



MPEG映像からの階層的クラスタリングによる時空間 インデキシング

堀江, 新
上原, 邦昭

(Citation)

神戸大学都市安全研究センター研究報告, 9:237-246

(Issue Date)

2005-03

(Resource Type)

departmental bulletin paper

(Version)

Version of Record

(JaLCD0I)

<https://doi.org/10.24546/00422539>

(URL)

<https://hdl.handle.net/20.500.14094/00422539>



MPEG 映像からの階層的クラスタリングによる 時空間インデキシング

Spatio-Temporal Indexing of MPEG Video Based on Hierarchical Clustering

堀江 新¹⁾

Arata Horie

上原 邦昭²⁾

Kuniaki Uehara

概要：近年、GPS（Global Positioning System）などの位置測定デバイスによって得られる空間上の軌跡から、ユーザの行動を分析しようとする研究が進んでいる。また、移動軌跡は多次元の時空間データであるため、その大量の情報から必要な情報を管理し、検索するための時空間データベース（Spatio-temporal database）の研究が進められている。しかしながら、誰もが GPS 端末を持っているわけではなく、すべての軌跡データを取得するのは困難である。そのため、映像の代用により時空間データを取得する試みがある。本研究では、MPEG 映像から抽出可能な動きベクトルを用いて、移動物体と軌跡の抽出を行う。また、時間を制約条件として階層的クラスタリングを行い、時間関係を保存したまま、時空間データを木構造によってインデキシングしている。さらに、Allen の時間関係を空間に拡張し、枝刈りを含めた木の探索を行い、高速かつ正確な検索を実現している。

キーワード：MPEG 映像、動きベクトル、時空間インデキシング

1. はじめに

近年、GPS（Global Positioning System）などの位置測定デバイスによって得られる、ユーザの空間内の移動軌跡から、ユーザの行動を抽出しようとする研究が進められている。図-1はその概略である。たとえば、コンビニエンスストア内を移動する客の移動軌跡データを分析すれば、効率的な店内のレイアウトのための情報が得られる。また、車の移動軌跡データを追えば、時刻ごとの交通量や交通渋滞の予測が可能になる。これらの分析はデータマイニングの分野として興味深い内容である。

従来から、時空間内での移動軌跡といった多次元、かつ大量のデータの中から必要な情報を管理、検索するための時空間データベース（Spatio-temporal database）の研究が進められている。インデキシングされた時空間データから、ユーザは任意時刻でのオブジェクトの位置や、特定の領域を移動する時刻といった「空間」、「時間」の範囲検索が可能で、データの分析に有効に利用できる。しかしながら、「時間」は一方向にしか進むことがなく、取り得る値も無限に増えるため、インデキシ

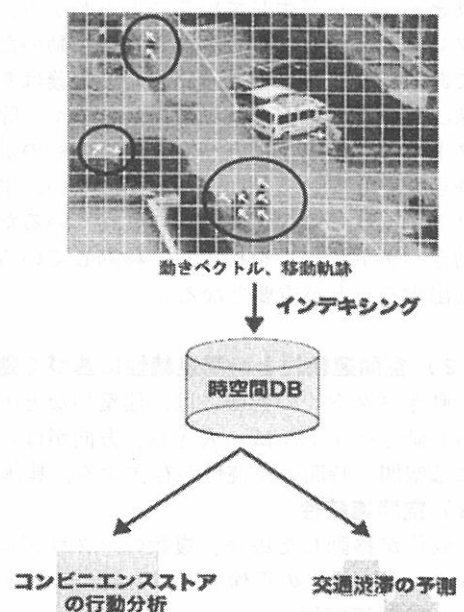


図-1 時空間インデキシングとその応用例

ングの際に時間と空間を統一的に扱うことは困難とされている。この問題に対する解決策として、空間に対して木構造を作り、空間検索を行ったのち、時間情報を保持した時間木による選択処理を行う手法や、重要と考えられる時刻のデータを保持する手法などが提案されているが、空間、時間を別にして検索をしているために、検索回数が増えること、「重要性」といった人間の主観を介したインデキシングになることなどの欠点がある。

これらの問題点を解決するために、映像形式のひとつである MPEG (Moving Picture Experts Group) の圧縮の際に利用する動きベクトルを用いて、GPS 端末を利用することなく、映像から自動的に移動オブジェクトを抽出する手法を検討する。MPEG は映像の圧縮形式のひとつで、動き補償 (Motion Compensation : MC) と離散コサイン変換 (Discrete Cosine Transform : DCT) を併用したハイブリッド符号化方式である¹⁾。また、時間情報を制約とした、空間の位置と動きベクトルの向きの類似度からの階層的クラスタリングを提案する。さらに、この結果からボトムアップ的に二分木を構成することにより、次元を維持したまま時間空間データを管理する手法を導入する。結果として、時間をキーに移動軌跡の形状や空間位置を含めた、高速かつ正確な検索が可能になるという特徴がある。

