

複数の順序情報を用いたコンテンツ識別手法

高田 智規 阿部 剛仁 川村 春美
日本電信電話株式会社 NTT サイバースペース研究所

デジタルコンテンツの利用頻度増加に伴い、コンテンツの持つ特徴量を用いてコンテンツに付随する情報を利用するサービスが要望されている。

我々はこれまでに、コンテンツ内局所特徴値の順序関係を用いてコンテンツが同一であるか識別する手法を提案した。この手法は既存の類似検索などと比べコンテンツを精度良く識別可能であったが、想定されるノイズが大きい場合に識別精度が劣化するという問題があった。

本稿では、同一種別の順序情報を複数個利用することで、ノイズ耐性と識別精度の双方を向上させる方式を提案する。また、実験を通じて有効性を示す。

Content Identification Method Using Multiple Order Information

TOMONORI TAKADA , TAKEHITO ABE and HARUMI KAWAMURA
NTT Cyber Space Laboratories, NTT Corporation

We have proposed a content identification method with conversion robustness. However, our method has a problem that identification accuracy is deteriorated when noise is big.

In this paper, we propose a new method that can improve identification accuracy by using multiple order information. The effectiveness of this method is shown through our experiment.

1. はじめに

ネットワークの発達と端末性能の向上は、デジタルコンテンツの普及に大きく貢献している。音楽や動画のダウンロードサービスや、デジタルカメラを用いた写真撮影など、一般の家庭でもデジタルコンテンツを利用するシーンが増えている。このように、情報伝達の幅が広がりエンドユーザの利便性が向上する一方、コンテンツの不正利用や増加していくコンテンツの管理方法など、新たな課題が発生している。

デジタルコンテンツから一意な特徴量を抽出し、識別子 (ID) 代わりに利用することで、コンテンツと付随する情報 (メタデータ) を結びつけることができる。これによって、著作権管理、コンテンツ管理、コンテンツから関連する情報に誘導するリファレンスサービスなどの新たなサービスを実現することが可能である。

従来、コンテンツの特徴量を抽出する手法として、類似検索技術やハッシュ関数が用いられてきた。類似検索は『似ている』コンテンツをいかに効率的に探索するか、という点に主眼が注がれ^{1)~4)}、ハッシュ関数は、デジタルデータが『同一である』ことを保証するために用いられている^{5)~7)}。コンテンツに対してサイズ変更やフォーマット変換などのコンテンツの本質的な価値を変えないような処理が行われた場合、コンテン

ツを識別するためには類似検索技術では曖昧性が大きすぎ、ハッシュ関数では厳密すぎるという問題がある。

我々はこれまでに、コンテンツ内局所特徴値の順序関係を利用し、フォーマット変換されたコンテンツを一意に特定するコンテンツ識別手法を提案した⁸⁾。この方法では、フォーマット変換時のノイズに耐性を持つように設計されているが、想定されるノイズが大きくなった場合に識別精度が劣化するという問題があった。

本稿では、提案済みのコンテンツ識別手法について、順序情報を複数用いることでノイズ耐性と識別精度を同時に向上させる方式を提案する。また、実験を通じ、提案手法の有効性を示す。

2. コンテンツ識別技術

デジタルコンテンツ、特に静止画・動画・音声などのメディアデータはそのデータ量の多さから、圧縮・符号化が行われて利用されることが多い。また、作成・配信・閲覧・保存といったコンテンツライフサイクルの各フェーズにおいて、コンテンツ品質・蓄積容量・ネットワーク速度など、それぞれのフェーズや環境に適したフォーマットが存在する。利用者 (エンドユーザ・コンテンツホルダ・配信事業者など) は、コンテンツを最適なフォーマットに変換し利用している。

コンテンツ識別技術とは、このように変換されたコ

コンテンツを、コンテンツ自身の持つ情報を用いて同一であると識別するための技術である。コンテンツを識別することができれば、コンテンツを介して付加情報（メタデータ）へアクセスするサービスが可能となり、著作権管理、コンテンツ管理、関連情報の表示、コンテンツ検索など様々な分野に適用できる。

