

動画像情報分類・抽出新手法に関する一考察

高須賀 義弘[†] 國井 利 泰^{††}

動画像情報について、その変化のみを元にした手法による情報の検索、分類を行なう方法を考案、実装を行なった。本システムは、離散値系ウェーブレット変換および、セル空間による Adjunction space の定義を利用して設計されている。

計算機の処理能力やネットワークの高速化にともない動画像情報を扱う頻度は、今後も増えていくことが予想される。今後増えることが予想される動画像情報の扱いを、より簡素に行うことを目指した。

Methodology of video information retrieval system

YOSHIHIRO TAKASUKA[†] and TOSIYASU L. KUNII^{††}

For video information processing, we have devised an efficient method to process only the incremental information. We have developed a prototype system that demonstrates the power of our method effectively. Our method is based on discrete-wavelet-transformation and adjunction-space-modeling.

Frequency analysis of video information is conducted effectively based on the proposed method, resulting in the improvement of the throughput of computers and the speed of computer networks. We also demonstrate that our method simplifies video information handling.

1. 背 景

いかなるデータであっても、データの中から、必要な情報を探すためには、データを分類し、検索・抽出するなどの作業が必要不可欠である。

インターネットのブロードバンド化、計算機の高性能化・大容量化にともない計算機やネットワーク上で扱われる情報は、従来のテキスト情報から画像情報・音声情報そして動画像情報へと複雑・多様化している。このような扱う情報の変化にともない動画像情報を機械的に分類し、検索・抽出していく手法が、必要になることは容易に想像できる。

動画像情報作成時にインデックスを付けることで、高速かつ確実に検索を行なうことは出来る。この方法は新規に作成される動画像情報に対しては適用できるが、すでにある動画像情報に対応することは困難である。情報を検索・分類する作業を自動化し、利用者は

最低限のパラメータ設定のみで処理を行なえるシステムが必要となると考えられる。

2. 離散値系ウェーブレット変換による多重解像度解析

離散値系ウェーブレット変換は、直交な基底関数を用いた線型変換の一種であり、フーリエ変換と同様に周波数解析などに利用されている。

$$S = WX$$

$$X = W^T S$$

X は入力データ、 W はウェーブレット変換のシステム行列、 S はウェーブレットスペクトラムである。上式に示すように、 $W^{-1} = W^T$ となるようなシステム行列を利用するため、逆変換を容易に行なうことが出来る。また、与えられたデータを、直交するデータの和に変換する機能があり、それは次式のような関係によって表現される。

$$S = \sum_{i=1}^n S_i$$

$$X = \sum_{i=1}^n X_i (X_i = W^T S_i)$$

通常、 X_1 は最も低い周波数情報からなる情報であ

[†] 法政大学大学院工学研究科電気工学専攻 IT プロフェッショナルコース

IT Professional Course, Graduate School, Hosei University

^{††} 法政大学大学院情報科学研究科
Hosei University

り、 X_n は最も高い周波数からなる情報である。 $X_2 \sim X_{n-1}$ はその中間の領域からなる情報である。

画像情報に対して、離散値系ウェーブレット変換を適用する場合、高い周波数とは、隣りあうピクセル間の差が激しい部分であり、低い周波数成分とはより小さい変化量を持つ部分を意味する。すなわち隣隔成分が高い周波数として抽出され、面成分が低い周波数成分として抽出される。図 1 は、画像をドビッシーの 2 次基底によって、多重解像度解析を行なった例である。この結果からもあきらかなように高い周波数で構成される高レベルでは、隣隔情報が残っており、逆に平面的な色情報の差があまり表われていない。逆に低い周波数で構成される低レベルでは、隣隔情報がほとんど残っていないが、モザイク状に平均的な色情報が表われている。小さな変化では、隣隔情報は変化するが、面となる平均的な情報は変化しにくい。逆に大きな変化では、面となる平均的な情報も変化する。したがって、低レベルの領域のみを比較すれば、面全体が動くような大きな変化の比較が行なわれ、高レベルの領域のみを比較すれば、些細は変化を比較することが出来る。

多重解像度解析を行なわない元の画像同士の比較では、小さな変化も大きな変化も同様の重みがあり、それらを区別するのは困難である。しかし多重解像度解析を行なった画像同士で比較を行なうことにより、大きな変化の中に隠されてしまう小さな変化だけを取り出した比較や、些細な変化を無視し、大きな変化だけを取り出すことが出来る。

3. セル空間によるモデル化

ある動画画像情報と、別の動画画像情報との関係をセル空間上で表現すると図 2 に示すように考えることが出来る。 Y_0 は、検索対象やグルーピングの中心となる動画画像情報 Y の部分集合である。関数 f は検索する内容と、対象とを関連づけるための接着関数である。等価関数 (商関数) g および、Adjunction space Y_f を

$$\begin{aligned} g: X \sqcup Y &\rightarrow X \sqcup_f Y \\ Y_f &= X \sqcup_f Y \\ &= X \sqcup Y / \sim \\ &= X \sqcup Y / (x \sim f(y) | \forall y \in Y_0) \end{aligned}$$

