

# 離散コサイン変換領域の特徴に基づいた 画像索引作成法<sup>†</sup>

岡本 一志 \*1・董 芳艶 \*1・吉田 真一 \*2・廣田 薫 \*1

JPEGやMPEG画像からの画像索引作成の高速化を実現するため、離散コサイン変換(DCT)領域から抽出した色特徴およびテクスチャ特徴に基づき画像索引を作成する手法を提案する。これらの画像に記録されているDCT係数から画像特徴を抽出し、復号処理省略と処理信号数の削減により、索引作成時の特徴抽出処理を1000枚のJPEG画像に対し2.86秒で実現している。現在提案されているSerata(2006)らの画像索引作成法に比べ116.17倍高速化している。両手法に基づいた画像索引の適合率と再現率をCorelデータベースを用いた画像検索実験により評価し、提案法では最大で適合率が22.7%、再現率が11.1%向上する索引を作成できることを確認している。検索精度を向上させ、現在処理できる画像数に対し100倍の画像数を持つデータベースの画像索引作成を可能にする。

キーワード：画像検索、離散コサイン変換、索引、画像特徴量、クラスタリング

## 1. はじめに

撮影した画像あるいはネットワーク上の画像を、携帯電話、デジタルカメラ、ポータブルマルチメディアプレーヤーなどの携帯端末で扱う機会が急増している。このような携帯端末においても要求する画像を検索する機能は必要とされており、キーワードや画像を用いる画像検索技術が利用されている。画像内容に基づいた画像検索[1]-[5]は、ユーザが保有している画像やシステムが提示する画像を検索に用いるため、画像内容をキーワードや文章で表現すること無く検索できる。

画像内容に基づいた画像検索のひとつとして、システムが提示する画像索引中のキーを検索に用いるビジュアルキー型画像検索[6][7]が提案されている。画像索引中のキーは、ビジュアルキーと呼ばれ、文書検索におけるキーワードの役割を果たす。システムからビジュアルキーが提示されるため、ユーザはクエリ候補画像を準備する必要がなく、十字キーやダイヤルのような入力装置のみで検索クエリを作成でき、これら

の入力装置を持つ携帯端末で画像検索が実現できる。また、検索に画像索引を用いるため、検索クエリとデータベース中の各画像との類似度を逐次比較する必要無く、検索時の計算量削減が可能である。画像索引は、データベース中の各画像を複数画像に分解し、各部分画像の画像特徴量に基づく類似画像集合への分類処理により作成される。画像索引作成の計算量は、画像数や画像解像度に依存するため、データベース中に大量の画像や高解像度画像がある場合や携帯端末のような処理性能に制約がある端末への実装には、索引作成に必要な計算量を可能な限り削減する必要がある。

本論文では、現在普及しているJPEGやMPEG画像に対する画像索引作成に必要な計算量削減を目的とし、離散コサイン変換(DCT: Discrete Cosine Transform)領域での画像特徴を用いた画像索引作成法を提案する。JPEGやMPEGではDCTに基づき画像情報を削減しており、Serataら[6][7]が適用している画像特徴[8][9][10]には復号処理が必要である。一方、提案法は、画像中のDCT係数から画像特徴を直接抽出し、復号処理を省略するため、画像索引作成時の計算量を削減できる。画像特徴には色特徴とテクスチャ特徴を用いる。色特徴はDC係数から求められ、DC係数中の $YCrCb$ 信号[11]から算出したHMMD信号[12]ヒストグラムを特徴量とする。また、圧縮時に色差信号がダウンサンプリングされることを利用し、HMMD信号取得時の計算量を削減する。テクスチャ特徴は $YCrCb$ 信号中の輝度信号から求められ、3つのAC係数から算出した15種類の輝度パターンヒスト

<sup>†</sup> DCT Domain Features Based Image Index  
Kazushi OKAMOTO, Fangyan DONG, Shinichi YOSHIDA and Kaoru HIROTA

\*1 東京工業大学 大学院総合理工学研究科 知能システム科学専攻  
Department of Computational Intelligence and Systems Science,  
Interdisciplinary Graduate School of Science and Engineering,  
Tokyo Institute of Technology

\*2 高知工科大学 情報システム工学教室  
Department of Information Systems Engineering, Kochi University  
of Technology

グラム、輝度パターン自由度ヒストグラム、輝度パターン変化強度ヒストグラムを特徴量とする。

画像索引作成時の画像特徴抽出に必要な演算数、計算機への実装による特徴抽出処理時間、および画像索引を用いた画像検索実験により提案法の有用性を確認する。画像検索実験では、1,000枚の画像を含む Corel データベース [4]-[7] 中の各画像に対し、画像索引中の各ビジュアルキーを検索クエリとした際の最大適合率と最大再現率を測定する。また、画像索引作成時の部分画像分類処理に Serata らが適用している階層的クラスタリングは、要素数  $n$  に対し  $O(n^2)$  の計算量を必要とするため、計算量が  $O(n)$  である  $k$ -means 法 [13] を分類処理に適用する。提案する DCT 領域の画像特徴と  $k$ -means 法により作成した画像索引の検索精度を測定し、JPEG や MPEG 画像に対する索引作成時の計算量削減を確認する。

2 では、画像索引作成に適用する DCT 領域の色特徴とテクスチャ特徴抽出法について述べる。3 では、画像索引作成時の特徴抽出処理に必要な演算数を求め、計算機への実装により処理時間を測定する。4 では、DCT 領域の画像特徴、Serata らが適用している画像特徴、 $k$ -means 法に基づいた画像索引を用い、要求画像に対し各ビジュアルキーを検索クエリとする画像検索実験を行い、検索精度を評価する。

## 2. 画像索引作成に適用する DCT 領域の画像特徴抽出法

提案する画像索引作成法では、JPEG や MPEG 画像中の DCT 係数から色特徴とテクスチャ特徴を抽出する。これらの特徴は、各要素が  $[0, 1]$  で正規化されており、21次元の色特徴と29次元のテクスチャ特徴で表現される。DCT に基づく画像圧縮処理は、入力画像が  $8 \times 8$  ピクセルのブロック単位で、(1)RGB 信号から  $Y_C_R C_B$  信号 [11] への変換、(2)DCT、(3)量子化、(4)エントロピー符号化の順に処理される。画像情報量削減のため、(1)の処理において、色差信号がダウンサンプリングされる場合がある。画像の復号は上記の逆順で処理され、DCT 領域での特徴抽出により、60の DCT 係数に対する逆量子化処理、逆離散コサイン変換処理、色空間変換処理を省略している。また、復号処理の省略に加え、画像の信号エネルギーが低周波数成分の係数に集中する DCT の特性を利用し、画像の大まかな特徴を抽出することで処理信号数を削減している。DCT 領域での画像特徴抽出について述べるにあたり  $8 \times 8$  点逆離散コサイン変換を

$$f_{yx} = \frac{1}{4} \sum_{v=0}^7 \sum_{u=0}^7 \alpha_{vu} F_{vu} \cos \theta_{yv} \cos \theta_{xu} \quad (1)$$

