I-071

動画検索のための MPEG-2 データ中の特徴量抽出方法の検討

Examination of Method of Extracting Feature in MPEG-2 Data for Movie Retrieval

小友 知己† 伊藤 慶明† 小嶋 和徳† 石亀 昌明†
Tomomi Otomo Yoshiaki Itoh Kazunori Kojima Masaaki Ishigame

1．はじめに

近年，パソコン，HDD レコーダの普及や，記録媒体の大容量化に伴い，長時間，高品質の動画を大量に保存する機会が増加した．ユーザは録画した動画中の特定のシーンのみを見たい場合，早送りで確認しながら探さなければならない．DVD の場合，チャプター分割がなされているため，早送りで探索する程の手間は要しないが，人手で動画内容を確認しながら探索する点は変わらず，簡便かつ迅速な検索とはいえない．このように，動画中の必要部分を簡便かつ迅速に検索する技術が求められている．

DVD や HDD，ブルーレイディスクレコーダは動画を長時間保存するために，MPEG-2 や MPEG-4 AVC 等の形式に圧縮して保存している．動画検索時には対象データの特徴量の抽出が必要であるが，色情報から特徴量を作成する場合，解凍処理を行った上で圧縮された動画特徴量を抽出する必要がある．動画の解凍処理には処理時間を要するため，ユーザは長い検索時間（解凍処理，特徴量抽出処理，照合処理）を待つことになる．特徴量作成までの処理時間を短縮することにより，検索時間の短縮につながると考える．

本研究では，高速な動画検索の実現を目指す．本研究では動画圧縮で多く用いられる MPEG-2 形式を対象に，解凍処理を行わずに抽出可能な特徴量として，I ピクチャ内の輝度直流成分を利用する検索方式を提案する．市販のキャプチャボードでアナログテレビ放送を一旦 MPEG-2 形式で録画した実データを用い，その中に含まれる特定の CM を検索する実験を行い，本方式の有効性を示す．

2．直流成分を用いた動画検索方式

MPEG-2 等の圧縮された動画を検索する場合，通常動画像の特徴を取得するために一度解凍処理を行う必要があり，解凍処理のため検索時に付加的な待ち時間が必要となる．そこで本研究では MPEG 圧縮データ中の I ピクチャを直接読み込み，その輝度直流成分を用いることで，解凍処理の時間の削減を図る．I ピクチャ以外の P,B ピクチャは使用するエンコーダによって値が変動するが，I ピクチャは変動が少ない．このことから，I ピクチャ内の値を用いた．

2.1 MPEG-2 の圧縮方式

MPEG-2 は主に離散コサイン変換 (Discrete Cosine Transform: DCT)とフレーム間の動き予測で動画像圧縮を行っている．以下に MPEG-2 の圧縮方式を概説する．

（1）DCT 処理

DCT 処理は，画像を 16×16 画素で分割し，各々をマクロブロックとして管理する．1 つのマクロブロックは 8×8 画素のブロック 4 個を輝度成分 Y，ブロック 2 個を 2 つの色差成分 Cr, Cb として管理する．輝度成分はブロックをそのまま取得するが，色差成分は縦横方向それぞれに対して 2 画素おきに値を取得し，ブロックを作成する．人間の視覚は色の変化よりも明るさの変化に敏感であるという特性がある．この特性により，色差成分の量が少なくても元の画像に近い表現ができることから，保存される色差成分の量が抑えられている．

（2）GOP 構造

MPEG-2 は I，P，B ピクチャの 3 種類の画像で構成されている．I ピクチャは動き補償を行わず，1 フレーム内のデータ全てを DCT 係数で保存する画像である．P，B ピクチャは他の画像を参照して符号化する．P，B ピクチャは次項に述べるフレーム間予測と動き補償を行い，現画像を圧縮した場合より少ないデータサイズで符号化する．これら 3 種類の画像を GOP (Group Of Picture)という 1 つのグループで管理する．I, B, B, P, B, B, P, B, …という順番で保存され，一般的に 1GOP につき 15 枚，I ピクチャは 1GOP につき 1 枚で構成される．

