

# MPEG-7 の実用化に向けたシーン分割技法・階層化アノテーション技術等の研究

藤本文彦

## ●要約

始めに本論文の目的と、意図を述べておく。映像データベース化の規格である MPEG-7 の規格に於いては、そのアノテーションの方法や、データの記述法に関して、非常に詳細にわたる定義がされている。そのため、実装が困難であることが、MPEG-7 の実用化を妨げてきたと考える。また、同様に、映像データベース化の必要が叫ばれているにもかかわらず、実際のデータベース化の技法は、まだ、研究段階である。

本論文では、映像のデータベース化は急務と考え、思い切って、MPEG-7 の詳細部分の実現を省略し、実用化に向けての簡便な映像データベースの作成と利用のための技法について、研究を行った。そのため、本研究は、実際の MPEG-7 の規格に忠実に準拠したものではなく、時に大胆な省略や変更を行ったため、もとの MPEG-7 規格とは、かけ離れたものになっている部分が多々存在する。そうした省略、変更を行ったのは、MPEG-7 の規格が、あまりに詳細多岐にわたるため、その実装の困難さが、大企業をしても、実用化開発を打ち切らせてきたと考えられる。個人に於いて同様なことを実現しようとするれば、容易に暗礁に乗り上げることが予測される。そこで、本研究では、敢えて、そうした規格の全てを取り込むことはせず、実現可能な範囲から、実用化にむけての研究を始めるべきであるという、大胆な発想の転換を行うこととした。MPEG-7 の規格制定に関わって努力された研究者の方々にはお叱りを受けるかも知れないが、こうした形での実用化研究が、将来の映像データベース化に役立つと考えて研究報告と提案を行う。

## ●キーワード

MPEG-7

アノテーション

映像データベース

カット、ショット、シーン分割

輪郭重心法

低画質映像によるアノテーション

サーバーとデータベースの分離管理

映像データベースの著作権的問題

シーンクラスタ化

階層化アノテーション技法

検索の揺らぎ

## はじめに

映像圧縮技術には、様々な技法があるが、その中でも、最も広く使われているのが、MPEG-1、MPEG-2、MPEG-4である。これらの圧縮技術の詳細については、成書に譲ることとする

MPEG-7は、こうした圧縮技術の発展型ではなく、映像情報のデータベース化についての規格を定めたものである。

爆発的に増大する映像情報の中から、目的とする映像をデータベース化し、検索する技法の開発は急務であり、こうした目的で定められたのが、MPEG-7規格である。このように映像をデータベース化する技法をアノテーションという。

本研究室では、過去5年ほどの間、学生の総合研究(卒業研究)(9)(10)を含め、様々な方向から、MPEG-7の研究を行い、その成果について、学会での発表なども行ってきた(5)-(8)。

しかし、実際には、MPEG-7の実用化は、現在頓挫しているように思われる。本論では、その理由について考察した上で、それを打開するための様々な技法について、提案を行う。そのため、本論は、実際に実験により確かめた研究報告から、将来実現可能と思われる方法についての考察まで、様々な観点からの記述を、織り交ぜた物となっている。

## 1. MPEG-7 規格について

### 1-1 MPEG-7 規格書

MPEG-7は、ISO/IEC 15938-1~11という規格文書(1)でその定義がなされている。全11パートから成り、英文3000ページに及ぶ物である。その内容は、映像圧縮の技術についてではなく、映像データベース(勿論音声含む)の構築法に関して詳細に定義された物である。現在、ようやく、若干の修正文書を除き、全パートの発行がなされた物であり、2002年から2005年までの間に主要な規格書が発行されている。

特に第3部の映像のデータベース化についての規格は詳細にわたり、映像の動きやカメラの移動のデータ化は勿論、そこに映っているオブジェクトの形状から、移動、顔の認識まで、およそ考えられるほとんどの映像をデータベース化する技法が掲げられている。

また、それを記述する方法として、言語を用いる方法と、数値化したデータを用いる方法の両者について述べられている。

これらのデータベースは、XMLで記述することを基本とするが、勿論、人手でXMLを入力する必要はなく、数社のコンピュータメーカーから、支援ツールが発行されていたが、次項に述べるように、これらのツールのサポートは、既に終了しており、現在、厳密にMPEG-7規格に準拠した入力支援ツールは発行されていない(参考文献補遺)(本研究室では、以前のバージョンをそのまま使っている)。

## 2. MPEG-7 の普及が遅れている理由

### 2-1 MPEG-7 規格の膨大さ

MPEG-7は、その必要性が急務とされているにも拘わらず、未だに、普及が遅れている。その理由として、いくつかの問題が考えられる。

まず、MPEG-7は、3000ページにも及ぶ規格書(1)からなり、その全ての規格を満たすシステムを

実現するためには、膨大な労力が必要となることが考えられる。

また、データの入力にも多くの人手が必要となり、自動化が困難であるのも大きな理由であると考えられる。映像をシーンに分割することは、半自動化が可能であるが、それに、意味付けを行うためには、どうしても、人手による作業が必要となる。

また、人間が介在すると言うことは、アノテーションを行う人により、異なるデータが付けられる可能性があることを意味する。さらには、こうした、「アノテーションの揺らぎ」は、検索を行う側にとっても望ましいことではない。そのため、アノテーションシステムの構築が行われても、検索システムは、未だに実用的な物が作られていないのが現状である。

## 2-2 大手メーカーの撤退

詳細な理由は明らかではないが、大手コンピュータメーカーも、数年前にアノテーションツールの開発を中止し、サポートも打ち切られている(参考文献補遺)。

おそらく、メカによって多大な労力が投入されたにもかかわらず、実用的な検索システムの構築はきわめて困難であり、商業的に、採算を取ることが難しいことが、メーカーが実用化に積極的でない理由と思われる。

また、個人的あるいは、大学等の学術研究機関等に於いても、2004年頃以降、新たな研究が行われた形跡が見あたらない。学会に於ける研究発表に於いても、MPEG-7に関する発表は、近年ほとんど存在していないと思われる。

このように、次世代を担う重要な規格として期待されていた MPEG-7 が、現在ほとんど研究されていないことは、憂慮すべきことである。大企業が採算ベースに乗らないので、研究を打ち切ったとするならば、そのことは理解できないわけではないが、学術研究機関等でも研究が行われていないと言うことは、他にも各種の要因があると思われる。おそらくその要因のひとつは、規格自体が膨大なものであり、一研究機関の手に余る物だからではないかと思われる。

本論ではこうした状況を打開するために、MPEG-7の全規格を忠実に実現することより、実現可能な部分から始めると同時に、独自の技法を開発し、また、こうした映像情報データベースが、商業的に採算レベルに乗るためには、どのような事を行えば良いかについて考察することにする。

## 3. MPEG-7の実用化に向けて

前記のように、多大な労力が必要とされているにも関わらず、商業的レベルに乗ることが困難であることが、MPEG-7の実用化の遅れとなっていると思われる。

そこで、MPEG-7の実用化に向けて、いくつかの技法を提案する。

- ・ MPEG-7の全規格を実現しようとする、大変な労力が必要であるので、実用化可能な部分から実現する。
- ・ 映像のシーン分割においては、極力、簡単なパラメータで比較する技法を開発し、長時間の映像の比較、クラスタ化を可能とする。
- ・ アノテーションを行う人による「揺らぎ」を極力減らすために、あらかじめ、アノテーションに

使用する用語を定義する。なお、この定義は、4階層各7項目とし、アノテーションの便宜を図ると同時に、検索を容易にする（なお、用語は、映像の種類によって、試行錯誤的に適切な用語を選択する）。

## 4. 映像のショット分割とシーン分割の新提案

### 4-1 カット、ショット、シーンという用語の定義

映像を分割するに当たり、まず、「カット」(2)「ショット」(3)「シーン」(4)の関係について、定義しておく。

連続した映像に関して、場面が転換する場所を「カット」と呼び、「カット」から「カット」までの連続した映像を「ショット」と呼ぶのが、一般的であるが、アニメなどの場合には、連続した映像を「カット」と呼ぶこともある。本論では、場面が転換する場所を「カット」と呼び、連続した(切れ目のない)映像を「ショット」と呼ぶことにする。「シーン」とは、密接な関連のある、一連の「ショット」をさすことと定義する(図1)。

また、この図には示していないが、動画中の1コマのことをフレームと称する。日本に於けるテレビ放送の規格であるNTSC方式の場合には、1秒の映像に、30フレーム(正確には29.97フレーム)が映し出されている。