コンテンツと付加情報を結びつけるための技術として、コンテンツ自身の情報を用いる方法のほかに、任意の情報をコンテンツに重畳する手法がある。この代表的な例として、電子透かし⁹⁾などがある。これらの手法は、コンテンツに対して事前に付加情報を重畳しておき、利用時にその情報を読み出して利用する。しかし、事前処理が必須であり、また、コンテンツ自身の品質が劣化するという短所が存在する。

2.1 関連技術

ここでは、コンテンツ識別に用いることのできる、2種類の特徴量抽出技術、類似検索技術とハッシュ関数を紹介する。

類似検索技術は、コンテンツの持つ特徴を数値化した特徴量を用いて類似コンテンツを検索する。この特徴量を用いれば、同じコンテンツや似ているコンテンツを識別することができる。しかし一般に、類似検索では似ているコンテンツを広く検出しすぎるといった問題がある。

一方、ハッシュ関数は、コンテンツデータからハッシュ値を求めることでコンテンツを一意に特定することができる。1bitでもデータが変更された場合に、ハッシュ値が異なる値になるという性質を持つため、ソフトウェアの偽装防止などの目的で広く使われている。コンテンツに対してフォーマット変換が行われた場合、ハッシュ値は異なる値となり、同一と判定されないという問題がある。

2.2 従来技術

上述のような問題を解決するため、我々はこれまでに、フォーマット変換やサイズ変換への耐性を備えたコンテンツ識別技術について提案を行なった。本技術では、コンテンツを複数のブロックに分割し、ブロック内局所特徴値の大小関係に着目し、ブロック列の順序情報をコンテンツの特徴量としていた。以下に、従来手法の概要を示す。ここでは、対象コンテンツを静止画としている。

(1) ブロック分割

コンテンツを固定数ブロックに分割する。図1-(1)に4×4ブロックに分割した例を示す。

(2) ブロック内局所特徴値の計算

分割した各ブロック中から、そのブロックの特徴値（局所特徴値）を求める。局所特徴値として、平均画素値・画素値の分散・周波数成分など、各

ブロックの特徴となる値を用いることができる。ここでは、局所特徴値としてブロック内平均画素値を用い、YCbCr各色成分の平均画素値を求める。図1-(2)に例を示す。この図では横軸をブロック番号、縦軸を平均画素値としており、図中< #1, 169 > はブロック1の平均画素値が169であることを表している。図には簡単のためY成分のみについて示すが、実際にはCb,Cr成分についても同様の計算を行う。

(3) ブロック列のソート

各色成分について、平均画素値をキーとして昇順ソートを行なう。ソート結果について、ブロック番号のみを取り出し、得られたブロック列 $Block_Y, Block_{Cb}, Block_{Cr}$ を求める。図1-(3)の例では、ブロック列 $Block_Y = [14, 13, 15, 16, 10, 12, \dots, 7, 1, 2]$ が得られている。

(4) 部分ブロック列の選択

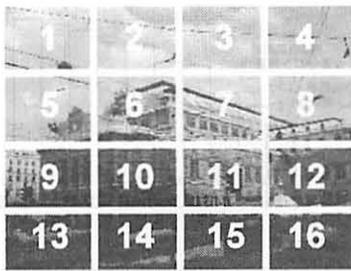
ブロック列から、初期ブロックを選ぶ。次に、選んだブロックの平均画素値がある閾値（抽出間隔）以上離れたブロックを選ぶ。これを繰り返し、各色成分全体に対してブロックを選択する。初期ブロックを最も局所特徴値の小さなブロックとし、抽出間隔を30とした時の例を図1-(4)に示す。まず、初期ブロックとしてブロック14が選ばれる。次に、ブロック14の平均画素値28より抽出間隔以上離れたブロック10（平均画素値70）を選択する。同様に、ブロック4（平均画素値122）、ブロック3（平均画素値153）を選択する。