（3）動き補償フレーム間予測

P ピクチャは，現画像以前の I，P ピクチャを，B ピクチャは，前後の I，P ピクチャを一時的に複合する．現画像内のマクロブロックと参照した画像のマクロブロックを比較し，対応するマクロブロックのズレを動きベクトルで保存する．動画の複合化時には，参照画像からマクロブロックを参照し，動きベクトルを用いて参照マクロブロックの位置をずらすことにより，画像の再現を行う． 図 1 は符号化時の流れである．I ピクチャは，そのまま DCT 処理を行い，エントロピー符号化する．P ピクチャは，入力画像以前に DCT 処理により圧縮された I，P ピクチャを複合し，動き補償フレーム間予測をした後，エントロピー符号化する．B ピクチャは，まず B ピクチャとなる画像をバッファに保存し，入力画像より後に出現する I，P ピクチャが DCT 処理で圧縮されるまで待つ．I，P ピクチャを圧縮したら，バッファに保存した入力画像と圧縮された前後の I，P ピクチャ間で動き補償フレーム間予測を行い，符号化する．

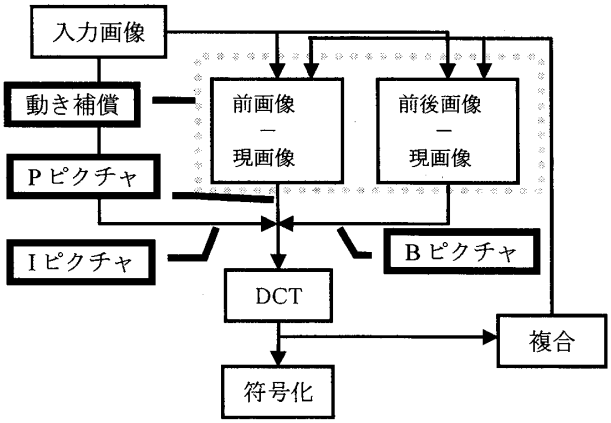


図1 動き補償の流れ

†岩手県立大学ソフトウェア情報学部ソフトウェア情報学研究科

2.2 Ｉピクチャ内輝度直流成分の抽出

本研究では動画像の特徴量として Ｉピクチャ内の輝度直流成分を用いる．圧縮を行っていない画像のブロックの座標を x, y，圧縮を行っていない画像のブロックの画素の値を f(x,y) とした時，DCT 処理による輝度の直流成分 Y の算出方法と 1 ブロック内の画素値の平均 Ave との関係は式 (1)のようになる [1]．

$$Y=\frac{1}{8}\sum_{x=0}^{7}\sum_{y=0}^{7}f(x,y)=8\cdot\frac{1}{64}\sum_{x=0}^{7}\sum_{y=0}^{7}f(x,y)=8Ave \quad (1)$$

式 (1)より，１ブロック内の平均値に比例しており，直流成分は 1 ブロック内全体の 1 つの特徴と考えることができる．P，B ピクチャの動きベクトルを求める処理は，標準として定められていないため，使用するエンコーダによって値が変動する[3]．このため，録画環境が異なった場合でも安定した特徴量とすべく，値の変動が少ない Ｉピクチャ内の直流成分を特徴量とする．Ｉピクチャ内には輝度 Y の他に，Cr，Cb の 2 つの色差成分がある．これらについても MPEG-2 の解凍不要の動画検索用の特徴量として考え，検索の際に利用しその有効性を評価する．直流成分の抽出には，MPEG-2 デコードツール[2]を用いる．動画を複合化する処理の途中で，直流成分をフレーム番号と共に出力し，直流成分がフレーム内のブロック数分出力したフレーム番号をＩピクチャとし，実験に用いる．

2.3 線形照合による類似区間探索

ある動画データと同じ内容の動画データでは，時系列データとしては時間伸縮のない同一と仮定できるため，線形照合による探索を行う．まず，参照動画の特徴量パターンを $ref(l, j)$，$1 \leqq l \leqq L$，$1 \leqq j \leqq J$ とする．l は参照動画に含まれる l 番目のＩピクチャ，L は参照動画中に含まれるＩピクチャの数である．j はＩピクチャ内の j 番目の直流成分を示す．表 1 より，1 フレームのサイズが 720×480，輝度 Y の場合，ブロックのサイズが最少で 8×8 画素となるため，J=5400 (720/8×480/8)となり，色差 Cr, Cb の場合，ブロックサイズが 16×16 画素であるから，J=1350 となる．入力動画の時刻 t における特徴量パターンを $inp(t, j)$とする．入力動画の始端から 1GOP ずつずらしながら参照動画との照合を行う．時刻 t を終端としたときの累積距離 distance(t)は式(2)にて算出する．即ち参照パターンは $ref(1, j)$から $ref(L, j)$まで，入力パターンは $inp(t\text{-}L+1, j)$から $inp(t, j)$までの時間区間での線形照合となる．図 2 にて，入力動画の探索のイメージを示す．