2. MPEG 映像からの動領域抽出とインデキシング

(1) 動き補償と動きベクトル

ビデオカメラなどから撮影された、オリジナルの映像は情報量が非常に大きい。これを手軽に伝送・保存するためには、画像情報圧縮を行って情報量を減らす必要がある。一方、映像は単位時間当たりいくつものフレームで構成されている。映像をフレームごとの静止画と見たとき、現在のフレームと直前のフレームの画像は非常に似ており、時間的な変化は限られていることが多い。つまり、符号化の終了した画像を次のフレームの予測画像とすることができ、予測誤差を求めれば時間的な冗長性を減らすことができる。このような時間的な冗長性を除去して画像圧縮を行うのが、フレーム間予測符号化と呼ばれる手法である。

また、映像における動きの性質を利用すれば、さらに効率のよい符号化ができる。たとえば、被写体が静止している場合、フレーム間予測を行えば差分はほぼ 0 になる。一方で、被写体がある程度動きを伴う場合、被写体の含まれる画像の一部を切り出し、動き量だけずらしたフレームを次のフレームの画像として予測することができる。これを動き補償と呼ぶ。動き補償の際に、ずらした方向と距離を用いて表されるベクトルを動きベクトル (Motion Vector: MV) という。

図-2 の(a)、(b)は、フレーム f_n 、 f_{n-1} における移動物体とマクロブロック(1,1)の動きベクトルを表している。フレーム f_n のマクロブロック(1,1)とフレーム f_{n-1} 中の 16×16 ブロックとの、画素値のずれを最小にする場所を探索し、マクロブロックの場所とのずれを動きベクトルとして求める事を示している。図中の(c)は、全マクロブロックに対して動きベクトルを抽出した結果である。移動したオブジェクトの周囲に右下方向の動きベクトルが生じていることがわかる。つまり、前フレームに対して移動物体が左上に動いたということである。ただし、動きベクトルの定義はあくまで「映像圧縮の際、時間の冗長度を最も多く、除去可能なベクトル」である。動きベクトルそのものが移動物体の動きを表しているということではない。また、動きベクトルには多くのノイズも含まれているため、必要な動きベクトルだけを抽出し、移動している物体のみを抽出することが重要となる。

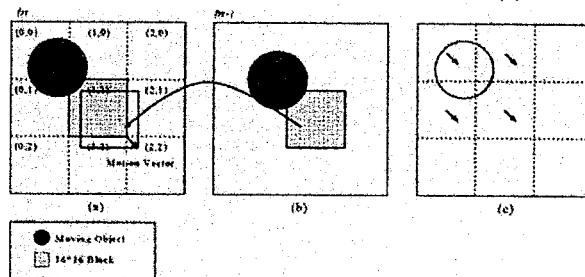


図-2 動きベクトルの例

(2) 空間連続性と時間連続性に基づく動領域抽出

動きベクトルの情報には、建築物などの影から発生する、細かなノイズの動きベクトルも含まれる。これらの動きベクトルは、大きさ、方向がばらばらで、連続性はない。しかしながら、移動物体の動きベクトルには空間、時間の連続性が存在する。具体的には以下の性質がある。

a) 空間連続性

物体が移動した場合、複数のマクロブロックにわたって移動が生じることが多く、動きベクトルが生じたマクロブロックの近傍でも似た向きの動きベクトルが発生する。

b) 時間連続性

微小時間で見た場合、物体の移動軌跡は直線で近似でき、動きベクトルの向きはフレーム間で連続した向

きを持っている。

図-3は、これらの連続性を図で示したものである。すなわち、フレーム内で空間連続性のある、似た方向の動きベクトルを持つマクロブロック同士、およびフレーム間で時間連続性のあるマクロブロック同士を結合している。これらの結合されたマクロブロックは、動領域として認識される。逆に、空間連続性がない動きベクトルや、時間連続性のない離散的な動きベクトルを含むマクロブロックは、動領域として認識されていない。したがって、単独の動きベクトルを分析するのではなく、物体の移動の空間連続性と時間連続性から結合した複数の動きベクトルを分析すれば、精度の高い動領域の抽出が可能となる。なお、カメラワークが固定された映像の連続性を利用して、マクロブロックを結合し移動物体のみを抽出している研究もある²⁾

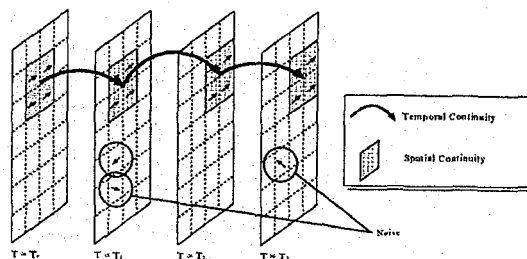


図-3 空間連続性と時間連続性

(3) R-treeによる空間インデキシングと検索

時空間インデキシングは、空間インデキシングに時間情報を加えたインデキシング手法である。まず、空間データベースにおけるインデキシング手法 R-tree³⁾について

説明する。R-treeはAntonin Guttmanによって提唱された、B木を二次元に拡張した空間インデキシングの手法である。葉節点にあたる空間オブジェクトは最小の長方形(MBR: Minimum Bounding Rectangle)で表現され、非葉節点を自分の子節点内のすべてのMBRで表現して、均衡のとれたm分木を構成している。