で定義する。このとき、 $F_{vu}$  を DCT 係数、 $f_{yx}$  を座標  $(x, y)$  ( $0 \leq x, y \leq 7$ ) の画素値、

$$\theta_{xu} = \cos \frac{2x+1}{16} u\pi \quad (2)$$

$$\alpha_{vu} = \begin{cases} \frac{1}{2} & u = v = 0 \\ \frac{1}{\sqrt{2}} & u = 0 \text{ or } v = 0 \\ 1 & \text{otherwise} \end{cases} \quad (3)$$

とする。

### 2.1 色特徴

DC係数はブロックの信号平均であり、DC係数からの色特徴抽出は、正確な色特徴を抽出できないものの、画像の大まかな色の傾向を取得でき、処理信号数を64分の1に削減することが可能である [14]-[17]。このため、色特徴はDC係数から求め、係数中の  $Y_C_R C_B$  信号 [11] から算出した HMMD (Hue-Max-Min-Diff) 信号 [12] の色相信号 (hue)、彩度信号 (diff)、明度信号 (sum) ヒストグラムを特徴量とする。HMMD 色空間では、入力信号が無彩色の場合、色相信号は定義できないため、色相信号ヒストグラムは無彩色信号および有彩色信号レベルを等間隔に10分割した計11ピンで、彩度信号と明度信号ヒストグラムはそれぞれ等間隔に信号レベルを分割した5ピンで構成される。

$Y_C_R C_B$  信号の色差信号は画像情報量削減のため圧縮時にダウンサンプリングされている場合があり、これを利用し、 $Y_C_R C_B$  信号から HMMD 信号を算出する際の計算量を削減する。 $Y_C_R C_B$  信号は輝度信号 (Y) と色差信号 (CR, CB) で構成されており、 $Y_C_R C_B$  信号から RGB 信号への変換 [11] は、

$$\begin{pmatrix} R \\ G \\ B \end{pmatrix} = \begin{pmatrix} 1 & 1.40200 & 0 \\ 1 & -0.71414 & -0.34414 \\ 1 & 0 & 1.77200 \end{pmatrix} \begin{pmatrix} Y \\ C_R \\ C_B \end{pmatrix} \quad (4)$$

で、RGB 信号から HMMD 信号への変換 [12] は、

$$\begin{aligned} \text{sum} &= \frac{\max(R, G, B) + \min(R, G, B)}{2} \\ \text{diff} &= \max(R, G, B) - \min(R, G, B) \end{aligned} \quad (5)$$

$$\text{hue} = \begin{cases} 60 \times \frac{G - B}{\text{diff}} & \max(R, G, B) = R \\ 60 \times \frac{B - R}{\text{diff}} + 120 & \max(R, G, B) = G \\ 60 \times \frac{R - G}{\text{diff}} + 240 & \max(R, G, B) = B \end{cases}$$

でそれぞれ定義される。このとき、 $\max(R, G, B)$ はRGB信号の最大値、 $\min(R, G, B)$ はRGB信号の最小値とする。diffが0の場合は入力信号は無彩色であり、hueが360を超える際は360を加える必要がある。式4より、 $\max(R, G, B) - \min(R, G, B)$ は、 $|R - G|$ 、 $|R - B|$ 、 $|G - B|$ のいずれかであることから、色相信号と彩度信号は色差信号のみで求めることができる。圧縮時に $2 \times 2$ 輝度信号に対し1色差信号を取得する場合、HMMD信号算出に必要な処理信号数は、色相信号と彩度信号において、それぞれ4分の1になる。DC係数による $YC_{R,C_B}$ 信号の取得と上記の関係から、ピクセル領域でのHMMD信号ヒストグラムの算出に比べ、処理対象信号数を、明度信号において64分の1に、色相信号および彩度信号において256分の1にそれぞれ削減できる。

2.2 テクスチャ特徴

テクスチャ特徴は輝度信号から求められ、3つのAC係数から算出した15種類の輝度パターンヒストグラム、輝度パターン自由度ヒストグラム、輝度パターン変化強度ヒストグラムを特徴量とする。この特徴は、既存のDCT領域の特徴[14]-[16][18]-[23]に対し、 $8 \times 8$ ブロックの輝度パターンと輝度パターン変化強度を同時に算出できる利点がある。輝度パターンは、 $8 \times 8$ ブロックの信号平均と、 $8 \times 8$ ブロックを水平、垂直方向に2分割することで得られる4領域の各信号平均 $a_{ji}$  ( $i, j \in \{0, 1\}$ )との差分 $d_{ji}$ により定まる。 $a_{ji}$ は、DCT係数により、

$$\begin{aligned} a_{ji} &= \frac{1}{16} \sum_{y=4j}^{4(j+1)-1} \sum_{x=4i}^{4(i+1)-1} f_{yx} \\ &= \frac{1}{64} \sum_{y=4j}^{4j+3} \sum_{x=4i}^{4i+3} \sum_{v=0}^7 \sum_{u=0}^7 \alpha_{vu} F_{vu} \cos \theta_{xu} \cos \theta_{yv} \quad (6) \\ &= \frac{1}{64} \sum_{v=0}^7 \sum_{u=0}^7 \alpha_{vu} c_{iu} c_{jv} F_{vu} \end{aligned}$$

で表現される。このとき、

$$\begin{aligned} c_{iu} &= \sum_{x=4i}^{4i+3} \cos \theta_{xu} \\ &= \begin{cases} 4 & u = 0 \\ \frac{(-1)^j}{\sqrt{2}} \cos\left(\frac{u\pi}{8}\right) \cos\left(\frac{u\pi}{16}\right) & u = 1, 5 \\ \frac{(-1)^{i+1}}{\sqrt{2}} \cos\left(\frac{u\pi}{8}\right) \cos\left(\frac{u\pi}{16}\right) & u = 3, 7 \\ 0 & \text{otherwise} \end{cases} \quad (7) \end{aligned}$$

