

MPEG-7 メタデータ付与の新規格と技法の研究

藤木文彦

稚内北星学園大学 情報メディア学部

MPEG-7による映像情報へのメタデータ付与に関して、従来は、カット・シーンの分割の後、人手によるデータ付与を行っていたが、この方法では、付与する人間によるデータの内容・形式のばらつきの他に、同じ人間が、類似の映像にデータを付与する場合であっても、その内容・形式にばらつきが生じることが避け難く、単に多くの労力が必要なだけでなく、後の検索における困難が伴っていた。ここでは、まず、映像の、カット・シーンの分割とクラスタ化の新技法を提案すると共に、メタデータの16ビット規格化と4段階の階層化により、入力の省力化と、データの安定性を増大させる方法と、検索の容易性への寄与などを提案する。

A study of new standardizing and metadata annotation for MPEG-7

Fumihiko FUJIKI

Wakkanai Hokusei Gakuen University

Metadata annotation of movie information was annexed by human hands, after cut and scene separation. But, by this method, even if the same human may annex other metadata to a similar image, so it was hard to avoid that variation arises in its contents. And the difficulty in data search. Here, we proposing the new technique of cut and scene separation and making cluster. Next, we proposing the stability and labor saving of the input data, by 16 bits standardizing and 4 stage hierarchizing.

1. MPEG-7普及における問題点

MPEG-7の規格制定がほぼ終了して、5年近くなるが、この間、MPEG-7の実用的な使用がなされていない。研究論文も散見される程度であり、商業レベルにのりような実用例は皆無である。

原因として、考えられることを列挙する。

- ・メタデータを付与するコストが大きい
映像に対するメタデータの付与は、現状ではほぼ人力に頼っているが、そのメタデータを付与するための時間がかかる。
- ・メタデータの品質が一定でない。
映像に対するメタデータの付与規格が明確でないために、たとえば、同じような画面でも、人によって付与するデータが違ったり、同じ人でも、同じデータが付与されるとは限らず、メタデータの品質が一定範囲におさまらない。
- ・実用的な長さの映像に対するメタデータ付与が困難である。

現在のアノテーションツールでは、10分程度の映像を扱うのがせいぜいであり、1時間、2時間あるような映像に、メタデータを付与する処理が困難である。

- ・メタデータの活用が確立していない。
メタデータをもとに、映像を検索し、表示するシステムなどが開発されていない。
- ・googleなどの検索システムの充実
googleなどのWeb検索システムにより、映像に関する概略のデータが検索できるようになり、多くのユーザにとって、それ以上の細かい検索がいなくなった。
- ・商業的メリットが考えにくい。
メタデータを付与するコストを、市販DVD等に乗せした場合、メタデータの無いものより、メタデータを付与したもののほうが、高価となるが、両者を比較したとき、購買者が、メタデータ付きのDVDを購入することは、考えにくい。上記に述べたように、メタデータの活用環境が充実して

いない段階で、メタデータ付きのDVDを買うメリットがない。

上記のような、各種の事情により、MPEG-7の実用化が遅れているが、増大する映像アーカイブに対応するためには、MPEG-7による、メタデータの付与が必要不可欠であると考える。

本稿では、これらのうち、メタデータの付与を容易にする、シーン分割、統合の新技法、メタデータの規格化の提案と、安定したデータの付与の方法の研究などについて論じることとする。

2. アノテーションのための技法 (1) シーン分割の新技法

2-1. シーンの分割技法

まず、用語を定義する。

カット：最小の連続した1つの映像の単位。

シーン：いくつかのカットをまとめた、意味的につながりのある、一場面。

現在、HDDレコーダなどでは、自動チャプター分割機能などが搭載されているものが多いが、その仕組みについては、公開されていない。映像が大きく変化する場所で区切っているものと思われるが、音声などで、注目すべき部分を検出する方法などもあり、各種の技法が使われているものと思う。いずれにしても、あまり細かな場面転換にチャプターを付けても、多くなりすぎて逆に不便である。

いくつかの開発されたアノテーションツールでは基本的に場面転換ごとにシーンを分割しているようであるが、どのような技法を使っているのかについての詳細は明かされていない。入力がMPEG映像であれば、Iフレームが、通常の場合以外に挿入されている場合、そこを、場面転換と見ることが出来ると思うが、いずれにしても、大きく映像の変化する場面を分割点と見なしていると推測される。

しかし、映像処理的には、前後の映像を比較して、大きく変化する点を見いだす作業は、かなりの時間を要する作業であると考えられる。



fig.1 同一のシーンと見なされるカットの例

2-2. 輪郭重心法

ここで、カットの分割点とを見いだす、新しい技法を提案する。

一般的には、前後の映像を比較して大きな変化の検出された場所をカット分割点と考えるのが妥当であるが、移動する映像であったり、登場する人物が移動していたりする場合には、前後フレームの比較を、まっとうな方法で行うと、連続した画面か否かを判断するためには、大きな処理能力が必要となる。

ここでは、より、大まかではあるが、カット単位に分割する技法として「輪郭重心法」という技法を提示する。この技法は、2枚の映像を比較する際に、映像全体を比較するのではなく、その映像の「重心」位置を比較し、変化の大きい場所を、カット分割点と見なす方法である。これにより、計算量の大幅削減、多量の映像の効率的なデータの蓄積と、比較を容易にするものである。

