

14 映像データの処理

14.1 映像データの取得

映像をコンピュータに取り込む場合、ビデオカメラやWebカメラなどを利用することができます。ビデオカメラの映像は、デジタル化される以前は、VHSやHi-8といったアナログ形式で、磁気テープに記録されていました。その後、デジタル化されたビデオの規格としてDV形式が提案されました。DV形式のビデオカメラでは、撮影した映像をカメラ内部でデジタル化し、DV形式に対応した磁気テープに映像データを記録します。DVカメラでは、IEEE 1394と呼ばれる端子が付いているパソコンがあれば、専用のケーブルを用いて接続することで、デジタル化された映像データを直接コンピュータ上に取り込めるようになりました。その後、ビデオカメラのデジタル化された映像データは、ハードディスクやSDカードなどに直接保存されるようになりました。この場合、ビデオカメラをUSBケーブルでパソコンにつなぐことで映像データをパソコン上に取り込むことができます。また、ビデオカメラから取り出したSDカード等をパソコンにつないだカードリーダーに挿して、映像ファイルをパソコン上にコピーすることもできます。

このような映像データを記録する際に考慮しなければならないことの1つとしてビットレートがあります。ビットレートとは、1秒間の記録に必要なデータ量です。例えば、1.5Mbpsのビットレートを考えてみましょう。1.5Mbpsとは、1秒間に1.5Mビットのデータになります。8ビットで1バイトなので、1秒あたりでは、(1.5/8) MBのデータになります。1時間は3600秒なので、3600倍すると、1時間あたり約700MB、すなわちCD1枚分程度のデータ量になります。ここで注意しなければならないのは、あまりビットレートを下げて高い圧縮を行うと画質が悪くなってしまうということです。従って、どのぐらいの時間撮影したいかや、映像のデータ量をどのぐらいにしたいか、利用にどのぐらいの画質が必要であるかなどによって、どのような画像サイズやビットレートで撮影するかを決める必要があります。

映像データの編集方法としては、大きく分けて、リニア編集とノンリニア編集があります。リニア編集とは、テープからテープへの編集です。映像をテープで編集するリニア編集の場合には、素材の再生用と録画用の2台のビデオデッキを用意し、再生用ビデオデッキの映像出力を録画用のビデオデッキに入力できるように設定します。そして、まず、編集先のテープのコピー先となる部分を頭出しします。次に、素材テープの再生、巻き戻しなどを行いながら、コピー元となる部分の頭出しを行います。次に、素材テープの再生を開始するとともに、編集先のテープの録画を開始します。これにより、再生用ビデオデッキで再生された映像が、録画用ビデオデッキに録画されます。最後に、コピー元となる部分の再生が終了したら、録画を終了させます。編集を繰り返す場合は、以上の手順を繰り返し行います。このような、テープからテープへの編集は、リニア編集と呼ばれます。

一方、映像データをハードディスク上に取り込み、コンピュータ上で行われる編集は、ノンリニア編集と呼ばれます。コンピュータ上の編集では、まず、素材映像中のコピー元となる部分を選択します。次に、必要に応じて選択した映像部分を並べ直します。映像の並べ替えなどが終わった場合、映像フォーマットを指定して、映像を書き出すことで映像の編集が完了します。ノンリニア編集の利点としては、必要な映像部分を選び出し、目的とする位置にコピーするといった操作を、時間順序に縛られず容易に行える点が挙げられます。また、後から再び、映像の入れ替えや、削除を行うといったことも容易にできるという利点があります。また、デジタル化された映像データを扱うため、編集により、画質や音質が劣化しにくいという利点もあります。一般に、映像はデータ量が大きく、その処理も重くなることから、ノンリニア編集では、高性能なコンピュータや、大容量の記録媒体が必要となります。

ここで、映像の選択について、もう少し詳しく見てみましょう。ノンリニア編集では、通常、素材映像は、1つの映像アイコンとして表示されます。この素材映像から、必要な映像部分を選択して新たなアイコンを作成し、それらを並べ替えることで映像の編集が行われます。素材映像から編集のコピー元となる部分を選択する方法には、大きく分けて2つの方法があります。1つは、素材映像で、コピー元となる映像部分の開始点と終了点の両方を指定する方法です。マウスやキーボードを用いて映像を再生したり巻き戻ししたりしながら、開始点と終了点を指定し、新たな映像アイコンを作成します。もう1つの方法は、映像の切れ目を指定し、映像を分割していく方法です。マウスやキーボードを用いて分割点を指定することで、映像を表すアイコンを複数に分割していきます。そして、不要なアイコンはゴミ箱に捨てて、残った映像アイコンを編集に利用します。映像の再生順序の編集は、このように選択した映像アイコンを並べ直すことで行われます。ただし、この編集段階では、編集ソフト上では、どの映像のどの部分を選び出し、どう並べるかといった情報のみを編集しており、実際の映像データに直接変更を加えているわけではありません。このため、編集後には、映像データに対して実際の変更を行うレンダリングと呼ばれる処理を行う必要があります。レンダリングは、編集集中にも明示的に行うことができますが、通常は、編集後にファイルの書き出しを行う際に同時に行われます。