とする。 $|\alpha_{vu} c_{iu} c_{jv}|$ は、添字 $u, v$ が大きくなるに従い $|\alpha_{vu} c_{iu} c_{jv}|$ は小さくなり[22]、DCT係数 $F_{vu}$ が0になる可能性が高くなるため、 $F_{01}, F_{10}, F_{11}$ の3つのAC係数のみを利用する。 $8 \times 8$ ブロックの信号平均 $\frac{1}{8}F_{00}$ [14]-[17]と $a_{ji}$ の差分 $d_{ji}$ は、 $c = \cos\left(\frac{\pi}{8}\right) \cos\left(\frac{\pi}{16}\right)$ としたとき、

$$\begin{aligned} d_{ji} &= a_{ji} - \frac{1}{8}F_{00} \\ &= \frac{c}{8} \left( (-1)^i F_{01} + (-1)^j F_{10} + (-1)^{i+j} c F_{11} \right) \quad (8) \end{aligned}$$

であり、2回の加算と2回の乗算で求めることができる。このとき、 $8 \times 8$ ブロックは、 $d_{ji} \leq 0$ となる領域と、 $d_{ji} > 0$ となる領域に分けられ、4自由度を持つ15種類(= $2^4 - 1$ )の輝度パターンで表現できる。図1に15種類の輝度パターンを示す。図1では、 $d_{ji} \leq 0$ となる領域を黒、 $d_{ji} > 0$ となる領域を白とし、上から順に自由度(DOF: degree of freedom) 0, 1, 2, 3とする。また、15次元の輝度パターンヒストグラムのみでは、画像回転時に画像特徴量が異なるため、回転に対し頑健な特徴として、輝度パターン算出時に取得可能な輝度パターン自由度と輝度パターン変化強度ヒストグラムを利用する。輝度パターン変化強度 $s$ は、 $d_{ji}$ が、

$$\sum_{j=0}^1 \sum_{i=0}^1 d_{ji} = 0 \quad (9)$$

という性質を有しているため、 $d_{ji} > 0$ となる $d_{ji}$ の総和

$$s = \sum_{j=0}^1 \sum_{i=0}^1 d_{ji} \quad d_{ji} > 0 \quad (10)$$

とする。輝度パターン変化強度ヒストグラムは9分割した $s < 63$ の強度レベルと $s \geq 63$ の強度レベルの計10ビンで構成される。

DOF

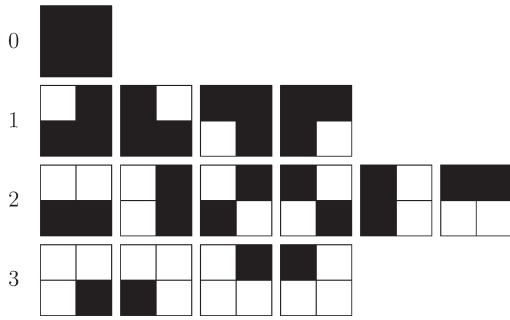


図1 テクスチャ特徴で利用される 15 パターン

### 3. 画像特徴抽出に必要な演算数と計算機への実装

Serata らが適用しているピクセル領域での画像特徴 [8][9][10]と 2 で提案した DCT 領域の画像特徴を、特徴抽出に必要な演算数と計算機への実装により評価する。評価では、JPEG 画像からの特徴抽出を想定しており、サンプリングファクタを  $Y : C_R : C_B = 4 : 1 : 1$  とし、4 輝度信号  $8 \times 8$  DCT ブロック、1 赤色差信号  $8 \times 8$  DCT ブロック、1 青色差信号  $8 \times 8$  DCT ブロックの計 6 DCT ブロックで構成される MCU (Minimum Coded Unit) 単位でハフマン符号化されるものとする。

#### 3.1 画像特徴抽出に必要な演算数

Serata らが適用している画像特徴は、JPEG や MPEG 画像からの特徴抽出時に、逆量子化処理、逆離散コサイン変換処理、色空間変換処理が必要になる。11回の乗算と29回の加算での8点高速逆離散コサイン変換 [24] から、 $176 (= 11 \times 16)$  回の乗算と  $464 (= 29 \times 16)$  回の加算処理で式 1 で定義されている  $8 \times 8$  点逆離散コサイン変換が実現できる。表 1 に、1 MCU 当たりの、Serata らが適用している画像特徴の抽出に必要な加算、乗算、条件処理、冪乗の演算回数を示す。表 1 の上段は JPEG 復号処理、中段は特徴抽出のための色空間変換処理、下段は特徴抽出処理に対応している。表 1 中の、*add.*, *mul.*, *cmp.*, *exp.* はそれぞれ加算、乗算、条件処理、冪乗の演算回数を意味している。

一方、提案する DCT 領域の画像特徴は、DCT 係数を利用して特徴抽出するため、JPEG や MPEG 画像からの特徴抽出時に、逆離散コサイン変換処理、60 の DCT 係数に対する逆量子化処理、 $Y C_R C_B$  信号から RGB 信号への変換処理を省略できる。表 2 に、1 MCU 当たりの、提案する DCT 領域での画像特徴の抽出に必要な加算、乗算、条件処理の演算回数を示

表 1 Serata らが適用している画像特徴量の抽出に必要な 1 MCU 当たりの最大演算数

	<i>add.</i>	<i>mul.</i>	<i>cmp.</i>	<i>exp.</i>
IDCT	2,784	1,056	-	-
inverse quantization	-	384	-	-
$Y C_R C_B$ to RGB	1,024	1,024	-	-
normalization	-	768	-	-
RGB to $L^*a^*b^*$	2,048	3,584	768	768
RGB to HSI	256	256	1,024	-
color moments [8]	3,063	1,536	-	-
texture neighborhood [9]	2,296	-	2,048	-
Sobel edge detector [10]	30,712	-	8,192	-
total	42,183	8,608	12,032	768

表 2 提案する DCT 領域の画像特徴の抽出に必要な 1 MCU 当たりの最大演算数

	<i>add.</i>	<i>mul.</i>	<i>cmp.</i>
inverse quantization	-	18	-
$Y C_R C_B$ to HMMD	10	7	8
pattern	40	8	40
hue histogram	2	2	1
diff histogram	1	1	1
sum histogram	4	4	4
pattern histogram	4	-	-
weight histogram	4	4	4
DOF histogram	4	-	-
total	69	44	58

す。表 2 の上段は特徴抽出に用いる DC 係数と 3 つの AC 係数の逆量子化処理、中段は HMMD 信号、輝度パターン、輝度パターン自由度、輝度パターン変化強度の算出処理、下段はヒストグラムの各ビンへの割当処理に対応している。表 2 中の、*add.*, *mul.*, *cmp.* はそれぞれ加算、乗算、条件処理の演算回数を意味している。提案する DCT 領域での画像特徴は、Serata らが用いているピクセル領域の画像特徴に対し、冪乗演算を省略し、かつ加算処理 0.16%、乗算処理 0.51%、条件処理 0.48% の演算量で JPEG や MPEG 画像から特徴を抽出できる。

#### 3.2 計算機への実装による特徴抽出処理時間の測定

提案する DCT 領域の画像特徴抽出法と Serata らが適用している画像特徴抽出法を、Intel 社製 CPU Core 2 Duo 1.86 GHz、FreeBSD 7.0 の計算機上に実装し、処理時間を測定する。C 言語で実装し、コンパ

イラには GNU C Compiler 4.4 を利用する. また, 特徴抽出は JPEG デコーダ上でを行い, デコーダは文献 [25] の仕様に従うものとする. 入出力処理, JPEG ヘッダ解析処理およびハフマンテーブル作成処理を除く 1,000 枚の  $384 \times 256$  ピクセルの JPEG 画像に対する特徴抽出処理時間は, 提案する DCT 領域の画像特徴抽出法において 2.86 秒, Serata らが適用している画像特徴抽出法において 332.25 秒であり, 116.17 倍の処理性能向上を確認している.