図-4、図-5にm=3のR-treeの例を示す。図-4がm分木であり、図-5が空間上でのMBRの配置の様子である。非葉節点のMBRの構成には、MBRの面積の総和が最小となるように構成する方法と、MBR同士の重なる部分の面積が最小になるように構成する方法があり、いずれかを用いてMBRが構成される。

R-Treeは空間オブジェクトを木状にして管理する。このため、木をたどりながら検索されるため高速な検索が可能である。また、R-Treeの最大の特徴として、範囲検索を可能にしていることがあげられる。多次元空間においては、検索のためのクエリが範囲を持つことが多い。

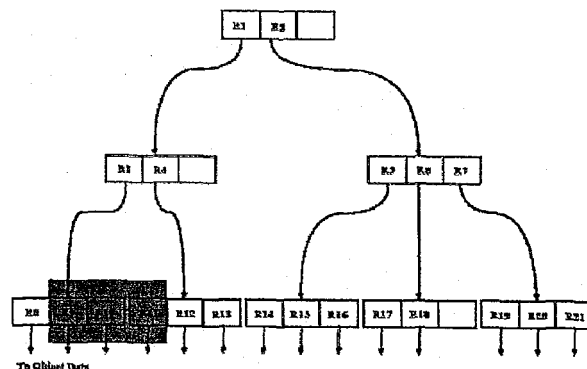


図-4 m=3のときのR-Tree

R-treeからの範囲検索の例を図-5に示す。図中の網掛け部分がクエリである。クエリが渡されると、まず、R1→R3とたどり、R9、R10を出力する。さらにR1→R4とたどり、R11を出力する。このように、R-treeは空間オブジェクトのような範囲のある多次元データのインデキシングに高い効果を発揮する。しかしながら、子節点の数mを事前に指定しておく必要がある。そのため、mの値によっては構成される木に偏りができ、検索効率が悪くなるという欠点がある。

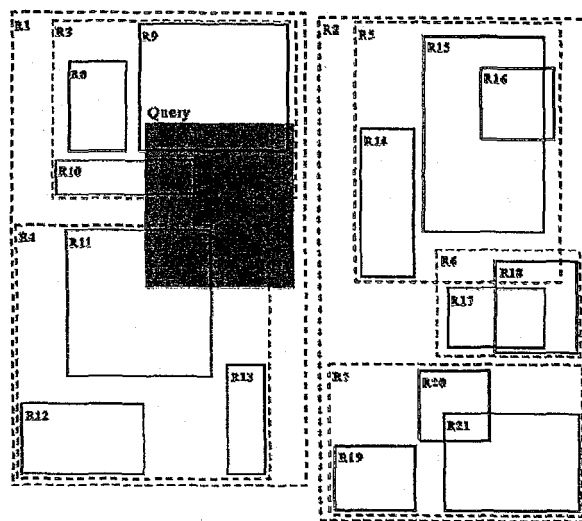


図-5 範囲検索の一例

3. 動領域の時空間インデキシング手法

動領域は空間情報と時間情報を持つため、R-treeをそのまま用いることはできない。本章では、R-treeの時空間への拡張について考え、時間制約による階層的クラスタリングから時空間インデキシングを行い、時空間データに対する効果的な検索手法を提案する。

(1) 時間情報を含めた時空間インデキシング

空間オブジェクトに対しては R-tree が効果的である。しかし、R-tree を時空間へと拡張した場合、空間の情報に加え、時間の情報も含まれるため、R-tree のインデキシングはうまく機能しない。

a) 時間情報と空間情報

映像中での時間情報には以下のような特徴がある。

- ・値が減少することではなく、常に一方向に増加し続ける。
- ・映像の空間は有限であるのに対して、時間は映像の長さに応じて無限に増加する。
- ・空間データと時間データの単位は異なる。

これらの特徴のため、「位置」、「軌跡」といった映像空間内の情報と「時間」という情報を同等に扱うことは困難とされている⁴⁾。さらに、検索が時間をキーとして利用される以上、時間の前後関係を無視することはできず、時間の関係を保存したままデータ構造を管理する必要がある。この解決策として、空間のみの検索を行った後に、時間による選択処理を行う手法、重要と考えられる時刻のデータを一次領域に保持しておいてから構造化する手法などが提案されている。しかしながら、空間と時間を別々に検索するために、計算量が増えてしまうという欠点がある。また、R-tree によるインデキシングでは、MBR の面積が最小になるように、もしくは MBR の重なる面積が最小になるように構成していた。R-Tree を時空間に適用する際には、時間と空間からなる直方体の体積が最小になるように拡張すればよいと考えられるが、空間情報と時間情報を体積として単純に掛け合わせることはできない。

b) 時間制約つき階層的クラスタリング

インデキシングの研究として、ドキュメント番組から内容記述のためのキーワードを用いた階層構造を作る手法がある⁵⁾。この研究では、手動で入力した、シーンの内容を記述するキーワードの時間存在率をキーとして、各シーンを階層化し、シーン記述モデルを作成している。また、隣接した映像の内容は相関が高いという相関性を利用して、時間的に隣接したシーンのみをクラスタ分析している。つまり、時間関係を保存したまま、階層化を行っている。本研究では、このような考え方を時空間データに応用し、時間情報を「隣り合うデータ集合同士としか結合しない」という制約条件と考えて、時間の前後関係を保存したまま、かつ次元を落とすことなく階層的クラスタリングを行う手法を提案する。

