevner の研究では，楽曲構造要素として調性 ($key$)・テンポ ($tempo$)・音高 ($pitch$)・リズム ($rhythm$)・和声 ($harmony$)・旋律 ($melody$) の 6 つを挙げている．Hevner は，この 6 つの楽曲構造要素と

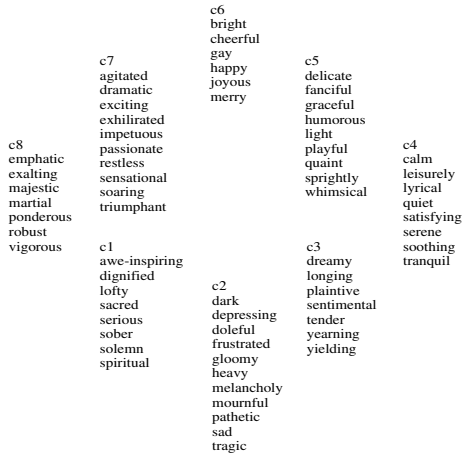


**図 3** Hevner による 8 つの印象語群リスト
Fig. 3 Hevner's 8 categories of impression words.

| | key' | tempo' | pitch' | rhythm' | harmony' | melody' |
|---|---|---|---|---|---|---|
| $\mathbf{c}_1$ | 4 | -14 | -10 | 18 | 3 | 4 |
| $\mathbf{c}_2$ | -12 | -12 | -19 | 3 | -7 | 0 |
| $\mathbf{c}_3$ | -20 | -16 | 6 | -9 | 4 | 0 |
| $\mathbf{c}_4$ | 3 | -20 | 8 | -2 | 10 | 3 |
| $\mathbf{c}_5$ | 21 | 6 | 16 | 8 | 12 | -3 |
| $\mathbf{c}_6$ | 24 | 20 | 6 | -10 | 16 | 0 |
| $\mathbf{c}_7$ | 0 | 21 | -9 | 2 | -14 | -7 |
| $\mathbf{c}_8$ | 0 | 6 | -13 | 10 | -8 | -8 |

**図 4** 変換行列 $T$
Fig. 4 Transformation Matrics $T$

8 つの印象語群 (図 3) によって表現される印象との相関関係を調べた．この相関関係から，変換行列 $T$ を作成する (図 4).

**Step 2** ：楽曲構造要素ベクトル **s** を抽出.

デジタル化された楽譜データとしてピアノ曲の *Standard MIDI File*(以下，*SMF*) を入力として与え，楽曲メディアデータを Hevner が挙げた 6 つの楽曲構造要素によって特徴づけした楽曲構造要素ベクトル **s** を生成する．楽曲構造要素ベクトル **s** を以下のように定義する.

$$\mathbf{s} = (key, tempo, pitch, rhythm, harmony, melody)^t. \quad (4)$$

**Step 3** :印象語群とその重みの抽出

Hevner による各印象に対する楽曲構造要素の相対重要性の表より構成された変換行列 $T$ と楽曲構造要素ベクトル $s$ によって楽曲を 8 つの印象語群 $c_1, c_2, \cdots, c_8$ の重み $v_{c_1}, v_{c_2}, \cdots, v_{c_8}$ によって特徴づけされた楽曲印象語群ベクトル **v** を生成す

る．楽曲印象語群ベクトル $\mathbf{v}$ を以下のように定義する．

$$\mathbf{v} = T\mathbf{s} \quad (5)$$

$$\mathbf{v} = (v_{c_1}, v_{c_2}, \cdots, v_{c_8})^t. \quad (6)$$

これにより印象語群とその重みが出力されたこととなる．

**Step 4** :楽曲印象語群ベクトル $\mathbf{v}$ の正規化

Step3 で楽曲に合致した印象語群とその重みが出力される．しかし，この重みは一般に正規化されておらず，印象語を合成するときに重みが有効に働かない場合がある．これらを補正するため重みの正規化を行う．詳細は文献 [7] に示す．

本方式では，文献 [7] の結果から，下記の正規化を行う．

$$f_N(v_{c_1}, v_{c_2}, \cdots, v_{c_8}) : (v'_{c_1}, v'_{c_2}, \cdots, v'_{c_8}) \mapsto (\frac{v_{c_1}}{max_1}, \frac{v_{c_2}}{max_2}, \cdots, \frac{v_{c_8}}{max_8}). \quad (7)$$

