

同期多視点映像の検索による ビデオハイパーメディアの動的構成

竇 勇 偉[†] 十 河 孝 至[†]
秦 淑 彦^{††} 田 中 克 己^{††}

デジタル技術の進歩とともに多地点で同時に撮影されたビデオの検索・呈示をいかに行うかが重要になりつつある。これまでの研究ではビデオごとが研究対象とされているため、同期多視点映像の検索にたいして有効な手段とは言えない。そのため本論文では、同期多視点映像の検索結果が存在する区間を求める方法と検索結果の呈示を代数的に記述する方法について述べ、記述された検索結果をSMIL基準に従ってビデオハイパーメディア化する方法を提案する。

Dynamic Structuring of Video Hypermedia by Synchronized Multi-Perspective Video Retrieval

YONGWEI DOU,[†] TAKASHI SOGO,[†] TOSHIHIKO HATA^{††}
and KATSUMI TANAKA^{††}

With the exponential growth of video data comes the challenge of managing this information in a controlled, organized and efficient manner. The database researchers have set their sights on addressing the fresh challenges offered by retrieving meaningful intervals of video database. However, for extension of the method to multi-stream video it will not be effective because of the more meaningful spatio-temporal features that be presented in multi-stream video. In this paper, we analyze the problem of multi-stream video retrieval and propose a methodology to construct video hypervideo dynamically as a meaningful unit for the end user. And also we try to reverse the result to SMIL file in the cause of the open hypermedia system.

1. はじめに

近年、デジタル化の技術が目覚しく進歩するにつれ、デジタルビデオは至る所に存在する。その中でスポーツ中継映像や監視ビデオなど同時間軸に対する複数映像も少なくない。本論文ではこのようなビデオデータを同期多視点映像と呼ぶことにする。

ビデオデータにおいて、幾つかのキーワードによりそれらのキーワードが出現する区間を問い合せるとき、ユーザとしては意味的にまとまりのある区間を答えとして望むのが一般的である。ビデオ検索モデルにはスジットらにより提案されたグルー演算²⁾がある。このグルー演算を用いることで、答えとなる区間を予めデータベースに格納しておくのではなく、ある問い合せに対して答えと成り得るすべての区間を動的に生

成することができる。また、このグルー演算のみでは不適切な解も生成してしまうので、そのような区間を解から取り除くフィルタも十河らにより提案されている³⁾。しかし、このような意味的にまとまりのある区間は同期多視点映像の場合では必ずしも一つのビデオに含まれるとは限らない。一方、我々のグルー演算は同時間軸に対する複数のビデオにまたがる問い合せに対して有効な検索結果が得られない。つまりグルー演算では複数のビデオに渡って区間を合成する方法は与えられていないためである。本論文ではグルー演算に基づいて同期多視点映像を取り扱うための拡張を行う。この拡張によって、指定されたキーワード群を全て少なくとも1ヶ所には含むようなマルチストリームビデオ区間集合が得られる。また、検索結果であるマルチストリームビデオ区間集合を代数的な演算により同期化コンテンツとして呈示する方法を提案する。

以下2章では、本論文の研究目的について述べ、3章では、マルチストリームビデオを操作するためのイ

[†] 神戸大学大学院自然科学研究科情報知能工学専攻

^{††} 神戸大学大学院自然科学研究科情報メディア科学専攻

インターバル演算と拡張グルー演算について述べる。また4章では、拡張グルー演算を用いた同期多視点映像に対する検索について述べ、5章では検索結果をSMILに基づいて呈示する方法を提案する。6章は結論である。

2. 研究目的

本論文では、二つの研究目的がある。

- 同期多視点映像の検索
- 検索結果の呈示

2.1 同期多視点映像の検索

例としてはサッカーの中継映像を考える。サッカーを中継するために複数のカメラが設置されている。またそれぞれビデオストリームにキーワードが記述されていると想定する。このような同期多視点映像に対して、幾つかのキーワードにより区間検索を行うと、スジツらが提案したグルー演算が必ずしもユーザが望む区間を生成することができるとは限られない。例えば、図1のようにビデオ1では選手Aを映している区間がある、ビデオ2では選手Bを映している区間