階層的クラスタリングに先立ち、初期のデータ集合を決定する必要がある。空間インデキシングでは、空間内のオブジェクトの位置を示すために MBR が効果的であった。このため、MBR に時間情報を加えて、時空間に拡張した直方体 MBC (Minimum Bounding Cube) を定義している。MBC は空間位置、方向の連続した一定のデータの集合から構成される。ここで動きベクトルの連続性の指標となる類似性は、内積を用いた式で判別している。

$$\frac{MV_n \cdot MV_{n-1}}{\|MV_n\| \|MV_{n-1}\|} > Threshold \quad (1)$$

MV_n はフレーム n の動領域の平均動きベクトル、 MV_{n-1} はフレーム $n-1$ の動領域の平均動きベクトルを表している。式 (1) の左辺は 2 つのベクトルのなす角度の余弦を表している。一定の閾値以上を持つベクトル、つまり互いの動きベクトルの角度差が少ない、方向の似ている領域同士を結合すれば、時間連続性を持つデータとして MBC にまとめることができる。図-6 に時空間上でのデータと MBC を示す。

構成された MBC には以下の情報が含まれている。

- ・ x 座標の区間 $X_{Range} = [X_{Start}, X_{End}]$
- ・ y 座標の区間 $Y_{Range} = [Y_{Start}, Y_{End}]$
- ・ 存続時間 $T_{Range} = [T_{Start}, T_{End}]$
- ・ MBC 内の平均動きベクトル MV

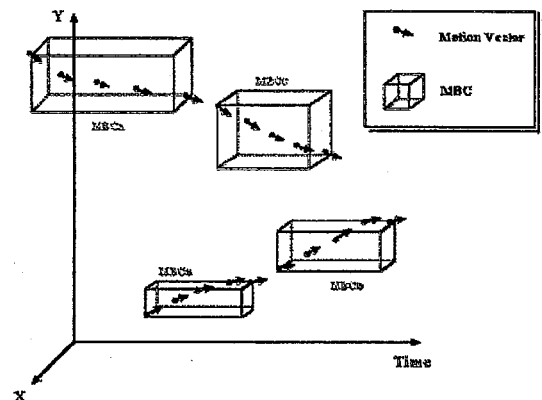


図-6 動きベクトルと構成される MBC

次に、これらの MBC をクラスタリングするための非類似度を定義する。空間連続性と時間連続性から、同一移動物体が構成する MBC の位置と、MBC 内の動きベクトルの向きは似ている。たとえば、図-6 において、 MBC_A と MBC_C は同一物体から発生する MBC であり、MBC の位置も MBC が持つ動きベクトルの向

きも似ている。このため、隣り合う MBC_i 、 MBC_j ($i < j$) の非類似度 $Dissimilarity(i, j)$ を MBC の位置、動きベクトルのなす角度からの標準化ユークリッド距離を用いて以下の式 (2) で定義している。

$$Dissimilarity(i, j) = \frac{(X_{jEnd} - X_{iStart})^2}{\sigma_{X_{Start} + X_{End}}^2} + \frac{(Y_{jEnd} - Y_{iStart})^2}{\sigma_{Y_{Start} + Y_{End}}^2} + \frac{(Rad_i - Rad_j)^2}{\sigma_{Rad}^2} \quad (2)$$

ここで、 Rad は動きベクトルのなす角度で式 (3) より定義される。とりうる値は $[0, \pi]$ である。 σ^2 はそれぞれの変数の分散を表している。

$$Rad = \cos^{-1} \left(\frac{MV_x}{\sqrt{MV_x^2 + MV_y^2}} \right) \quad (3)$$

標準化ユークリッド距離は、各次元の値を分散 1 の区間に標準化するため、単位の影響を受けないという特徴がある。非類似度の値は MBC 同士の空間距離が近いほど、そして MBC 内の軌跡が似ているほど小さい値を示す。したがって、非類似度を用いて階層的クラスタリングを行っている。

アルゴリズム 1

1. 初期クラスタの数を N とする。隣り合う MBC の組 $N-1$ 個に対して $Dissimilarity(I, j)$ を計算する。
2. 最小の $Dissimilarity$ を持つ MBC の組を結合して新たな MBC とする。
3. MBC が 1 個になるまで 1~2 を繰り返す。