累積距離が事前に設定した閾値より下回り，一定検出窓内で局所最小を示す区間を同一または類似した動画区間と判断する．

図 3 にて，累積距離 distance(t)から，一定窓内の局所最小の箇所を検出するイメージを示す．

$$\mathrm{distance}(t)=\sum_{l=1}^{L}\sum_{j=1}^{J}\left|inp(t-L+l,j)-ref(l,j)\right| \quad (2)$$

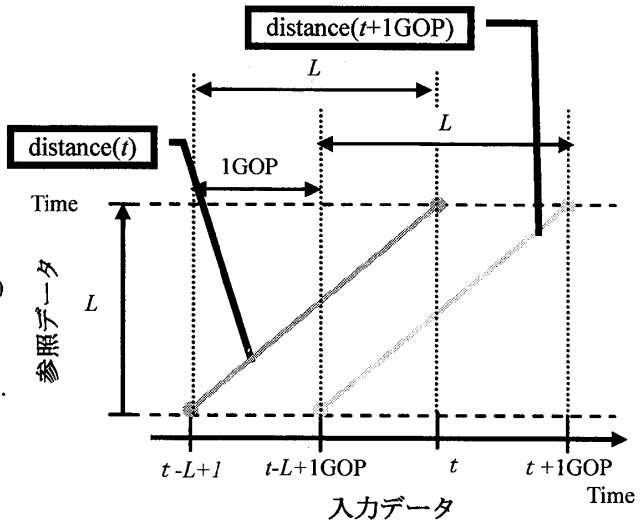
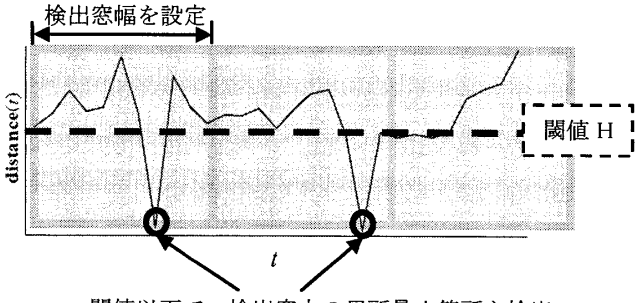


図 2　線形照合による探索



閾値以下で，検出窓内の局所最小箇所を検出

図 3　検出窓設定による局所最少箇所の検出

3. 評価実験

3.1 実験データ

実験用ビデオデータは，SONY のノート PC，VAIO typeA VGN-AS34B モデル内蔵の，MPEG ハードウェアエンコーダボードを用いて，アナログテレビ放送を高画質モードで一旦録画しノート PC に MPEG-2 データとして保存した．高画質モードの仕様は，表 1 に示す．

表 1　高画質モード仕様

録画形式	MPEG2
ビットレート	8Mbps
画像サイズ	720×480
フレームレート	30fps

録画したデータの中から CM を 220 本抽出し，ツール[2]を用いて直流成分を抽出して検出すべき参照データとした．220 本の CM の長さは 4 秒が 4 本，15 秒が 202 本，30 秒が 14 本であった．入力データは上記の参照 CM220 本を繋ぎ合わせたもので，58 分の動画である．このデータを用いた場合の正解数は 542 件である．

3.2 実験条件

評価指標は検索で一般的な適合率 (Precision)・再現率 (Recall)グラフと，調和平均（F 値）を用いる．適合率，再

現率，F 値は式(3)(4)(5)にてそれぞれ算出する．評価時の正解判定は以下の基準で判定する．

(1)正解箇所に時系列的に 75%以上一致している

(2)テロップ以外は同じ動画内容

輝度 Y，色差 Cr,Cb の各特徴量の検索精度を確認するためめために，(Y),(Cr),(Cb)の各特徴量を単体で用いた場合と，(Y, Cr, Cb)，(Y, Cr)，(Y, Cb)，(Cr, Cb)の組み合わせ，検出窓長を 20 秒に固定した場合の評価を行う．