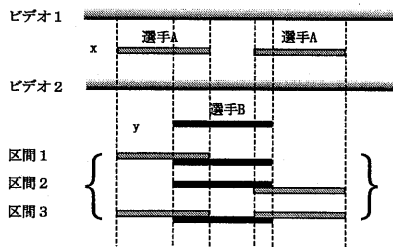


図1 同期多視点映像の検索

がある。ここで、ユーザが「選手A」と「選手B」というキーワードを用いて検索を行うと、「選手A」と「選手B」のシーンが二つの異なるビデオに撮られているため解が存在しない。しかし、区間1、区間2、区間3はユーザの問い合わせに対する適切な答えである可能性がある。

本論文では、このように、検索対象のキーワードが別々のビデオに存在していても問い合わせの解と成り得る検索方法を提案する。

2.2 検索結果の呈示

同期多視点映像は複数のビデオが同時に撮られているため、検索結果は一つのビデオの連続的なシーンとは限られず、各ビデオにおいて問い合わせたキーワードが振られている区間である。あるいは、時間の関連性を持った区間集合である。このような結果を呈

示する方法には、マルチ画面で映す方法があるのを考えやすい。しかし、マルチ画面には画面数が多くなるほど見にくくなる危険性がある。そのため、本論文では、検索結果をシングル画面領域とマルチ画面領域で同時に表す方法を提案する。シングル画面領域で検索結果となる区間集合は、エフェクトにより一本のメディアにし、シングル画面で映す。マルチ画面領域で検索結果になる区間集合は、撮影場所の3Dモデルに配置し、マルチ画面で映す。

例として、図1の区間1を呈示する方法について説明する。区間1はビデオ1の「選手A」が記述されている区間とビデオ2の「選手B」が記述されている区間からなる。図2に示すように左上の領域はシングル画面領域である。二つの区間がかさなっている区間では選手Aの画面をフェードアウトし、選手Bの画面をフェードインして画面を切り替える。右下の領域はマルチ画面領域である。背景としては撮影された場所を示すためのサッカー場である。二つの区間がかさなっている区間ではそれぞれ撮影された場所に画面を配置し、マルチ画面で映す。

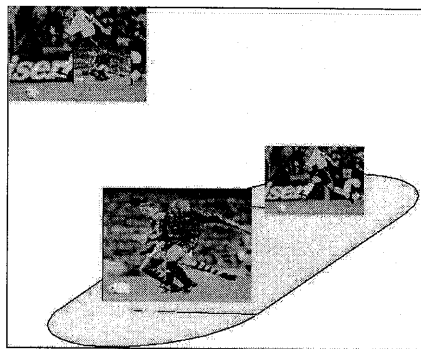


図2 検索結果の呈示1

また、マルチ画面領域では同時に映されている画面の間のリンクにより視点を変えることができる。例えば、図2のマルチ画面領域で手前の画面をクリックすると、ユーザはサッカー場全体を見る視点から、リンクしている画面が配置された場所に視点を移すことにより、図3のような画面になる。このような同期多視点映像の検索結果の呈示方法は、我々のもう一つの目的である。

3. 拡張グルー演算

スジツらが提案したグルー演算は一本のビデオに対して意味的なまとまりを求めることができる。しか

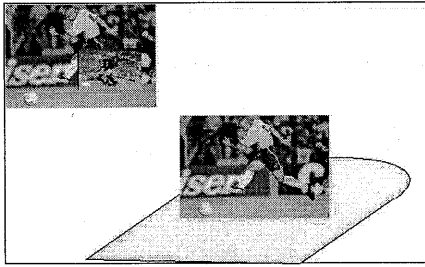


図3 検索結果の表示2

し、グルー演算は同時間軸に対する複数のビデオにまたがる問い合わせに対して有効な検索結果が得られない。それは、グルー演算では同期多視点映像の取り扱いについて述べられていないためである。本論文では、同期多視点映像を扱えるように拡張を行う。次に、拡張したグルー演算の定義について説明する。

3.1 拡張グルー演算の定義

拡張グルー演算の対象としては複数ビデオの区間であるため、スタートフレーム時刻とエンドフレーム時刻という情報しか持っていない区間の定義ではビデオに対するどのビデオの区間であるか識別できなくなる。そこで、我々はビデオ識別番号を保持しているビデオ区間を以下のように定義する。