(5) 部分ブロック列の抽出

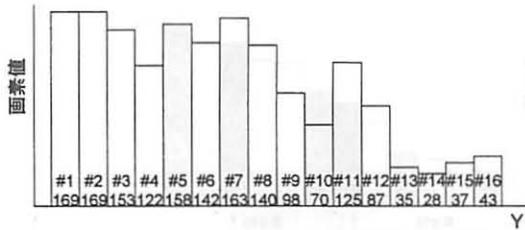
選択した部分ブロック列 $Part_Y, Part_{Cb}, Part_{Cr}$ を出力する。図1の例では、 $Part_Y = [14, 10, 4, 3]$ となる。

従来手法では、2種類の特徴量、ブロック列と部分ブロック列が出力される。1つ目のブロック列は、コンテンツ内ブロックの画素値分布を表す情報である。この情報は全ブロックの並び順を含んでいる。2つ目に出力される特徴量は部分ブロック列であり、前述のブロック列と同じくコンテンツ内ブロックの画素値分布を表しているが、一部のブロックの順序のみを表している。

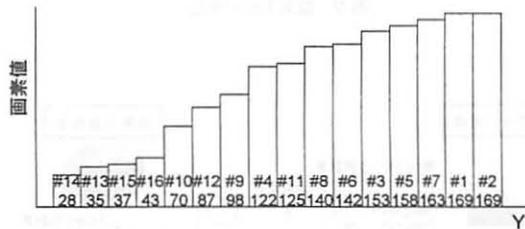
2つのコンテンツA,Bが同一であるかどうかを識別するには、コンテンツAの部分ブロック列の順序とコンテンツBのブロック列の順序を比較する。これらの順序が同じ場合（コンテンツAの部分ブロック列において、ブロック*i*<ブロック*j*が成立するなら、コンテンツBのブロック列においてもブロック*i*<ブロック*j*が成立する場合）、コンテンツAとコンテンツBは同一コンテンツであると言える。



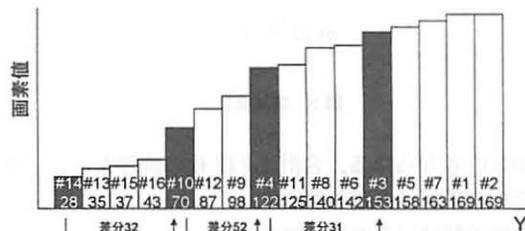
(1) 従来手法～ブロック分割



(2) 従来手法～ブロック内局所特徴値計算



(3) 従来手法～ブロック列ソート



(4) 従来手法～部分ブロック列選択

図 1 従来手法の概要

2.3 従来手法の特徴と問題点

従来手法では、コンテンツを固定数ブロックに分割することでサイズ変換に対する耐性を持たせている。また、部分ブロック列を抽出する際に、部分ブロック中のブロック内局所特徴値が、想定されるノイズの量以上の画素値幅を持つように抽出間隔を設定することで、フォーマット変換時に発生するノイズに対する耐性を持たせることができる。

文献⁸⁾において、従来手法は既存の類似検索手法の1つである MPEG-7 の特徴量抽出手法 (Edge Histogram, Color Layout)¹⁰⁾ に比べ、高い精度でコンテンツの同一性を識別できることがわかっている。

しかし、従来手法では、想定されるノイズが大きい場合に、識別精度が劣化するという問題があった。これは、部分ブロック列の抽出時に、想定されるノイズ分の幅を持たせる必要があるため、ノイズが大きい場合に部分ブロック列の列長が短くなることに起因する。

従来手法のコンテンツ識別可能数は部分ブロック列の列長に依存し、例えば、各色成分の部分ブロック列長が5の場合には、 $(5!)^3 = 1,728,000$ 個のコンテンツしか識別できず、現実のコンテンツ識別に用いることが困難となる。

