

## 映像データの収集利用方式の開発

星野 聰

京都大学大型計算機センター

人文科学の研究者が、自分自身で映像データを経済的に収集利用するためのシステムの開発を行っている。画像はビデオカメラから入力する。コンピュータによってビデオカメラを制御し、録画開始時のビデオテープの位置やセンサーによる付加情報を自動的にコンピュータファイルに記録し、これらの付加情報を用いて映像の頭出しをすることにより、速やかにビデオ画像を検索できる。センサーとしては、撮影場所の地理的位置や地磁気による方位のセンサーを使用する。

このシステムでは、映像ファイルのサブファイルの作成やマージなども可能で、ワークステーションの表示画面上で操作できる特徴がある。また、カメラのパンの方向やズームの検出、ビデオ画面の字幕の検出も試みている。

The System Development for Image Data Collection and its Use

S. Hoshino

Data Processing Center, Kyoto University

Yoshida Honmachi, Sakyo-Ku, Kyoto 606, Japan

This paper deals with a system developed to collect and use image data economically for researchers engaged especially in humanity studies.

Images are input by an inexpensive camera, which is controlled by a computer. The video-tape position at the start of each recording and the information read from sensors detecting the geographical location and camera direction at the instant are transmitted into a file of the computer. The additional information enables to make image directories by which the required video image can be retrieved effectively by forwarding and positioning the video-tape to the beginning position of the required image.

The recognition of camera operations such as the panning and zooming of camera and the detection of the appearance of characters on video image have been tried, which will be efficient for retrieving image data.

## 1 はしがき

歴史や文学などの人文科学の研究者による文字テキストの利用のためのシステムについては、先に報告したが、「<sup>(1)</sup>引き続いて、ここでは、研究者、特に人文科学の分野の研究者が自分自身で映像データを収集し、利用するのに適したシステムを開発し、その実用化を目指している。<sup>(2,3,4,5)</sup>

さて、研究者を対象とする場合には、商業的な映像作品の作成利用とは異なる面が多い。例えば、(1) 内容が専門的であり、特定の関心を有しない限り、興味のあるものではない。利用も研究者に限定されることが多い。そこで、目的意識を有している研究者自身が作成するのが効果的である。(2) 資料を持ち帰ることができないときでも、現場で撮影できるかも知れない。

(3) 研究者が収集する映像は大量になると思われる所以、映像の収集・入力・利用が経済的で労力も少ないことが望ましい。また、個人ベースで研究が行われることが多いので、負担し得る経費は限られている。

画像・映像情報の重要性は文字情報に比して認識が薄いように思われる。しかし、歴史史料は文字でなく、画像や映像として残っているものも多く、この価値は高く、もっと有効に利用すべきである。そこで、情報処理技術を用いて、画像・映像情報を研究に活用することを目標としている。

さて、映像は、個人によって興味の対象が大いに異なる。従って、一般に流通している映像で満足できるとは限らない。そこで、個人的に作成できる必要がある。画像はビデオカメラから入力するのが経済的であり有効である。ただし、後で映像のデータベース作成や検索に利用するための情報(付加情報)をできるだけ取り込んでおくべきである。これは文献データベースの場合に、文献を特定するために書誌情報を付加することに相当する。但し、画像、特に動画像では、いちいち文字情報を付すのはわずらわしい。そこで、付加情報を、出来れば自動的に作成し、記録することを望ましい。そのため、各種のセンサーを用いることを提案する。この情報には、カメラの状態や、種々のセンサーが検出する撮影地点の地理的位置や撮影方位などの環境情報が含まれる。他の問題点として、テープを全部再生するのではなく時間がかかる。そこで、付加情報を用いて映像の頭出しをすることによって、速やかに映像を検索できる。

このシステムでは、映像ファイルのサブファイルの作成やマージなども可能で、ワークステーションの表示画面上で種々の操作ができるので能率がよい。また、ビデオ画像を効果的に検索するた