定義1 (ビデオ区間) ビデオ V^a の中の任意のビデオ区間を x^a と書く

$$x^a = [s, e, a]$$

ここで、 s はビデオ区間 x^a のスタートフレーム時刻である。 e はビデオ区間 x^a のエンドフレーム時刻である。 a はビデオの識別番号である。

定義2 (並列化ビデオ区間) スタートフレーム時刻とエンドフレーム時刻が等しいビデオ区間 $x_1^{a_1}, x_2^{a_2}, \dots, x_n^{a_n}$ があるとする。並列化ビデオ区間 x は以下のように表される。

$$x = x_1^{a_1} | x_2^{a_2} | \dots | x_n^{a_n}, (n \geq 1)$$

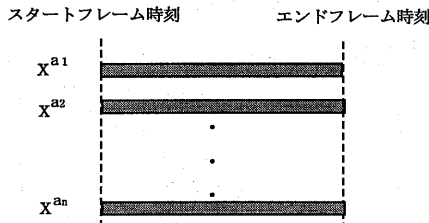


図4 並列化ビデオ区間

並列化ビデオ区間の性質

並列化ビデオ区間は以下のような性質を持っている

- $x_1^a | x_2^b = x_1^a, (a = b \text{ の場合})$
- $x_1^a | \phi^b = x_1^a$
- $\phi^a | \phi^b = \phi^a | \phi^b$

ただし、ビデオ a においてあるビデオ区間 x^a が出現していない区間を ϕ^a で表す

並列化ビデオ区間の呈示

二つの並列化ビデオ区間呈示方法を考える。一つはマルチ画面呈示で、並列化ビデオ区間の各区間をマルチ画面で映す方法である。もう一つの手法は並列化ビデオ区間の各区間をエフェクトにより重畳し、単一の画面領域でユーザに呈示する方法である。

定義3 (マルチストリーム型ビデオ区間) 連続している並列化ビデオ区間 v_1, v_2, \dots, v_n があるとする。このとき、マルチストリーム型ビデオ区間は以下のように表される。

$$v = v_1 \cdot v_2 \cdot \dots \cdot v_n, (n \geq 1)$$

定義4 (区間射影) ビデオ区間 $x^a = [s, e, a]$ の時間区間 $[t_i, t_j]$ に対する射影 $x^a[t_i, t_j]$ を以下のように定義する。

$$x^a[t_i, t_j] = \begin{cases} [t_i, t_j, a] & s \leq t_i, \text{ 且つ } t_j \leq e \\ \text{undefined} & \text{otherwise} \end{cases}$$

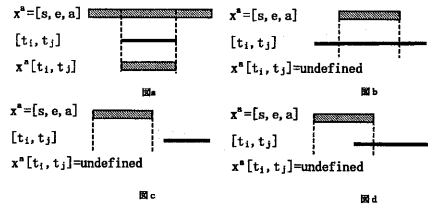


図5 区間射影

定義5 (分割区間列) ビデオ区間集合 $X = \{x_1, x_2, \dots, x_n\}$ に対して、 X の分割区間列 $pseq(X)$ を以下のように定義する。

$$pseq(X) = [t_1, t_2][t_2, t_3] \dots [t_{m-1}, t_m]$$

ただし

$$\text{sort}(\{f_s(x_i) | x_i \in X\} \cup \{f_e(x_i) | x_i \in X\})$$

$$= t_1, t_2, \dots, t_m \text{ となるもの}$$

定義6 (ビデオ区間集合の射影演算 Π) ビデオ区間集合 X の要素 $x^a = [t_i, t_{i+k}, a]$ の分割区間列 $pseq(X)$ に対する射影演算 Π は以下のように定義

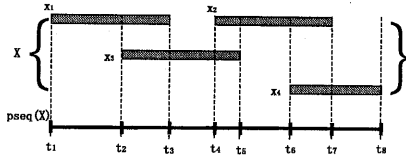


図6 分割区間列

する。ただし、 $pseq(X) = [t_1, t_2][t_2, t_3] \dots [t_{m-1}, t_m]$ とする

$$\begin{aligned} \prod(x^a, pseq(X)) = & \phi^a[t_1, t_2] \cdot \phi^a[t_2, t_3] \dots \phi^a[t_{i-1}, t_i] \cdot \\ & x^a[t_i, t_{i+1}] \cdot \dots \cdot x^a[t_{i+k-1}, t_{i+k}] \cdot \\ & \phi^a[t_{i+k}, t_{i+k+1}] \cdot \dots \cdot \phi^a[t_{m-1}, t_m] \end{aligned}$$