と書くことが出来る。 y は、 Y_0 上の任意の成分であり、 x は y と対応可能な X_i 上の任意の成分である。

すなわち、検索する内容 Y の持つ特徴量である Y_0 によって、 X の中にある X_i と対応づけられるならば、検索結果として X_i を取り出すことが出来る。 X_i と同様に Y_0 によって対応づけが可能な動画画像情報が

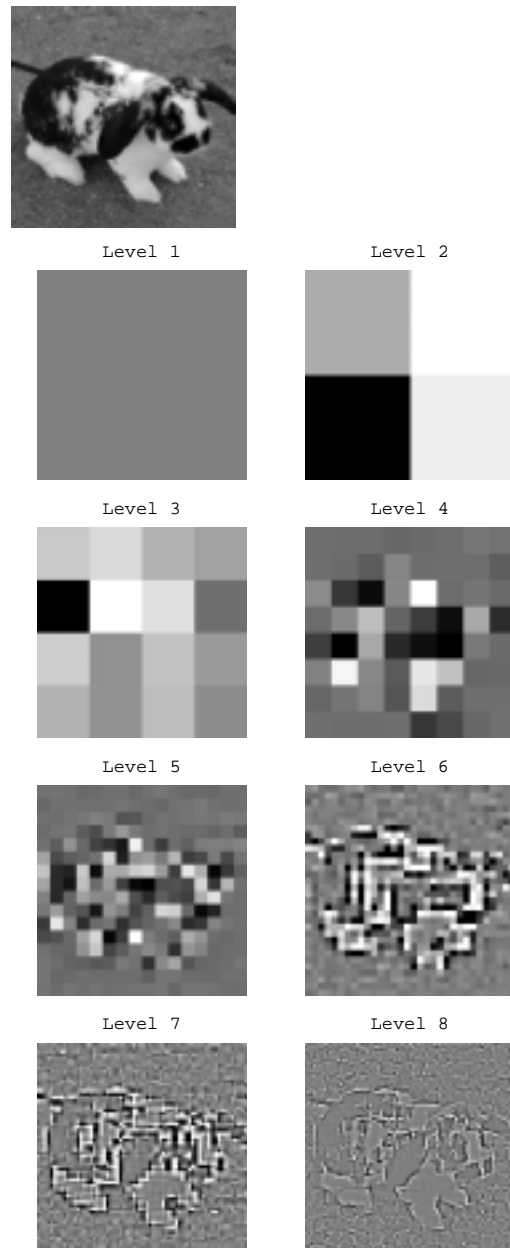


図 1 多重解像度解析による直交分解

あるならば、それらを同種類の画像として同じグループに属すとみなす事が出来る。また Y_0 の条件である次数や要素が異なる Y の部分集合 Y_k 、 Y_l を考えた場合、 $f(Y_k)$ 、 $f(Y_l)$ によって対応づけ可能な X_k 、 X_l 等が存在するならば、それらを別のグループであるとし、グループ分けを行なうことが出来る。

本システムにおける、 Y および、 X_n はウェーブレット変換によって求められた多重解像度解析の各レ

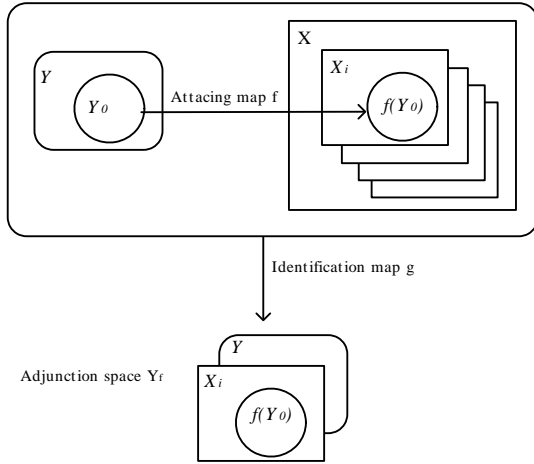


図 2 Adjunction space による関係

- Y : 検索内容の画像情報
- X : 検索対象の動画画像情報全体の集合
- X_i : 検索内容の動画画像情報と対応付け可能な検索対象の動画画像情報の部分 ($X_i \in X$)
- $f : Y_0 \rightarrow X$
- $g : X \sqcup Y \rightarrow X \sqcup_f Y$