3. 提案手法

次に、本稿で提案する手法について説明する。従来手法で識別精度が劣化する原因は、部分ブロック列の列長が短くなることであった。しかし、列長を長くするには限界があり、例えば 8bit の色空間では 256 階調の画素値しか持たないため、想定されるノイズが 30 の時には、列長を 9 以上にする事ができない。そこで、列長はそのまま、列の数を増やすことでこれを解決する。つまり、提案手法では同一種別の特徴量を複数利用する。静止画の場合では、同一色成分のブロック列から、 $Party_1, Party_2, \dots, Party_n$ というように複数の異なる部分ブロック列を抽出する。

以下に提案手法の手順を示す。

- (1) ブロック分割
- (2) ブロック内局所特徴値の計算
- (3) ブロック列のソート

ここまでの処理は従来手法と同様である。

- (4) 部分ブロック列初期位置の選択

提案手法では、複数の部分ブロック列を抽出するため、それぞれの部分ブロック列の初期ブロックを選択する。部分ブロック列になるべく重複が生じないように選ぶことが望ましいが、ここでは単純な方法として、初期位置選択のための閾値を設定し、各列の初期値の差がこの値を超えるように初期ブロックを選択していく方法を紹介する。図 2-(1) に、列数 3、初期選択閾値を 10 とした場合の例を示す。この例では、まず列 1 の初期ブロックをブロック 14 (平均画素値 28) とする。列 2 の初期ブロックは、平均画素値が 38 以上であるブロック 16 (平均画素値 43)、同様に列 3 ではブロック 10 (平均画素値 70) となり、初期ブロックとして 14, 16, 10 の 3 つのブロックが選択される。

(5) 部分ブロック列抽出

得られた初期位置を基に、ブロック列の選択と抽出を行う。従来手法と同様の処理を複数のブロック列に対して行うこととなる。図2-(2)では、抽出間隔を50としてブロック抽出を行っている。この例では、まず列1について初期ブロック(14)から平均画素値の差が抽出間隔以上になるブロックを選び、ブロック12(平均画素値87)、ブロック8(平均画素値140)が選ばれる。列2,3についても同様の処理を行い、その結果、Y成分のブロック列として $Part_{Y_1} = [14, 12, 8]$, $Part_{Y_2} = [16, 9, 3]$, $Part_{Y_3} = [10, 4]$ の3つの部分ブロック列が抽出される。

(6) 重複情報削減処理

この処理は必須ではないが、特徴量のデータ量を削減したい場合(携帯端末でデータ量を削減したい場合や通信コストを削減したい場合など)に必要である。抽出された複数の部分ブロック列に重複する情報がある場合、これを削減する。例えば部分ブロック列 $Part_{Y_1}$ が [1,2,3,4] で部分ブロック列 $Part_{Y_2}$ が [2,3,4,5] である場合、[2,3] という順序情報が重複しており、いずれかの部分ブロック列に含まれていれば良い。そこで、 $Part_{Y_1} = [1,2,3,4]$, $Part_{Y_2} = [4,5]$ のように重複している部分の情報を削除することで、識別精度を保ちつつデータ量を削減することができる。

(7) 特徴量情報出力

抽出された部分ブロック列を特徴量として出力する。

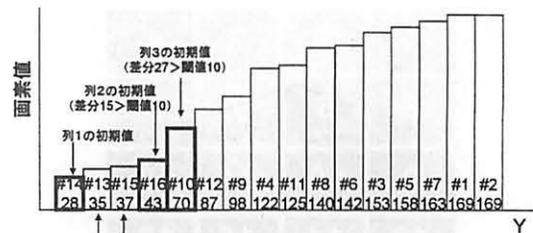
以上のように、提案手法では、同一種別の特徴量から複数の部分ブロック列を抽出する。これによって、大きなノイズが予想される場合でも、列数を増やすことで識別精度の劣化を防ぐことができる。例えば列長5となる場合、各色空間の列数を3にすることで、 $(5!)^{3 \times 3} = 5.2 \times 10^{18}$ 個のコンテンツを識別できる。