めの手掛かりとして、カメラのパンの方向やズームの検出、更にビデオの字幕の検出についても試みている。

## 2 システムの特徴

システムに対する要件を、さらに詳しく述べる。

(1) 研究の性格上、研究の過程でシステムの機能の追加や変更などができる必要がある。

(2) 検索のキーは自動的に付加されることが望ましい。研究者は多忙であるから、撮影後に文字データを映像に対応させるような作業は避けたい。検索のための情報として、例えば撮影地点の地理的位置に、目印となるものがない地域の場合や、あるいは写っているが、名称不明の寺院や門などを特定したい場合などもある。あるいは、写っている事物を特定するために撮影方位を知りたい場合もある。このためには、種々のセンサーを使用し、カメラとセンサーを共にコンピュータで制御すればよい。ただし、携帯して行きたいので、軽量・小型・小消費電力でなくてはならない。

(3) 映像を編集して、同じ主題について撮影している部分だけの映像を抽出し、それらをつなぎ合わせられることが望ましい。編集された映像に対して、検索できるべきである。

(4) 一定時間ごとに短時間づつ撮影する。あるいは被写体に変化があった時など、あるイベントが生起した時にだけ撮影するなど、カメラの外部から記録の開始・終了を制御できることが望ましい。

(5) 映像に対して、文字情報を付加して、検索・表示できることが望ましい。

(6) テキストを撮影する際には、研究の目的や資料に記されている文字などの細かさ、資料の性格(例えは、手書きの古文書、手紙など)で決まる、あるレベル以上の分解能がないと、研究に役立たない。古文書では濃度も表現しないと文字が判読できない可能性がある。また、横長の資料や大きな地図なども入力したい。

## 3 開発中のシステムの方式

これらの要件を満たすために、図1に示すような方式を採用している。即ち、

### 3.1 カメラによる入力

安価な8mmビデオカメラを用いて、映像を記録したビデオテープを使用する。操作が容易であるから、研究者自身で撮影できる。カメラは外部から

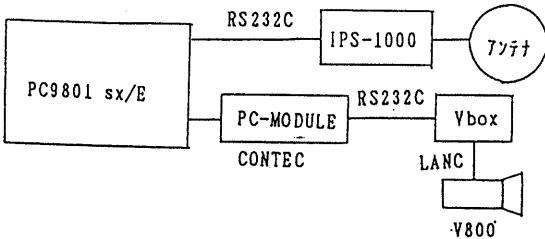


図1 カメラ及びセンサーのPCによる制御

制御し、またカメラの状態を問い合わせる機能が必要である。またカメラは、ビデオテープの再生・録画機能も有しているものとした。

### 3.2 画像検索機能

映像のなかから、その内容を推測できる可能性がある場面を静止画像としてWSに読み込んで、画像ファイルとして格納し、さらにそれらの静止画像データを間引いて縮小した64個の静止画像から成る画像ディレクトリ（静止画像ファイル）を、自動的に作成する。図5は、ワークステーションの表示画面で、中央上部が画像ディレクトリである。このような処理を行うには、静止画像を抽出するタイミングを与える情報（頭出し情報）が必要である。この抽出の方式は後に述べるが、この頭出し情報に従ってビデオテープは起動・早送り・停止を自動的に繰り返して、静止画像をファイルに転送する。これを基に作られる画像ディレクトリを用いて、ビデオテープから所要の画像を検索するには、まずユーザが画像ディレクトリ内にある所要の縮小画像をマウスで指定すると、その画像が間引かれない以前のサイズで表示画面の上部に大きく表示され、（図6参照）画面のplayボタンをピックすると、ビデオテープはその画像が格納されている位置まで送られてから、再生を始める。再生は画面のstopボタンによって、任意の処で停止させることができる。

画像ディレクトリを用いると、静止画像を一覧できるので、映像を効果的に検索できる。また、この静止画像に文字情報を付加することもできる。これらの情報は、ハードディスク、光磁気ディスク又は安価なCD-ROMに格納して保存し利用する。