ここで，$max_1, max_2, \cdots, max_8$ は，各語群の重みの最大値を表す．

### 3.3 感性作用素

物理的な刺激と人間の感覚の関係を調べた研究で Fechner の法則 [17] がある．本機能では，各特徴について印象語の重みを合成するときに，各特徴の総和を刺激の強さと位置付け，その刺激に対応する感覚の大きさを合成後の値として求めるために，Fechner の法則を用いた感性作用素を実現する．感性作用素によって，人間の感覚に合致した値に変換するだけでなく，メディアデータやユーザの個々の感覚の違いを調整することが可能である．

詳細は，文献 [7),8)] に示している．

本方式では，文献 [7),8)] の結果から，感性パラメータ $k$ と $\alpha$ を，画像メディアデータの場合は $k = 1$, $\alpha = 8$，楽曲メディアデータの場合は $k = 1, \alpha = 6$ に設定した．

## 4. 楽曲メディアデータと画像メディアデータ間における連想検索方式

本節では，楽曲メディアデータと画像メディアデータ間における連想検索方式について示す．4.1 節では，異種メディア間検索の全体的な概要について示す．4.2 節では，楽曲メディアデータと画像メディアデータを統一的に扱うことができる連想方式について示す．

### 4.1 異種メディア間検索方式

本節では，異種メディア間連想検索 [10] について示す．異種メディア間連想検索の概要図を図 5 に示す．

本方式は，様々な異種のメディアデータから特徴量としてメディアデータの印象を表す印象語を各メディアデータにおいて実装された Media-lexicon Transformation Operator で抽出することにより，言葉の関係を計量できる空間でそれらの印象の相関を計量をすることを目的とする．全てのメディアデータの印象を印象語とその感覚の大きさで表現し，その印象語を計量できる意味の数学モデルによる意味的連想検索機構によって計量することによって，異種メディア間の印象の相関を計量する．

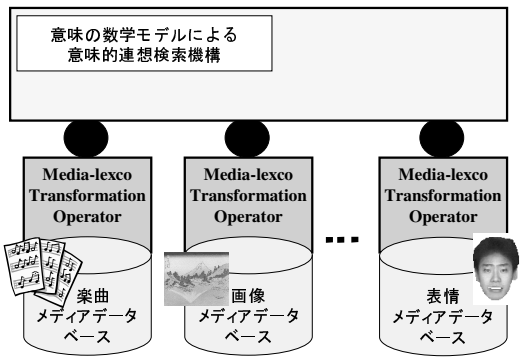


**図 5** 異種メディア間検索の概要．
Fig. 5 Fundamental framework for heterogeneous media to media search.

本方式は，印象語を意味の数学モデルで計量することによって，異種メディア間連想検索を実現している．そのため，異種メディア間の関連を知ることなく，メディアデータ毎に Media-lexicon Transformation Operator を実現できれば，あらゆるメディアデータ間とも感性に基づく連想検索が可能となる．

### 4.2 楽曲メディアデータと画像メディアデータ間における連想検索の実現

本節では，楽曲メディアデータと画像メディアデータ間における異種メディア間連想検索の実現方式について示す．

本方式は，以下の手順で実現される．

**Step 1** ：各メディアデータから印象語を抽出．

3 節で示した，各メディアデータ毎に実装された Media-lexicon Transformation Operator により，メディアデータをメタデータとして複数の印象語で特徴づけする．

