

## マルチエンジン対応型 MSP 動画像検索システムの提案

高橋 秀和 児玉 明 金田 和文 山下 英生

広島大学

〒 739-8521 東広島市鏡山 1-7-1-C112

TEL : (0824) 24-6542

FAX : (0824) 31-4855

E-mail : hide@eml.hiroshima-u.ac.jp

あらまし 我々は、統合的マルチメディアサービスを目指して、マルチメディアスケーラビリティパッケージ (MSP) というパッケージデータを用いた動画像検索システムを提案している。本稿では、本 MSP 検索システム上において、複数の検索エンジンを有するデータベースサーバへの検索要求に対し、MSP データおよび、動画像の時間的あるいは空間的な特徴量を利用した検索エンジンを用いて、効率的な動画像検索を可能とする方法について提案する。また、シミュレーション実験により、本提案システム上で、入力データの特徴に応じて検索エンジンを選択することにより、効率的な検索が行えることの一例を示す。

キーワード MSP 検索システム, 動画像検索, マルチエンジン, エンジンマネージャ, 時間特徴量, 空間特徴量

## A Proposal of MSP System for Video Search Multi-Engine

Hidekazu TAKAHASHI Mei KODAMA Kazufumi KANEDA Hideo YAMASHITA

HIROSHIMA University

1-7-1-C112, Kagamiyama, Higashi-Hiroshima,  
739-8521 JAPAN

TEL : +81-824-24-6542

FAX : +81-824-31-4855

E-mail : hide@eml.hiroshima-u.ac.jp

Abstract We had indicated video searching system using "Multimedia Scalability Packages (MSP)" in integrated multimedia services. In this paper, we propose MSP system which has video search multi-engine. In addition, we indicate the efficient searching method using MSP data and search engine using the temporal characteristics and the spatial characteristics of query video data. Finally, by simulation results, we show an example of the efficient search process which is the suitable search engine for the query video data is selected based on the characteristics of the query data on this proposed MSP system.

key words MSP Search System, Video Search Method, Multi-Engine, Engine Manager, Temporal Characteristics, Spatial Characteristics

## 1. はじめに

我々は、統合的な次世代マルチメディアサービスの実現を目指し、機能化パッケージデータであるマルチメディアスケーラビリティパッケージ (MSP) を利用した通信システム の概念を提案している [1] .

現在、マルチメディアサービスとして、情報検索機能に着目し、その中でも特に、画像検索を中心に検討している [2] . MSP の特徴は、データ自身で動作機能を有することであり、よって、画像検索システムにおいて、画像特徴量抽出、マッチング処理などの検索処理や情報配信などの自動的な処理動作を行うことが可能である。また、画像の検索エンジンとしては、動画像データの時間特徴量、および、空間特徴量を利用した手法を提案している [3][4] .

ただし、従来のシステムでは、1つの検索エンジンで検索を行っており、その検索エンジンに適さない入力データに対しては、マッチング精度が低下するという問題点があった。

そこで本稿では、本 MSP 検索システム上に複数の検索エンジンを有する場合において、MSP データおよび、動画像の時間的あるいは空間的な特徴量を利用した検索エンジンを用いて、効率的な動画像検索を可能とする方法について提案する。そして、入力データの特徴に応じて検索エンジンを選択することにより、効率的な検索が行えることの一例をシミュレーション実験により示す。

## 2. MSP 検索システム

### 2.1 MSP データ構造と検索システム の概念

本 MSP 検索システムで用いる、検索処理用の MSP データの構成の概略を図 1 に示す。

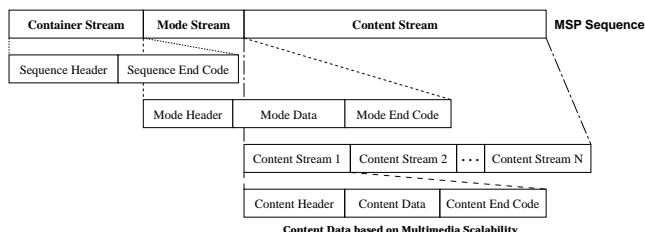


図 1: MSP データ構成の概要

”MSP”は、”共通フォーマット化されているメディア情報と操作情報とその包み紙情報から構成される情報群”を示し、MSP データ構造は図 1 に示

すように、コンテンツ情報 (Content Stream)、振舞情報 (ModeStream)、コンテナ情報 (Container Stream) の大きく 3 つの情報により構成される。コンテンツ情報は、マルチメディアソース情報或は MSP 自身である。ソース情報とは、画像、音声、文書などの表現メディア情報を示す。ある MSP をまた一つのコンテンツ情報の一部として扱うこともある。また、振舞情報 (動作情報) は、MSP 内部の各コンテンツ情報がどのような動作を行うかという情報及び MSP データの振舞の条件に関する情報が記述してある。コンテナ情報は、コンテンツ情報をまとめて作成した MSP データの概要と作成者に関する情報について記述してあり、パッケージデータのヘッダ情報と考えることができ、本のカバーに相当する情報である。以上の MSP データを用いて、検索を行う。

