データベースシステム 114－6
（1998. 1. 19）

映像データベースのための異種メディア間の演算

吉山雅彦　田中秀明　植村俊亮

奈良先端科学技術大学院大学 情報科学研究科

〒630-0101 奈良県生駒市高山町 8916-5
{masah-yo,hideak-t,uemura}@is.aist-nara.ac.jp

計算機技術の発達により，映像データをディジタル化して，映像データベースを構成することが可能になりつつある．映像データベースを検索するために，画像や音声に索引付けすることが行なわれる．しかし，画像だけ，音声だけに対する索引付けでは，映像データベースからまとまった意味を持つ情報を適切に取り出すことは困難である．本研究では，こうした索引付けに用いる時区間演算を整理し，意味のある映像情報を検索するための新しい演算を提案する．

Operations among Heterogeneous Media for Video Databases

Masahiko YOSHIYAMA, Hideaki TANAKA, and Shunsuke UEMURA

Graduate School of Information Science
Nara Institute of Science and Technology

8916-5 Takayama, Ikoma, Nara 630-0101,JAPAN
{masah-yo,hideak-t,uemura}@is.aist-nara.ac.jp

Traditional indexing methods for visual data or audio data are not sufficient to retrieve semantically coherent video objects. This paper first reconstruct traditional time interval operations, and then proposes a new operator to integrate time intervals on different media (audio and visual). This operator enables us to retrieve semantic information unit (video object) from video databases.

1. 研究の目的

近年コンピュータ能力の進歩により,動画像をホストコンピュータに格納し,ネットワーク上の端末で再生することが実現可能になってきた.これにともないデジタル化された映像データが飛躍的に増大し,多くのデータ中から希望する映像データを検索する要求が高まってきている.

映像データベースを作成する際には,映像を検索するための機構が必要となるが,その検索手法としては索引付けによる方法が一般的である.

索引付けでは,まず,「画像」に写っているものによる索引付けをすることが考えられる.これは画像認識の立場による研究が進んでおり,画像に写っている背景,人物,物体などを認識し,これらをそれぞれ索引付けする.「音声」による索引付けをする場合でも,音声認識の分野で研究が進んでおり,個人の音声の識別が可能になってきている.これらのことから将来的には映像データの中から「画像」に誰が映っていて,「音声」には誰が喋っているかをより手軽に索引付けすることができるようになるものと考えられる.

しかし画像や音声の索引を単独で用いると,利用者の要求する映像と関係のない映像も多く検索されがちである.また映像データベースから返された映像に意味のまとまりがないことが多い.

本稿では,画像や音声の索引からまとまった意味内容を持つ映像を得る手法の一つとして,映像演算を用いた映像データベースの検索,索引付け手法を提案する.これにより意味のまとまりによる索引付けを半自動化し,利用者の要求に合致した映像検索を可能にしたい.

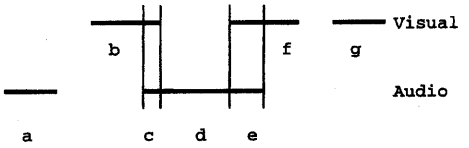

図 1 映像オブジェクトの抽出

例えば図1において,ある映像オブジェクトを抜き出す演算を考える.映像オブジェクトとは,意味的にまとまった映像(画像+音声)をいう.画像と音声が図のような関係にある時に,画像と音声の重なり(and)部分は[c,e]であるが[c,d,e]が音声で繋がっているので一連の内容とみなし,また[bとc],[eとf]が画像で繋がっているので[b,c,d,e,f]のすべてを一連の内容とみなしたい.しかしorでは含まれるはずのa,gは含めない.これにより意味のある映像オブジェクトの抜き出しが可能となる.本稿の目的は,こうした演算を定義し,その意味付けを行うことである.

2. 時区間データモデル

映像オブジェクトすなわち意味のある場面を時間軸上の閉区間として時区間を用いて表現することを考える.