### 4-2 カット、ショットシーンの関係の定義

通常の映像に於ける1ショットは、数秒から数十秒であり、100~1000フレーム程度で1ショットが成り立っている。通常、一連のシーンは、数ショットから成り、連続するショットは、別のアングルで撮影されている。シーンは、これらのショットを、意味的に結合して構成される。

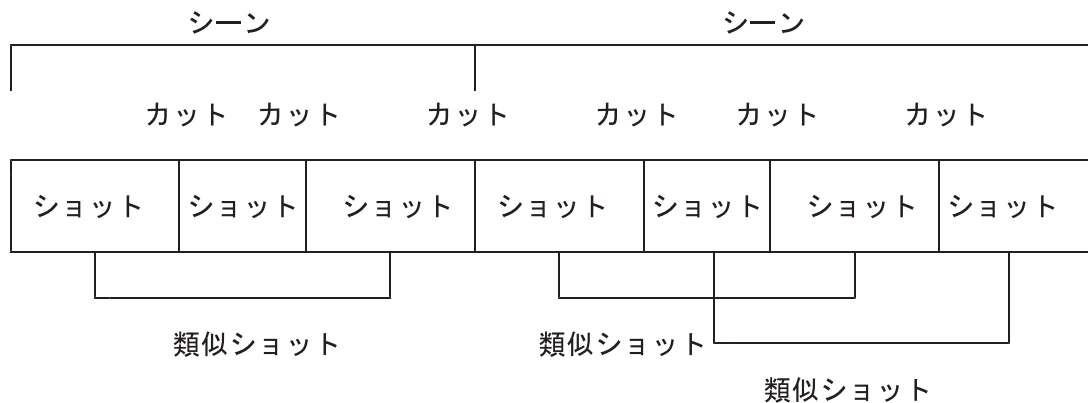


図1 カット、ショット、シーンの関係

### 4-3 半自動的分割法の探求

後述する従来のショット分割法においては、ショット間のカットの検出に関しては、半自動的に行うことが出来たが、意味的つながりのある、シーンの分割は、人手によるしかなかった。

本論の後半に述べるが、今回提案の方法では、こうした類似ショットの検出を行い、シーンを半自動的に行う方法を提案する。勿論、意味的に異なるシーンを完全に自動的に分離することは、困難で



あるが、少なくとも、今回提案の方法は、そうしたシーン分割の助けになるものとする。

なお、実際の MPEG-7 の規格では、このようなシーン分割を行う方法だけでなく、一定時間ごとのフレームに対するアノテーションの方法が定義されている。しかし、本論では、そうした方法は採用せず、あくまで、意味的なシーンの分割法を提案する。

## 5. 映像のショット分割の各種技法

### 5-1 従来の分割技法（画素データの比較法）

映像をショットに分割するには、各種の技法が提案されている。代表的な方法としては、隣り合う画面の画素データを比較し、急激な変化の起こった場所を、カットと見なす方法である。ただし、この方法は、カメラが移動した場合や、ズームした場合などに対応するために、単に同一点の画素データを比較するだけでなく、移動を考慮した周囲の画素データも比較しなければならないなど、計算量が極めて多くなる事が考えられる。尤も、このようなカメラの移動に関しては、MPEG における各種の動き予測により、あらかじめ行われている事を利用することにより、計算量を減少させることが可能であると考えられる。

しかしながら、このような動き予測は、隣接するフレーム、あるいは、数フレームの範囲内での比較しか行うことができず、前図に示すような、別のショットが挿入された場合などには、有効ではない。従って、図 1 に示すような、数十フレームから数百あるいは、数千フレーム離れた場所に、類似のショットが撮影されていたとしても、それを、類似ショットとして検出することが困難である。具体例としては、2人の人物が、交互に映されているシーンにおいて、同一人物が映されて居るのか否かを判別するためには、遠く離れたフレーム間の類似性の計算を行わなければならないが、實際上、現状のコンピュータの能力では、不可能である。

しかし、ショットからシーンを再構成するためには、このような、離れたフレーム間の比較は必須である。本論では、以下に、そのための新技法を示す。

### 5-2 従来の分割技法（色要素の分布比較法）

また、もう一つの技法としては、フレームの色の変化を検出する方法がある。この方法では、色の分布の類似度から、カット位置を求める。この技法を用いれば、計算量を減らすと同時に、画面の移動に対して、連続的な色変化が、起こると考えられるため、従来のフレーム間画素の比較に比べて、容易にカット位置を検出する事が可能と考えられる。しかしながら、この方法も、急激な光源の変化や、蛍光灯下での撮影の様な、フリッカーが問題となる場合などには、誤検出の可能性がある。また、映像内に光源があり、その色彩が変化する場合などにも、カットを誤検出することが知られている。さらに、洋画のように、ほとんどのシーンが比較的暗い映像の場合には、シーンのカット点を求めることが困難であると予想される。また、この方法も、前記の方法に比較すれば、計算量的には有利であるとはいえ、同様、離れたフレーム間の比較においては、計算量の爆発的増大の問題は、依然残ることになる。

### 5-3 従来技法によるフレーム間類似度計算の問題点

長時間の映像のアノテーションに於いて、実用的なアノテーションを行うためには、最低でも、30分程度の映像に関して、シーン分割が出来なければならないと考えられる。

しかし、30分の映像は、30フレーム×60秒×30分=54000フレームから成り立つことを考え、その各フレーム間の類似度を計算するためには、54000<sup>2</sup>のフレームの類似度を計算しなければならない。これは、現在、リアルタイムにMPEG圧縮を出来るコンピュータにおいても、実時間の数万倍の時間がかかることになる。計算量を減らすために、各ショットの代表的なフレームを使う方法を考えた場合、1ショットを10秒程度と仮定して、毎分6ショットの代表フレーム間の相互関係を考えれば、6フレーム×30分=180フレームとなり、そのフレームの全類似度の計算には、180<sup>2</sup>=32400回の比較が必要となる。これは、実時間の半分強の比較時間が必要であることを意味する。この程度の比較時間であれば、実的に可能な範囲と思われるが、そのためには、各ショットを代表するフレームを適切に選ぶ必要がある。しかし、実際には、1ショットの間にもカメラが移動することは、十分に考えられ、代表フレームの選択は容易ではない。勿論、カメラが固定ショットである場合には、類似ショットの検出は可能であるが、カメラが移動している場合には、とたんに類似ショットの検出精度が悪化することが考えられる。

上記のような理由から、長時間の実用的なアノテーションを、半自動的に行うことは困難であり、人力に頼らなければならない。

しかし、人力による場合、例え、シーン分割が出来たとしても、長時間の映像の類似シーンの判別は、容易ではない。これは、長時間映像の中のどこに類似シーンがあるかを、検索するための効率的な方法が無いからであると考えられる(5)(6)。

### 5-4 類似シーンへの同一アノテーションの困難性

また、類似シーン(たとえば、登場人物の部屋や事務所など)には、同じアノテーションが行われなければならない。しかし、そうした場合も、アノテーションを行う人間による意味づけの「揺らぎ」が存在することが考えられる。これは、別人であればもちろんであるが、同一人物が行うアノテーションでも十分に考えられる事である。もしも、類似シーンに同じアノテーションを行うとすれば、毎回、過去のアノテーションで付随したデータを調べて行わなければならないと、多大な労力の必要とされる作業となることは容易に想像できる。

### 5-5 コストパフォーマンスの問題と、入力・検索の「揺らぎ」問題

もちろん、そのような労力を払っても、商業的に成り立つものであれば、企業が行うことは考えられないことではないが、そうしたアノテーションを必要とするユーザ層がどのようなものであるのか、商業的に成り立つものかどうかについて、おそらく、多くの企業が検討の結果、アノテーションツールの開発が打ち切られているのではないかと考えられる。

こうした理由から、MPEG-7の実用化が遅れているものと思われるが、以下に、実用化を主眼とした。シーン分割及び、類似シーンの抽出、さらには、アノテーションの「揺らぎ」を提言し、検索の便宜を図ると同時に、商業的にも成り立つであろう方法について、検討することにする。

なお、シーンより、さらに大きな単位としてのセグメント分割に関しては、HDDレコーダなどで、大手電機メーカーが既に実用化を行っているが、それらの技法については、詳細が公表されておらず、また、必ずしも人間の思惑通りの分割が行われているとは思えない事から、今回の研究では、そうした、HDDレコーダ等のシーン分割技法に関する研究は見送ることとした。