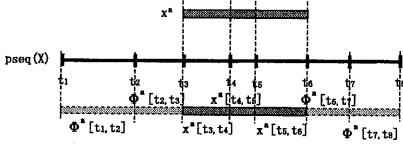


図7 ビデオ区間の集合の射影演算

定義7 (拡張グループ区間演算 \oplus) 拡張区間グループ演算 \oplus はビデオ区間 x^a, y^b に対して、以下のようにマルチストリーム型ビデオ区間 v を求める。

$$\begin{aligned} x^a &\subseteq V_a, \quad y^b \subseteq V_b \\ x^a \oplus y^b &= v \\ \text{ただし} \\ v &= \prod(x^a, pseq(\{x^a, y^b\})) \mid \prod(y^b, pseq(\{x^a, y^b\})) \end{aligned}$$

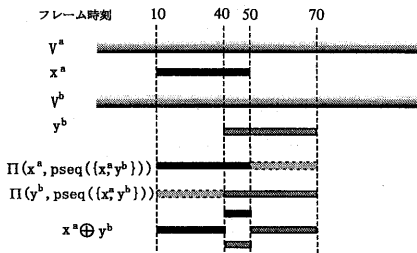


図8 拡張グループ演算

図8のように $x^a = [10, 50, a]$ はビデオ a のビデオ区間であり、 $y^b = [40, 70, b]$ はビデオ b のビデオ区間である。 x^a と y^b の拡張グループ演算を行うと、ビデオ区

間集合 $\{x^a, x^b\}$ の分割区間列は以下ようになる。

$$pseq(\{x^a, y^b\}) = [10, 40][40, 50][50, 70]$$

ビデオ区間 x^a, y^b の分割区間列 $pseq(\{x^a, y^b\})$ に対する射影演算は以下ようになる。

$$\begin{aligned} \prod(x^a, pseq(\{x^a, y^b\})) &= x^a[10, 40] \cdot x^a[40, 50] \cdot \phi^a[50, 70] \\ \prod(y^b, pseq(\{x^a, y^b\})) &= \phi^b[10, 40] \cdot y^b[40, 50] \cdot y^b[50, 70] \end{aligned}$$

また、ビデオ区間 x^a, y^b のグループ演算は射影演算により得られた結果を用いて並列化ビデオ区間を生成する。

$$\begin{aligned} x^a \oplus y^b &= \prod(x^a, pseq(\{x^a, y^b\})) \mid \prod(y^b, pseq(\{x^a, y^b\})) \\ &= x^a[10, 40] \cdot x^a[40, 50] \cdot \phi^a[50, 70] \mid \\ &\quad \phi^b[10, 40] \cdot y^b[40, 50] \cdot y^b[50, 70] \\ &= (x^a[10, 40] \mid \phi^b[10, 40]) \cdot (x^a[40, 50] \mid y^b[40, 50]) \cdot \\ &\quad (\phi^a[50, 70] \mid y^b[50, 70]) \\ &= [10, 40, a] \cdot ([40, 50, a] \mid [40, 50, b]) \cdot [50, 70, b] \end{aligned}$$

定義8 (拡張ペアワイズグループ) 拡張ペアワイズグループ演算 \oplus は任意のビデオ区間集合 X, Y に対して、以下のようにマルチストリーム区間集合を作り出す。

$$X \oplus Y = \{i \mid i \in X, i \in Y, i = x \oplus y\}$$

定義9 (拡張パワーセットグループ) 拡張パワーセットグループ演算 \otimes は任意のビデオ区間集合 X, Y に対して、以下のようにマルチストリーム区間集合を作り出す。

$$X \otimes Y = \{i \mid X' \subseteq X, Y' \subseteq Y, X' \neq \emptyset, Y' \neq \emptyset, i = \bigoplus(X' \cup Y')\}$$

$$\text{但し、} \bigoplus(\{i_1, \dots, i_n\}) = i_1 \oplus \dots \oplus i_n$$

4. 同期多視点映像の検索

同期多視点映像に対する検索を行う場合に、断片的にキーワードがふられている複数のビデオストリームから、幾つかのキーワードにより問い合わせを行う。十河らによりグループ演算の不適な解を取り除くために最長ノイズフィルタ³⁾が提案された。この最長ノイズフィルタを用いることで、キーワード同士の時間関係により、問い合わせ処理することができる。