#### 3.2.1 ハードウェアによる方式

撮影の際ににおいて、カメラの録画ボタンを押して撮影を開始する毎に、その瞬間のビデオテープ上の録画位置が自動的にカメラから読み出され、

```

0001 テ-7' 仔 hour=00 min=00 sec=06
0001 シ-カン Wed Mar 31 00:13:48 1993
0001 エイ 35 ト' 01 7' 24.7 ピ-ヨウ
0001 トケイ 135 ト' 47 7' 05.4 ピ-ヨウ
S
0072 テ-7' 仔 hour=00 min=09 sec=22
0072 シ-カン Wed Mar 31 00:59:20 1993
0072 エイ 35 ト' 01 7' 15.0 ピ-ヨウ
0072 トケイ 135 ト' 47 7' 15.1 ピ-ヨウ

```

図2 フロッピーディスクへの  
画像の付加情報

コンピュータのフロッピーディスクに、他のセンサーからの情報と共に記録される。その内容を図2に示す。録画ボタンを押す毎に、新しい対象物の撮影が始まることが期待されるので、その瞬間を頭出しの時点として利用するのである。以下において、ポケット型コンピュータ、パーソナルコンピュータ、ワークステーションをそれぞれ、PKC、PC、WSと略記する。

さて、カメラによる撮影が終了すると、研究室でWSにVbox (SONY製：ビデオ／コンピュータインターフェースアダプタ) CI-1000<sup>(6)</sup> 経由でカメラを接続し、WSによってカメラを制御する。ビデオカメラには、SONY製の8mmビデオカメラ CCD-V800及びCCD-VX1を用いた。カメラは、図1に示すように、Vbox を経由して、コンピュータとRS-232Cにより接続している。カメラはVboxからの制御信号 (LANC信号) より制御される。但し、IPS-1000とアンテナは、GPS (Global Positioning System) で、<sup>(7)</sup> 地理的位置 (東経、北緯) を知るためのセンサーで、年月日・時間の情報も得られる。

付加情報はPCのフロッピーディスクからWSのファイルにコピーして利用する。この方式によって、録画を開始する瞬間の画像を確実に捉えることができる点で優れている。図5は、筆者がカメラとGPS、PCを携帯して撮影した景観から成っている。

#### 3.2.2 ソフトウェアによる方式

従来、映像の頭出しの位置として、カットの変わり目を探るものがあり、これを自動的に見い出すには、ソフト的に色濃度の変化をもとにして、カット変わりを検出するものがある。<sup>(8)</sup> ここでは、景観ビデオ画像を対象としたが、カットの切れ目（ここでは、録画ボタンを押した瞬間）で、景観の配色があまり変わらないことがあり、確実にカ

ットの変わり目を捉えられるとは限らない。しかし、ハードウェアによる頭出し情報が得られない場合に、頭出しを自動的に行うにはソフト的に判定する必要がある。

ここでは、画像資料が多量であるとしているので、ビデオを通常の速さで再生しながら処理を行うものとし、また画面はゆっくり変化したものとした。さらに撮影は、同時に次の三つの撮影動作（モードと呼ぶ）の一つのみから成ると仮定した。即ち、

- (1) 水平方向のパン（左または右）
- (2) 垂直方向のパン（上または下）
- (3) 一定方位に対するズーム（拡大または縮小）

ビデオ映像は、(280\*210)ピクセルのサイズに縮小してディスプレイに表示する。従って、細部は表示されないが、大体の画像の変化（移動・拡大・縮小）を把握するのが目的である。このため、ある時点に於ける画面 $I_0$ と、短い時間後( $I_0$ に対しての、ここで述べる処理が終了した直後の) $I_1$ とを比較する。 $I_0$ 上で、その中心を原点とするとき、ピクセル位置が( $n*i, n*j$ )の点 $c_{ij}$ , ( $i, j=-4, -3, \dots, 3, 4$ ),  $n=20$ に就いて、それらの近傍で間隔が $\delta=4$ （ピクセル）の点も考える。操作は以下の通りである。