## 6. 輪郭重心法の提案

### 6-1 フレーム間比較法としての輪郭重心法の提案

映像のカット点を見いだすには、隣り合うフレームを比較するしか方法がない。前記に述べたように、フレームを比較しカット点を見いだす技法に関しては、様々な方法が研究されているが、ここで新しく「輪郭重心法」という技法を提案する。この技法は、映像からその輪郭を白黒のデータとして求め、その輪郭の画素の重心を求めて、そのフレームを代表する数値とする。連続した1ショットの間にカメラの移動があっても、重心の位置は、連続的に変化すると考えられる。この場合、不連続な重心位置の変化がある点を、カット点と見なすことが出来ると考えられる。

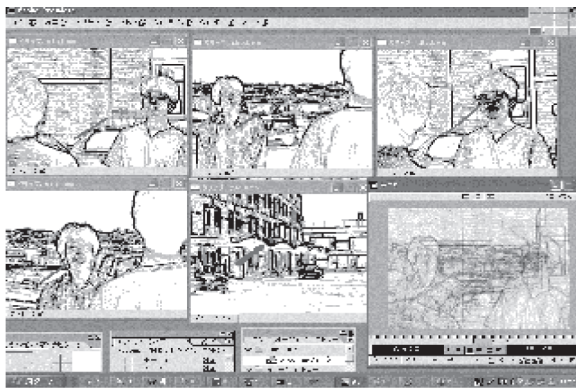


図2 輪郭重心を求める例

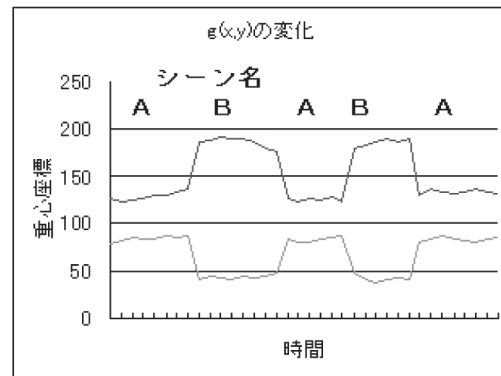


図3 重心の時間変化の例

### 6.2 輪郭重心法の問題点

ただし、この方法には、いくつかの問題点もある。

第1には、輪郭線の太さが、フレームによって変化することである。ただし、この技法は、全体としての輪郭線を構成する画素の重心を求めるものであるため、輪郭線の太さが一様に変化する限り、突然の重心の移動は起こらないと考えられる。また、この方法においては、輪郭線が不連続で間が開いたとしても、そこを補充する必要が無いので、一般の輪郭線を求める際の困難性を排除することが出来る。

第2には、フレームの外周近くにある輪郭線がフレーム外に出た場合、あるいは、外から、フレーム内に入ってきた場合に、大きく重心位置が変わる可能性があることである。しかし、これも、以下のような方法で回避できると考えられる。まず、フレーム内に一回り小さな内周フレームを設け、その範囲内での重心位置を求める。次にその内周フレームの位置を周囲8方向に数画素ずらして、重心位置を求める。そして、直前のフレームと近似の位置に重心があった場合、そのフレームは、わずかに移動し、前のフレームと連続であると考えられる。もし、周囲8方向に内周フレームをずらしても、前

フレームの重心と大きくずれた場合には、そこをカット点とすることにより、カメラのわずかな移動により、重心位置に対して大きな影響を持つ輪郭線がフレーム外に移動した場合の影響を抑えることが出来る。ただし、移動した内周フレームからさえも、輪郭線がはみ出した場合や、逆に内周フレーム内に輪郭線が入ってきた場合には、重心位置の移動が大きくなるが、その場合には、次のような方法で、重心の不連続性を補完して、あたかも、連続したショットであるかのように処理することが出来る。

まず、内周フレームをずらして計算した重心位置を、記録しておく、さらに、次のフレームとの比較の際には、その重心を使用して、近似した位置であれば、中央に配置した、内周フレームではなく、新しくずれた重心の位置を、次の重心位置として使用する（この場合、カメラが、その方向に移動したと考えることができる）。

この場合、前後3フレームの中で、中心にあるフレームの重心位置が、不連続に飛躍することになるが、この場合は、連続した変化と見なすことにする。

例えば、建物の風景を、横にパンしたような映像では、こうした処置を行わないと、元の映像から、あるところ（例えば、建物の輪郭が消失した場合など）で、不連続と判断されてしまう。

なお、この場合も、カメラが、高速で移動した場合や、オブジェクトが高速で移動した場合には、検出が困難となるので、周囲8方向に何画素分内周フレームを移動させたら良いのかについては、現在試行錯誤で検討中である。なお、一般的なカメラの移動は、縦方向のチルトや、斜め方向より、水平方向へのパンがほとんどであることや、オブジェクトの移動も、横方向が多いことを考慮すると、より良いカット検出が出来るものと考えられる。ただし、アニメやSF等のように、斜め方向にも少なからず移動する映像の場合には、別途、適切なパラメータを決める必要がある。

### 6-3 未解決の問題

従来研究されてきた各種の技法においては勿論、今回提案の技法に於いても、未解決の問題がある。例えば、アクション映画やアニメの爆発シーンなどの場合、瞬間的に強い光が発生したり、点滅したりする場合がある。このような場合、人間にとっては、連続したシーンと理解できても、機械的なカット検出では、この部分を、別のショットとしてとらえてしまう。現状では、このような瞬間的な光の明滅のような映像を、正しく連続したショットととらえることは困難であると考えられるが、別の技法と組み合わせる等して、認識率を上げることが今後の課題である。

## 7. 離れたフレームにある類似ショット、類似シーンの検出

### 7-1 計算量の問題

「輪郭重心法」は、他のカット検出法に比べて、特に優れているとは言えないものではあるが、先に述べたように、この技法は、類似ショット、類似シーンの検出に於いて、能率的な手段を与えるものであると考えられる。

この手法の特徴的な点は、1フレームの画像を、重心という(x, y)のただ1つのパラメータ組で表す点である（周辺8方向も考慮すれば、9つのパラメータ組となる）。

このように少ないパラメータで画像の特徴を表すため、長時間の映像の全フレームの比較が可能となると考える。実際には、 $54000^2 = 2916000000$ 回の単純な数値の比較をすれば良いので、現在のコン



コンピュータでも充分計算が可能な回数である。周囲8方向を考慮しても、この10倍の計算量で済むと考えられる。

ただし、この方法にも、いくつかの問題点が考えられるが、どれも、改良の方法が考えられ、逆に、問題点を長所に転換することも可能であると考えられる。

第1に、事前の計算として、輪郭を求めなければならないが、この計算量は、それほど多いとは考えにくい。

第2に、重心という、単純で、近似の値が出やすいパラメータで比較する関係上、類似フレーム以外のフレームを誤検出する可能性が高い事が挙げられる。しかし、この方法には逆に長所もある。機械による、半自動的検出を行うのであれば、検出された類似フレームに関して、画面を4つあるいは、それ以上に分割し、それぞれの重心を比較する事で、より精度を高めることが出来る。あるいは、より精度を高めるためには、全画面比較などの方法を用いても良い。なお、この場合、1ショットの中には、数10から数1000フレームが含まれるので、その全部のショットを検出に用いることは、能率的でない。輪郭重心法による比較の場合には、最も重心位置の近いフレーム同士の比較を行うことで、計算量を減らす事が出来る（一連のフレームは、同一ショットに属することがあらかじめわかっているため、その中の代表フレームを用いる事には問題はないと考えられる）。もしここで、人手を使って類似フレームの検出を行うことが許されるのであれば、より精度良くまた、能率的な検出が可能である。人手を使う場合には、最初に選択したフレームに対して、輪郭重心法で、検出されたフレームをサムネイルで表示し、類似のフレームのみを候補として選択することが出来る。半自動的に行うにしても、人手で行うにしても、類似フレームを含むショットには、ユニークなID番号を付けることにより、今後のアノテーションの際に互いに参照し合う事が可能となり、言葉による意味づけの「揺らぎ」を低減することが出来る。