本手法では 3.1 で述べたように閾値 H 以下で累積距離が一定窓内で局所最少の場合出力を行う．この際の窓幅により性能が影響されるため，次の窓幅による性能評価を行う．

与える窓幅は 4 秒，15 秒，20 秒，30 秒，参照動画長を L とし，L，$L+L/4$，$L+L/3$，$L+2L/3$ の時間でそれぞれ評価を行う．累積距離の閾値を最小値（1 つのみ検出）から最大値（全て検出）まで変化させてその時の適合率，再現率で評価する．

$$\text{Precision}=\frac{\text{正しく検出された数}}{\text{全検出数}} \quad (3)$$

$$\text{Recall}=\frac{\text{正しく検出された数}}{\text{全正解数}} \quad (4)$$

$$\text{F値}=\frac{2\times\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}} \quad (5)$$

3.3 特徴量(Y, Cr, Cb)の評価結果と考察

輝度・色差の検索性能の評価を行った．輝度 Y，色差 Cr, Cb の各特徴量を単体で用いた場合(Y),(Cr),(Cb)と，(Y, Cr, Cb)，(Y, Cr)，(Y, Cb)，(Cr, Cb)の組み合わせた場合について，F 値の最大値を表 3 に，適合率・再現率のグラフを図 4 に示す．この際，検出窓長を 20 秒に固定した．

輝度 Y のみを用いた場合，F 値は 97.8%，他の組み合わせでは F 値が 95.2%以下となり，色差 Cr, Cb を用いない方が良いという結果になった．理由として，2.1 の（1）項より，人間の視覚の特性による圧縮処理より，色差成分は抑えられているため，色差成分は異なる動画間でも大きな特徴の差異が出なかったと考える．以降の実験では輝度のみを特徴量として用いる．

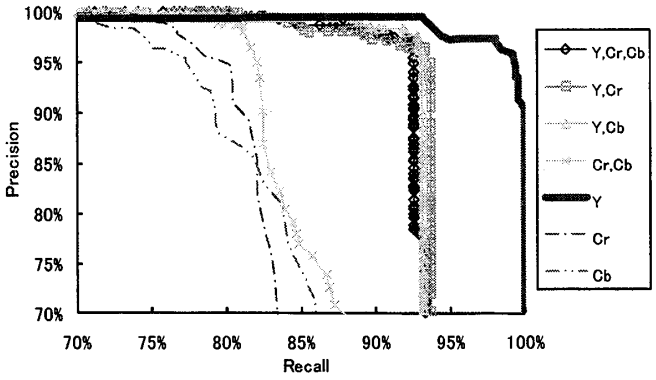


図 4　各特徴量での検索性能

表 3　特徴量組み合わせによる F 値の比較（単位:%）

	Y	Y,Cb	Y,Cr	Y,Cr,Cb	Cr,Cb	Cr	Cb
F 値	97.8	95.2	94.9	94.8	89.1	86.9	85.4

3.4 検出窓長による評価

（1）固定長検出窓による評価

輝度 Y のみを用いて，検出窓長を 4 秒，15 秒，20 秒，30 秒に固定させた場合の適合率，再現率の F 値の最大値を表 4 に，適合率，再現率のグラフを図 5 に示す．表 4 より 20 秒の窓幅で 97.82%と最も良い精度となり，図 5 でも窓長が 20 秒のとき，全体として高い検出精度を示した．以下この理由を考察する．

CM の 220 本のうち，15 秒の CM が 202 本と全体の約 90%を占めていたため，20 秒の窓を用いることにより，高い検索性能になった．窓幅が 4 秒の場合，正解ではない箇所が多く検出されたために，適合率が低くなったと考える．同様に，窓幅が 15 秒の場合も，20 秒の固定窓よりも正解ではない箇所を若干検出したために，適合率が低下したと考える．窓幅が 30 秒の場合，表 4 では 97.80%と窓幅を 15 秒に固定した場合と同じとなったが，図 5 を見ると，再現率が 100%に到達せず，全ての正解箇所を検出できなかった．30 秒の窓長が大きすぎたために，4 秒の CM と 15 秒の CM が連続した場合には全てを検出できず，検出漏れが発生し，精度が低下したと考える．以上より，窓長は参照動画の中で最も多い 15 秒にその長さの 1/3 程度（5 秒程度）を付加した長さである，20 秒で良い精度が得られたといえる．