#### 4. Corel データベースを用いた画像検索実験

提案する DCT 領域の画像特徴と Serata らの適用している画像特徴に基づいた画像索引の検索精度を画像検索実験により評価する. 実験では, それぞれ 10 カテゴリ, 各カテゴリ  $384 \times 256$  ピクセルの JPEG 画像 100 枚で構成される Corel1000-A [4]-[7] と Corel1000-B [5] データベースを用いる. 検索精度には, 情報検索システムの評価指標である適合率と再現率を適用する. 画像索引による検索では, ユーザは検索システムにより提示された  $n$  枚のビジュアルキーから, データベース DB 中の画像  $D \in 2^{DB}$  を検索するための検索クエリ  $q_D = (q_1, q_2, \dots, q_n)$  ( $q_i \in \{0, 1\}, 1 \leq i \leq n$ ) を作成する. 検索システムは, 検索クエリと画像索引から, コサイン尺度に基づき検索結果画像をユーザに出力する.  $A(q_D) \in 2^{DB}$  を検索クエリ  $q_D$  に対応する検索結果画像集合,  $\#\{\cdot\}$  を集合中の画像数とした場合, 検索クエリ  $q_D$  に対する適合率  $p(q_D)$  と再現率  $r(q_D)$  は,

$$p(q_D) = \frac{\# \{A(q_D) \cap D\}}{\# \{A(q_D)\}} \quad (11)$$

$$r(q_D) = \frac{\# \{A(q_D) \cap D\}}{\# \{D\}} \quad (12)$$

で定義される.

データベース中の各画像はビジュアルキーを索引として持ち, 各ビジュアルキーに対応する検索結果画像は定まっているため, 1 ビジュアルキーを検索クエリとした際の検索精度に検索システムの検索性能が依存する. そのため, Corel データベースの各カテゴリを要求画像集合  $D_j$  ( $1 \leq j \leq 10$ ) とし,  $D_j$  に対する各ビジュアルキー  $q_{D_j}^i$  ( $\|q_{D_j}^i\| = 1, \|\cdot\|$  はノルム) の適合率と再現率を測定し,  $D_j$  に対する最大適合率

$$\max_{v_i} \left\{ p \left( q_{D_j}^i \right) \right\} \quad (13)$$

と最大再現率

$$\max_{v_i} \left\{ r \left( q_{D_j}^i \right) \right\} \quad (14)$$

を求め, 平均最大適合率

$$\frac{1}{10} \sum_{j=1}^{10} \max_{v_i} \left\{ p \left( q_{D_j}^i \right) \right\} \quad (15)$$

と平均最大再現率

$$\frac{1}{10} \sum_{j=1}^{10} \max_{v_i} \left\{ r \left( q_{D_j}^i \right) \right\} \quad (16)$$

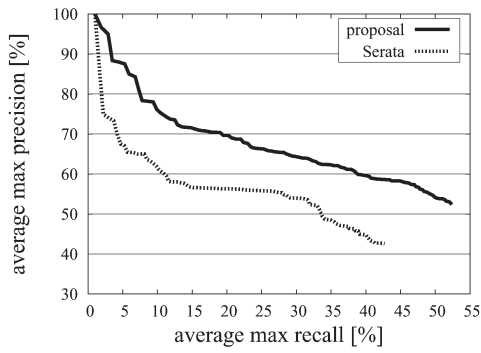
を算出する.

##### 4.1 実験環境

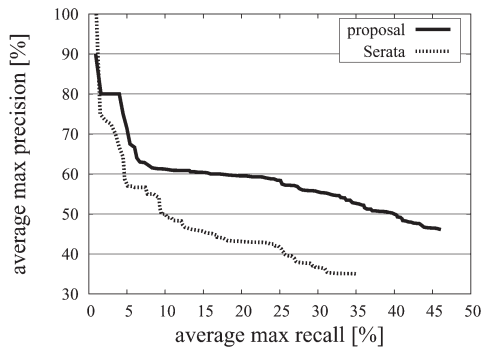
画像検索実験に用いる Corel1000-A データベースには, Africa, Beach, Buildings, Buses, Dinosaurs, Elephants, Flowers, Foods, Horses, Mountains のカテゴリが, Corel1000-B データベースには, Autumn, Desert, Flowers, Mountains, Night, Perennial, Roses, Seaside, Sunset, Winter のカテゴリがある. 画像索引作成には, 索引候補を得るためのデータベース中の各画像に対する複数画像分解処理と, 各部分画像を類似画像集合に分類する処理が必要である. JPEG や MPEG では一定の大きさの矩形領域単位で画像を符号化しており, 符号の解析により矩形領域情報を取得できる. このため, 各画像の複数画像分解には,  $128 \times 128$  ピクセル単位での矩形領域分割を用いており, Corel1000-A および Corel1000-B データベースからそれぞれ 6,000 の部分画像を得ている. 6,000 の部分画像を, クラスタリングアルゴリズムを適用することで類似部分画像集合に分類し, 各クラスターの重心に最も近い要素をビジュアルキーとする. そして, 類似部分画像集合の要素を持つ画像をビジュアルキーで索引付けする. クラスタリング処理における距離関数には, Serata らは重み付きマンハッタン距離とユークリッド距離を適用しているが, 提案法では, 単純化のため, ユークリッド距離を用いる.