このように、輪郭重心法は、長時間の映像アノテーションに威力を発揮するため、映像アノテーションの実用化に向けて大きな推進力となると考えられる。

## 8. シーンのクラスタ化

### 8-1 1段階のクラスタ化

本研究のアノテーション技術に於いて、重要なことは、類似のシーンには、(ほぼ)同じアノテーションを行うことが出来ることを目的としていることである。そのために、時間的に遠く離れた場所に存在するシーンに対して、(ほぼ)同じアノテーションを行う事が出来なければならない。このような状況は、次のような場合に見られるものであり、むしろ、通常の映像ならこのようなシーンの並びが起らないことの方が希である。



図4 類似シーンの検出と1段階のクラスタ化例

図4のように、「自宅」-「学校」-「自宅」のように、ほぼ同じシーンが、同一映像内に出現することは、むしろ当然である（そうでなければ話しの展開が分けがわからなくなってしまう）。

## 8-2 類似シーンのアノテーションの「揺らぎ」を低減する方法

このような場合、ほぼ同じシーンを、類似シーンとして、半自動的に検出、ないし、人手によって検出する技法については、前項で述べた。図の場合、最初に自宅シーンAが出現した後に、その類似シーン検索を、全映像について行う。検出の結果類似シーンA'として特定されたシーンに関して、始めに検出されたシーンに対して行ったアノテーションを類似シーンA'に対して行う。場合によって、同じシーン内であっても、ほとんど異なる状況が映し出されている場合もある。例えば、自宅でも、応接間のショットもあれば、台所のショット、玄関のショットなど、映像的には異なるショットが複合してシーンが構成されている場合がある。従って、類似シーンの中でも、シーン全体に共通するアノテーション項目もあれば、ショット単位で異なるアノテーション項目も存在する。従って、細かく分ければ、ショット単位でアノテーションを行うことが望ましいと考えられるが、そこまで行うのは、大変である。そこで、第1段階では、シーン単位のアノテーションを行い、全体のアノテーションが終了した後、第2段階として、ショット単位のアノテーションを行うことが望ましいと考えられる。本研究の目的の重要な項目として、映像全体のアノテーションを行うことが掲げられるが、始めから、ショット単位でアノテーションを行った場合、全体のアノテーションには多大な時間と労力を必要とするために、アノテーション作業がなかなか進まないことが考えられる。また、全体を通して話しの内容を理解した上で、細かなアノテーションを行う方が、より、適切なアノテーションが出来ると思われる。

## 8-3 類似シーンのアノテーションに於ける前後の変化

さて、ここで、始めに自宅のシーンAが現れたときに、それ以降の自宅シーンA'等には、同アノテーションを行うことで、シーン間のアノテーションの「揺らぎ」を押さえることが出来ることがわかった。そこで、次のシーンBを経過後、再び自宅シーンA'が現れたときには、既に、アノテーションが行われているので、このシーンのアノテーションは楽であるように思われる。しかし、実際には、同じ自宅シーンA'に戻ったとはいえ、必ずしも全く同じアノテーション項目が、そのまま使えるわけではない。例えば、最初のシーンAには居なかった友人が、シーンA'には存在しているかも知れない。また、そうした変化以外にも、学校シーンBの間に不快な思いをしたことで、シーンA'では、シーンAと異なる感情状態にあることも考えられる。また、シーンAでは気がつかなかったあるいは誤解していた事項が、シーンA'で判明することもあるため、前のシーンに戻って、アノテーションの修正を行う必要がある場合もある。このような場合にも、前項で述べた、類似シーンのサムネイル表示機能は、有効である。

## 8-4 「ネタバラシ」レベル設定の必要性

映像の中には、推理ドラマの様に、犯人特定のために必要な情報が、以前のシーンに映し出されている物もあり、このような場合、前に戻ってのシーンアノテーションの修正が、いわゆる「ネタバラ



し」になる場合が考えられる。ここで、本研究の想定している、シーンアノテーションは、既に、映像を1度以上見ている場合を暗黙に仮定しているが、初めて見る人に対しては、こうした「ネタバレ」が、望ましくないことも考えられる。そのため、こうした場合の、シーンアノテーションについては、どこまで視聴者に表示するかのレベルを設定することが必要な機能ではないかと考えられる。現在のところ、アノテーションされた項目に、このような機能を実現する方法について有効な手段を模索中である。「刑事コロボ」のように、最初から、犯人がわかっている場合ですら、犯人の、「いつ、私が犯人だと疑ったのだ。」といった質問に対して、コロボが「あの時、あなた〇〇と仰いましたよね。それで、怪しいと思ったのですよ。」というような、巧妙な「落ち」が存在する場合がある。

### 8-5 多段階のクラスタ化

1段階のクラスタ化に於いては、場所的に同一シーンを結びつけることを目的としたが、映像は、場所的な関連だけでなく、内容的な関連を持って進行していくものである。そこで、物語的に、関連性の大きさに従って、シーンをクラスタ化することも、映像の理解を深めるためには有効であると考えられる。このような技法は、まだ、研究途上であり、MPEG-7の規格書でも、具体的にそのような技法について記述された部分を見いだすことは出来なかったが、第1段階で場所的同一シーンを結びつけたのに対して、内容的な関連性を、多段階にクラスタ化することは、映像の理解に、新しい視点を導入することになるのではないかと考える。

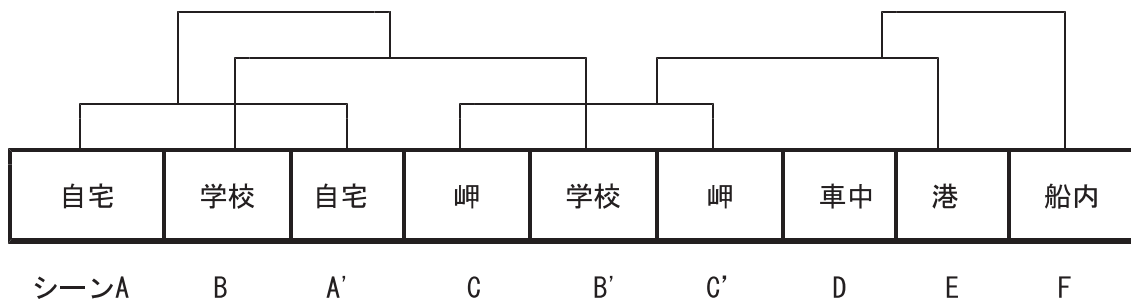


図5 多段階のシーンのクラスタ化例

### 8-6 特殊な意図で時間軸を変更した映像への対応

なお、映像の中には、敢えて物語上の時間軸と、映像の時間軸を交錯させることにより、特殊な意図を表現する物も存在する。こうした映像に於けるストーリー展開を時間軸上でクラスタ化することは、なかなか困難な作業である。回想シーンの様に、明確に時間がさかのぼったことがわかるような演出をしている映像であれば、それほど困難ではないが、物語全体が、ストーリー上の時間軸を分割し、映像の時間軸上に再構成して配置している様な場合、あるいは、同一時間の複数の登場者の視点で制作された映像の場合には、こうしたクラスタ化は、困難な作業となることが予測される。

こうした映像の場合には、ここで提案したようなクラスタ化の必要性について、意見の分かれるところであると思うが、それでもなお、製作者の特殊な意図を解析する事により理解を深めるというような目的のためには、映像の多段階クラスタ化は、有効な手段であると考えられる。

なお、シーンのクラスタ化については、さらなる研究を行っているが、誌面の都合上、詳しく述べる事ができないので、詳細は参考文献を参照されたい(5)(6)(7)(8)。

## 9. 縮小画像と映像サーバー、検索クライアント

### 9-1 アノテーションに必要な画質

映像アノテーションの便宜を図る目的を考えたとき、映像をショット、シーンに分割し、アノテーション作業を行うことになるが、実はその際に、フルサイズの映像を用いる必要は無い。通常のテレビ画質あるいは、映画画質の映像を、そのままアノテーションツールに入力し、全ての画素に関する計算を行うとなれば、その計算量は、大きなものとなる。しかし、実際に、カットポイントを求め、シーン分割を行う事だけを考えれば、縮小した映像を用いても、十分な分割を行うことが出来る。このような縮小映像を用いることにより、画質は犠牲になるものの、アノテーションには、必要な限度の画像が得られるばかりでなく、計算量を、大幅に減らすことが可能となる。勿論、その映像は、実際の鑑賞用としては、不十分な物と考えられるが、そこで、鑑賞用のフル規格の映像と、アノテーション用の縮小画像とを分けて管理し、アノテーションは縮小映像で行うことを考える。