表 4　各固定の検出窓長の F 値の比較（単位:%）

	4sec	15sec	20sec	30sec
F 値	97.26	97.80	97.82	97.80

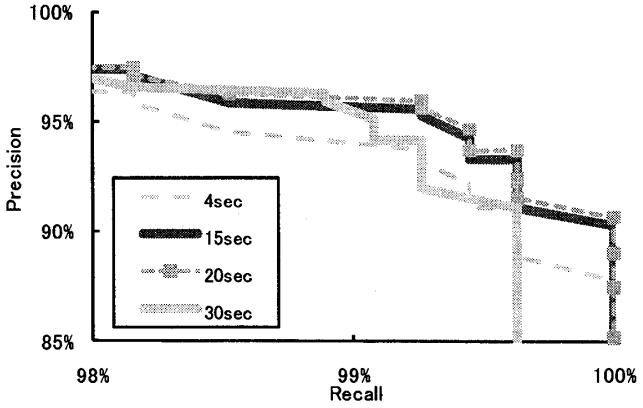


図 5　固定窓長による検索性能

（2）可変長検出窓による評価

3.4 の（1）項の結果に基づき，参照動画長 L として，窓長を $L/4$，$L/3$，$2L/3$ を付加した時間長とした場合の F 値の最大値を表 5 に，適合率・再現率のグラフを図 6 に示す．

表 5　各可変窓長の F 値の比較（単位:%）

	20sec	$L+L/4$	$L+L/3$	$L+2L/3$	L
F 値	97.82	97.71	97.82	97.06	97.80

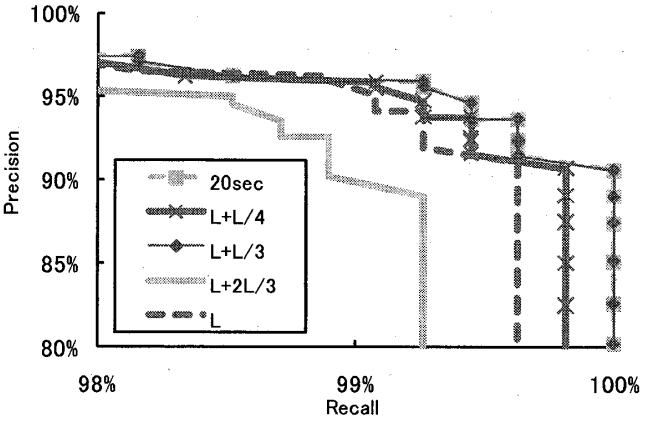


図 6　窓長を参照動画長に合わせた場合の検索性能

図 6 を見ると，窓長を 20 秒に固定した場合と，参照動画長 L に $L/3$ を付加した時間が同じく最も良い精度が得られた．この原因を以下分析する．

参照信号長 L を 4 秒，15 秒，30 秒とし，窓長を $L+L/3$ として，それぞれの信号長毎の検索性能を調査した．その結果を表 6，図 7 に示す．

表 6　各 CM 長の F 値の比較（単位:%）

	CM_4sec	CM_15sec	CM_30sec
F 値	100.0	98.5	100.0

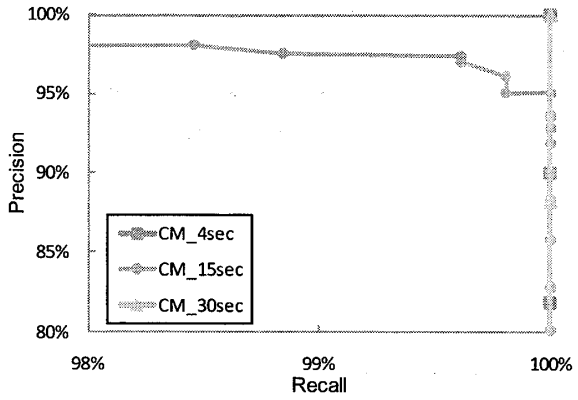


図 7　各 CM 長別の検索性能

4 秒，30 秒の CM では誤検出は発生せず，15 秒の動画は誤検出が見られた．3.1 でも述べたとおり，今回使用した CM のうち，4 秒と 30 秒の CM は少なかったため，誤検出がなく，高い精度が得られたと考える．