##### 4.2 提案する DCT 領域の画像特徴に基づいた画像索引の平均最大適合率と平均最大再現率

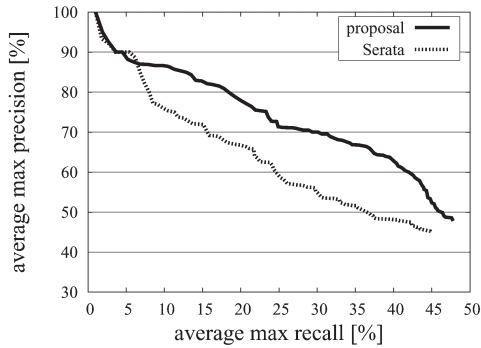
提案する DCT 領域の特徴と Serata らの画像特徴に基づいた画像索引に対する平均最大適合率と平均最大再現率を測定する. 両画像特徴に基づく画像索引の検索精度への影響を評価するため, 部分画像の分類処理には階層的クラスタリングを適用している. 文献 [6] ではビジュアルキー数が 80 程度が望ましいという結果が報告されており, 携帯端末で表示可能なビジュアルキー数にも限られるため, 画像索引中のビジュアルキー



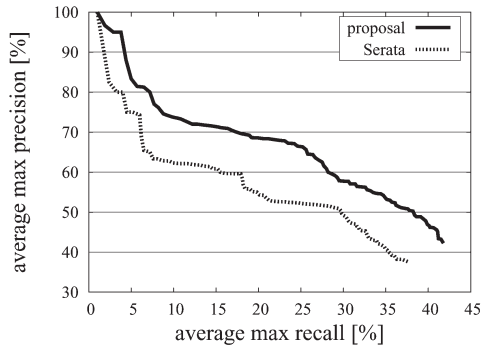
(a1) Corel1000-A, ビジュアルキー数 20



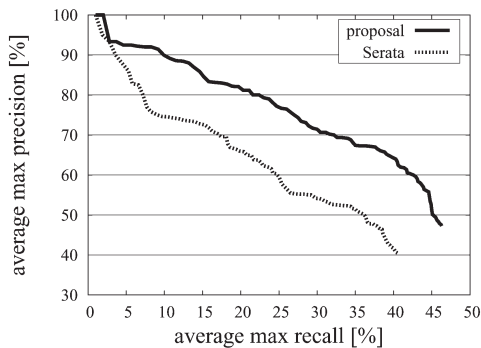
(b1) Corel1000-B, ビジュアルキー数 20



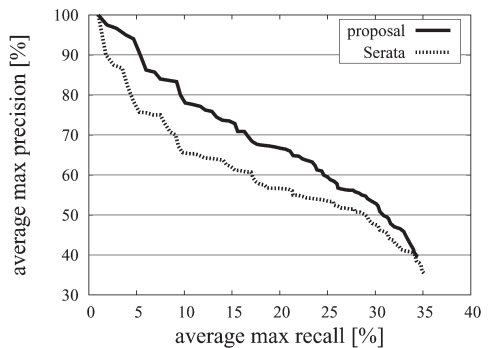
(a2) Corel1000-A, ビジュアルキー数 40



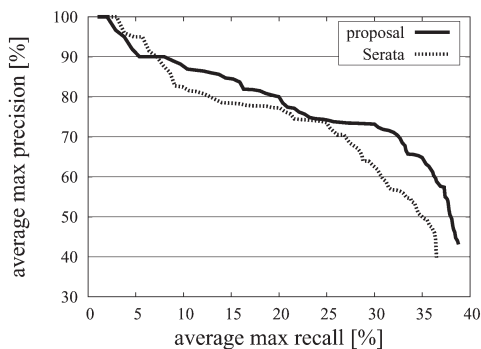
(b2) Corel1000-B, ビジュアルキー数 40



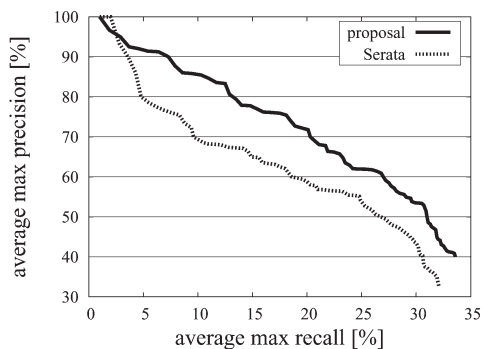
(a3) Corel1000-A, ビジュアルキー数 60



(b3) Corel1000-B, ビジュアルキー数 60

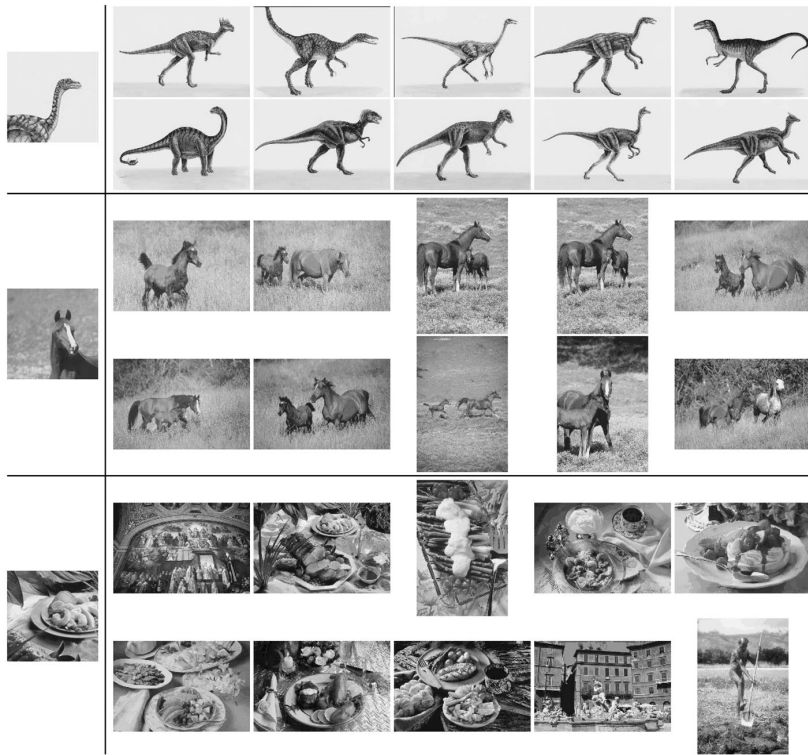


(a4) Corel1000-A, ビジュアルキー数 80

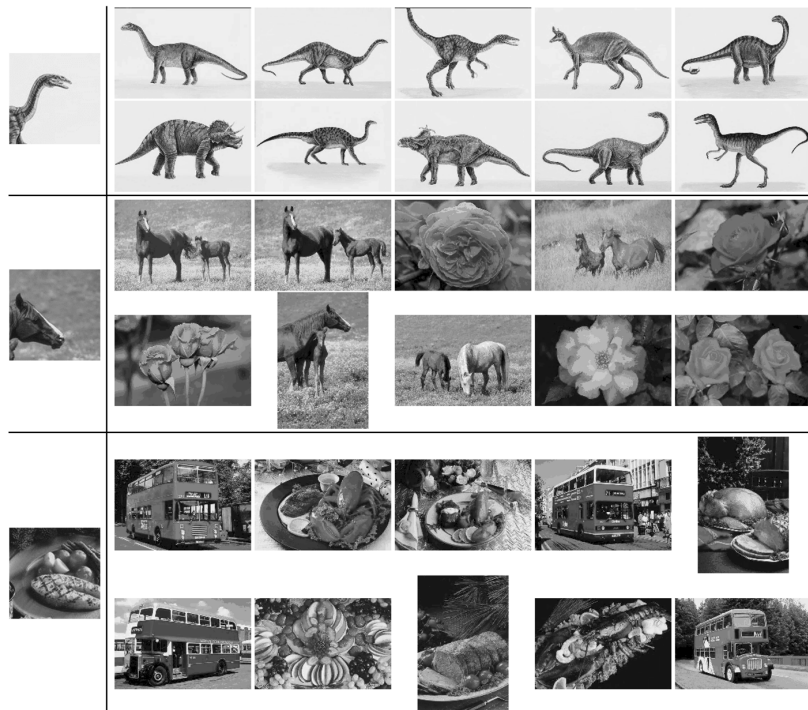


(b4) Corel1000-B, ビジュアルキー数 80

図2 Corel1000-AとCorel1000-Bデータベースにおける平均最大再現率-平均最大適合率グラフ



(a) 提案する DCT 領域の画像特徴に基づく画像索引の検索結果例



(b) Serata らが適用している画像特徴に基づく画像索引の検索結果例

図3 提案する DCT 領域の画像特徴と Serata らが適用している画像特徴に基づいた画像索引による検索結果例

数を20, 40, 60, 80とする. 図2に, Core1000-AとCore1000-Bデータベースに対し, 画像索引中のビジュアルキー数を20, 40, 60, 80とした際の平均最大再現率-平均最大適合率グラフを示す. グラフの横軸は平均最大再現率とし, 縦軸は各再現レベルにおける平均最大適合率とする. 図2の上段から順に, ビジュアルキー数が20, 40, 60, 80の場合であり, 各段において左側がCore1000-Aデータベース, 右側がCore1000-Bデータベースにおける平均最大再現率-平均最大適合率グラフである.

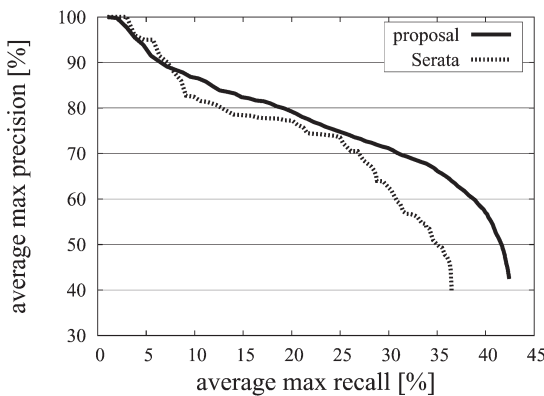