Allenは二つの時区間に対し13種の時間的関係[1]が存在することを示した.また天笠ら[2]は時区間モデルとして実時区間,空時区間の二つを用い,事象をこれらの集まりとして表現した.

ここでは,映像データベースを対象とするので,事象は実時区間の集まりによって表現されるものと考える.また映像データベースの扱う時区間の最小単位は画像の1フレーム分に相当するため,時区間の端はすべて等号を含んだ閉時区間と考える.

数学的には時区間の否定などの結果として得られる時区間に等号があると矛盾するわけだが,映像データベースの場合では,等号がなければ画像が途中で途切れたことになり矛盾する.このため,常に等号を含んだ時区間を結果として返すものとする.

3. 単純時区間演算

3.1 単純時区間

一つの時区間をA,その開始時刻,終了時刻をそれぞれ $a_s, a_e (a_s \leq a_e)$, $A(a_s, a_e)$ は時区間Aが時刻 a_s で開始して時刻 a_e に終了することを示す.これを単に (a_s, a_e) と書くこともある. $A(a_s, a_e)$ を単純時区間とよぶ.

なお映像情報を扱うので,時間軸の最初と最後がある.それをそれぞれstart,endと表す.また時区間がないことを (ϕ) と表す.

3.2 単純時区間演算

単純時区間演算は,二つ以下の単純時区間を対象とする演算,すなわち2項演算または単項演算である.演算の結果(返り値)もまた二つ以下の単純時区間になる.

3.3 和(union, ∪)

AとBの時区間の和は, $A \cup B$ と表し,次の時区間を返す.

$A \cup B =$

$$\begin{cases} (a_s, a_e), (b_s, b_e) & ((a_e < b_s) or (b_e < a_s)) \\ (a_s, b_e) & ((b_s \leq a_e)\ and (a_e \leq b_e)) \\ (b_s, a_e) & ((a_s \leq b_e)\ and (b_e \leq a_e)) \\ (a_s, a_e) & ((a_s \leq b_s)\ and (b_e \leq a_e)) \\ (b_s, b_e) & ((b_s \leq a_s)\ and (a_e \leq b_e)) \end{cases}$$

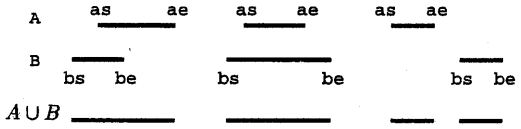

図 2 A ∪ B

3.4 積 (intersection, ∩)

A と B の時区間の積は, $A \cap B$ と表し, 次の時区間を返す.

$A \cap B =$

$$\begin{cases} (a_s, b_e) & ((b_s \leq a_s)\ and (b_e \leq a_e)) \\ (b_s, a_e) & ((a_s \leq b_s)\ and (a_e \leq b_e)) \\ (a_s, a_e) & ((b_s \leq a_s)\ and (a_e \leq b_e)) \\ (b_s, b_e) & ((a_s \leq b_s)\ and (b_e \leq a_e)) \\ (\phi) & ((a_e \leq b_s)\ or (b_e \leq a_s)) \end{cases}$$

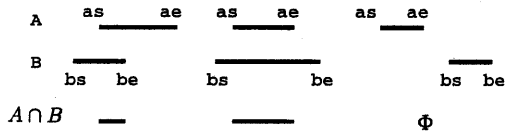

図 3 A ∩ B

3.5 否定 (negation, ¬)

時区間 A の否定は,$\neg A$ と表し, 次の時区間を返す.

$\neg A = (start, a_s), (a_e, end)$

start,end, とくに end は現実には特定が困難である.

3.6 差 (difference, −)

$A - B$と表し, A に含まれるが B には含まれない時区間を返す. これは非可換の演算子である.