緩い時間関係を持つクエリ: キーワード k_1, k_2 が緩い時間関係を持ち、 k_1, k_2 が少なくとも一度は出現

するシーンを探すという問い合わせである。

X, Y はキーワード k_1, k_2 が記述されているビデオ区間集合である。 $F_{N, \{k_1, k_2\}}(X)$ は最長ノイズ幅が N のノイズフィルタである。検索の結果は以下のように表す。

$$Ans(k_1, k_2) \leftarrow \{F_{N, \{k_1, k_2\}}(X \otimes Y)\}$$

図9ではサッカー中継映像において選手A(k_1)と選手B(k_2)が少なくとも一度は出現するシーンを探すという問い合わせの処理を示している。

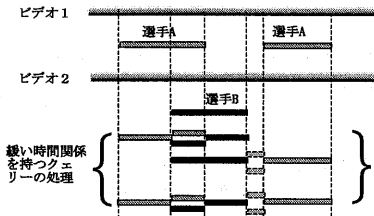


図9 緩い時間関係を持つクエリーの処理

強い時間関係を持つクエリー キーワード k_1, k_2 が強い時間関係を持ち、 k_1, k_2 が少なくとも一度は出現するシーンを探すという問い合わせである。

X, Y はキーワード k_1, k_2 が記述されているビデオ区間集合である。 $F_{0, \{k_1, k_2\}}(X)$ は最長ノイズ幅が0のノイズフィルタである。検索の結果は以下のように表す。

$$Ans(k_1, k_2) \leftarrow \{F_{0, \{k_1, k_2\}}(X \otimes Y)\}$$

例としては選手Aがゴールするシーンを探すという問い合わせである。このような問い合わせの処理を図10に示す。

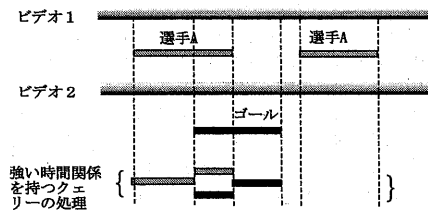


図10 強い時間関係を持つクエリーの処理

多視点映像において、各カメラの注目点が違うので、同じ時刻に同じ場が映されている区間であっても同じキーワードが記述されているとは限らない。そのため、多視点映像においては問い合わせのキーワードの補完が

重要である。補完機能を持っている問い合わせ処理が問い合わせパターン3になる。

補完機能を持つクエリー キーワード k_1 に関連しているシーンを探すという問い合わせである。

定義10 (区間からキーワードを取り出す) 区間 x に対して、スタート時間 $f_s(x)$ からエンド時間 $f_e(x)$ まで、ビデオに記述されたキーワードが関数 $k(x)$ を用いて返される。

また $K(x)$ を用いてキーワード $k(x)$ が記述されている区間集合を求める。

X はキーワード k_1 が記述しているビデオ区間集合である。 $F_{0, \{k_1, k_2\}}(X)$ は最長ノイズ幅が0のノイズフィルタ。

$$Ans(\vec{k}_1) \leftarrow \forall x(x \in X \wedge \{F_{0, \{k_1, k(x)\}}(\{x\} \otimes K(x))\})$$

例としてはキーワード「ゴール」(k_1)を用いてゴールシーンを探すという問い合わせである。関数 $k(x_1)$ が選手Aとゴール二つのキーワードを返す。この二つのキーワードが記述されているビデオ区間集合が $K(x_1)$ である。検索結果は図11の答え1である。

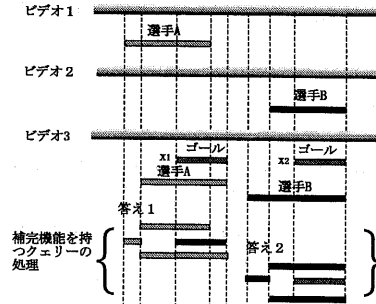


図11 補完機能を持つクエリーの処理

5. SMIL ファイルへの変換

5.1 SMIL とは

SMIL (Synchronized Multimedia Integration Language) はインターネット上で、マルチメディアの同期をとるための言語である。SMILはXML (eXtensible Markup Language) と呼ばれる特定の用途に応じた記述言語を開発するためのメタ記述言語を使って作られている。