4. 実験

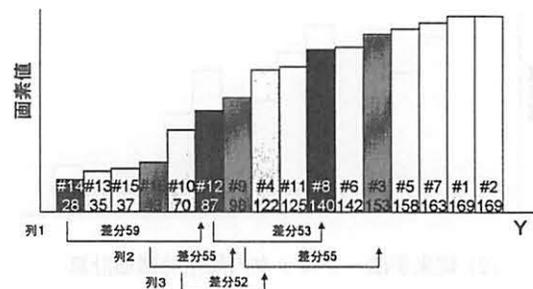
提案手法の有効性を示すために、フォーマット変換済みコンテンツの同一性識別実験を行った。図3に実験の概要を示す。

本実験では、問合せ画像として、表1に示す3つの画像セットを用意し、1枚の画像についてそれぞれ、オリジナル・GIF・JPEG(QF=40)の3種類のフォーマットを作成した。例えば画像セット1(570枚)では、それぞれの画像に対して3フォーマット、合計1,710枚の画像を用意した。

検索対象画像は、画像セット1~3、およびその他の



(1) 初期位置の選択 (初期選択閾値=10)



(2) 部分ブロック抽出 (抽出間隔=50)

図2 提案手法の概要

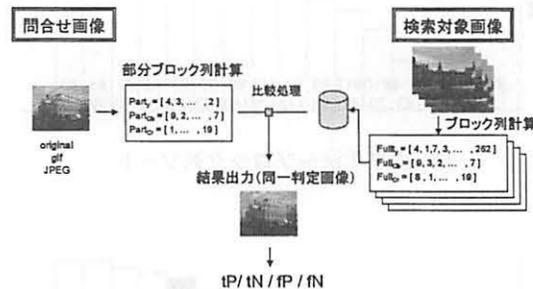


図3 実験概要

画像336枚からなる、合計2,247枚の画像を用い、それぞれのブロック列を求めた。

実験手順は以下の通りである。

- (1) 問合せ画像を1枚選び、 32×32 合計1024のブロックに分割し部分ブロック列を計算する。列数は1(従来手法)、3, 10の3種類とし、抽出間隔を変動させた。
- (2) 得られた部分ブロック列と、検索対象画像のブロック列との比較を行う。
- (3) 同一と識別された画像を出力する。
- (4) 出力結果を、 tP, tN, fP, fN の4種類に分類する(表2)。

これら4つの分類から、識別精度の指標となる適合率・再現率を求めることができる。適合率(Precision)

表 1 画像セット

No.	種別	枚数
1	デジタルカメラ写真	570 枚
2	オブジェクト写真, 周囲は白色	1,244 枚
3	イラスト・書・オブジェクト写真など	97 枚

表 2 識別結果の分類

真陽性 (true Positive, tP): 同一と判定され, 問合せ画像と検索対象画像が同一の画像である場合
偽陽性 (false Positive, fP): 同一と判定され, 問合せ画像と検索対象画像が同一の画像でない場合
真陰性 (true Negative, tN): 同一でないと判定され, 問合せ画像と検索対象画像が同一の画像でない場合
偽陰性 (false Negative, fN): 同一でないとして判定され, 問合せ画像と検索対象画像が同一の画像である場合

は適合していると判定された結果のうち, 正しい結果の割合を表し, $precision = tP/(tP + fP)$ にて計算できる. 再現率 (Recall) は正しい結果のうち, 適合していると判定された割合を表し, $recall = tP/(tP + fN)$ にて計算できる. 適合率と再現率を同時に高めることができれば, コンテンツを間違いなく・漏れなく識別可能となる.

本実験ではパラメータごとに, 識別結果の tP, fP, fN の 3 種類の頻度を求め, 適合率・再現率を計算した. この結果を表 3,4,5 に示す.

5. 考 察

表 3 からわかるように, 従来手法では, 抽出間隔を大きくしていくにつれ偽陽性 (fP) の数が増え, 適合率が悪化していることがわかる. 画像セット 1,2 については小さな抽出間隔で高い適合率と再現率が同時に達成されているため問題ないが, 画像セット 3 について小さな抽出間隔では再現率が非常に低くなっている. これは, 画像セット 1,2 が自然画であるのに対し画像セット 3 がイラストなどの圧縮時にノイズの発生しやすいコンテンツであり, 変換時のノイズによりコンテンツを異なるものと判定していることが原因と考えられる. この場合に抽出間隔を大きくすると, 偽陽性の数が飛躍的に増加し適合率が大幅に低下している.