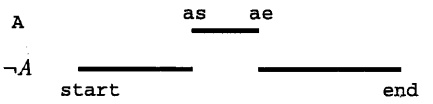

図 4 ¬ A

$A - B =$

$$\begin{cases} (a_s, b_s) & ((a_s \leq b_s)\ and (b_s \leq a_e \leq a_e)) \\ (b_e, a_e) & ((b_s \leq a_s)\ and (a_s \leq b_e \leq a_e)) \\ (a_s, b_s), (b_e, a_e) & ((a_s \leq b_s)\ and (b_e \leq a_e)) \\ (a_s, a_e) & ((a_e \leq b_s)\ or (b_e \leq a_s)) \\ (\phi) & ((b_s \leq a_s)\ and (a_e \leq b_e)) \end{cases}$$

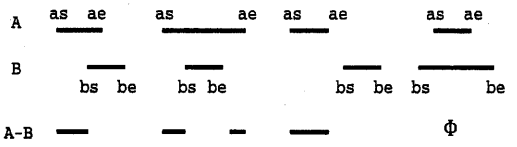

図 5 A − B

この演算は $A \cap (\neg B)$ とも表される.

3.7 含有 (contain, ⊃)

$A \supset B$と表し, B が A に含まれるときに A の時区間を返す.

$A \supset B =$

$$\begin{cases} (a_s, a_e) & ((a_s \leq b_s)\ and (b_e \leq a_e)) \\ (\phi) & ((b_s \leq a_s)\ or (a_e \leq b_e)) \end{cases}$$

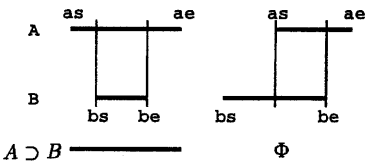

図 6 A ⊃ B

3.8 共有 (joint, ⊎)

$A \uplus B$と表し, A と B とが共有部分を持っているとき論理和を返す.

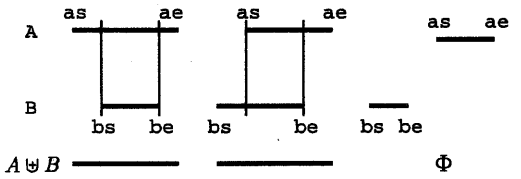

図 7 A ⊎ B

$$A \uplus B = (A \cup B) \supset (A \cap B)$$

この演算は和演算と類似した2項演算であるが,共有部分のある時区間を抽出する.

3.9 単純時区間演算の検討

3.9.1 異なるメディア上の演算

同じメディア上の二つの時区間 A (a 氏が写っている時区間) と B (b 氏が写っている時区間) との積 (図 8) を考える.

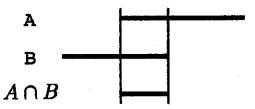

図 8 a 氏と b 氏が写っている

この場合,$A \cap B$は「a 氏と b 氏の両方が写っている時区間」と考えるのが自然である.またこのメディアから,該当する時区間を取り出すだけでよい.

しかし A が「a 氏が写っている時区間」,B は「a 氏が話している時区間」であれば,$A \cap B$(図 9) の意味はそう単純ではない.

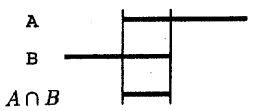

図 9 a 氏が話して写っている?

この場合 $A \cap B$は, おそらく「a 氏が写っており, かつ話している時区間」を意味するが, 具体的に演算結果がなにであるか (画像, 音声, 両方) はよく検討しなければならない.

3.9.2 単純時区間演算の連結

単純時区間演算を施した結果の時区間に, また演算を施すことができる.

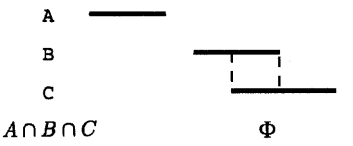

図 10 $A \cap B \cap C$

しかし時区間演算としては, 図10における $A \cap B \cap C$ の結果はϕではなくて,$B \cap C$と同じとした方が実用的である.