映像の書き出し形式（フォーマット）に関しては、映像をどのように利用するかに応じて決めることになります。映像ファイルは、一般に、1つの入れ物の中に映像データや音声データを格納した構造を持ち、この入れ物のことをコンテナと呼びます。コンテナとしては、Windowsなどで利用されるAVI、macOSなどで利用されるMOV、DVDなどで利用されるMPEG-2システム、携帯電話などで利用される3GPPなどがあります。

映像データは、一般に、文字や画像に比べてデータ量が大きくなることから、通常は、圧縮と呼ばれる処理により、データ量の削減が行われます。データの圧縮は、人が違いを感じにくい画像中の小さな構造を省略したり、多少の色の変化を許容するなどして行われます。また、映像の場合、映像を構成する1枚1枚の画像が、前後で似ているという性質を用いることで、圧縮を行うこともできます。例えば、最初に背景だけが映っていて、そこに人が入ってきた場合、背景画像がほとんど同じであれば、人の部分のみをデータとして新たに保存すれば良い訳です。このような、さまざまな手法を用いることで、映像データの圧縮が行われます。

映像データの圧縮、伸張方式はコーデック（CODEC）と呼ばれます。例えば、MPEG-1 Videoは、1枚のCDに1時間程度のビデオ映像を記録するVideo CDなどで利用されます。MPEG-2 Videoは、DVDなどで利用されます。MPEG-4 Visualは、低速な回線などでの映像の利用を想定して開発されました。このように、映像のフォーマットにはさまざまなものがあり、映像をどのように利用したいかなどを考慮して、どのフォーマットを利用するかを決めることになります。

14.2 シーンチェンジの検出

映像は情報量が大きく、そのまま扱うのは必ずしも容易ではありません。このため、映像の変わり目であるシーンチェンジの検出を行い、映像を代表するような画像の検出が行われる場合があります。シーンチェンジを検出することにより、検出した画像を映像検索のための手がかりとして利用したり、検出した画像をアイコンとして利用することで映像編集などの処理を効率的に行うことができるようになります。

シーンチェンジ検出には、輝度、色、動きなど、さまざまな情報を利用することができます。輝度とは明るさ情報で、輝度を用いる方法としては、連続する2枚の画像間の輝度の差分を用いる手法があります。例えば、背景だけの画像と、背景の中に人が映っている画像があったとします。このとき、それぞれの画素の明るさの差の絶対値を求めて、その大きさが一定値を超えた場合に、シーンの変わり目として検出することができます。また、差分が大きくなる部分の面積を求めて、面積が大きくなった場合に、シーンチェンジとして検出する手法もあります。色情報を用いる手法としては、例えば、画像間のRGBの差分を用いる方法があります。これは、RGBの値の差の絶対値を足し合わせ、その値がある値以上になった場合にシーンチェンジとするものです。

また、RGBを色相、明度、彩度に変換し、色相のヒストグラムの差を用いてシーンチェンジを検出する手法などもあります。色相のヒストグラムの差は、カイ二乗検定に類似した式で計算することができます。図14-1(a)「ヒストグラムの差」は、連続する2枚の画像ごとに画像間の距離を求めた結果です。この結果から、ヒストグラムの差が大きい順に10枚の画像を検出すると、図14-1(b)「検出されたシーンチェンジ」のような結果になります。この結果を見ると、ほぼ編集時のシーンの切れ目に対応しており、おおむね正しく、場面の切り替わりを検出できていることが分かります。