この方法は、単にアノテーションの便宜を図るだけでなく、後述するように、著作権上の問題も回避することが可能で、多くの企業が映像のデータベース化に取り組むことが可能となると考えられる。

### 9-2 画像のデータ量の検討

例えば、テレビ画質の640×480ドットの映像をそのまま処理するのに比べて、縦横1/4にすることにより、160×120ドットの映像を使用してアノテーションを行う方法を考える。この程度の画質では、詳細な文字などを判別することは困難であるが、どこで、誰が何をしているか、というような情報を認識するには、十分と考えられる。これにより、映像のデータ量は、単純計算で、 $1/4 \times 1/4 = 1/16$ となるため、それにより、画像表示の負担を大幅に減らす事ができ、ストレスの少ないアノテーションが行えると考えられる。

アノテーションをこのように元映像と分離して作ることは、映像の制作を専門とする者と、アノテーションを行う者が分業で作業を行うためにも有効である。

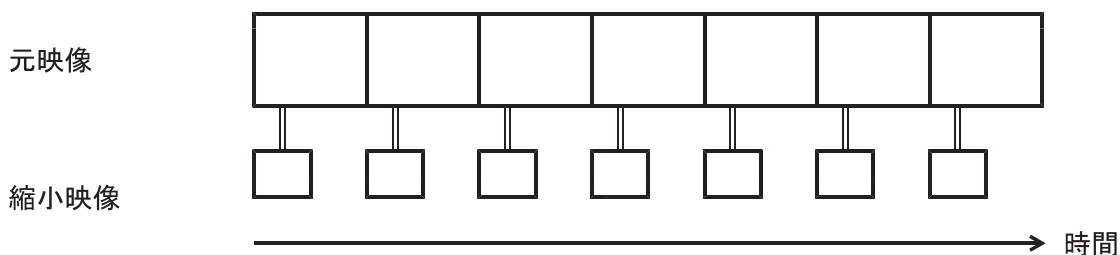


図6 元映像と縮小映像の時間軸によるリンク

### 9-3 データベースと元映像の分離と合成

アノテーションによって作られたデータベースは、映像とは独立したデータとして存在するが、元

映像との間では、映像の頭からの時間でリンクしているために、独立して入手(購入)した元映像と、データベースをリンクして検索・表示するブラウザを利用することで、映像検索を行う事が出来る。

#### 9-4 アノテーションの分離と映像・データの管理

このシステムは、元映像を置くサーバー、アノテーションを行うクライアント兼サーバー、検索映像表示を行うクライアントの3台のコンピュータをネットワークで接続して行うことを想定している。勿論、これらの作業は、全て1台のコンピュータ上で行うことが可能であるが、実際の利用形態を考えると、上記のように、3台のコンピュータに分けて行う事が望ましい。

それぞれのコンピュータは、次のような関係にある。

- ・映像サーバー：映像の著作権を保持する会社、販売権を持つ会社等
- ・アノテーションクライアント：著作権者の依頼により、映像にデータを付加する業者  
あるいは自主的にアノテーションを行う個人等
- ・検索及び映像表示クライアント：映像を検索、鑑賞する一般ユーザ

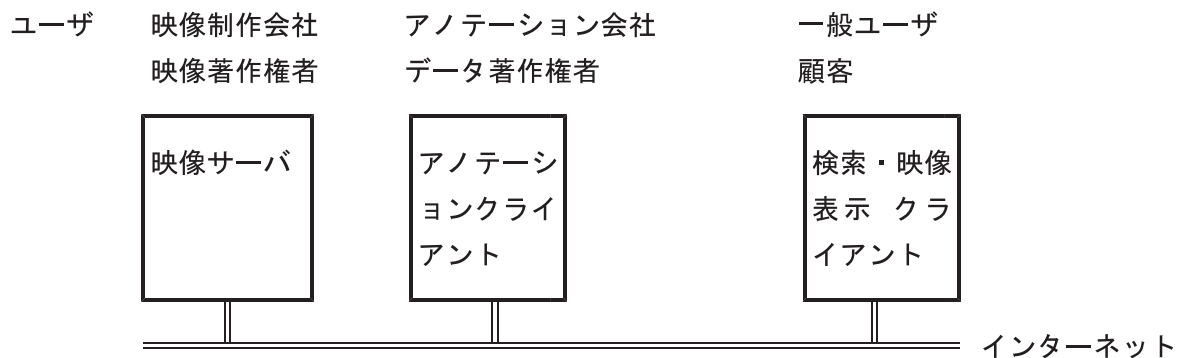


図7 映像アノテーションネットワークの構想

実際の流れは、次のようになる。

- ・映像制作会社(著作権者)が、低画質の縮小映像を制作、アノテーション会社に送付し、データベース作成を依頼。
- ・アノテーション会社は、縮小映像を見ながら、映像アノテーションを行い、データベースを作成する。
- ・一般ユーザは、アノテーション会社より、データベースを購入する。データを検索の上、必要と判断した場合は、映像制作会社から、高品質な映像を購入する(順序は同時の場合もあれば、逆の場合もある)。

こうした流れにより、一般ユーザは、探したい映像を効率的に探すことが可能となり、また、映像制作会社は、大規模な宣伝活動を行う事無く販売促進が出来る。また、アノテーション会社は、映像データベースの販売により利潤を上げることができる。さらに、多数の映像制作会社と提携を結び、

作成したデータベースを、自前のサーバ上に置いて有料で提供することにより、より多数の顧客を獲得でき、大幅な利潤を上げることが可能である。

## 9-5 アノテーション専門業者の必要性

ここで、疑問に思われるかも知れない点は、映像制作会社（販売会社）が、社内でアノテーションを行えば、別途、アノテーション会社にデータベースの作成を依頼する必要は無いのではないか、ということである。しかし、実際、映像制作会社は、制作作業に専念するのに忙しく、アノテーションを行う余裕はない。映像制作会社は、アノテーションに必要なシステムの構築、マーケティングを行うスタッフを雇用する余裕が無いし、もし、そのようなスタッフを雇用したとしても、実際には、撮影、編集などの映像スタッフの不足を補うための業務を行わなければならない事が十分に考えられる。このような状況では、アノテーションに専念する事は、容易ではない。こうした場合は、外注の専門業者に依頼するのが、最も効率的な方法であると考えられる。

また、発売時期の問題もある。もし、映像制作会社がアノテーションを行ったデータベースを発売するとすれば、映像の発売と同時に進行する必要がある。しかし、実際には、映像制作会社の発売直前に、そのような事が行える時間があるはずはなく、アノテーションのために、発売時期が遅れることは許されない。また、映像購入者の全てがアノテーションデータを必要とするわけではなく、データの不要な顧客に対しては、アノテーションの価格が上乘せされることは、好まれないと考えられる。それに対して、外注のアノテーション業者に依頼すれば、業者はその業務に集中することが出来ると同時に、データの発売時期が映像の発売時期より、若干遅れても構わないと考えられる。こうした分業体制により、データの管理は専門業者に全て依頼する事が出来、必要な顧客だけがデータを購入する事が可能となる。

このように、分業体制を取ることににより、それぞれの立場の会社、個人が、有用な情報を入手し、利潤を上げることができる。このような方向性は、現在のインターネット上の検索エンジンと類似の構造であり、提供者、利用者共に、利益を得る事が出来る事から、今後、発展が期待される。ただし、検索エンジンが運営初期にそうであったように、最初から利益を上げることは困難であり、ある程度の世間の理解が得られるようになって、はじめて利益を上げることが出来ると考えられることから、初期には、その重要性、将来性に関する十分な理解を得るための活動が必要と考えられる。

## 10. 著作権的な問題

### 10-1 アノテーションと現行著作権法

このような分業体制を取るためには、映像著作権者（販売会社）は、アノテーション会社に対して、低画質の映像、及び、設定資料、台本などを提供しなければならない。その際には、守秘義務や著作権に関する詳細な取り決めが必要となるであろう。

ただし、以下に述べるように、映像アノテーションが著作権法上問題が無いと考えて良いのであれば、こうした契約は不要となり、会社は勿論個人でも、自由にアノテーションを行い、配布することが可能と考えられる。

また、必要な場合には、アノテーション会社は、数値や文書のデータの他に、低画質の映像をあわ

せて提供する事も考えられる。この場合は、著作権的な問題が生じるはずであるが、ここで、提供される映像は、低画質で、一般的な鑑賞に堪える物ではないために、正規の画質のビデオ・DVD等の販売に大きな影響を持つことは無いと考えられる。むしろ、正規品の購買意欲を高めるために貢献するのではないかと考えられる。