次に、MSP 検索システム の概念を図 2 に示す。本検索システムは、利用者の持っている情報から、利用者の意志を受け、情報検索という機能を持った MSP 情報を生成し、データベースへ送信する。また、データベースにおいて、スケーラビリティを利用した情報検索を行い、その結果を、MSP 検索結果データとして返信してもらい、データを解凍した際に、検索結果を知るというシステムである。ただし、検索データの生成に予め利用者側でスケーラビリティを利用することも可能である。

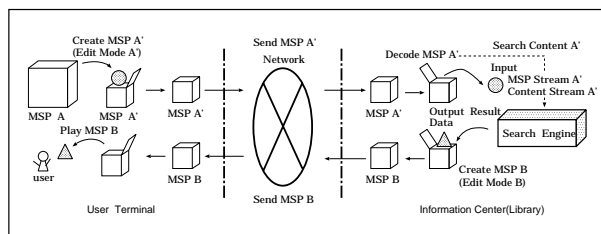


図 2: MSP 検索システム の概念

具体的には、利用者が検索対象とするコンテンツ MSP A から利用者が検索対象とするコンテンツ MSP A から部分的な MSP データ (MSP ヘッダ情報とコンテンツ) を抽出して検索用の MSP データを作成する。ここで新たに生成された MSP データを検索用の MSP A' とする。MSP A' は検索動作情報と利用者の意図する検索項目、検索数や検索時間などの付加的情報を有する。次に、この MSP A' を検索サーバへ配送する。検索サーバでは、受信した MSP A' を復号して、検索指令のデータである

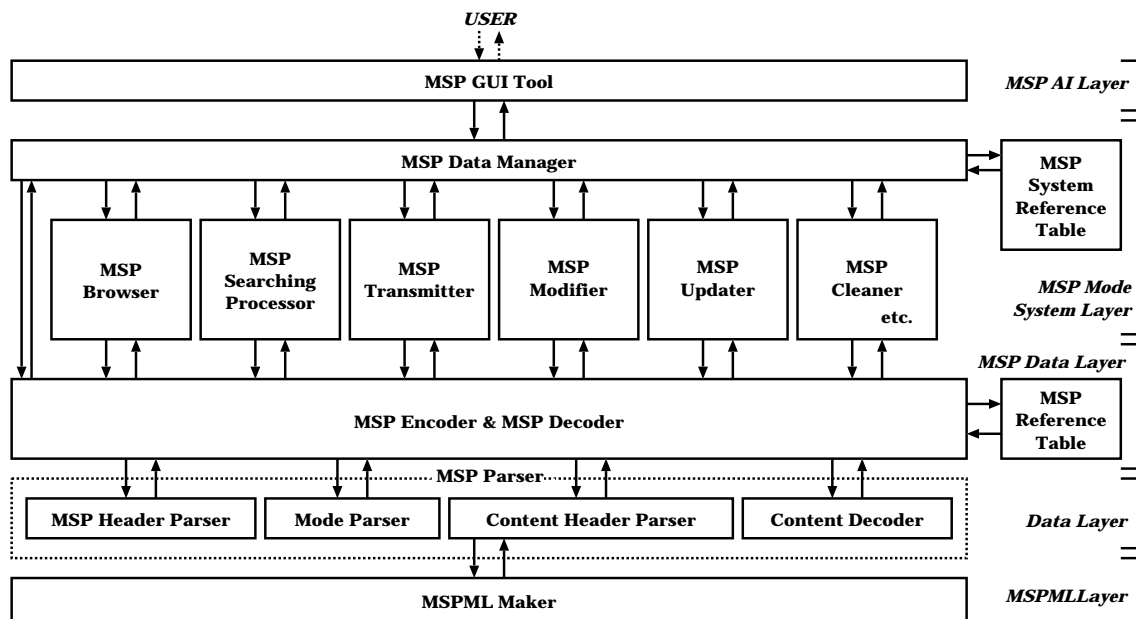


図 3: MSP データ処理の概要

ことを認識する。そこで、検索処理の詳細項目及び対象コンテンツデータを MSP A から抽出し、検索エンジンにかける。検索エンジンからの結果(出力データや対象情報、あるいはネットワークアドレスなど)を元に、新たな MSP データとして MSP B を生成し、検索サーバから利用者へ返送する。ユーザは記述された MSP B のモードデータに基づき、スケーラビリティ機能を利用して、検索結果あるいは要求するデータを取得することができる。