#### (i) 水平方向へのパンの検出

$I_0$ の $c_{ij}$ 点より右に $\delta*p$ だけ, ( $p=-4, -3, \dots, 3, 4$ ) 移動した各点での画像のカラー値と $I_1$ の $c_{ij}$ における画像値との差の絶対値の平均 $a_h$ を計算する。

#### (ii) 垂直方向へのパンの検出

同様に、上方に $\delta*p$ だけ移動させて(i)と同様のことを行い $a_v$ を求める。

#### (iii) ズームの検出

$p=-4, -3, \dots, 3, 4$ に対して、 $I_0$ の $(0, (n+\delta*p)*i)$ ,  $((n+\delta*p)*i, 0)$ ,  $((n+\delta*p)*i, (n+\delta*p)*i)$ の各点の画像値を、 $I_1$ の $(0, n*i)$ ,  $(n*i, 0)$ ,  $(n*i, n*i)$ , 但し $i=-4, -3, \dots, 3, 4$ と比較して差の絶対値の平均が最小になる $p$ とその時の平均値 $a_z$ を求める。

(i) (ii) (iii) で得られた $a_h$ ,  $a_v$ ,  $a_z$ のうちで最小をとるケースから、パン（上下左右）やズーム（拡大・縮小）をしているか、していないかを判定する。

筆者が撮影した景観のビデオを用いたテストでは、ゆっくり水平又は垂直方向のパン又はズームをしていると、この判定は正しく行われていることが多い。しかし、誤って判定することがあり、二つ以上の撮影モードが同時に行われた場合が含まれるものと思われる。画面にコントラストが乏しい場合、例えば山が霞んでいるような場合にも、

有効な判定が行われないか誤るのが見られる。

この方式はさらに調査と改良を要し、定量的な評価をしなければならないが、動作モードが変化する時点で、映像の頭出しをしようとするのが最終目標である。

ディスプレイ上の表示を図6に示す。動画は左下の窓に表示され、その右の窓には、上記の処理が行われている画像が写されている。左端の下から二番目のボタンの←印は、カメラが左へパンしていることを示している。

### 3.2.3 その他の方

頭出し情報が記録されていない場合には、WSからビデオカメラを制御して、テープの前進・後退・停止などを指示して、隨時、利用者が頭出し情報を記録できる。但し、時間と労力がかかる。もし、他に附加情報を得る方法がなければ、一定時間毎に頭出しをすることが考えられる。

### 3.3 センサーについて

地理的位置のセンサーであるGPSの情報をパソコン側に取り込むことができる。磁気方位センサーも、コンピュータとRS232Cで接続している。通常カメラは前方に構えるであろうから、このセンサーを撮影者が装備していると、カメラの撮影方向を知る手掛かりになるだろう。

### 3.4 画像ファイルの作成

画像ディレクトリを利用して、必要な部分を選択し、サブファイルを作成できる。即ち、静止画像ファイルと、その縮小画像から成る画像ディレクトリが収められているディレクトリのサブディレクトリを作り、その上に、利用者が指定した範囲の画像のみから成る画像ファイル、即ち画像ディレクトリや、附加情報などを格納したファイルを作成する。これを画像サブファイルと呼ぶ。静止画像ファイルはサイズが大きいので、必要ならば、元のディレクトリにあるから、それをアクセスすればよい。従って、静止画像ファイルはサブディレクトリ上にはコピーしていない。

### 3.5 画像サブファイルのマージ

指定した二つの画像サブファイルをマージし、それらに対する静止画像ファイル・画像ディレクトリ・附加情報ファイルを合成することができる。映像データは、ソースになる映像のテープを装着