## 10-2 アノテーションデータの独立販売

さらに、アノテーション会社が制作したデータは、現行の著作権法の保護する、アイデアを表現したものであるものとしての著作物には当たらないと考えられるので、あくまでも映像とは独立したデータとして販売することが可能であると考えられる。そして、そのデータ自身の著作権は、アノテーション会社に属すると考えられる事から、アノテーション会社が、複雑な著作権上の処理を行わなくても、自社の判断で、販売できる物と考えることも出来る。

もちろん、このような判断が有効であるか否かは、今後の著作権法の改定により変化することも考えられ、さらに、アノテーションの方法によっては、2次著作物として、オリジナルの映像著作権に抵触する可能性も捨てきれないことから、実際の運用に当たっては、細心の注意を払わなければならない。

## 10-3 一般ユーザの便宜

また、一般ユーザにとって見れば、始めから映像にバンドリングされたデータベースを購入する必要は無く、あくまで、映像は映像、データはデータとして別個に購入することができるので、映像だけで、データを希望しない者に余分な価格の上乗せが行われることもないというメリットが考えられる。また、データベースは、元映像を購入することなく、独立して購入することが出来るため、あらかじめ、関連のありそうな、データを購入することで、その中から、希望するデータを含む映像のみを購入することが出来るというメリットもある。

さらに、将来的には、個別のデータの販売という形態ではなく、アノテーション会社がサーバを設置し、そこに置かれたデータベースを、ユーザが検索し、必要なデータは、アノテーション会社から、映像は、映像販売会社から購入する事も可能になると考えられる。具体的な見通しを建てることは困難であるが、本論に述べたようなアノテーション技術が実用化されれば、インターネット及びコンピュータの技術的発展の状況を鑑みるに、数年以内に可能であると考えられる。

# 11. 映像の4階層7項目の分類法の考察

## 11-1 アノテーションを行う際の「揺らぎ」の防止法

次に、映像アノテーションを行うために、シーンに、映像データを付加する技法について考察する。ここで、最も重視するのが、アノテーションの容易さ、人による表現の「揺らぎ」を極力抑えること、及び、検索の際に、容易に検索できることである。

MPEG-7の規格では、映像シーンにアノテーションを行う方法として、言葉による方法と、数値化されたデータを付加する方法とが用意されている。

どちらの表現も、多様な表現が用意されており、特に言葉による方法では、表現の自由度が高いた



めに、どのような言葉を用いてアノテーションを行うのかの判断が、アノテーションを行う人間の裁量にゆだねられていると考えられる。多人数でアノテーションを行う際には、個人の表現の差が出ないことが望ましいが、そのためには、事前に十分な用語の定義など打ち合わせを行っておかなければならない。

そのためには、あらかじめ映像全体を全員で通して見た上で、討論を行って、表現の揺らぎが起こらないような言葉を選ばなければならないことがあり、これは非常に時間と労力のかかる作業である。また、このような形での言語によるアノテーションは、連続ドラマのように、途中の段階までしか映像が発表されていない物の場合、途中で新しい状況が現れるたびに、用語の定義を行わなければならない。1人ならいざ知らず、多人数で行う場合には、打ち合わせの時間などを含め統一した表現を作ることが困難になる。

また、せっかくデータを付加したとしても、検索の際に、どのような用語を使用したら良いのかが明確でないと、思い通りの検索が困難である。しかし、そのための用語を、データベースに別途付加するようなことが必要であるとすると、検索の便が悪く、実用的に使用することが困難となる。こうした、不便なシステムは、せっかく開発しても、結局ユーザに受け入れられずに、お蔵入りになってしまう可能性が高い。

今日、インターネットの検索エンジンなどでは、類似語を自動的に推測し、検索してくれるなど、利便性の高いシステムを使用するようになって来たが、同じシステムをそのまま、映像データベースの検索に使用するのは困難であると思われる。

## 11-2 4階層7項目の分類法の提案

そこで、今回の提案では、アノテーションを行う側も入力容易で、検索用語の入力の際の「揺らぎ」が小さくなり、また、検索する側も、特別な予備知識無く、容易に、目的のシーンを検索できるような方法を考案した。

日常我々が目にする映像を表現するのに、どのような方法が必要十分な方法であるかを検討の結果、次のように、大項目を7項目に分類しさらにその下に、各7項目ずつ3階層の分類を構築する。この大項目は、日常的に物事を表現する際に使われている、5W1Hを再検討し映像表現に必要な項目として、8項目に整理した。

具体的には、「いつ」「どこで」「誰が」「なぜ」「何を」「どうして」「どうなった」「感情」という8項目が、最上位階層の分類である。この各項目の下に、さらに、7つの項目を設ける。例えば、「どこで」という項目の下には、「自宅」「学校」「会社」「駅前」「商店街」「公共の場」「不明」というような、7項目を設ける。さらにそれぞれの項目の下にも、7項目の分類を設け、さらに必要であれば、7項目を、さらに細分化が必要な場合には、各項目の下に、7項目を設けることにする。これにより、最上位8項目+3階層7項目という分類ができる。

このような分類を行ったのには、いくつかの理由がある。

まず、最上位項目に関しては、映像表現である以上、必ず必要となる項目であると考えられる。もちろん、人物が登場していないシーンなどでは、「誰が」の項目は、空欄になるが、その下の第2階層の項目に「無人」という項目を作ることによって記述することができる。



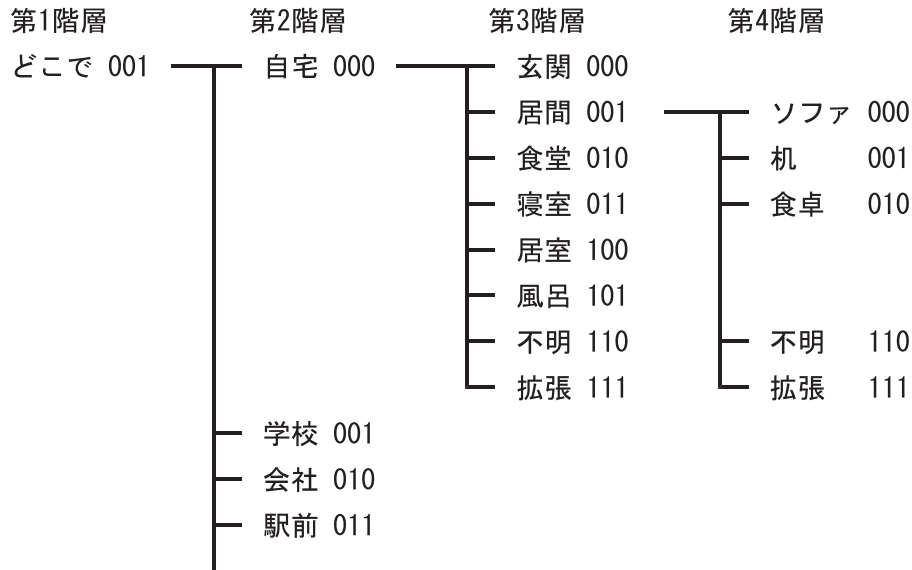


図8 4階層7項目分類と数値化の例（一部）

### 11-3 7項目に分類を制限した5つの理由

第2, 3, 4階層については、それぞれ、7項目に分類することにしたが、これは、次の理由による。

第1に、状況表現する際に、最終的には、数値データとすることを念頭に置いた。この場合は、3ビットで表すことの出来る、8項目から、1項目を、拡張既述用として確保し、残りの7項目を状況既述に用いることで、数値化が容易となることを目標とした。どのような状況既述に於いても例外的な既述が必要な場合が存在する可能性があるため、1項目は、こうした場合に言語を用いた記述を可能とするための、予備項目として予約した。

第2に、状況表現する場合に、大分類-中分類-小分類のような、階層構造で表した方が、平面的に多数の項目を羅列するよりも、わかりやすいと考えた。

第3に、数値化を考えるのなら、3ビットではなく、4ビット(15項目)あるいは、8ビット(255項目)で表す事も出来るのではないかと考えられるが、日常目にする映像を見る限り、7項目で不足することは少ないし、不足する場合でも、その下の階層で分類することで可能と考えた。全ての項目が均等に割り振られる事は希であり、大抵の場合に、分類に「空き」が出来ることがわかっている。特に、連続ドラマやアニメのような場合には、ほぼ決まった状況しか現れないので、こうした空白が出ることはめずらしくない。仮に、理想的に、7項目3階層の全ての項目が埋まったとすると $7 \times 7 \times 7 = 343$ 項目の記述が可能であるが、最上位分類1項目に対して、343項目の小項目の記述が可能であり、實際上、1つの映像で、これだけ複雑多岐にわたる状況が現れることは希であろう。