次に、MSP データを動作させる際の処理手順の概要を図 3 に示す [5]。

ユーザは GUI ツールを介して MSP 管理ツールへアクセスする。MSP を解凍する際に、システムの参照情報(動作可能な振舞情報、アクセス許可情報など)を見て、データに記述してあるモード情報(動作情報)にしたがって各々の処理器を動作させる。検索システムの場合、検索処理器、送信器、再生器などを利用する。

具体的に検索システムにおける処理手順について説明する。本検索システムにおいて、MSP 管理ツールは、MSP データの生成を監視する。MSP データが存在した場合、直ちに、MSP データ内の動作情報を解釈する。

まず、クライアント側から考えて行くと、検索用 MSP データを作成すると、検索サーバへデータを送信することを解釈して、MSP 送信処理器を動作させる。サーバ側では、クライアント側と同様に

管理ツールが動作しており、データを受信した後、管理ツールは送信されたデータが検索指令を受けていることを解釈して、検索情報(MSP データ、或は、入力情報)から、利用者が検索するコンテンツ、パラメタなどの項目を選び、検索エンジンに検索対象情報を渡す。

検索エンジンから出力された情報を、検索結果情報として作成し、クライアントへ送信する。クライアントは、検索結果情報を管理ツール、GUI ツールを介して、MSP ブラウザで再生・表示する。また、検索エンジンが備えている検索手法を、クライアント側で検索用の参照テーブルとして用意することでユーザは検索の指定ができる。さらに、本検索システムは、ローカルデータベースか、ネットワークを介したデータベースで検索を実行するかを自動的に判別したり、自動配送するといった、柔軟的な情報検索システムを実現できるのが特徴である。

## 2.2 サーバにおける MSP 検索処理手順の概要

以上の MSP 検索システムにおいて、本稿では、サーバにおける MSP 検索処理手順に着目し、また、サーバが複数の検索エンジンを有する場合の検索システムを提案する。

サーバが複数の検索エンジンを有することにより、ある 1 つの特徴量を用いる検索エンジンでは精度のよい検索が行えないような入力データに対して、その入力データに応じた特徴量を用いる、他の