索引中のビジュアルキー数が少なくなるにつれ, 各ビジュアルキーを索引として持つ画像が増加するため, 1ビジュアルキー当たりの再現率は増加する. ビジュアルキー数を20とした場合, 図2(a1)と図2(b1)より, 平均最大再現率はSerataらが適用している画像特徴に基づいた画像索引に対し, Core1000-Aデータベースにおいて9.7%, Core1000-Bデータベースにおいて11.1%向上している. また, 平均最大適合率はSerataらが適用している画像特徴に基づいた画像索引に対し, 図2(a4)において再現レベルが8.0%未満の場合, 最大で5.0%低下するものの, それ以外の各再現レベルにおいて同等以上の平均最大適合率を有していることが図2より確認できる. 図2(a3)の40.0%の再現レベルにおいて, 平均最大適合率は最大で22.7%向上している. Serataらが適用している画像特徴に基づいた画像索引に対し, 提案するDCT領域の画像特徴に基づいた画像索引は, 同等以上の1ビジュアルキー当たりの平均最大適合率と平均最大再現率を有しており, AND検索やOR検索時の検索精度向上も期待できる.

Core1000-Aデータベースに対し, ビジュアルキー

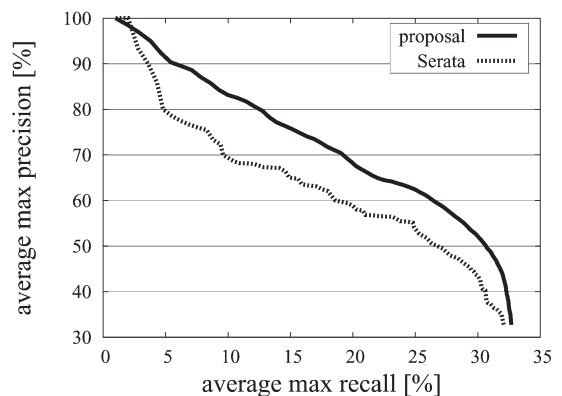
数を80とした際の, 両画像特徴に基づく画像索引による検索結果の例を図3に示す. 図3では, 画像索引中に現れた同様のビジュアルキーにより検索している. 図3内にある各画像は, 左から順に, ビジュアルキーとビジュアルキーに対する1位から10位までの検索結果である. 図3(a)の索引は提案するDCT領域の特徴に基づき, 図3(b)の索引はSerataらの適用している画像特徴に基づきそれぞれ作成している. 図3(a)と図3(b)の上段の検索結果については, 同じカテゴリ(Dinosaurs)の画像が同じビジュアルキーを索引として持っている. 中段において, 提案するDCT領域の画像特徴に基づいた索引では, 同じカテゴリ(Horses)の画像が同じビジュアルキーを索引として持つ一方, Serataらの画像特徴に基づいた索引では, Horses以外の異なるカテゴリの画像が同じビジュアルキーを索引として持つ. 下段については, 両者ともFoods以外の異なるカテゴリの画像も同じビジュアルキーを索引に持つ.

#### 4.3 提案するDCT領域の画像特徴とk-means法に基づいた画像索引の平均最大適合率と平均最大再現率

Serataらが部分画像の分類処理に適用している階層的クラスタリングは, 要素数 $n$ に対し $O(n^2)$ の計算量が必要であることから計算量が $O(n)$ のk-means法[13]を適用する. 提案するDCT領域の画像特徴とk-means法に基づく画像索引およびSerataらが適用している画像特徴と階層的クラスタリングに基づく画像索引による平均最大再現率-平均最大適合率グラフを図4に示す. グラフの横軸は平均最大再現率とし, 縦軸は各再現レベルにおける平均最大適合率とする.