ベルが要素となる。したがって次元数はレベル数と同じである。レベル数を z とすると、

Y : 検索内容の画像情報 = $B^{z_{sample} Y}$

X_n : 検索対象の動画画像 = $B^{z_{target} X}$

また、 Y_0 によって対応づけられる X_i は、 Y_0 の次元数を z' とすると

Y_0 : 検索内容の画像情報 = $B^{z'_{sample} Y}$

X_i : 検索対象の動画画像 = $B^{z_{target} X}$

$f(Y_0) : X_i$ 内で Y_0 と対応可能な要素の集合 = $B^{z'_{target} X}$

と書ける。

4. システムの実装

計算機上で本稿の方法論を用いたシステムを実装を行なった。動画画像情報を入力し、各フレームを切り出す処理や、結果の表示方法については、本稿の主旨ではないので詳細は省略する。

4.1 動画画像情報 Y および、 X の生成

動画画像情報 Y は動画画像情報から切り出された複数枚のフレームを元に作成される。 X は Y と同様の手順によって生成された X_n の集合である。本システムでは、何枚のフレームを、一つのシーンとするかは任意に変更可能である。 Y は以下の手順によって生成される。

(1) 各フレームのウェーブレット変換による多重解

像度解析

各フレームの画像に対して前述のウェーブレット変換によって多重解像度解析を行ない、この結果を保持する。以下、多重解像度解析を行なった結果を R とし、 n フレーム目のレベル m は $R_{m,n}$ と表記する。

(2) R より、 Y' を生成する。

$R_{m,x}$ と $R_{m,x+1}$ との相関係数 $Y'_{m,x}$ を計算する。 Y' は計算された $Y'_{m,x}$ の配列である。これにより、保持する情報は、各フレーム間での変化の大きさだけとなる。

(3) Y' から Y を生成する。 Y' はシーンの開始位置によって大きく異なる結果を導くおそれがある。そのため、本システム上では、生成された Y'_m についてフーリエスペクトラムを求め、その結果を Y_m として処理を行なう。フーリエスペクトラムは、ウェーブレットスペクトラムと異なり位相情報が失われ、周波数領域が残る。従って、シーン内のどの位置から始まった変化であるかという情報を取り除き、シーン内で起こった変化の内容だけを取り出すことが出来る。これにより、シーンの開始位置ずれを補正することが出来る。

検索する動画画像全体を Y と同じフレーム数単位で区切り、 Y と同様の手順により X_n を生成する。 X_n 全体を集め X とする。

4.2 Attaching map f の実装

前述のとおり、本システム上では Y の次元は多重解像度解析のレベル数と同じである。 Y_0 については、 Y より任意の要素を任意の量、選択できる。

Y_m に対して $f(Y_0)$ によって対応づけられる可能性がある X_i の要素は、 X_{im} である。本システムでは、 X_{im} と、 Y_m との相関係数を求めることにより対応づけが可能かどうか判定する。つまり X_{im} と、 Y_m とが一定以上の相関を持つことにより似た変化量であるかどうか判定する。本システムでは閾値を任意に設定し、その閾値以上であれば対応づけが出来るものとしている。本システムでは、この判定により接着関数 $f : Y_0 \rightarrow X$ を実装している。そしてこの結果により等価関数 (商関数) g となる抽出結果を導く。

本システムでは、入力として与えられている X 全てについて確認を行ない全ての結果を出力する。

Y を構成するフレーム数が少ない場合、FFT による実際の計算では、位相のずれによる影響が大きくなってしまふ。本システムではフレーム数が十分に少ない場合などのために、任意に $Y = Y'$ とする事も可能にしている

5. 今後の課題

本システムでは、各フレームの情報を離散値系ウェーブレット変換による多重解像度解析により直交分解を行なっている。しかし、この分離は必ずしも人間が意図する分類方法、たとえば「手を振っている」「ボールを蹴っている」などのような情報とは合致しない。人間が意図する抽象的な動作としての分類方法と、自動化のために行なわれる数学的な直交分解とを合致させることは、一般に困難である。現時点で両者を合致するためには、大量のサンプルデータによる統計的な調査が必要になり、膨大な実験が必要である。両者を合致させるための一般的な解決策を用意することが必要になる。

また、本システムでは、単純な計算ではあるが処理量が非常に多く、計算機のメモリ使用も膨大である。そのため高速な計算機を使用しても、演算結果が得られるまで長時間が必要になる。実用的に利用するには、計算量とメモリ使用量を抑えるための最適化が不可欠である。

6. ま と め

本稿では、多重解像度解析によって分割された画像を取り扱うことで、全体の動きではなく微細な動きのみを識別することが可能である事を示し、セル空間による動画像情報のマッチングを記述する方法について示した。また、これらの方法論を計算機上で実装する具体例を示した。

本稿で示した多重解像度解析し、Adjunction Spaceを定義するという方法論を用いることで表面的には微細で他の情報に隠されてしまう部分を抽出し比較することが出来る。これは動画像情報にとらわれず多様な情報の比較・検索、分類に用いることが可能な方法論であり、多くの応用が考えられる。

参 考 文 献

- 1) Toshiyasu L. Kunii: A Memo on a Cellular Model for Cyberworlds Design (1999-2002).
- 2) 斎藤兆古: Mathematica によるウェーブレット変換, 浅倉書店 (1996).