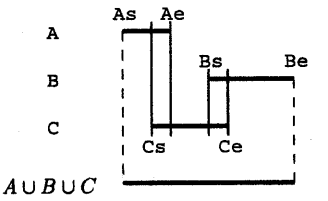

図 11 $A \cup B \cup C$

また図11の場合,$A \cup B \cup C$は (a_s, b_e) であってほしいが,3.3の和の定義ではこれを正しく計算できない.

4. 複合時区間演算

3.9の考察にもとづき, 複合時区間演算を定義する. $A_i, B_j (1 \leq i \leq n, 1 \leq j \leq m)$ が単純時区間を表すとき, $S = \{A_1, A_2, A_3, ..., A_n\}$, $T = \{B_1, B_2, B_3, ..., B_m\}$ はそれぞれ複合時区間を表すものとする. すなわち複合時区間は単純時区間の集まりである. さらに

$$\bigcup_{i=1}^{n} X_i = X_1 \cup X_2 \cup \cdots \cup X_n$$

$$\bigcap_{i=1}^{n} X_i = X_1 \cap X_2 \cap \cdots \cap X_n$$

とする.

4.1 和 (union, ∪)

$S \cup T$と表し, S,T の各要素のどちらかに存在する時区間の集まりを返す.

$$S \cup T = \bigcup_{i=1}^{n} \left(\bigcup_{j=1}^{m} (A_i \cup B_j) \right)$$

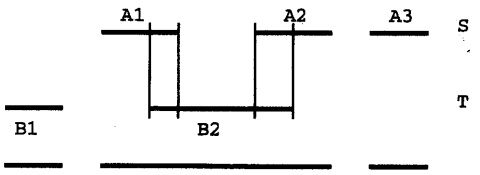

図 12 S ∪ T

4.2 積 (intersection, ∩)

$S \cap T$と表し, S,T の各要素の両方に存在する時区間の集まりを返す.

$$S \cap T = \bigcup_{i=1}^{n} \left(\bigcup_{j=1}^{m} (A_i \cap B_j) \right)$$

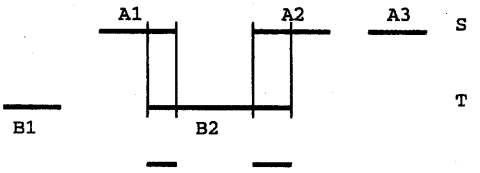

図 13 S ∩ T

4.3 否定 (negation, ¬)

$\neg S$と表し S のどの要素にも存在しない時区間の集まりを返す.

$$\neg S = \bigcap_{i=1}^{n} (\neg A_i)$$

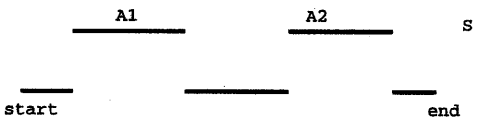

図 14 ¬ S

これも単純時区間演算の時と同様に単独で用いることは困難である

4.4 差 (difference, −)

$S - T$と表し, S の時区間には存在し T の時区間に存在しない時区間の集まりを返す.

$$S - T = S \cap (\neg T)$$

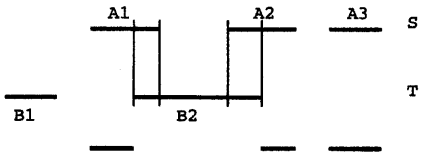

図 15 S − T

4.5 含有 (contain, ⊃)

$S \supset T$と表し, T の要素 B_jが S の要素 A_iに含まれるとき A_iを要素とする時区間の集まりを返す.

$$S \supset T = \bigcup_{i=1}^{n} \left(\bigcup_{j=1}^{m} (A_i \supset B_j) \right)$$

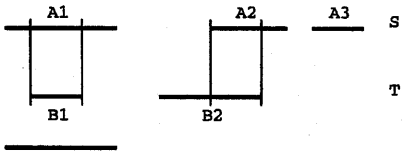

図 16 S ⊃ T

この演算を用いると誰かが話している場面を含む映像などの検索が可能となる. また次に示す共有演算を定義する際にも用いている.