(2) Allen の時間関係を用いた階層化

マルチメディアデータ間の時間関係を表す基本となる Allen の時間関係⁶⁾ を用いて、 MBC の時間関係を記述し、「時間関係の隣り合った MBC 」を定義している。Allen の時間関係は 13 種類あり、2つの存続時間を持ったデータ間の時間関係を表している。なお、Allen の時間関係のうち、equal 以外の 6 組はそれぞれ対となる関係である。そこで、 MBC 同士の関係を分かりやすく記述するため、以下の 3 種類の時間関係にまとめている。

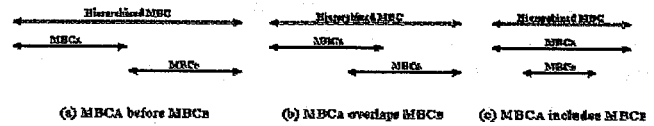


図-7 MBC の 3 つの時間関係

before, meets $\Rightarrow$ *before*

overlaps $\Rightarrow$ *overlaps*

equal, during, starts, finishes $\Rightarrow$ *includes*

図-7 に 3 種類の時間関係で記述された 2 つの MBC を示す。この 3 つの時間関係を導入するため、アルゴリズム 1 を以下のように変更する。

アルゴリズム 2

1. 初期クラスタの数を N とする。3 種類の時間関係 *before*、*overlaps*、*includes* について、それぞれ該当する MBC があれば $Dissimilarity(I, j)$ を求める。
2. もっとも小さな $Dissimilarity$ を持つ MBC の組を結合して新たな MBC とする。
3. MBC が 1 個になるまで 1~2 を繰り返す。

これにより、Allen の時間関係を考慮した階層化が実現される。時間関係を保存したまま、連続性のある MBC が構成される。図-8 (a) は、アルゴリズム 2 にしたがって図-6 の MBC を階層化したものである。図-8 (b) に構成された二分木を示す。検索の際には二分木をたどれば無駄のない検索が実現できる。

(3) Allen の時間関係の拡張による検索手法

階層的クラスタリングにより、時間をキーとしてボトムアップ的に二分木が構成される。階層化された MBC がそれぞれの節点となる。各節点は以下の情報を含んでいる。

- ・ x 座標の区間 $X_{Range} = [X_{Start}, X_{End}]$
- ・ y 座標の区間 $Y_{Range} = [Y_{Start}, Y_{End}]$
- ・ 存続時間 $T_{Range} = [T_{Start}, T_{End}]$
- ・ 動きベクトル MV
- ・ 左右の子供の時間関係 $R_T(L, R) = \{ \text{before} \mid \text{overlaps} \mid \text{includes} \}$

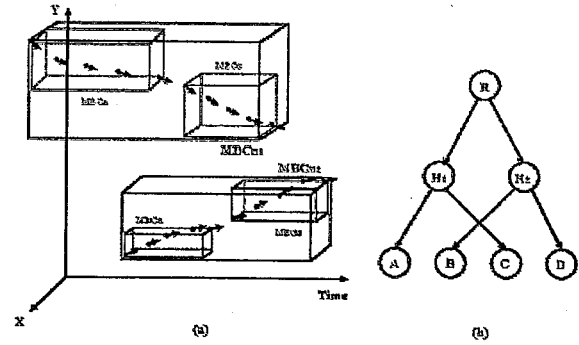


図-8 MBC の階層化と構成される二分木

$R_T(L, R)$ は左の子供の MBC_L と右の子供の MBC_R の間に成立する時間関係の記述である。また、各節点 Node と左右の子供 Child の間には、以下の時間関係が成立している。

$Node.T_{Range} \text{ includes } Child.T_{Range}$

本手法によって構成される二分木の高さは一定ではない。したがって、検索の効率は一定とならず、検索方法に工夫が必要である。そこで、先ほどの Allen の時間関係を MBC 全体に拡張し、検索の効率化をはかることを検討する。

a) Allen の時間関係の MBC への拡張

Allen の時間関係を空間に拡張した場合、x 座標、y 座標それぞれに対しても、before、overlaps、includes が記述できる。X、y それぞれの関係から xy 空間上の空間関係を記述することが可能である。表-9 は MBC_A 、 MBC_B の x 座標の区間、y 座標の区間から記述した MBC の空間関係 $R_S(A, B)$ の値を表している。

また、時間関係同様、親節点 Node とその子供 Child は階層構造を持つため

$Node.X_{Range} \text{ includes } Child.X_{Range}$

$Node.Y_{Range} \text{ includes } Child.Y_{Range}$

表-9 MBC の空間関係

Relation y \ Relation x	before	overlaps	includes
before	 before	 before	 before
overlaps	 before	 overlaps	 overlaps
includes	 before	 overlaps	 includes