**Step 2** ：メディア間の相関計量．

印象語で特徴づけされたメディアデータのメタデータをを言葉の相関を計量できる意味の数学モデルのメタデータ空間に写像し計量する．

なお，本方式での意味の数学モデルのメタデータ空間は”Longman Dictionary of Contemporary

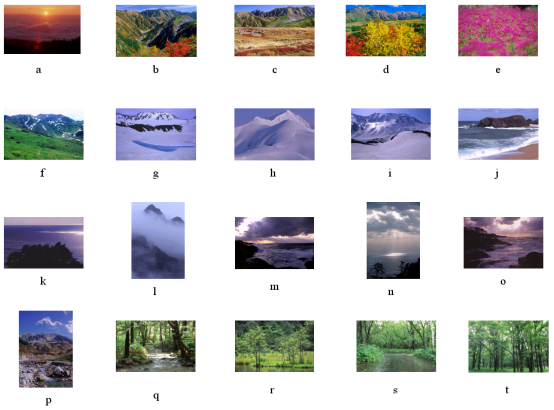


**図 6** 対象とした静止画像メディアデータセット.
Fig. 6 Image data set for experiments.

English”[11] という英英辞書を使用することとする．同辞書は，約 2000 語の基本語だけを用いて約 56000 語の見だし語を説明している．ここで基本語を特徴とみなし，各見出し語を説明する基本語が肯定の意味に用いられていた場合“ 1 ”，否定の場合“ −1 ”，使用されていない場合“ 0 ”，見出し語自身が基本語である場合その基本語の要素を“ 1 ”として，データ行列 $M$ を作成し，固有値分解をする．これより，約 2000 次元の正規直交空間であるメタデータ空間を生成する．これにより，様々な印象を表す印象語を含む一般的な語を計量する空間を実現している．

## 5. 実　　験

本方式の有効性を検証するため，本方式に基づく実験システムを構築し，検証実験を行った．

### 5.1 実験方法

印象が比較的分かりやすい楽曲メディアデータを与え，どのような画像メディアデータかを実験した．

楽曲メディアデータとして,「幸せなら手を叩こう」,「四季の歌」を用いた.「幸せなら手を叩こう」はうれしい，楽しい楽曲である.「四季の歌」は暗く，静かな楽曲である．

また，検索対象の画像メディアデータとして，図 6 に示す 20 の風景画像を用いた．

### 5.2 実験結果

楽曲メディアデータとして,「幸せなら手を叩こう」,「四季の歌」を与えた結果をそれぞれ，図 7，図 8 に示す．

図 7 から「幸せなら手を叩こう」の場合，明るい色合いの画像が上位に出力される傾向にある．これは，

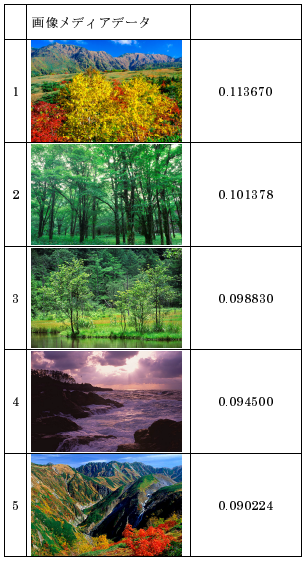


**図 7** 「幸せなら手を叩こう」の場合.
Fig. 7 A case of ”Clap your hands”.

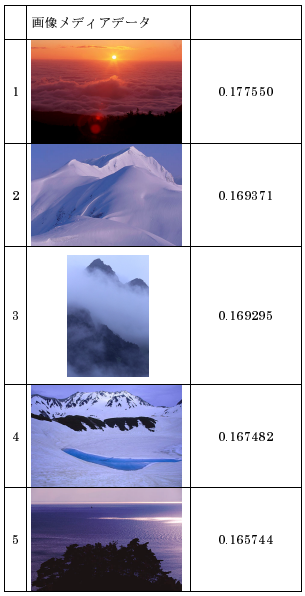


**図 8** 「四季の歌」の場合.
Fig. 8 A case of ”Songs of four seasons”.

明るい色は”bright”,”active”などの印象語が割り当てられているからである．一方，4 位に赤みを帯びた黒いの画像メディアデータが出力されている．これは，表 1 が示すように,「幸せなら手を叩こう」から出力された印象語群は，”happy”などを含む C6 語群のほかに，”light”などを含む C5 語群，”exciting”などを含む C7 語群の重みが大きい．C7 語群は激しい印象を表す語群であり，この語群の印象を反映された結果と

**表 1** 「幸せなら手を叩こう」の抽出された印象語群と重み
Table 1 Impression word categories for "Clap your hands"

| | |
|---|---|
| C6 | 0.776570 |
| C5 | 0.564565 |
| C7 | 0.254095 |
| C8 | -0.141027 |
| C4 | -0.250397 |
| C3 | -0.312459 |
| C1 | -0.480369 |
| C2 | -0.651811 |

思われる．しかし，この画像の印象は描かれている風景から,「幸せなら手を叩こう」からは外れている．

これは，画像メディアデータから印象語を取り出す操作が色彩情報によるものである．メディアデータの多面性から，他の要因による印象が大きいことが考えられる．そのため，他の要因からも印象語を取り出す機構や学習機構を実現する必要があると考えられる．しかしながら全体的には，概ね楽曲「幸せなら手を叩こう」の印象に合致した画像メディアデータが出力されている．