するカメラの他に、別のカメラを用意する。ソースになるテープを順に装着すると、受け側のカメラにビデオがコピーされる。

### 3. 6 文字列の付加

画像ディレクトリ上で、指定した画像に対して、説明文を WS の画面上で入力できる。

### 3. 7 字幕の検出

現在の情報はやがて歴史史料となる。前述のように映像情報も重要な史料となるから、例えば、現在のニュースを記録し、利用できるようにすべきである。ただし大量になるから、有効に利用するための方式を開発しなければならない。そこで、字幕が出ると、字幕部分を画像として抽出し、多くの字幕を一覧できるようにすると、ニュースの内容を能率よく検索できるであろう。もしニュースの時間がわかっていると、カメラを WS から制御して、ニュースだけを収録することが出来るであろう。また所要の字幕を指定すると、その字幕の位置からビデオを再生することが出来るであろう。

字幕の文字は白色で明るく、周辺に黒い縁があるものとした。ところが、実際には、文字に薄くブルーやオレンジ色に着色されており、文字と同色の濃い色の縁が付いていたり、あるいは文字とは別の色の太いアンダーラインが付いていて、文字のアンダーライン部分には黒の縁が付いていないこともある。これらは字幕の認識には不利であり、字幕が見分けやすい様に考慮されることが望まれる。字幕は一般に、

(1) 文字の意味が認識されるよう、ある時間は同一の文字列が表示されている。

(2) 内容を伝えるために、一定以上の文字から成る。

これらの特徴は、字幕の認識で有利になる。

字幕の検出も、ビデオを通常の速さで再生しながら行いたい。そこで、検出の手続きは簡単でなければならない。

筆者は、WS として NEWS3720 を使用し、画像はビデオカメラから Vbox 経由で、WS にビデオマップボード NWB254 によって入力している。ディスプレイ上のビデオ画面（図 7 の中央上部）は、(640\*480) ピクセルで、この内で字幕が現れると予想される部分として、上から (375~444) ピクセルの部分（サイズは、(640\*70) の画像を抽出し、図 7 の中央下部に読み出して表示している。また、画面左下端のボタンに、判定の結果を示すカウント c を表示した。字幕が現れると、c の値が大き

くなると期待される。

字幕が表示されたことを検出するには、最初に  $c = 0$  としてから、上記の字幕が現れると予想される部分の画像に、垂直または水平の辺がどれ程あるかを調べる。実際に文字が書かれていないので、文字の辺であると誤って判定されることもあり得る。もし、辺と判定されたケースの個数が多ければ、字幕と判断する。

垂直の辺を判定するには、字幕を調べる範囲の各行に就いて、白色に近く、明るい部分が見つかるたびに、次の処理を行う。即ち、そのコラム  $x_{\min} < x < x_{\max}$  が辺にふさわしい巾  $d = x_{\max} - x_{\min}$ ,  $0 < d < 8$  を満足するものが見つかったとし、 $x_{\min}, x_{\max}$  の、その上の行に於ける位置からのずれが一定値（テストでは 2）を超えるとき  $e=0$ 、その一定値以下で、しかもそのような行が一定個数（テストでは 3）以上続くとき、縦の辺とみなし、 $e$  に 1 を加える。但し、一旦、辺の巾が大きくなり、横方向の辺と判定された後、元の辺の延長上に近い位置に戻った場合には、もとの  $e$  の値を生かして、縦方向の辺は続いていると判定する。もし、この条件を満たさなければ  $e=0$  とする。

横方向の辺は、横巾が 2 以上、24 以下であるとした。もし、これを満たさなければ  $e=0$  とした。もし横巾の条件が 5 回以上続いて満たされると、辺より太いと判断して、辺ではなかったものと判定し、 $e$  の値を補正する。