(a) Core1000-A, ビジュアルキー数80



(b) Core1000-B, ビジュアルキー数80

図4 提案するDCT領域の画像特徴とk-means法に基づいた画像索引および従来法による画像索引の平均最大再現率-平均最大適合率グラフ



$k$ -means 法は、クラスタリング結果が初期値に依存するため、初期値を20パターン用意しており、図4の平均最大適合率と平均最大再現率は、20パターンの平均である。図4から、階層的クラスタリングを適用した結果である図2と  $k$ -means 法を適用した結果が同様の傾向を示していることが確認できる。したがって、提案する DCT 領域の画像特徴と  $k$ -means 法の適用により、検索精度を向上させかつ JPEG や MPEG 画像に対する画像索引の作成時間短縮が可能である。

## 5. おわりに

本論文では、JPEG や MPEG 画像に対する画像索引作成の計算量削減を目的とし、DCT 領域での画像特徴を用いた画像索引作成法を提案している。提案法は、画像中の DCT 係数から画像特徴を直接抽出し、復号処理を省略するため、索引作成時の画像特徴抽出に必要な計算量を削減している。画像特徴には色特徴とテクスチャ特徴を用いている。色特徴は DC 係数中の  $YCrCb$  信号から算出した HMMD 信号ヒストグラムを特徴量としている。また、圧縮時に色差信号がダウンサンプリングされることを利用し、HMMD 信号取得時の計算量を削減している。テクスチャ特徴は輝度信号から求められ、3つの AC 係数から算出した15種類の輝度パターンヒストグラム、輝度パターン自由度ヒストグラム、輝度パターン変化強度ヒストグラムを特徴量としている。

画像索引作成時の画像特徴抽出に必要な演算数、計算機への実装による特徴抽出処理時間、および画像索引を用いた画像検索実験により提案法の有用性を確認している。提案する DCT 領域での画像特徴は、Serata らが用いている画像特徴[8][9][10]に対し、乗算演算を省略し、かつ加算処理0.16%、乗算処理0.51%、条件処理0.48%の演算量で JPEG や MPEG 画像から特徴抽出が可能である。また、Intel 社製 CPU Core 2 Duo 1.86GHz 計算機における1,000枚の  $384 \times 256$  ピクセルの JPEG 画像に対する特徴抽出処理時間は、提案する DCT 領域の画像特徴抽出法で2.86秒、Serata らが適用している画像特徴抽出法で332.25秒であり、116.17倍の処理性能向上を実現している。画像検索実験では、1,000枚の画像をそれぞれ含む Corel1000-A[4]-[7]と Corel1000-B[5]データベース中の各画像に対し、画像索引中の各ビジュアルキーを検索クエリとした際の最大適合率を測定して、その結果、従来法に対し提案法は、同等以上の1ビジュアルキー当たりの平均最大適合率と平均最大再現率を有しており、AND 検索や OR 検索時の検索精度向上も期待できる。また、画像索引作成時の類似部分画像分類処理に Serata ら

が適用している階層的クラスタリングは、要素数  $n$  に対し  $O(n^2)$  の計算量を必要とし、画像数や画像解像度の増加に伴い計算量が増加する。そのため、提案する DCT 領域の画像特徴と計算量が  $O(n)$  である  $k$ -means 法[13]を組み合わせることで、従来法に対し、平均最大適合率と平均最大再現率を向上させかつ JPEG や MPEG 画像に対する索引作成時間の短縮が可能であることを確認している。

提案する DCT 領域に基づく画像索引作成法では、画像索引作成の計算量を削減しているため、既に普及している携帯電話、デジタルカメラ、ポータブルマルチメディアプレーヤー、デジタルビデオレコーダーなどの端末でも、ソフトウェア追加により画像検索を利用できる可能性がある。ディスプレイやデジタルフォトフレームのように処理装置の小型化が必要な機器への実装も期待できる。また、処理性能に制約のある端末だけでなく、大量の画像や高解像度画像のあるネットオークション、写真共有サービス、動画像配信サービスなどの高処理負荷なシステムへの拡張も可能である。

## 参考文献

- [1] M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, and P. Yanker, "Query by Image and Video Content: The QBIC System," *IEEE Computer*, vol.28, no.9, pp.23-32, 1995.
- [2] A. Pentland, R. W. Picard, and S. Sclaroff, "Photobook: Content-Based Manipulation of Image Databases," *J. of Computer Vision*, vol.18, no.3, pp.233-254, 1996.
- [3] C. Carson, S. Belongie, H. Greenspan, and J. Malik, "Blobworld: Image Segmentation using Expectation-Maximization and Its Application to Image Querying," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.24, no.8, pp.1026-1038, 2002.
- [4] J. Z. Wang, J. Li, and G. Wiederhold, "SIMPLcity: Semantics-Sensitive Integrated Matching for Picture Libraries," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.23, no.9, pp.947-963, 2001.
- [5] Z. Stejic, Y. Takama, and K. Hirota, "Relevance Feedback-Based Image Retrieval Interface Incorporating Reagon and Feature Saliency Patterns as Visualizable Image Similarity Criteria," *IEEE Trans. on Industrial Electronics*, vol.50, no.5, pp.839-852, 2003.
- [6] M. Serata, Y. Hatakeyama, and K. Hirota, "Designing Image Retrieval System with the Concept of Visual Keys," *J. of Advanced Computational Intelligence and Intelligent Informatics*, vol.10, no.2, pp.136-144, 2006.
- [7] 世良田 学, 島山 豊, 延原 肇, 廣田 薫, "ビジュアルキー木に基づく効率的な画像検索手法の提案," 日本知能情報ファジィ学会, vol.18, no.6, pp.892-899, 2006.