一方で, 表 4,5 からは, 提案手法では, 抽出間隔を大きくした場合にも偽陽性の増加を抑えることができ, 飛躍的に高い適合率が得られており, 従来手法の問題点が改善されていることがわかる.

画像セット 1,3 における適合率と再現率の関係を図 4,5 に示す. 従来手法と提案手法に加え, MPEG-7 で用いられている Edge Histogram, Color Layout の 2 手法¹⁰⁾ の結果も比較のために図示した. 画像セット 1 (図 4) では, 従来手法と提案手法は理想値 (適合率=1, 再現率=1) に近い値を実現できており, MPEG-7 の各手法と比較して優れた値を示していることがわかる. ま

表 3 実験結果～従来手法

画像セット	抽出間隔	tP	fP	fN	適合率	再現率
1	5	1708	0	2	1	0.999
	10	1709	17	1	0.990	0.999
	15	1710	466	0	0.786	1
	20	1710	1982	0	0.463	1
2	5	3724	0	8	1	0.998
	10	3732	14	0	0.996	1
	15	3732	278	0	0.931	1
	20	3732	2195	0	0.630	1
3	5	129	0	162	1	0.443
	10	189	0	102	1	0.649
	15	249	21	42	0.922	0.856
	20	262	204	28	0.562	0.903

表 4 実験結果～提案手法 (列数 3)

画像セット	抽出間隔	tP	fP	fN	適合率	再現率
1	5	1708	0	2	1	0.999
	10	1709	2	1	0.999	0.999
	15	1710	12	0	0.993	1
	20	1710	137	0	0.926	1
2	5	3724	0	8	1	0.998
	10	3732	0	0	1	1
	15	3732	16	0	0.996	1
	20	3732	197	0	0.950	1
3	5	129	0	162	1	0.443
	10	166	0	125	1	0.570
	15	225	0	66	1	0.773
	20	253	5	38	0.981	0.869
	25	272	31	19	0.898	0.935

表 5 実験結果～提案手法 (列数 10)

画像セット	抽出間隔	tP	fP	fN	適合率	再現率
1	5	1708	0	2	1	0.999
	10	1709	0	1	1	0.999
	15	1710	5	0	0.997	1
	20	1710	19	0	0.989	1
2	5	3724	0	8	1	0.998
	10	3732	0	0	1	1
	15	3732	4	0	0.999	1
	20	3732	83	0	0.978	1
3	5	129	0	162	1	0.443
	10	149	0	142	1	0.512
	15	196	0	95	1	0.674
	20	227	0	64	1	0.780
	25	257	2	34	0.992	0.883
	30	275	9	16	0.968	0.945
	35	284	59	7	0.828	0.976

た, 画像セット 3 (図 5) では, 提案手法 (列数 10) と MPEG-7 Color Layout が良い値を示している.

次に, 画像セット 3 について, 各手法で偽陰性の数が 0 (再現率=1) となる場合における偽陽性の数と適合率を比較した (表 6). この値は, 検出漏れを無くした場合の識別精度を表しており, サービスによっては非常に重要な値となる. また, 提案手法では, 従来手法

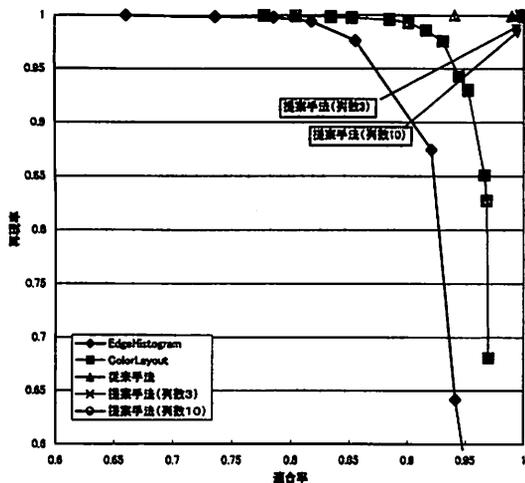


図4 実験結果 (画像セット 1)