が成立する。したがって、時間情報に加えて、MBC の空間情報についても関係を記述することが可能となる。

b) 左右の節点とクエリの関係の推論

検索の際には、自分の左右のノードとクエリの関係から、探索する必要のないノードを枝刈りをして探索効率を向上させることができる。そのための Allen の基本アルゴリズムを導入する。Allen の基本アルゴリズムは、3 オブジェクト間の関係のうち、2 オブジェクト間の関係がわかっているときに、残りの関係を推測する方法である。以下に、 $R(A, B)$ 、 $R(B, C)$ の際に $R(A, C)$ を求める Allen の基本アルゴリズムを示す。

Allen の基本アルゴリズム

1. $R(A, C) = \varepsilon$ とする。
2. (A, B) の満たす関係のうち、いずれかの関係を取り出す。同様に、 $R(B, C)$ の満たす関係からいずれかを取り出し、それぞれ $r1$ 、 $r2$ とする。
3. Allen の推移表より、 $r1$ 、 $r2$ から推測される関係集合を導出し、これを $R(A, C)$ に加える。
4. $R(A, B)$ 、 $R(B, C)$ に含まれるすべての関係について、2、3 を繰り返す。
5. $R(A, C)$ が出力となる。

探索する際には、ノード間の時間関係 $R_T(L, R)$ 、空間関係 $R_S(L, R)$ が分かっているので、左右一方の子供とクエリとの関係 $R_T(L(R), Q)$ 、 $R_S(L(R), Q)$ が分かれば、もう一方の子供とクエリとの関係を推測でき

る。

o) 左右の節点とクエリの関係の推論

二分木に対する問い合わせは、点検索ではなく、結果が範囲を持つ範囲検索として行われる。つまり、クエリは MBC と同様に、 x 、 y 、 t のそれぞれが範囲を持つ直方体であり、クエリと MBC の重なる部分を結果として返すことになる。時空間に対する検索は、以下の 5 ステップからなる。

ステップ 1：時間関係に基づく検索位置の決定

表-10 は、クエリと節点の時間関係から決定される、探索すべき節点の場所を示した表である。この表を用いて、クエリの時間情報と、対象節点の左右の子の時間情報を比較して、探索する場所を決定する。具体的には Allen の基本アルゴリズムより、木の左の節点とクエリとの時間関係 $R_T(L, Q)$ 、左右の節点の時間関係 $R_S(L, R)$ から、クエリと右の節点の関係を判断し、左右の節点を調べるかどうかを決定する。

表-10 MBC の時間関係と探索すべき節点

$R_S(L, R) \backslash R_T(L, Q)$	before	overlaps	includes	includes inv
before				
overlaps				
includes				

Node Time: Query Time:

ステップ 2：時間関係に基づく検索位置の決定

探索する節点とクエリの空間関係から、それ以降の探索を決定する。クエリの空間が節点の空間の外にあった場合、つまり、表-10 において、探索する節点とクエリの空間関係が $R_S(Node, Query) = \text{before}$ である場合、以下の節点を探索する必要はないため、枝刈りを行い、その節点に対する探索を中止する。

ステップ 3：出力の決定

枝刈りが行われることなく、葉節点まで達したのならば、MBC はクエリの要求を満たしていることになる。しかしながら、MBC がクエリに対して、空間的にどれほど一致しているかはわからない。そこで $R_S(Leaf, Query)$ を調べて出力区間を調整する。

- $R_S(Leaf, Query) = \text{overlaps}$ のときは、クエリと葉のデータが重なっている部分のみ出力する。また、他の葉部も出力になり得るので、探索を継続する。
- $R_S(Leaf, Query) = \text{includes}$ のときは、クエリが該当ノードに含まれているために、該当ノードのみ出力する。MBC が見つければ、探索を終了する。
- $R_S(Leaf, Query) = \text{includes inv}$ 、つまり、 $R_S(Query, Leaf) = \text{includes}$ のときは、MBC のデータすべてが出力される。また、他の葉部も出力になり得るため、探索を継続する。

ステップ 4：軌跡の選択処理

ステップ 3 で得られたデータは、平均の動きベクトルを持っているため、向きによる選択処理を行う。式 (1) を用いて、角度処理に基づいた選択を行い MBC を限定する。

ステップ 5：軌跡の選択処理

以上の処理で出力されたデータに対して、時間的に連続している区間、つまり時間差がほとんどない区間同士を結合し、最終的な出力とする。

4. 時空間インデキシングによる検索の評価実験

実験に用いた映像は MPEG 形式で、人の往来が多い場所をカメラ固定で 9 分 30 秒間撮影した映像である。映像サイズは 352×240 であり、1 秒間のフレーム数は 30 frame/s (30fps) である。実験に用いたのは、CPU が PowerPC970 搭載の 2GHz、メモリ 1GB のデスクトップパソコンであり、OS は Apple 社の MacOSX である。