第4に、7項目という数は、人間にとって、一度に認識できる「マジックナンバー」と言われる数であることが掲げられる。例えば、画面上に一度に7項目の表示がなされた場合、自分の検索したい項目が、その中のどの項目に属するかを判断するのは、難しいことではない。これが、一度に、100もの項目が画面一杯に表示されたとすると、その中から、自分の検索したいものがどこにあるのかを見つけるためには、それなりの時間が必要である。さらに、該当するのではないかと考えられる項目が多数ある場合、どちらの項目に属するのかが、すぐには判断できないという問題が起こりやすい。

このような場合、アノテーションを行う人間による判断の「揺らぎ」が生じることが考えられると同時に検索する側も、どの項目を検索したら良いのか、判断に迷うことになり、検索の利便性を損なう可能性が高い。

第5には、これも視認性の問題であるが、画面上に「WindowsのExplorer」のような階層表示を行う場合に、1階層の項目数が多くなると、画面をスクロールしなければ見えない表示が生じることになる。入力、検索の便宜を考えたときには、スクロールという操作は、存在しない方が望ましい。確かに、4ビット（15項目）程度であれば、スクロール無しに一画面上に表示することが出来るが、ある項目を選択して、下の階層を調べた場合の事まで考えると、7項目であれば、 $7+7=14$ 行の表示で収まるので、1画面内に収まる。仮に検索項目が下位階層に存在しないことが判明し、一段階上の階層に戻る場合の事を考えるならば、このくらいの項目数が適切ではないかと考えられる。1階層15項目でも、下位階層まであわせて $15+15=30$ 行の表示画面で済むので、ぎりぎり1画面に収まる範囲であるが、上記の各種理由を考慮すれば、1階層15項目の表示より、7項目表示のほうが、適切なのではないかと、総合的に判断することが出来るであろう。

#### 11-4 4階層に制限した理由

第6に、階層を4段階としたのは、こうした階層化検索に於いて、階層の深さが不定であるとする、ユーザにとって見れば、どこまで検索すれば済むのかがわからないという、心理的不安を与えると考えからである。必ず、どの項目も、4段階の階層の中に入っていることが、あらかじめ明確になっていることは、ユーザに、少なくとも4階層検索して見つからない場合には、検索が失敗したのであって、元に戻ってやり直す必要があるのだ、という安心感を与えることができると同時に、検索の能率化にも寄与すると考えるからである。そうした意味では、事実上無限の階層構造を可能とするような技法は、望ましくないと考える。

#### 11-5 7項目に収まらない場合の例外的記述法

ただし、例外的に、多数の項目を並列に並べなければならない場合も起こることを考える必要があり、このような場合の例外的表示法も考慮しなければならない。例えば、色の「和名呼称」のようなものは、全てが、同一階層に存在すべき性質のものであるが、微妙な色合いの違いによる和名呼称の違いを一覧とすると、100種類以上の名称があることから、こうした場合の表現法については、例外的に5階層目を設け、125項目あるいは、32766項目の記述を可能とした。

#### 11-6 分類項目の個別化から統合化に

なお、こうした項目は、最上位項目を除いて、個別それぞれの映画、ドラマ、アニメなどによって、別の項目を設ける必要がある。例えば、SFアニメーションであれば、「どこで」の下位分類の項目には、「宇宙空間」「スペースコロニー」「月面」というような項目が入るかも知れない。ただし、項目の作り方を、試行錯誤で練り上げていくことで、類似のジャンルの映像に関しては、共通の分類項目を使用することが可能になると考えられる。例えば、「パニック映画」「恋愛ドラマ」「SFアニメ」というような、大まかなジャンルには共通の項目を作ることが出来ると期待される。現在のところ、個

別の映像に対して、おのおの別の項目建てを行っている段階であるが、各種の映像に関する分類が蓄積して来た段階で、項目建ての整理・再構築を行うことで、次第に広い範囲の映像に対応出来る項目分類法が構築できると考えられる。ただし、個人の氏名や、地名などは、例外的に文字による記述を行わなければならない場合も起こるだろうが、現在までの検討の結果では、「誰が」の分類法をうまく構築することで、個別の登場人物の氏名を数値化された項目に分類することは可能であると考えられる（実際、群衆などを除いて数百人もの人物が登場する映像は、ほとんど存在しないと考える良いだろう）。

### 11-7 階層化表記の本質的問題

ここで、言語での状況の表現の際に常につきまとう問題であるが、分類の上下関係が明確にならない場合が存在することについて、触れておく。このことは、特に階層化関係を構築する際に注意しなければならない。

例えば、「赤い花」「猫の人形」の様な例である。後者の場合であれば、猫という分類の下に人形を置くのか、人形という分類の下に猫を置くのかを決めることは困難である。ただし、これは、言語表現には、常につきまとう問題であり、各所で議論されているが、本論の範囲を超える問題であるので、これについては、今後の研究の進展を見たい。なお、本研究に於いては、猫の下の人形と、人形の下の猫と、両方とも記述することで、当面の対応策とする（本来このように同一のものが、複数の場所に存在することは好ましいことではない）。

### 11-8 検索性の向上法

こうした形で、4階層7項目（最上位階層のみ8項目）に、分類することにより、アノテーションを行う人間が項目を選ぶ作業が容易になるばかりでなく、データ入力「揺らぎ」が減少すると同時に、検索を行う側にとっても、特段の予備知識や、別途の一覧表などを用意することなく、画面上のクリックのみで、希望の項目を選択することが可能となると考えられる。なお、入力する側は、最上位項目については、全ての項目を必ず入力することとするが、（該当項目無しの場合は、2階層目に「無し」の項目を用意しておく）検索する側は、全ての項目を入力する必要はなく、不明あるいは、不要の最上位階層については、入力を省略することで、その分類に何が記述されていても該当するシーンを検索できるようにする。こうすることにより、例えば、検索者が、「どこで」だけを検索項目として入力した場合に、そこに、誰が居ても、何が映っていても、その場所の映像だけを選択して表示することが可能となる。勿論、複数分類を入力することで、絞り込み検索が出来ることは、インターネット上の検索エンジンと同様である。

### 11-9 階層化項目分類の再整理

階層化項目分類法では、新項目の追加、移動を誰がどのように行うかが大きな問題であるが、ここでは、個別のアノテーション入力者にゆだねることにした。時には階層化がきちんと出来ない場合などもあり得るが、その場合、階層や項目の移動を行っても、今まで映像に付加されたデータが無効にならないように自動的に変換、移動する方法も考える必要がある。MPEG-7では、ここで定義したような階層構造と項目数に制限を設けることはしていないが、このような、項目（ノード）間の関係が移

動した場合にも、対応できるような規格が制定されている。

本論で述べた、制限された階層化分類法の長所は、無秩序に階層や項目を増やすことで、全体の関係が複雑でわかりにくくなることを防止し、常に、定型の階層と項目内で事項を記述していくことにより、自然に（なかば強制的に）分類の方法が明確化し、整理され、わかりやすいものとなっていくことが期待されることである。

また、連続ドラマやアニメのように、途中までしか映像が存在しない場合においても、空白項目が少なからず存在することが考えられるため、必要に応じて、そこに、新項目を設けていくことが出来る事もこの手法の有効な点である。万一項目が足りなくなった場合にも、拡張予約項目を用いれば、ほとんどの場合対応可能であり、また、現在の案では、4階層を超える階層、7項目を超える項目の必要となった場合について、4階層目の拡張項目を使用して、その下に、7ビット（127項目）ないし、15ビット（32766項目）を設ける事で対応可能な設定を行う予定である（現在のところまだ、そのような拡張階層、項目の必要な場合は、存在していない）。

## 12. まとめ

本論文では、始めに述べたように、MPEG-7が、規格制定されて何年も経つにもかかわらず、なぜ実用段階に至らないのかの理由に関して考察し、その規格が複雑で、また、アノテーションにかかる費用に対して、採算が取れないであろう事について考察した。