図 8 から「四季の歌」の場合暗い色合いの画像が上位に出力される傾向にある．特に 2 位から 5 位は暗い青色である．これらの色からは"chic","quiet"などの静かな印象語が抽出されている．これらから，楽曲「四季の歌」の印象に合致した画像メディアデータが出力されている．

### 5.3 考　　察

この実験から，概ね楽曲メディアデータから楽曲の印象に合致した画像メディアデータを検索することが可能であることを示した．しかしながら，メディアデータの多面性やメディアデータ間における独特な関連は本方式では加味されていない．これらの問題は，Media-lexicon Transformation Operator の他の要因に対する実現と学習機構の実現が重要となってくると考えられる．

## 6. おわりに

本稿では，楽曲メディアデータと静止画像メディアデータ間の異種メディア間連想検索の実現方式について示した．

本方式により，異種メディアデータ間において，統一的に扱うことが可能となる．これにより，異種メディアデータを統合による新しい情報生成が可能となり，情報資源の有効利用になると考えられる．

今後の課題は，本方式における学習方式の実現，メタデータ自動抽出方式への個人差の計量方式の導入，大規模データベースへの適用，有効性の検証が挙げられる．

## 参 考 文 献


1) M.W.Berry, S.T.Dumains, G.W.O'Brien, "Using linear algebra for intelligent information retrieval", SIAM Review Vol. 37, No.4, pp.573-595, 1995.
2) T.Kitagawa, Y.Kiyoki, "Fundamental framework for media data retrieval system using media lexco transformation operator", Information Modeling and Knowledge Bases, IOS Press, 2000.
3) T.Kitagawa, Y.Kiyoki, "The Mathematical Model of Meaning and its Application to Multiodatabase Systems", Proceedings of 3rd IEEE International Workshop on Research Issues on Data Engineering: Interoperability in Multidatabase Systems, pp.130-135, April 1993
4) Y.Kiyoki, T.Kitagawa, H.Takanari, "A Metadatabase System for Semantic Image Search by a Mathematical Model of Meaning",Multimedia Data Management– using metadata to integrate and apply digital media –, McGrawHill, Amit Sheth and Wolfgang Klas(editors), Chapter 7, 1998.
5) 串間和彦, 赤間浩樹, 紺谷精一, 山室雅司: "色や形状等の表層的特徴量のもとづく画像内容検索技術", 情報処理学会論文誌：データベース, Vol.40, No.SIG3(TOD1), pp.171-184, 1999.
6) 清木康, 金子昌史, 北川高嗣, "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, D- , Vol.J79-D- , No.4, pp.509-519, 1996.
7) 北川高嗣, 中西崇文, 清木康, "楽曲メディアデータを対象としたメタデータ自動抽出方式の実現とその意味的楽曲検索への適用," 電子情報通信学会論文誌 Vol.J85-D-I,No.6,pp.512-526, 2002.
8) 北川高嗣, 中西崇文, 清木康, "静止画像メディアデータを対象としたメタデータ自動抽出方式の実現とその意味的画像検索への適用," 情報処理学会論文誌：データベース, VOl.43,No.SIG12(TOD16), pp38-51, 2002.
9) 中西崇文, 北川高嗣, 清木康, "任意の印象語による顔の表情の自動合成方式の実現," 情報処理学会論文誌：データベース, Vol.44,No.SIG8(TOD18), pp21-36, 2003.
10) Takafumi Nakanishi, Takashi Kitagawa, Yasushi Kiyoki, "An Implementation Method of Associative Search for Heterogenous Mediadata Utilizing the Mathematical Model of Meaning and its Application to Image Data and Facial Expression," 2003 IEEE Pacific Rim

Conference on Communications, Computers and Signal Processing (PACRIM '03), pp.613-618, August(2003).
11) "Longman Dictionary of Contemporary English",Longman, 1987.
12) 小林重順, "カラーイメージスケール", 講談社, 1984.
13) K.Hevner "expression in music: a discussion of experimental studies and theories". *Psychological Review*, Vol. 42, pp. 186–204, 1935.
14) K.Hevner "experimental studies of the elements of expression im music". *American Journal of Psychology*, Vol. 48, pp. 246–268, 1936.
15) K.Hevner "the affective value of pitch and tempo in music". *American Journal of Psychology*, Vol. 49, pp. 621–630, 1937.
16) 梅本堯夫 (編) "音楽心理学", 誠信書房, 1966.
17) "新編　感覚・知覚心理学ハンドブック", 誠信書房, 1994.