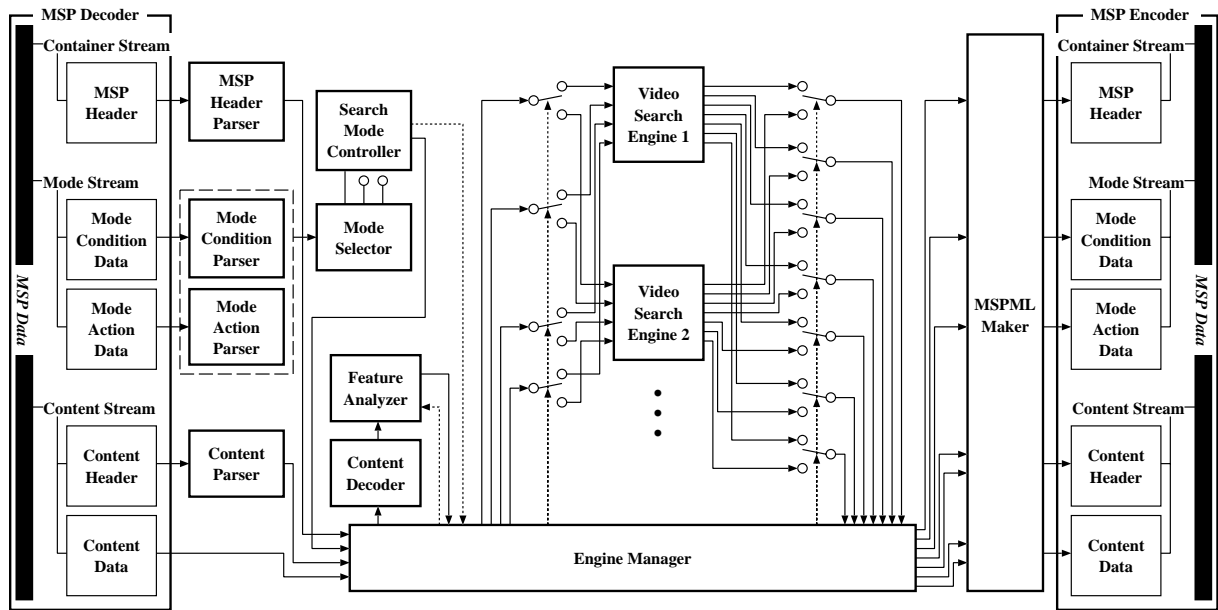


図 4: マルチエンジンを有するサーバにおける MSP 検索処理手順の概要

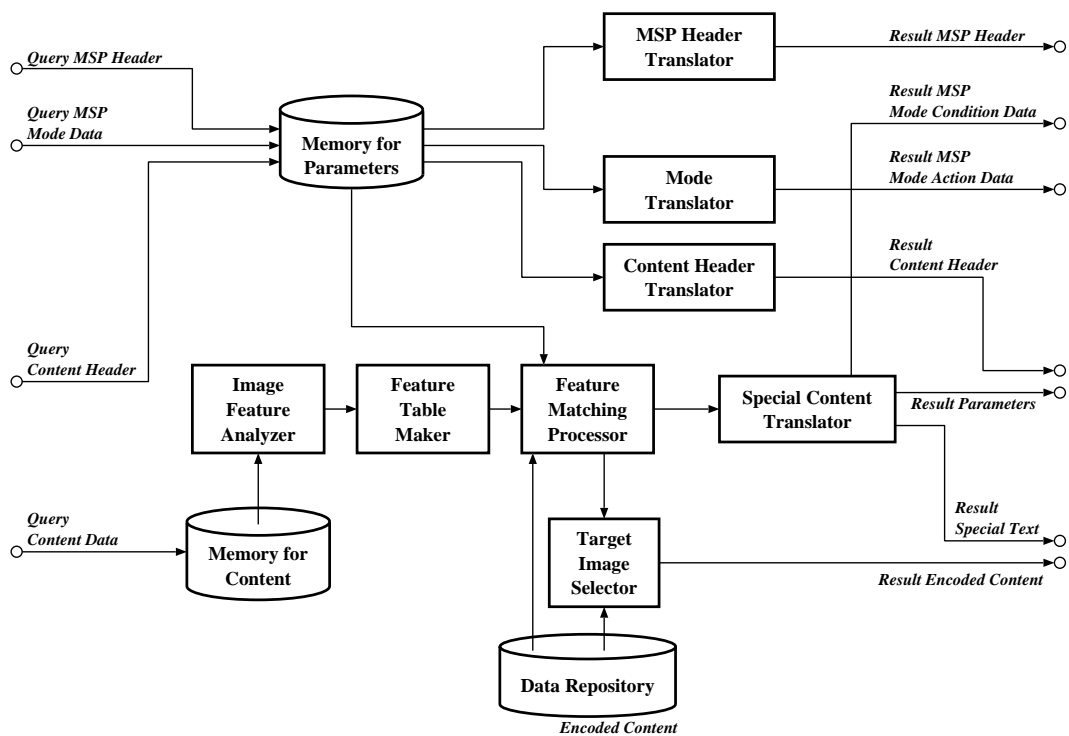


図 5: MSP 検索システムに基づく検索エンジン処理手順の概要

検索エンジンを使用することができる。本稿で提案するシステムは、入力データを判定し、最適な検索エンジンを選択する機能をもつ MSP 検索システムである。本章では、この MSP 検索システムの処理手順について述べる。

マルチエンジンを有するサーバにおける MSP 検

索処理手順の概要を図 4 に示す。入力された MSP データに対して MSP デコードを行い、Mode 情報の解析により、Mode Selector が検索処理器を選択するとともに、各種パーサの解析により、検索に必要な情報を取り出す。そして、Search Mode Controller が、Engine Manager を呼び出す。Engine

Manager は、MSP Header, Mode Data (Mode Condition Data & Mode Action Data), Content Header の情報および、Content Data を入力とし、それらの入力データにより、検索エンジンを制御する役割を持つ。具体的には、コンテンツのデコードを行い、検索用のパラメタとして入力画像データのサイズ、信号種類や閾値などの入力情報により、エンジンを選択するための特徴量を算出し、用いるシステムに登録されている検索エンジンを制御決定する。検索実行後は、検索エンジンから、検索結果用パラメタおよび、検索結果コンテンツが出力される。ここで、検索結果によっては、検索結果コンテンツおよび、必要なパラメタを入力とし、異なる検索エンジンで、再度検索を行うことができる。以上の処理により、最終的に得られた検索エンジンからの出力 (検索結果用パラメタおよび、検索結果コンテンツ) をもとに、MML<sup>[2]</sup> を作成し、MSP エンコードを行い、検索結果用 MSP データが生成される。