(1) 動領域の抽出

以下の手順で動領域の抽出を行った。

1. KDDI 研究所の MP-Factory⁷⁾ を用いて、素材映像から B ピクチャ、P ピクチャの各マクロブロックの動きベクトルを抽出した。また、動きベクトルを物体の移動方向としてとらえるため、B ピクチャの順方向

予測動きベクトル、Pピクチャの動きベクトルの向きをそれぞれ反転させた。

2. サイズの小さい動きベクトルを除去した。これは、カメラのわずかな揺れから生じたものと判断したためである。今回の実験では $|MV| < 1$ となる動きベクトルをすべて除去した。

3. 空間連続性、時間連続性より、動領域の抽出処理を行った。なお、角度処理の閾値には $\cos\left(\frac{\pi}{2}\right)$ を用いた。

4. 以上の処理により、抽出された連続データ 100 個を初期 MBC とした。

表-1 1 動領域抽出の結果

Factor	Number
Extracted moving object(completely)	28
Extracted moving object(partially)	7
Over Extracted	0
Can't Extracted	0
Total	35

78 個の初期 MBC によって動領域を抽出した。抽出結果を表-1 1 に示す。移動物体に対して、完全に MBC として抽出できた数、部分的に抽出できた数、まったく認識できなかった数をカウントしたものである。移動物体を完全に抽出できないことや、移動物体と間違えて影などを抽出することはなかった。とらえられなかった軌跡を調べてみると、ほとんどが空間内の同じ場所、ちょうど建物の影の部分で起こっていた。これは、移動物体と影の動きベクトルが重なって生じた結果、連続性が不明瞭になり、不連続データとして除去されてしまったためである。

図-1 2 に、素材映像のうち、1800 フレームまでの移動軌跡とそれに対応した 20 の MBC の様子を示す。映像の X 軸、Y 軸、そして時間軸について、抽出した動領域を MBC で表現している。図中の各点が、抽出した動領域の重心の位置で、それを囲む直方体が MBC である。

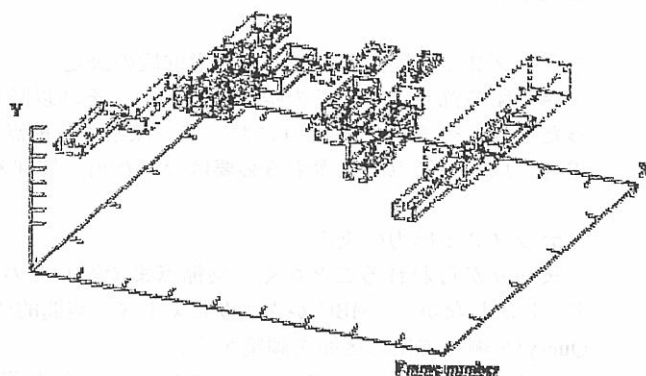


図-1 2 実験における MBC の分布

図-1 3 は、マルチメディアデータを組み合わせたインタラクティブ表現ができる SMIL⁸⁾を用いて、MBC と映像を同期させ、同一画面に表示させたときのスクリーンショットである。MBC に当たる部分は濃い長方形に囲まれており、動きベクトルは矢印で表示されている。この映像では、複数の人間のおおまかな軌跡と移動範囲が MBC によって表示されている。



図-1 3 映像と MBC の同期

(2) 階層的クラスタリングによる時空間インデキシング

78 個の MBC について、時間情報を制約とした階層的クラスタリングを行った。なお、クラスタリングの際に用いた類似度の成分には、マクロブロックの x 座標、マクロブロックの y 座標、動きベクトルの cos 成分がある。

(3) 時空間範囲検索に対する評価

MBC の検索を行い、高速性、正確性の両面から評価を行う。

a) 検索の高速性の評価

本手法は、時空間データに対し、時間幅、空間幅をもったクエリを与えることができるという利点がある。そこで、従来手法として、範囲検索ではない、点検索の線形探索を比較対象にあげ、その計算速度を比較する。表-1 4 は、x の幅と y の幅を固定したまま、時間の区間の長さを変えたクエリの線形探索と、本手法の計算量を表にしたものである。線形探索は時間幅が大きくなるにつれ、計算量が增大しているのに対し、本手法では時間

表-1 4 線形探索と提案手法との実行時間

T _{range}	Linear search(ms)	Proposed Method(ms)
100	940	216
500	2416	545
1000	4378	615
5000	19788	752
10000	39218	933
15000	58352	1125

幅の影響を受けることなく検索できている。大容量のデータを扱うときには、広い範囲に対する検索が行われることが多いため、本手法の範囲検索はマルチメディアデータに適しているといえる。

b) 検索の正確性の評価

次に、クエリに対する検索結果の内容から、正確性を評価する。時間幅、空間幅がそれぞれ異なる5種類のクエリの精度を求めた。ここで、C を検索対象の正解フレーム区間、D を検索できなかった正解フレーム区間「未検出フレーム区間」、E を正解でないものを過剰に検出したフレーム区間「過剰検出フレーム区間」とするとき、再現率 Recall と、適合率 Precision は式(4)で定義される。

$$\begin{aligned} \text{Recall} &= \frac{C}{C+D} \times 100 \\ \text{Precision} &= \frac{C}{C+E} \times 100 \end{aligned} \quad (4)$$