4.6 共有 (joint, ⊎)

$S \uplus T$と表し, S の要素 A_iと T の要素 B_jとが共有部分を持っているとき,A_i, B_jの論理和を返す.

$$S \uplus T = (S \cup T) \supset (S \cap T)$$

4.7 共有演算の適用例

例えば画像, 音声だけの索引であれば索引付けは一意に行なうことが可能であるが, 人物の抽出となると面倒な索引付けになり, 一意的に索引付けすることも困難となる. ここで共有演算の適用例をあげる.

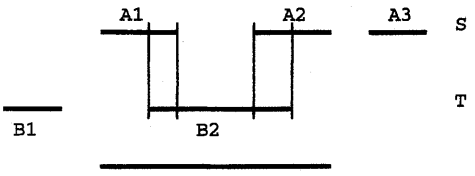

図 17 S ⊎ T

4.7.1 適用例 1

共有演算は映像データベースを対象に考えており, 映像データベースにおいて人物の抽出などに役立つ.

画像や音声に対する要求に合致した検索結果は索引付けにより返すことができるが, 映像の内容は意味のまとまりを持っていない場合がある.

ここで共有演算を用いると索引付けしたい人物が画像と音声の両方に関連する映像を取り出すことが可能になり, 利用者の望む場面映像により近い映像を検索結果として得ることができる.

4.7.2 適用例 2

また適用例 1 のようにして得られた人物と, 同様に他の人物の共有演算を行なった二人の人物の抽出結果に共有演算を施すことにより, 二人が会話をしていると思われる場面映像が得られる.

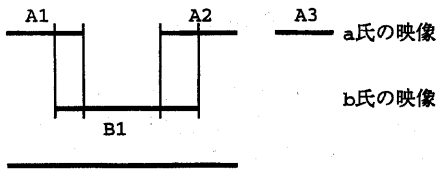

図 18 a 氏と b 氏の会話?

この方法は全然関連のない場合もあるが, 会話をしている場合には会話の場面が抽出される.

これらのように共有演算を用いると特別な索引付けを行なわなくてもある程度の索引付けを自動的に生成することが可能となる.

またこの共有演算の性質を利用者が理解し利用することにより所望の映像に近い映像を得ることが可能となる.

4.8 適用できない例

画像と音声で結合する際に映像と映像を音声が繋いでいるというのは, かなりの場合において意味のまとまった映像を得ることができるが, 音声と音声を映像が繋ぐ場合は会話の途中で相手に切り替わった場合が考えられるので, この時は共有演算ではあまり意味のまとまりのない映像を得ることが多くなると思われる.

5. まとめ

今回導入した共有演算などの映像演算を用いることにより検索の際に画像, 音声による検索だけでなく, さまざまな検索法により結果映像を得ることができるものと考えられる. すなわち, この演算はメディアに依存する.

6. これからの課題

提案した共有演算がどの程度意味のまとまりのある映像を検索結果として返すことができるか, いろいろな場合について試行する必要がある.

ある人物の画像と別の人物の画像に対して共有演算を行うことにより会話の映像を抽出できると思われるが, どのような場合に共有演算を用いれば会話の抽出が可能になるかを検証し, また田中ら [3] の提案する同種メディアや複数のメディアに対し共有演算を応用して, 別の意味のまとまりの抽出を行いたい.

参考文献

[1] J. F. Allen. Maintaining Knowledge about Temporal Intervals. *Comm. of the ACM*, Vol. 26, No. 11, pp. 832–843, November 1983.

[2] 天笠俊之, 鈴木邦彦, 有次正義, 金森吉成. 時区間概念モデルを実装した時区間クラス. In *Advanced Database Symposium'97*, pp. 59–66, December 1997.

[3] 田中秀明, 吉山雅彦, 植村俊亮. 映像データベースのための同種メディアの統合. 情報処理学会第 114 回データベースシステム研究会研究報告, January 1998.