### 3. 検索エンジン

#### 3.1 検索エンジン処理手順の概要

前節で述べた、マルチエンジンを有するサーバにおける MSP 検索処理手順の内、検索エンジンについて述べる。

MSP 検索システムに基づく検索エンジン処理手順の概要を図 5 に示す。検索エンジンには、入力 MSP データの各パラメタ (MSP Header, Mode Data, Content Header) と、Content Data が入力される。入力 MSP データの各パラメタは、それぞれのパラメタ変換器により、結果 MSP データ用のパラメタに変換される。

一方、Content Data に対するマッチング処理器は、画像特徴量算出器、特徴量テーブル作成器、特徴量マッチング判定器、対象画像選択器の 4 つから構成され、これら各処理器により、マッチング処理が行われる。まず、画像特徴量算出器は、入力画像よりその特徴量を算出し、算出結果として、特徴量テーブル作成器が入力データ特徴量テーブルを出力する。特徴量マッチング判定器は、入力された入力データ特徴量テーブルに対して、データベースからデータベース特徴量テーブルを取得し、Mode Condition Data より得られる閾値等のパラメタを利用して、マッチング処理を行い、マッチング処理の結果、生成されたマッチング数やマッチング処理

時間等のパラメタを出力する。そして、検索結果により対象のコンテンツを決定し、対象画像選択器が結果用コンテンツをデータベースから取得する。

また、マッチング結果のパラメタは、Special Content Translator により、Mode Condition Data, Content Header のパラメタに変換され、さらに、Special Text として、結果 Content Data に加えられる。

#### 3.2 各検索エンジンにおけるマッチング手法

本稿では、検索エンジンを 2 つと仮定し、それらの検索エンジンにおけるマッチング手法として、時間特徴量のグループ化マッチング手法<sup>[3]</sup>(検索エンジン 1) と、空間特徴量の時間拡張マッチング手法<sup>[4]</sup>(検索エンジン 2) を用いることとする。

##### 3.2.1 時間特徴量のグループ化マッチング手法

時間特徴量のグループ化マッチング手法における、画像特徴量算出器を図 6 に示す。

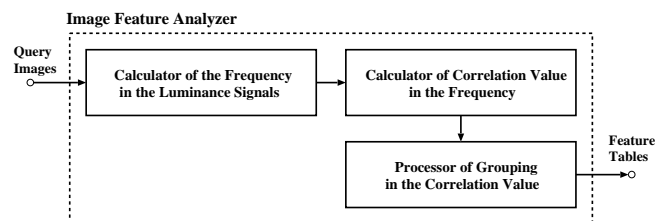


図 6: 画像特徴量算出器

本手法では、1 フレームの特徴量として、画像フレーム内の輝度の頻度分布を用いたフレーム前後間ヒストグラム相関値に着目し、各フレームごとの特徴量のグループ化を行い、特徴量テーブルを作成する。

まず、入力データの輝度ヒストグラム算出を行い、その算出結果を基にフレーム前後間のヒストグラム相関値を算出し、この相関値を特徴量として用いる。特徴量グループ構造の作成手順を図 7 に示す。横軸は動画像の時間方向のフレーム数を示し、縦軸は時間的に連続するフレーム前後間のヒストグラム相関値を示す。本特徴量算出手法は、動画像のヒストグラム相関値が変化が激しいところを除き、1.0 に近い傾向を利用し、それ以外の部分を特徴量として抽出する方法である。この特徴量を動画像データのマッチングに利用する。具体的には、ヒストグラム相関値のある閾値  $Th_{tbl}$  より高いフレーム