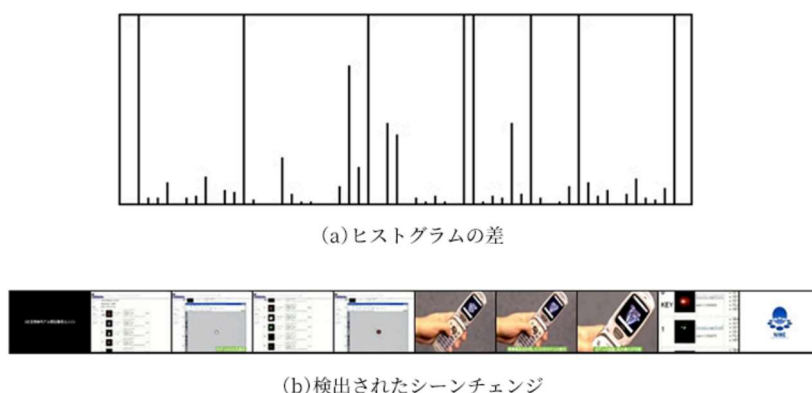


図14-1 シーンチェンジの検出

14.3 ドラマ映像の処理

ドラマ映像のデータベース化手法として、ドラマ映像に付随する音声やシナリオ文書を関連メディアとして利用する手法が提案されています。この研究では、映像を検索する際の手がかりとして、ドラマ映像に付随する音声やシナリオ文書を利用します。しかしながら、映像、音声、シナリオ文書といったメディアは必ずしも時間的に同期されていないという問題があります。このため、関連メディアの情報を利用するためには、映像と関連するメディアの同期をいかに行うかが問題となります。では、同期をどのように実現するかですが、この研究では、各々のメディアから、複数のメディアから参照できるパターンの抽出を行い、それらのパターンを対応付けすることにより、複数メディアの同期を行っています。ここでいうパターンとは、横軸は時間であり、縦軸は、各時間における「台詞の有無」や「場面の変わり目」といった、パターンの存在の有無や強度を表す系列のことです。ただし、横軸の時間とは、必ずしも何時間何分何秒といった絶対的な時間を表すものではなく、おおまかな時間順序を表す場合もあり得ます。

メディアの対応付けに利用できるパターンの1つとして、台詞の有無のパターンがあります。ドラマ映像は、通常、十分にコントロールされた環境の下で作られる映像ですので、台詞は明確に聞き取れるように調整されています。このため、大まかな近似として、音声レベルが十分高い部分は台詞が話されている部分と考えることができます。このため、音声から、音声レベルの大小のパターンを抽出することで、音声から抽出した台詞パターンとすることができます。一方、シナリオ文書は、それぞれの台詞がいつ始まり、いつ終わるかといった

時間情報を明示的には持っていませんが、台詞中の文字数を利用することで、大まかな台詞の有無のパターンを抽出することができます。具体的には、(台詞中の文字数)×(1文字を話すのに必要な時間)を計算することにより、1つの台詞を話すのに必要な時間を大まかに推定することができます。

女性の存在パターンもメディアの同期に利用できます。男性の声と女性声を比べた場合、一般には、女性の方がピッチが高い傾向があります。このため、音声からピッチの抽出を行うことで、大まかに、女性の存在パターンを抽出することができます。一方、シナリオ文書中のそれぞれの台詞の前には、その台詞を話す人物の名前の記述が存在します。このため、この名前の情報を使うことで、シナリオ文書からの女性の存在パターンの抽出を行うことができます。

場面の変わり目パターンも複数のメディアから抽出できるパターンの1つです。映像から場面の変わり目パターンを検出するためには、先ほどお話した、シーンチェンジの検出手法を利用することができます。一方、シナリオ文書は、警察署のシーン、裁判所のシーンといったように、意味的なまとまりごとに、シーンが分かれています。このシナリオ文書からの場面の変わり目パターンの抽出は、前述のシナリオ文書からの台詞パターンの抽出結果を利用することができます。すなわち、それぞれのシーンの最初の台詞の開始時に1、それ以外の場合に0のパターンをつくることにより、シナリオ文書から抽出した場面の変わり目パターンとすることができます。

このようにして抽出したパターン間のマッチングにはDPマッチングと呼ばれるアルゴリズムを利用することができます。DPマッチングは、2つのパターンが与えられた場合、それらのパターンを非線形に伸縮し、それらのパターン間の最適なマッチングを行う手法です。非線形とは、パターンをある部分は伸ばしたり、ある部分は縮めたりしてマッチングを行うということです。