このように白色に近く、明るい部分のすべてにつき計算される  $e$  値を  $c$  に加算する。これを、字幕を調べる範囲に就いて行う。

NHK ニュースをビデオに録画したものに就いてテストした結果では、白色で黒い縁があり、文字サイズが小さくないときは、カウント  $c$  は 100 ~ 250 程度になり、文字がないと、0 ~ 50 程度になるので字幕の検出に有効である。しかし、背景を文字の一部と見誤ることがある。色がついた文字に同色の縁がある場合も文字と判定されないので、改良すべき点が多い。

### 3. 8 その他の操作

その他の操作として、

- 一定時間毎に映像をカメラに短時間づつ録画させると、長時間にわたる画像のサンプリング記録ができる。例えば、五分毎に五秒間だけ録画すると、一日の雲の変化を観察することができる。

- ビデオ画像をデジタル的に取り出して、WS で若干個の点で画像の変化が認められるときだけカメラを起動することができる。

- 複数のカメラを使用すると、画像入力の分解

能を向上させることができる。二台のカメラを使用して、テキストを上下に分けて撮影し、各々ハードディスクなどのファイルに記録する。WSでは、上下のテキストをつなぎ、さらに左右にスクロールして見ることができる。(図4参照) (4)

これらの操作をすべて WS の画面上で操作できるので、作業の能率を高められる。

#### 4 システム開発の状況

上述のように、録画ボタンを押したことを検出すると、自動的にカメラやセンサーからPCのフロッピーディスクに、時間・テープ位置・地理的位置などを読み出して記録する。撮影後にカメラを WS に VBOX 経由で接続して、カメラを制御して、静止画像のファイルへの取り出しと画像ディレクトリの作成に利用している。

この方式のテストに、ノート型PC (PC9801 sx/e) と、RS232C拡張ユニットを使用して、VboxとGPSを接続した。(図1参照) カメラの制御用端子 (LANC端子) に、Vboxへのコードを接続しておく必要があるが、カメラ操作は通常と変わらない。

しかし、これは携帯には重すぎるので、カシオ製ポケットコンピュータ FX-890P (16ビットCPU) を使用できるようにした。但し、RS232Cの出口は一つしかないので、これを分岐して使用するためのインターフェースを自作した。(図3) (4)

PKCとセンサーを含む総重量は約3kgである。(PKC 650g、分岐用インターフェース 270g、Vbox 525g、GPS 600g、バッテリ類 630g、定電圧ユニット 340g) 磁気方位センサー (TMC-2000, トヨタ製) とRS232Cインターフェイス (自作) は、計800gである。

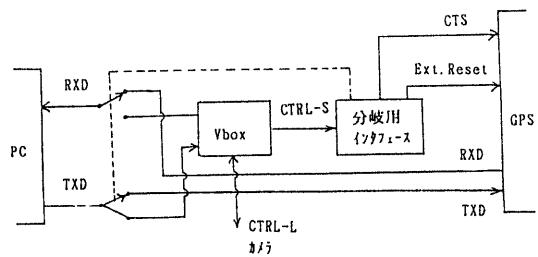


図3 カメラとセンサーへの分岐

頭出し情報を用いた画像ディレクトリ作成は、ビデオテープの内容を検索する上で非常に有効であることがわかった。図6はワークステーションの表示画面で、左端はコマンド発信用(画像ディレクトリ表示、画面のコピー、ビデオ表示関係、カメラの前進・後退・停止と、頭出しの指示、カメラの方位変化の検出、字幕検出)のボタン、下の三つの窓はカメラのビデオ動画表示用である。右側は上からディレクトリ表示(画像に対して付加した文字情報の一覧にも用いられる。)、その下が文字列入力エリヤ、コマンド発信用(静止画像ファイルの作成、画像サブファイル作成、ファイルのマージ、文字情報の付加など)のボタンを配置している。画像ディレクトリで、ある画像を選択するとき、縮小する前の画像が中央上部に表示されている。ただし、カメラから画像を入力して、画像ディレクトリを作成している間は、カメラからのビデオ動画が表示されている。