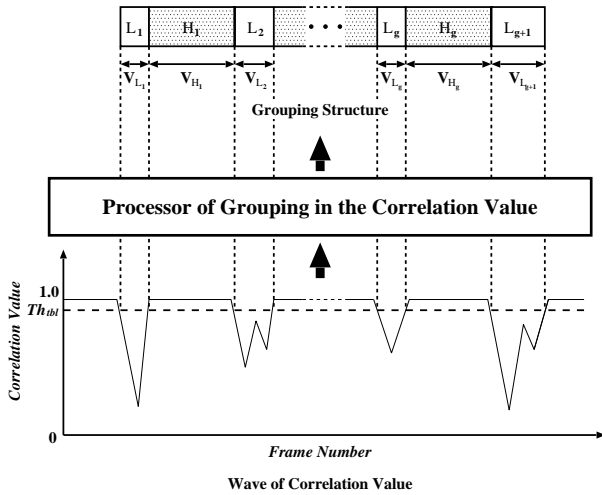


図 7: 特徴量グループ構造の作成

構造を H または、低いフレーム構造を L としてグループ化し、この各構造の長さや総構造数により判定する方式である。

本手法の特徴としては、特徴量のグループ化を行うことにより、1 フレーム 1 特徴量を用いたマッチング手法より、高速にマッチングを行うことができる。しかし、画像サイズや符号化劣化などのデータベーステーブルとの誤差があるコンテンツに対して、マッチング精度が低下するという特徴を持つ。

本手法は、相関値が大きく変化する点を特徴としてグループ化を行っている。データベースに用意されているオリジナル画像の特徴量テーブルとの比較を行う際、符号化劣化の誤差をもつ入力データに対しては、相関値の変化が小さい、つまり、動きの変化が少ない場合、データベーステーブルと入力テーブルの少しの相関値の違いがグループ構造の違いにつながり、マッチング精度が低下してしまう。しかし、相関値が大きく変化する部分においては、両者の違いは少ないため、符号化劣化の誤差をもつ入力データに対しても精度を低下させることなく、マッチングを行うことができる。

一方、画像サイズの誤差をもつ入力データの場合、画像サイズが小さくなるにつれて、全体の相関値が小さくなるため、相関値が大きく変化する点を特徴とすることが難しくなり、検索目標にマッチングすることができなくなるという問題点がある。

### 3.2.2 空間特徴量の時間拡張マッチング手法

本手法は、画像の特徴量として、一枚の画像フレーム内の空間的特徴量を、連続するフレーム間で

利用する、動画検索のための時間拡張検索方式である。検索に使用する空間的特徴量として、本手法では単純な検索用特徴量算出を行うため、画像の輝度値の平均値に着目した。

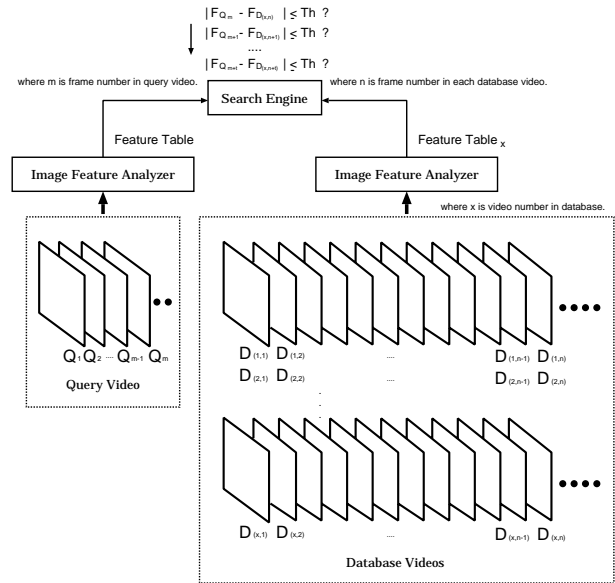


図 8: 空間特徴量の時間拡張動画検索方式

具体的には、図 8 に示すように、特徴量を入力画像、データベース画像それぞれで算出し、特徴量テーブルを作成し、動画単位で、特徴量テーブルを用いてマッチング判定を行う。マッチング判定は、まず、入力動画(検索対象動画)の 1 フレーム特徴量のマッチングを行い、判定式を満たした場合、入力動画の次フレームとデータベース動画の次フレームと時間方向に対して拡張する方法である。用いるマッチング判定式を式 (1) に示す。

$$|F_{Q_{m+t}} - F_{D_{(x,n+t)}}| \leq Th \quad (1)$$

ここで、 $F_{Q_m}$  は入力動画(検索対象動画)の  $m$  番目のフレームの特徴量を表し、また、 $t$  は時間方向に拡張するフレーム数を表す。判定に  $Th$  の閾値を利用する。

本手法の特徴としては、画像の空間サイズ(空間スケーラビリティ<sup>[6]</sup>)に対する特徴量誤差が発生した際、誤差の影響を受けにくい平均値を用いていること、および、時間拡張を行うことにより精度を低下させることなく、マッチングを行える。しかし、1 フレームに 1 つの特徴量を持つため、マッチング時間がフレーム数に依存するという問題点がある。