図14-2は、ドラマ映像を対象とした映像、音声、シナリオ文書の対応付け結果を示したものです。1番上のパターンがシナリオ文書から抽出した台詞パターンであり、上から5番目のパターンが、音声から抽出した台詞パターンです。上から2番目のパターンがシナリオ文書から抽出した女性の存在パターンであり、上から6番目のパターンが、音声から抽出した女性の存在パターンです。上から3番目のパターンがシナリオ文書から抽出した場面の変わり目パターンであり、上から7番目のパターンが、映像から抽出した場面の変わり目パターンです。そして上から4番目の枠には、DPマッチングを用いた映像、音声、シナリオ文書の対応付け結果を示しています。このメディア間の対応付け結果を用いることで、シナリオ文書中に記述されている人物名や場所といった情報の利用が可能となり、人物名による映像の検索や、場所による映像の検索など、映像の高度な検索を実現することができるようになります。

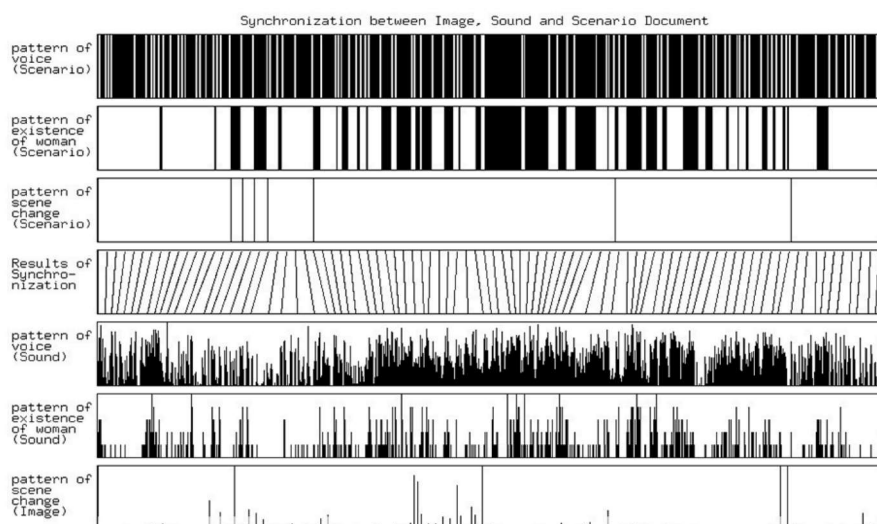


図14-2 映像、音声、シナリオ文書の対応付け結果

また、この結果を用いることで、シナリオ文書を用いた、映像の編集といった処理も実現することができます。すなわち、メディアの対応付け結果をもとに、編集されたシナリオ文書に対応する映像部分を選び出し、時間軸方向に並べ直すことで、編集された映像を作り出すことができます。具体的には、図14-3に示すように、まず、通常のテキストエディタを用いて、台詞や場面ごとの削除や入れ替えといったシナリオ文書の編集を行います。次に、映像とシナリオ文書の対応付け結果をもとに、編集されたシナリオ文書に対応する映像部分を選び出し、時間軸方向に並べ直すことで、編集された映像を作り出します。

このようなシナリオ文書を用いた映像編集は、別途、映像データに対して実際の変更を行うレンダリング処理が必要になりますが、映像に比べ情報量の少ないテキスト情報を用いることで、映像を容易に編集できるという利点があります。また、シナリオ文書という抽象度が高く、かつ、意味内容を反映した記述を利用することで、映像を意味的な内容に立ち入って編集しやすいという利点があります。

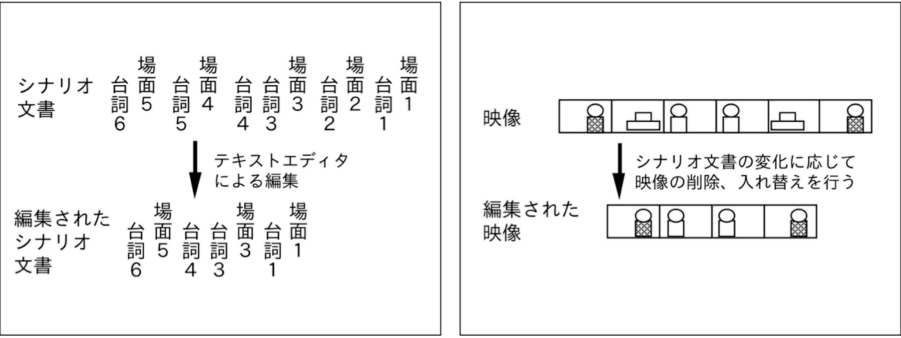


図14-3 同期したシナリオによる映像編集