なお、このシステムでは、利用者が適当にボタンを作って独自のコマンドを作成できる。

#### 5 今後の課題と可能性

近現代研究には、大量の映像情報の検索が必要である。そこで、価値が低い部分を如何に検索対象からはずすかが問題であるが、この目的にも頭出しの機能は有効である。

GPSを用いると、ある緯度経度の範囲を指定した検索ができる筈である。これは、種々のセンサーを付加すれば、今までになかった検索ができる可能性を示している。

センサー付きのカメラの開発、カメラその他の機器を制御するに適したコンピュータの小型軽量化、ワークステーションの低廉化などが望まれる。また、映像機器のデジタル方式が普及することが望まれる。ビデオ機器は機能を簡素化する傾向があるが、高度な利用が行われるには、コンピュータによって制御されるための機能など、むしろ

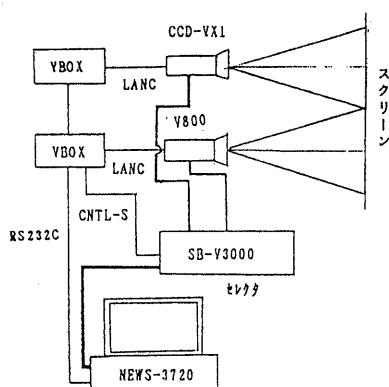


図4 カメラ2台を用いた画像入力

機能の活用を計るべきである。今後、歴史情報としてのサウンド情報も重要である。そこで、その収集・利用などに関する情報処理技術の適用について研究を行いたいと考えている。

## 6 あとがき

映像データの利用方式として、映像はビデオテープに、検索のための静止画像は光磁気ディスク、CD-ROMなどに格納して、経済的に収集・利用を行うための方式を述べた。自動的に頭出しを行うために、カメラやセンサーをコンピュータと接続して必要な情報を収集する。この情報は検索にも有效地に利用できる。種々の操作を、ワークステーションの画面上で行うことができる。開発は、C言語とOSF/Motifによった。

この研究は、試験研究(B)「東洋学研究における研究者用マルチメディア情報 CD-ROM の実用化」の下において進められているものである。

## 参考文献

(1) 星野聰：歴史学研究支援のためのコンピュータ利用、情報処理学会研究報告、89-CH-1、No. 1-4 (1989)

(2) 星野聰：ビデオ機器を用いた画像データベースの作成、研究発表報告集、p. 115-122、京都大学大型計算機センター研究開発部 (1993)

(3) 星野聰：カラー画像データベースの作成と利用、第39回研究セミナー「東洋学へのコンピュータ利用」、p. 21-29、京都大学大型計算機センター研究開発部 (1993)

(4) 星野聰：画像データのCD-ROM化と歴史・文学への利用、情報知識学会研究発表会 (1993)

(5) S. Hoshino: Multimedia Database for Research and Its Use, Presented at the Fifth ROC-Japan Information Symposium on Modern Information Services, Taipei (1993)

(6) VISCA ディベロッパズマニュアル、SONY

(7) 新訂版 GPS—人工衛星による精密方位測位システム、日本測地学会編著、日本測量学会 (1989)

(8) 長坂晃、田中謙：カラービデオ映像における自動索引付け法と物体探索法、情報処理学会論文誌、Vo. 83, No. 4, pp. 543-550 (1992)

(9) 映像データの収集利用方式の開発、研究開発論文集、pp. 60-65、全国共同利用大型計算機センター (1993)