## 4. シミュレーション実験

### 4.1 実験条件

表 1: データベース用動画データ

シーケンス	アニメ, スポーツ 1, スポーツ 2, 音楽 1 音楽 2, 映画, ニュース, スポーツ 3 スポーツ 4, スポーツ 5, Bus Flower Garden, Mobile & Calendar Popl, Table Tennis 計 15 シーケンス
データベース フレーム数	9000 フレーム (アニメーション, スポーツ 1, スポーツ 2 音楽 1, 音楽 2, 映画, ニュース, スポーツ 3 スポーツ 4, スポーツ 5) 150 フレーム (Bus, Flower Garden Mobile & Calendar, Popl, Table Tennis) 計 90750 フレーム
画像 フォーマット	ITU-R BT.601 4:2:0 704[pe]×480[line]

表 2: 入力用動画データ

入力 データ	シーケンス	画像サイズ	データベース テーブル との誤差
A	Bus	704[pe]×480[line]	なし
B	Bus	704[pe]×480[line]	あり
C	アニメ	704[pe]×480[line]	なし
D	アニメ	704[pe]×480[line]	あり
E	アニメ	352[pe]×240[line]	あり

以上の提案 MSP 検索システムおよび、2つの検索エンジンを用いて、本システムの評価実験を行った。今回シミュレーション実験に用いたデータベース動画データのデータ諸元を表 1 に示す。表 1 のシーケンスの順で、各検索エンジンに対するデータベース特徴量を、オリジナル動画データから作成した。また、入力動画データとして、データベース内の部分動画データの 150 フレーム (約 5 秒相当) を表 2 に示す各条件のもとで用いた。ここで、データベーステーブルとの誤差がある入力データを、B と D は 9.0[Mbps] の MPEG2 データ、E は非圧縮の縮小画像データとした。なお本実験には、Fujitsu GP 7000SM25 400MHz UltraSPARC-2、メモリ 512MB の計算機を用いた。

本実験では、検索エンジンを選択する基準となる、入力データの特徴量抽出について、以下のような方法を用いた。本実験で用いる時間特徴量を用いる検索エンジン 1 は、3.2.1 節で述べたように、相関値が大きく変化する点を特徴としており、相

関値の変化が小さい入力データに対しては、特に、符号化劣化誤差をもつ場合にマッチング精度が低下してしまう。以上の理由から、符号化劣化誤差をもつ入力データにも対応できるように、相関値の変化が大きい入力データに対してのみ、検索エンジン 1 を用い、相関値の変化が小さい入力データに対しては、検索エンジン 2 を用いることとする。そこで、エンジン選択に用いる特徴量として、図 9 に示すように、フレーム前後間のヒストグラム相関値のダイナミックレンジ  $d$  を利用する。動画のフレーム前後間相関値は、シーンチェンジなどを除いて、0.9 以上の 1.0 に近い値をとるため、すべての相関値が 0.9 以上である、つまり、ダイナミックレンジが 0.1 以下である入力データに対しては、検索エンジン 1 は適さない。よって、ダイナミックレンジ  $d$  が 0.1 より大きいかどうかをエンジン選択の判定基準とする。

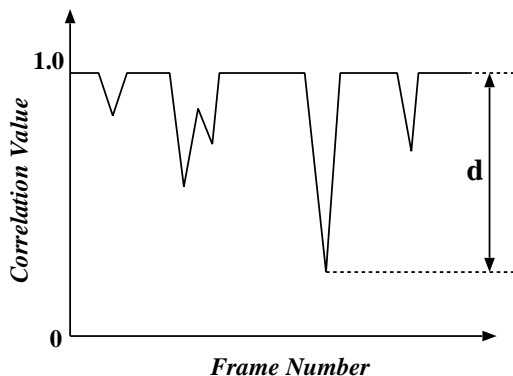


図 9: エンジン選択に用いる特徴量

また、検索エンジン 1 が、画像サイズの誤差を持つコンテンツに対しても、マッチング精度が低下するという特徴を持つことから、縮小サイズの入力データの場合には、エンジン選択のための特徴量判定を行うことなく、検索エンジン 2 を選択し、検索を行うこととする。

### 4.2 実験結果および考察

4.1 節の実験条件のもとで、本 MSP 検索システムのサーバにおける以下の各処理時間を測定した。

- 入力 MSP データの MSP デコード時間:  $T_{md}$
- 検索用符号化コンテンツのデコード時間:  $T_{cd}$
- エンジン判定用の特徴量抽出時間:  $T_{fe}$
- 入力動画コンテンツの特徴量算出時間:  $T_{fc}$