- [ 8 ] M. Stricker and M. Orengo, "Similarity of Color Images," *Proc. of Storage and Retrieval for Image and Video Databases III*, vol.2420, pp.381-392, 1995.
- [ 9 ] J. Laaksonen, E. Oja, M. Koskela, and S. Brandt, "Analyzing Low-Level Visual Features using Content-Based Image Retrieval," *Proc. of Int. Conf. on Neural Information Processing*, pp.1333-1338, 2000.
- [10] S. Brandt, J. Laaksonen, and E. Oja, "Statistical Shape Features for Content-Based Image Retrieval," *J. of Mathematical Imaging and Vision*, vol.17, no.2, pp.187-198, 2002.
- [11] ITU-R, "Studio Encoding Parameters of Digital Television for Standard 4:3 and Wide-Screen 16:9 Aspect Ratios," Recommendation BT.601-6, 2007.
- [12] B. S. Manjunath, J. R. Ohm, V. V. Vasudevan, and A. Yamada, "Color and Texture Descriptors," *IEEE Trans. on Circuits and Systems for Video Technology*, vol.11, no.6, pp.703-715, 2001.
- [13] J. A. Hartigan and M. A. Wong, "A K-Means Clustering Algorithm," *Applied Statistics*, vol.28, pp.100-108, 1979.
- [14] C. W. Ngo, T. C. Pong, and R. T. Chin, "Exploiting Image Indexing Techniques in DCT Domain," *Pattern Recognition*, vol. 34, no. 9, pp.1841-1851, 2001.
- [15] G. Feng and J. Jiang, "JPEG Compressed Image Retrieval via Statistical Features," *Pattern Recognition*, vol.36, no.4, pp.977-985, 2003.
- [16] C. Theoharatos, V. K. Pothos, N. A. Laskaris, G. Economou, and S. Fotopoulos, "Multivariate Image Similarity in the Compressed Domain using Statistical Graph Matching," *Pattern Recognition*, vol.39, no.10, pp.1892-1904, 2006.
- [17] J. Jiang, Y. Weng, and P. J. Li, "Dominant Colour Extraction in DCT Domain," *Image and Vision Computing*, vol.24, no.12, pp.1269-1277, 2006.
- [18] S. Climer and S. K. Bhatia, "Image Database Indexing using JPEG Coefficients," *Pattern Recognition*, vol.35, no.11, pp.2479-2488, 2002.
- [19] J. Jiang and Y. Weng, "Video Extraction for Fast Content Access to MPEG Compressed Videos," *IEEE Trans. on Circuits and Systems for Video Technology*, vol.14, no.5, pp.595-605, 2004.
- [20] C. C. Chang, J. C. Chuang, and Y. S. Hu, "Retrieving Digital Images from a JPEG Compressed Image Database," *Image and Vision Computing*, vol.22, no.6, pp.471-484, June 2004.
- [21] J. Jiang, A. Armstrong, and G. C. Feng, "Web-Based Image Indexing and Retrieval in JPEG Compressed Domain," *Multimedia Systems*, vol.9, no.5, pp.424-432, 2004.
- [22] H. S. Chang and K. Kang, "A Compressed Domain Scheme for Classifying Block Edge Patterns," *IEEE Trans. on Image Processing*, vol.14, no.2, pp.145-151, 2005.
- [23] K. J. Qiu, J. Jiang, G. Xiao, and S. Y. Irianto, "DCT-Domain Image Retrieval Via Block-Edge-Patterns," *Lecture Notes in Computer Science*, vol.4141, pp.673-684, 2006.
- [24] C. Loeffler, A. Ligtenberg, and G. S. Moschytz, "Practical Fast 1-D DCT Algorithms with 11 Multiplications," *Proc. of Int. Conf. on Acoustics, Speech, and Signal Processing*, vol.2, pp.988-991, 1989.
- [25] ITU-T, "Information Technology-Digital Compression and Coding of Continuous-Tone Still Image-Requirements and Guidelines," Recommendation T.81, 1992.

(2009年3月8日 受付)  
(2009年8月24日 採録)

[問い合わせ先]

〒226-8502 横浜市緑区長津田町4259G3-49

東京工業大学 大学院総合理工学研究科 知能システム科学専攻 廣田研究室

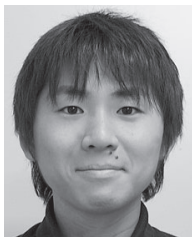
岡本 一志

TEL : 045-924-5686

FAX : 045-924-5676

E-mail : okamoto@hrt.dis.titech.ac.jp

## 著者紹介



おかもと かずし  
岡本 一志 [学生会員]

東京工業大学博士後期課程在学中(大学院総合理工学研究科知能システム科学専攻)。2006年、高知工科大学工学部情報システム工学科卒業。2008年、高知工科大学大学院工学研究科基盤工学専攻修士課程修了。修士(工学)。現在に至る。画像検索の研究に従事。



とう ほうえん  
董 芳艶 [正会員]

東京工業大学助教(大学院総合理工学研究科知能システム科学専攻)。2003年、東京工業大学大学院総合理工学研究科知能システム科学専攻博士後期課程修了。博士(工学)。東京工業大学特別研究員・産学官連携特別研究員(NEDO)を経て、2006年5月より同大学助手、2007年4月より現職。物流関連の組合せ最適化問題、メタ戦略、音声/画像による個人認証、ロボットなどの研究に従事。



よしだ しんいち  
吉田 真一 [正会員]

高知工科大学講師(情報システム工学教室)。1996年、中央大学理工学部電気・電子工学科卒業。2001年、東京工業大学大学院総合理工学研究科知能システム科学専攻博士課程修了。博士(工学)。2001年、東京工業大学大学院総合理工学研究科助手。2004年、青山学院大学附置情報科学研究センター助手。2007年より現職。ソフトウェアコンピューティングの検索等への応用に関する研究に従事。



ひろた かおる  
廣田 薫 [正会員]

東京工業大学教授(大学院総合理工学研究科知能システム科学専攻)。1979年、東京工業大学大学院理工学研究科電気工学専攻修了。工学博士。相模工業大学専任講師、法政大学工学部専任講師・助教授・教授を経て、1995年より現職。計算知能の研究に従事。当学会会長を務め、現在IFSA President、Banki Donat Medal、Henri Coanda Medal、Grigore MOISIL Award、当学会論文賞、日本音響学会論文賞などを受賞。de La Salle University・長春理工大学などの名誉教授、Bulacan State University 名誉博士(Honoris Causa)。

## DCT Domain Features Based Image Index

by

Kazushi OKAMOTO, Fangyan DONG, Shinichi YOSHIDA and Kaoru HIROTA

### Abstract :

An image index, extracting DCT domain color and texture features in indexing process, is proposed for JPEG and MPEG fast image retrieval. Feature extraction from 1,000 JPEG images requires 2.86 seconds, which is 116.17 times faster than previously reported by Serata (2006) because of efficient compressed domain feature extraction. Retrieval performances of the proposed and Serata's indexes are evaluated on Corel database, and the results suggest that the maximum precision and recall of the proposed index are improved by 22.7% and 11.1% with reduction of indexing times. Due to reduced computational times, the index is applicable to databases which include 100 times larger data sizes than current one.

**Keywords :** image retrieval, discrete cosine transform, index, feature, clustering

Contact Address : **Kazushi OKAMOTO**

*Hirota Laboratory, Department of Computational Intelligence and Systems Science, Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology*  
G3-49, 4259 Nagatsuta, Midori-ku, Yokohama 226-8502, JAPAN  
TEL : 045-924-5686  
FAX : 045-924-5676  
E-mail : okamoto@hrt.dis.titech.ac.jp