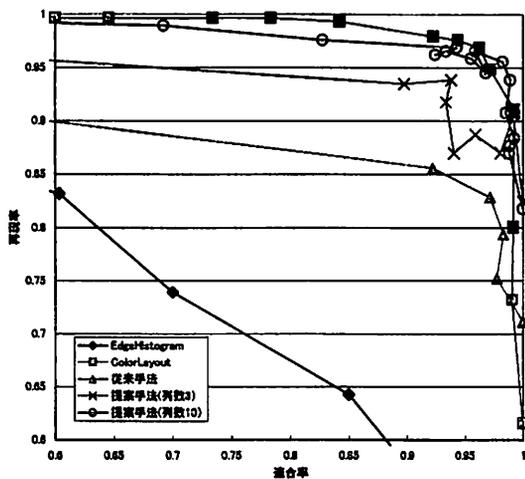


図5 実験結果 (画像セット 3)

表6 画像セット3で $fN = 0$ となる際の fP と適合率

手法	fP	適合率
Edge Histogram	21588	0.013
Color Layout	298	0.494
従来手法	7934	0.035
提案手法 (列数 3)	1219	0.193
提案手法 (列数 10)	790	0.269

の約 1/10 に偽陽性の数を減らすことができているが、Color Layout 手法に比べて多い値であり、さらなる削減が必要である。

6. まとめ

本稿では、コンテンツ識別技術について、複数の順序情報を用いることで、従来手法の課題であったノイズが大きい場合の識別精度劣化を抑える方法について述べた。また、実験を通じ、提案手法ではノイズの大きな画像に対して、適合率の劣化を大幅に抑えること

が可能であることを確認した。

課題としては、識別精度のさらなる向上、識別可能なコンテンツ数の検討、識別結果の信頼度に関する検討などが挙げられる。今後はこれらの課題を解決するとともに、コンテンツ識別を利用したサービスの具現化について検討を行っていく予定である。

参考文献

- 1) 村瀬 洋, V.V. Vinod : “局所色情報を用いた高速物体検索- アクティブ探索法”, 電子情報通信学会論文誌 D-II, Vol. J81-D-II No.9, pp.2035-2042 (1998).
- 2) 串間 和彦, 赤間 浩樹, 紺谷 精一, 木本 靖夫, 山室 雅司: “オブジェクトに基づく高速画像検索システム: ExSight”, 電子情報通信学会論文誌, Vol.40, No.2, pp.732-741 (1999).
- 3) M. Flickner, et al. : “Query by Image and Video Content: The QBIC System”, IEEE Computer, Vol. 28, No. 9, pp.23-32 (1995).
- 4) 呉 君錫, 金子 邦彦, 牧之内 顕文, Sang-Hyun Bae : “Wavelet-SOM に基づいた類似画像検索システムの設計・実装と性能評価”, 情報処理学会論文誌: データベース, Vol. 42, No. SIG1 (TOD8), pp.1-11 (2000).
- 5) The MD5 Message-Digest Algorithm, RFC 1321 (1992).
- 6) US Secure Hash Algorithm 1 (SHA1), RFC 3174 (2001).
- 7) Bruce Schneier: “Applied Cryptography”, John Wiley & Sons Inc (1995)
- 8) 高田 智規, 阿部 剛仁, 川村 春美: “変換耐性を備えたコンテンツ識別手法”, 画像の認識・理解シンポジウム, IS2-55, pp.685-690 (2005)
- 9) 画像電子学会編: “電子透かし技術”, 東京電気大学出版局 (2004)
- 10) MPEG-7 Part 1, ISO/IEC 15938-1:2002, MPEG-7 Part 3 Visual, ISO/IEC 15938-3:2002, MPEG-7 Part 6 Reference software, ISO/IEC 15938-6:2003