表 3: サーバにおける MSP 検索システムの各処理時間とマッチング数

入力データ	$T_{md}$ [s]	$T_{cd}$ [s]	$T_{fe}$ [s]	$T_{fc}$ [s]	$T_m$ [s]	$T_{me}$ [s]	マッチング数	使用エンジン
A	5.840	-	14.999	18.214	0.427	0.0223	1	エンジン 2
B	0.418	19.663	14.995	18.230	0.428	0.0217	1	エンジン 2
C	5.757	-	14.884	13.631	0.00492	0.0222	1	エンジン 1
D	0.414	20.374	14.908	13.630	0.00439	0.0228	1	エンジン 1
E	1.831	-	-	4.965	0.427	0.0216	1	エンジン 2

- マッチング処理時間:  $T_m$
- 検索結果 MSP データの MSP エンコード時間:  $T_{me}$

表 3に上記の各処理時間, および, 使用した検索エンジンの結果を示す. 表 3より, 入力データに応じて異なったエンジンが選択されていることが分かる. これは, 今回用いた入力データが, Bus は動きの少ないシーンであるのに対し, アニメはシーンチェンジを含むシーンであり, それぞれの入力動画の特徴がエンジンの判定に用いられていることを示している.

また, A から D の入力データに対して, エンジン 1 を用いた場合とエンジン 2 を用いた場合の各処理時間を比較してみると,  $T_{fc}$  で約 1.3 倍,  $T_m$  で約 87 倍から 97 倍, エンジン 1 の処理時間が高速であることが分かる. マッチング時間の点から見ると, エンジン 1 の方が優れているが, 上で述べたように, エンジン 1 はシーンの動きが少ない入力データには適していないため, そのような入力データに対しては, エンジン 1 よりマッチング時間を要しても, シーンの動きに影響を受けないエンジン 2 が選択されており, 本システムが, マッチング時間とマッチング精度の両方の面を考慮したエンジン選択を行っていることを示している. このエンジン選択を行うことにより, 符号化劣化をもつ入力データに対しても, 本システムでの最適な検索が行えていることが分かる.

ただ, エンジン選択のための特徴量抽出時間に, B の場合で, 全体の処理時間の約 28%, C の場合には, 約 43% を要しており, より短時間で効率的にエンジン選択を行うための特徴量抽出処理の検討については, 今後の課題である.

また, 入力データ E については, 画像サイズが小さいことから, 検索エンジン 2 による検索が行われている. 画像サイズが 4 分の 1 であることから, 同じエンジン 2 を用いた場合と比較して,  $T_{fc}$  が約

4 分の 1 になっていることが分かり, 画像サイズの誤差を持つコンテンツに対しても, 本システムで検索可能であることの一例を示した.

## 5. まとめ

MSP 動画検索システムとして, サーバが複数の検索エンジンを有している, マルチエンジン対応型の MSP 検索システムについて, その処理手順の提案を行った. また, 本システム上において, 入力データの特徴量を判定することにより, 入力データに適した検索エンジンを用いて検索が行えることの一例を, シミュレーション実験により示した. 今後の課題としては, エンジン選択のための特徴量抽出処理の高速化, クライアント側の処理や, 通信処理を考慮した, MSP 検索システム全体の評価などが挙げられる.

## 参考文献

- [1] 児玉 明, 笠井 裕之, 村井 正人, 富永 英義: “マルチメディアアブローディングとその情報構成に関する検討”, 信学技報, IN 96-122, OFS 96-10 (1997).
- [2] 児玉 明, 竹本 正行, 池田 朋二, 真崎 剛: “MSP 画像検索システムにおける情報構成法の考察”, PCSJ 99, pp. 53-54 (1999).
- [3] H.Takahashi, M.Kodama, K.Kaneda and H.Yamashita: “The Evaluation of Video Retrieval Methods from Temporal Characteristics”, PCS2001, TP4-2, pp. 251-254 (2001).
- [4] M.Kodama, T.Ikeda and H.Takahashi: “The Evaluation of Video Matching Methods Using Spatial Scalable Characteristics in MSP Search System”, PCS2001, TP4-3, pp. 255-258 (2001).
- [5] 児玉 明: “MSP 交換型高性能データ検索システムに関する研究開発”, 通信・放送機構, 産学連携支援型研究開発制度最終成果報告書, pp. 301-352 (2000).
- [6] ISO-IEC 13818-2, International Standard: “Information Technology – Generic Coding of Moving Pictures and Associated Audio Recommendation H.262” (1995).