再現率は検索対象をどれだけ漏れなく検出できたかという完全性を表現し、適合率は検索結果の中に必要な対象がどれだけ存在するかという正確性の指標である。検索システムの性能を評価する場合、漏れがなく、クエリに適した対象を検索することが目的となるため、再現率と適合率がともに高い値を示すことが求められる。表-15は、5種類のクエリに対して Recall、Precision を求めたものである。5つのクエリを統合すると、再現率、適合率ともに90%を超えており、全体として高い精度を出していることが分かる。それぞれのクエリの精度も高い数値を示しており、検索システムとして十分機能しているといえる。

表-15 線形探索と提案手法との実行時間比較

	Query ₁	Query ₂	Query ₃	Query ₄	Query ₅	Total
Correct(frame)	406	661	251	390	330	2038
Undetected(frame)	31	45	18	43	37	174
Over detected(frame)	19	17	3	41	0	80
Recall(%)	92.9	93.6	93.3	90.0	89.9	92.1
Precision(%)	95.5	97.4	98.8	90.4	100.0	96.2

5. おわりに

時空間における移動物体のインデキシングのために、MPEG 映像から抽出した動きベクトルと、空間連続性や時間連続性の概念を用いることにより、動領域を自動で抽出する手法を提案した。本手法では、MPEG 映像から動きを抽出して、GPS 端末を必要としない移動物体の追跡と軌跡の発見を可能にした。さらに、連続的なデータから MBC を作成して階層的クラスタリングを行い、時空間データの正確な検索ができることを示した。

一方、時間と空間を統一的に扱うことは、従来から困難とされている。そのため本研究では、時間情報を制約条件として扱い、時間関係を保存したインデキシングを実現し、次元を落とすことなく、時空間データを扱う手法を提案した。さらに、Allen の時間関係に基づいた階層化と空間への拡張、探索節点とクエリの関係の推論を行い、枝刈りを含めた無駄のない検索を可能にした。評価実験により、再現率、適合率ともに90%を超える精度で検索を行うことができた。しかしながら、MBC の階層化は平衡木を構成しないため、一定の計算量で検索を行うことはできない。さらに、データ処理の段階で損失した情報もあり、初期 MBC もすべての移動物体をトラッキングできなかったため、完全に検索を行えたとは言えない。初期 MBC の精度を上げることが今後の課題である。そのために

- ・ MPEG 映像に圧縮する際、利用される離散コサイン変換の係数(DCT 係数)から得られる色情報を考慮した処理
- ・ 連続性を想定した動きベクトルからの動き予測モデルの作成

などによる精度の改善を検討している。

6. 謝辞

本研究は平成 16 年度特別プロジェクトによるものである。

参考文献

- 1) 酒井 善則, 吉田 俊之, “映像情報符号化”, オーム社 (2001).
- 2) 岸野 文郎, 佐藤 隆夫, 横矢 直和, 相澤 清晴, 有川 正俊, “画像と空間の情報処理”, 岩波書店 (2000).
- 3) Antonin Guttman, “R-Trees: A Dynamic Index Structure for Spatial Search”, Proc. of the 1984 ACM SIGMOD Int'l Conf on Mgmt of Data, pp.47-57 (1984).
- 4) 堀之内 浩征, 黒木 進, 牧之内 顕文, “時空間データベースインデックス正規化 R*-tree の実装と性能テスト”, 情報処理学会論文誌, Vol.40, No.3, pp.1225-1235 (1999).
- 5) 柴田 正啓, “映像の内容記述モデルと映像構造化への応用”, 電子情報通信学会論文誌, Vol.J78-D-II, No.5, pp.754-764 (1999).
- 6) James F. Allen, “Maintaining Knowledge about Temporal Intervals”, Communications of the ACM, Vol.26, No.11, pp.832-843 (1983).
- 7) MP-Factory : <http://www.mpeg.co.jp/products/mpf/>
- 8) SMIL : <http://www.w3.org/AudioVideo/>

筆者： 1) 堀江 新、自然科学研究科、学生；2) 上原 邦昭、自然科学研究科、教授

Spatio-Temporal Indexing of MPEG Video Based on Hierarchical Clustering

Arata Horie
Kuniaki Uehara

Abstract

Recently, there have been many studies on analyzing human behavior from its spatial trajectory measured by location sensors such as GPS (Global Positioning System). Although we need location sensors to measure each trajectory, it is useful for us to analyze spatio-temporal data and to know how people act. However, as trajectory data is multi-dimensional data, we need to index such a large amount of data to get useful information easily.

In this paper, to solve these problems, first, instead of location sensors, we use motion vectors obtained from MPEG video compression. To consider two consecutiveness in moving object, we extract the region where movement occurs in each frame. Second, in order to access required data easily, we apply hierarchical clustering to index these regions and to construct binary tree. Spatio-temporal data include time attribute, so we regard time attribute as a restriction for hierarchical clustering. To extract region of moving in each frame and index these data, we can get trajectory easily and exactly.