そこで、MPEG-7規格を全て実現することは、将来の課題としつつ、その部分的実現の方法と、新技法による、映像分割、及びアノテーション、データベース化の技法について実験結果に基づき考察した。そして、アノテーション及び検索を能率化し、「揺らぎ」を無くすことが、実用化に於いて重要であると分析した。また、現在のMPEG-7の技法では長時間の映像のアノテーションが困難であることが、実用化に至らない大きな理由であるという仮定の下に、30分から2時間という、実際の映像にアノテーションを行う技法を提案した。そして、シーンのクラスタ化の技法と、それによってアノテーションが大きく効率化、安定化、省力化することを示した。

次いで、MPEG-7を商業レベルで実用化する方法について考察を行った。ここでは、新たに、MPEG-7アノテーションの専門業者を設立することで、映像制作会社と分業体制を作り、さらに顧客の求めるものが提供できる方法を考察、商業レベルに乗せる道筋を示した。

さらに、映像アノテーションに於いて、4階層7項目のデータを付加する方法を新たに提案した。これにより、アノテーションの能率化と、「揺らぎ」を少なくおさめることが出来ること、そして、検索に於いても、能率的な検索が可能であるという道筋を示した。

最初にも述べたように、本研究は、MPEG-7の規格を忠実に実現する物ではないが、少しでも、MPEG-7の実用化の助けとなるような、新しい技法の提起を行った。この研究が、将来の映像データベース化の普及の道筋を示すものとなれば幸いである。

### ●参考文献

- (1) ISO/IEC 15938 日本規格協会  
<http://www.jsa.or.jp/>

- (2) カット  
<http://ja.wikipedia.org/wiki/%E3%82%AB%E3%83%83%E3%83%88>
- (3) ショット  
<http://ja.wikipedia.org/wiki/%E3%82%B7%E3%83%A7%E3%83%83%E3%83%88>
- (4) シーン  
<http://ja.wikipedia.org/wiki/%E3%82%B7%E3%83%BC%E3%83%B3>
- (5) 「MPEG7によるCM・ドラマのデータベース化と、シーン分割法の研究」藤木 文彦 電子情報通信学会総合大会講演論文集 Vol.2005年\_情報・システム, No.2 p.176 社団法人電子情報通信学会  
<http://ci.nii.ac.jp/naid/110004746356/>  
<http://www.wakhok.ac.jp/~fujiki/ronbun/ieice/2004/mpeg7-2.html>
- (6) 「MPEG-7アノテーションのためのシーン分割・クラスタリングの研究」藤木 文彦 映像情報メディア学会2005年度年次大会論文  
<http://www.wakhok.ac.jp/~fujiki/ronbun/ite/2005/mpeg7-1.html>
- (7) 「MPEG-7メタデータ付与の新規格と技法の研究」藤木 文彦 情報処理学会2007年度春季全国大会論文  
<http://www.wakhok.ac.jp/~fujiki/ronbun/ipsj/2007/mpeg7-2.html>
- (8) 「MPEG-7のメタデータ付与に関する規格拡張の考察」藤木 文彦 映像情報メディア学会2006年度年次大会論文  
<http://www.wakhok.ac.jp/~fujiki/ronbun/ite/2006/mpeg7-1.html>
- (9) 「MPEG-7による映像のデータベース化」浅利 純、金子 隼之 2002年度稚内北星学園大学総合研究論文  
<http://www.wakhok.ac.jp/~fujiki/semi-kenkyu/2002/asari/asari.pdf>
- (10) 「MPEG-7による映像データベース化と検索エンジン」石川 健太、蔵 知晃、山根 康裕 2003年度稚内北星学園大学総合研究論文  
<http://www.wakhok.ac.jp/~fujiki/semi-kenkyu/2003/isikawa/isikawa.pdf>
- (11) 「半自動ビデオアノテーションとそれに基づく意味的ビデオ検索」山本 大介 名古屋大学工学部電気電子・情報工学科、長尾 確 名古屋大学情報メディア教育センタ 情報処理学会第65回全国大会  
[http://www.nagao.nuie.nagoya-u.ac.jp/papers/yamamoto\\_ipsj65.xml](http://www.nagao.nuie.nagoya-u.ac.jp/papers/yamamoto_ipsj65.xml)
- (12) 「半自動ビデオアノテーションとそれに基づく意味的ビデオ検索に関する研究」山本 大介 名古屋大学工学部電気電子・情報工学科 卒業論文  
[http://www.nagao.nuie.nagoya-u.ac.jp/papers/yamamoto\\_sotsuron03.xml](http://www.nagao.nuie.nagoya-u.ac.jp/papers/yamamoto_sotsuron03.xml)
- (13) 「マルチメディアコンテンツの高度利用のためのアノテーション技術の基礎と応用」山本 大介 名古屋大学 情報科学研究科 メディア科学専攻  
[http://www.nagao.nuie.nagoya-u.ac.jp/papers/yamamoto\\_thesis05.xml](http://www.nagao.nuie.nagoya-u.ac.jp/papers/yamamoto_thesis05.xml)
- (14) 「MPEG-7と映像検索」國枝 孝之/脇田 由喜/高橋 望 共著 CQ 出版社
- (15) 「技術開発 (MPEG 技術解説)第5章」Pioneer corporation  
<http://pioneer.jp/crdl/tech/mpeg/5.html>
- (16) 「ドコモとIBM、「MPEG-7」を利用した携帯電話向け動画配信技術」  
<http://wince.goo.ne.jp/snap/cnBoard.asp?PID=281>
- (17) 「MPEG-7を利用したハイパーメディア・システム」T. Kaneko, K. Masukura and O. Hori, "Description Method for Spatio-temporal Regions in a Video and Its Application," International Journal of Imaging Systems and Technology (IJIST), Vol.13, Issue 5, pp.257-266, 2003  
<http://www.toshiba.co.jp/rdc/mmlab/tech/w15.htm>
- (18) 「MPEG-7 Japan Domestic Pages」 Information Processing Society of Japan  
<http://www.itscj.ipsj.or.jp/mpeg7/>
- (19) 「マルチメディア内容記述標準 MPEG-7 とその応用」, 「電子情報通信学会誌 84巻9号」, (2001年9月1日)、



堀修著、電子情報通信学会発行、636頁 図2 ハイレベルな記述の概要

[http://www.jpo.go.jp/shiryousonota/hyoujun\\_gijutsu/bidirectional\\_video/111\\_03.htm](http://www.jpo.go.jp/shiryousonota/hyoujun_gijutsu/bidirectional_video/111_03.htm)

(20) 「動画像におけるシーン分割ツールの作成」 木村 聡一郎、阪田 泉 中部大学情報工学科

<http://www.sk.cs.chubu.ac.jp/abst7/abst-motomura.pdf>

(補遺) MPEG-7 アノテーションツールは、日本では「リコー社」が、海外では「IBM社」が開発していた。リコー社は、有料で販売・サポートを行っていたが、2004年にサポートが打ち切れ、現在、ホームページが見あたらない。IBM社は、 $\beta$ バージョンを無償配布していたが、こちらも、製品化されず、配布も打ち切れ、ホームページが見あたらなくなっている。

### ● 英文タイトル

Research of the scene division technique aims to make utilization of MPEG-7, and division-by-class annotation technology.

### ● 英文要約

At first, we will describe the purpose of this paper. In the standard of the MPEG-7 annotation which is a standard of image database creation, the definition over which it goes very in detail is carried out about the describing method. Therefore, that mounting make difficulties to make utilization of MPEG-7. Moreover, similarly, in spite of claims for the necessity for image database creation, the technique of actual database creation is still a research stage. In this paper, database creation of the image considered pressing need, so we decisively, omitted realization of the detailed portion of MPEG-7 and inquired about creation of the simple image database towards utilization, and the technique for use. Therefore, this research is not what was faithfully based to the standard of actual MPEG-7, and since a sometimes bold abbreviation and change were made, the portion which has been MPEG-7 standard of the basis to what was widely different exists plentifully. In order that the standard of MPEG-7 may go over having performed such an abbreviation and having made a change variably too much in detail, even if the difficulty of the mounting carries out a major company, it is considered to have closed utilization development. If it is going to research the same thing in an individual, coming to a deadlock will be easily predicted. So, in this research, we did not dare carry out taking in such all standards, but decided to convert the bold way of thinking. Although reproach may be received from many people concerned with standard establishment of MPEG-7, and researchers, utilization research in such a form will be useful for future image database creation. Then we performs a report of research and a proposal.