SMILは、2つの主要な要素「seq」と「par」からなる。要素「seq」は2つのマルチメディアファイル

を順に再生する。要素「par」は2つのメディアを同時に再生する。この2つ要素を組み合わせることにより、呈示を構成する。

5.2 SMIL ファイルへの変換

SMIL の規準に従って、拡張グルー演算結果を SMIL ファイルに変換する利点は検索結果の保存、公開の利便さである。一つの問い合わせに対して、二つの SMIL ファイルを動的に生成する。

- シングル画面 SMIL ファイル: マルチストリーム型ビデオを単一の画面に映すことを表現している SMIL ファイル。
- マルチ画面 SMIL ファイル: マルチストリーム型ビデオをマルチ画面に映すことを表現している SMIL ファイル。

シングル画面 SMIL ファイルとマルチ画面 SMIL ファイルの違いはレイアウトである。シングル画面 SMIL ファイルのビデオ再生エリアは一つしかないのに対してマルチ画面 SMIL ファイルはビデオ再生エリアが空間的に配置されている。SMIL ファイルの節点としては解区間集合の一つ要素である。例を挙げて見ると図 1 のような区間 1 に含まれている区間 x, y のグルー演算結果は以下に示す。

$$x \oplus y = x_1 \cdot (x_2 | y_1) \cdot y_1$$

この結果を SMIL ファイルに変換すると以下のファイルになる。

```
< par >
  < video src = "x" >
    < anchor id = "x2" begin = "f_s(x2)"
      end = "f_e(x2)" / >
    < anchor href = "y1" begin = "f_s(x2)"
      end = "f_e(x2)" / >
  < / video >
  < video src = "y" begin = "f_s(y1)" >
    < anchor id = "y1" begin = "f_s(y1)"
      end = "f_e(y1)" / >
    < anchor href = "x2" begin = "f_s(y1)"
      end = "f_e(y1)" / >
  < / video >
< / par >
```

例に上げた節点を「seq」という要素を用いてつなげれば、ある問い合わせに対しての検索結果である SMIL ファイルとなる。

6. 結 論

断片的にインデックス付けされたビデオデータに対

して、キーワードにより一連のシーンを検索するとき、一般的にユーザは、意味的にまとまりのある一連のシーンとしての結果を求めている。スジットらにより提案されたグルー演算は、一本のビデオに対して、このような意味的な一連のシーンを求めることができる。しかしながら、グルー演算では同時時間軸の複数のビデオを取り扱う方法は述べられていないため、同期多視点映像に対する検索には有効な手段ではない。本論文では、グルー演算を同期多視点映像を扱えるように拡張し、検索結果の呈示を代数的演算で行う、また演算結果を SMIL ファイルに変換する方法について述べた。

また、今後の課題の一つとしては検索結果のハイパーリンクである。本論文では、拡張グルー演算結果を構成するインターバル間にハイパーリンクをはりし、マルチ画面間を辿ることを実現する方法を提案した。今後は同期多視点映像で撮影されたビデオの空間位置により、拡張グルー演算結果をグループ化し、ハイパーリンクを利用してグループ化されたグルー演算結果間の相互参照リンクについて研究する。

謝 辞

この研究は部分的に文部省科学研究費特定領域研究「高度データベース」の援助を受けています。また、部分的に日本学術振興会未来開拓学術研究推進事業における研究プロジェクト「マルチメディア・コンテンツの高次処理の研究」によっています。ここに記して謝意を表すものとします。

参 考 文 献

- 1) Ramesh Jain. "Data Organization Issues in Presence System". The 5th International Conference on Foundations of Data Organization page 173, November 1998.
- 2) S.Pradhan, K.Tajima, and K.Tanaka. "Interval Glue Operations and Answer Filtering for Video Retrieval". *IPSJ Transactions on Databases*, vol.40 no.(SIG3(DBS & FI1)), 1999.
- 3) 十河 孝至, プラダン スジット, 田島 敬史, and 田中 克己. "Glue/Filter 演算によるビデオ検索モデルと質問処理について". 情報処理学会データベースシステム研究会研究報告 99-DBS-117. pages 59-64. January 1999.