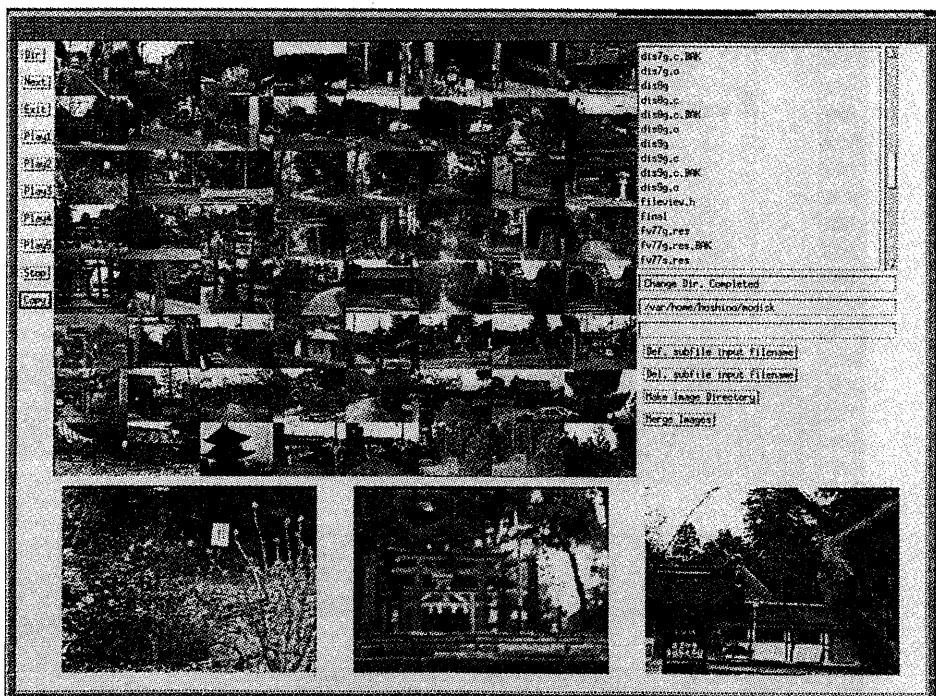


図5 WS の表示画面（ビデオ検索）



図6 WS の表示画面（カメラの方位など検出）

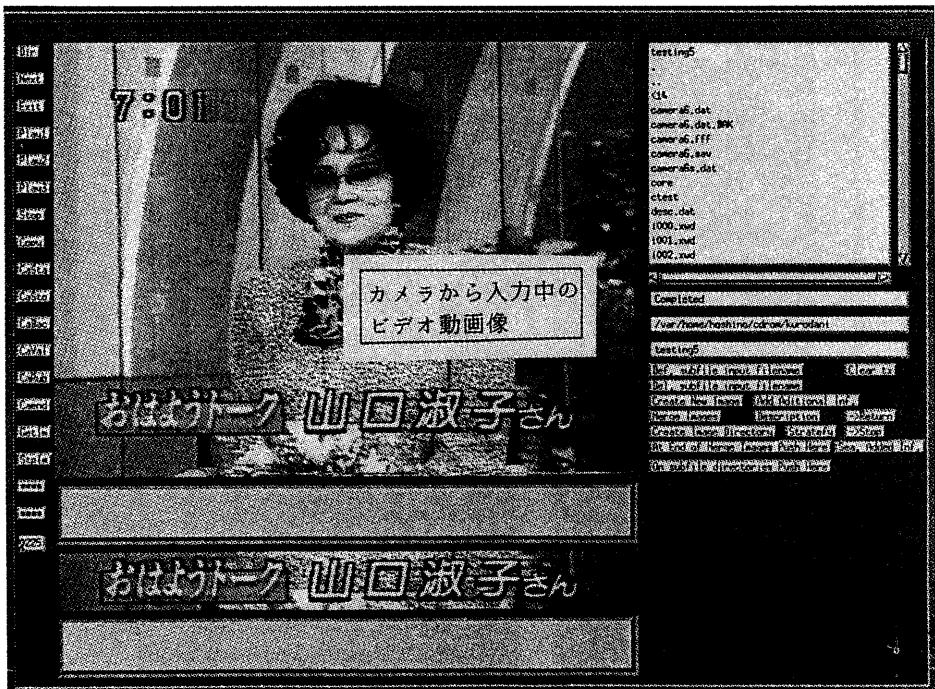


図7 WS の表示画面（字幕の検出）