次に，L，$L+L/4$，$L+L/3$，$L+2L/3$ の各窓長において，参照信号長 L を 15 秒とした場合の適合率・再現率のグラフを図 8 に示す．図 6 と同様に，15 秒の動画長においても，窓長 $L+L/3$ が最も良い精度を得られた．以上のことから，15 秒の動画において，窓長を $L+L/3$ の長さで検索した場合，高い性能を得られることが分かった．しかし，15 秒の動画長以外での検証は不十分であるため，今後はデータ収集を行い，評価実験を行う必要がある．

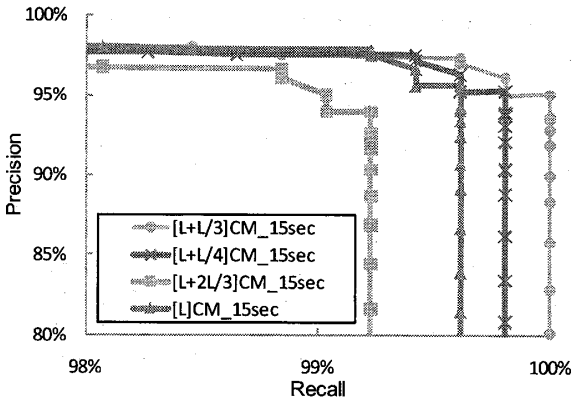


図 8　各可変窓長の CM15 秒の検索性能

4 秒の CM が連続して出現した場合，窓長を 20 秒に固定すると全ての CM を検出できない．そのため，窓長を参照動画長にあわせて図 6 に示すように $L+L/3$ と設定する．

3.5　処理時間

現在の処理時間を個々の処理時間毎に表 7 に示す．当初の目的通り解凍時間と特徴抽出の時間を削減することができた．

一方，MPEG からの読込が総処理時間の 8 割を占めている．現時点では解凍用のツール[2]をそのまま用いているため，読込処理の高速化は実現していないが，I フレームのみを読み込む処理を埋め込むことにより，総処理時間の大幅な短縮が可能であり，本提案方式の有効性をさらに主張できると考える．

表 7　処理時間の比較（単位:秒）

	入力	解凍	検索	総時間
解凍処理あり	540	180	120	840
本研究	540	0	120	660

4. 結論

本研究では MPEG-2 内における I ピクチャ内輝度直流成分を特徴量として動画検索を行う方式を提案した．現画像をそのまま圧縮する I ピクチャを用い，その中の直流成分を特徴量とすることで，解凍処理を必要としない特徴量の抽出を実現した．I ピクチャ内の輝度 Y，色差 Cr, Cb の直流成分で検索実験を行った結果，輝度のみを使用することで高い精度で動画検索が可能であることを確認した．また，窓長を参照動画長 L に，$L/3$ を付加した値で検索実験を行った結果，F 値 97.8%の精度で動画検索が可能であり本手法の有効性を示した．

また，本研究では解凍処理を必要としない特徴量を用いることにより，解凍処理および特徴の削減に成功した．今後は読込処理の高速化により，総処理時間の大幅な短縮を実現し，本方式の有効性を確認したい．今回使用した CM は動画長が 15 秒のものが 90%を占めていたため，15 秒以外の動画で詳細に評価実験を行う予定である．本特徴量が映画だけでなく音楽等の検索にも応用可能であるか，研究を進めていきたい．

参考文献

[1]高橋克直, 寺島信義, 富永英義: 画紋情報を用いた動画像検索方式に関する検討, 電子情報通信学会技術研究報告 画像工学, Vol.98, No.422, pp.1-8 (1998)
[2]Aaron Holtzman eds: libmpeg2, http://libmpeg2.sourceforge.net/
[3]マルチメディア通信研究会（編集）, 藤原洋, 安田浩: ポイント図解式 ブロードバンド+モバイル標準 MPEG 教科書, 株式会社アスキー, (2003)
[4]亀山渉, 花村剛: 改訂版 デジタル放送教科書 上, インプレス R&D, (2006)
[6] 古井貞熙, 酒井 善則: ねっとテクノロジー解体新書 5 画像・音声処理技術, 株式会社電波新聞社, (2004)