輪郭重心法は、次のようなステップからなる。

- ・映像を白黒映像に変換する。
- ・変換された映像から、輪郭線を抽出する。
- ・輪郭線の各点から、その重心位置を求める。
- ・重心位置が、連続して変化している部分を同一カットと見なし、不連続な点をカット分割点と見なす。

実際には、連続した映像であっても、登場者が移動したり、画面上に大きな変化があった場合などは、分割点と見なされる場合もあり、また、輪郭線の検出も、どの程度の変化があった場所を輪郭線と見なすか。輪郭線の太さをどのように取るかなどのパラメータに関しては、適切な閾値の設定が実験的に必要である。また、映像の種類によって適切な閾値の設定が異なるので、多数の映像によって検証していく必要がある。映像の種類によってアニメと、映画と、ドラマでは異なるし、また、ドラマでも、日本の映像と、海外の映像では、色調などに大きな違いがあるために、全てに有効なパラメータの設定は不可能と思われるので、映像の種類とパラメータのデータの蓄積も必要となる。



fig.2 輪郭重心の求め方

2-3. シーンの構成

いくつかのカットを集めて、意味的にまとめた場面を構成することをシーンの構成と呼ぶ。例えば、部屋の中で、2人が会話している場合、カメラアングルを変えるなどしたいくつものカットがあっても、それらをまとめて、一つのシーンとすることが望ましい。

アノテーションに当たっては、以前は、カット単位でアノテーションを行っていたが、分量が膨大になることと、類似カットが何度も出てくる場面（2人で会話しているところなど）では、それぞれにメタデータを付与する必要性が少ないことなどから、アノテーションを付けるにあたっては、まず、シーン単位で大きなアノテーションを行い、その上で、カット単位のアノテーションが必要なら、付けるようにした。

- ・シーン単位でのアノテーション
- ・展開の必要性がある場合、カット単位のアノテーションを行う。

たとえば、2人の人物が、話しをしているうちに、だんだん話しの様子が変わってくる、というような場合は、場所や、人物、という情報は変わらないとしても、話しの内容と、感情の変化などについては、個別のカットごとのアノテーションが必要と考えられる。

ここで、アノテーションの階層化が、必要となる。ここでは、シーンの中に、カット単位のアノテーションが必要と述べたが、実は、大きく見ると、シーンがいくつかまとめて、さらに大きな状況を構成している場合もあり、こうした階層化が、従来のMPEG-7アノテーションでは、述べられて居なかった部分である。

| | | | |
|-----|-----|-----|-----|
| 状況 | | 状況 | |
| シーン | | | |
| カット | カット | カット | カット |

fig.3 状況、シーン、カットの関係

2-4. 人手による補正

どのようなシーンも、意味的なつながりの有無については、最終的には人間の判断が必要となる。

例えば、地上と、宇宙船の中、というような場面では、明らかに、シーンが異なると考えられる。しかし、

地上の司令室と、宇宙船との交信、というような場面であれば、映像的には、異なる場所であっても、意味的には、連続した1シーンと見なすのが自然である。また、走りながら、相手を追跡する場合など、自動的な検出では、連続したシーンとみなすかどうか、判断させるのは困難である。

このように、意味的なつながりの判断は、機械には不可能であるので、人間が判断して、シーンの再構成をしなければならない。

2-5. シーンのクラスタ化

映画やドラマなどの映像では、時間的に離れた所に、同じ場所、同じ人物のシーンが現れる事が多い。特にアニメなどでは、「使い回し」の必要性から、そのような場面の出現する場面が多い。こうしたシーンでは、登場人物の心理や状況などが異なっても、場所や人物に関するアノテーションはほぼ同一となることから、類似度の高いシーンとみなすことができる。

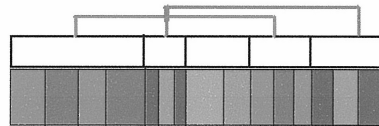


fig.4 離れた場所のシーンのクラスタ化

こうしたシーンは、類似シーンとして、クラスタ化することが望ましい。従来の技法であれば、このようなクラスタ化は人力で行うしか無く、記憶に頼らなければならない部分が多い。あるいは、機械的に画面全体の比較を用いるとすると、長時間にわたる多数の映像の比較が必要となり、事実上計算が困難である。今回提案の輪郭重心法では、1カットに付けられた特徴データが、少ない数値で表されているために、長時間の映像内での比較が容易である。こうして、全映像内のカットを比較し、データの類似度によって、クラスタ化することで、類似場面の存在場所を一覧化し、検索を、より効率的に行う事が出来る。

なお、この方法も、異なる場面を同一クラスタと認識したり、同じ場面を同一クラスタに出来ない場合など、誤判断の可能性があるので、最終的には、人間による調整が必要となる。

また、どこまでを類似クラスタと見なすか等についても、その閾値の設定は、試行錯誤によるところがあり、適切なパラメータは、映像の種類によって異なる。ただ、連続ドラマ、アニメのような場合、回を重ねる

に従って、同じ場所、同じ登場人物が出現する事が頻発する。むしろ、回を追うごとに新規の状況の出現確率は減少していく。したがって、このような映像に対しては、分析経験の積み重ねが効果を持つようになり、自動的な分類が威力を発揮すると思われる。

3. アノテーションのための技法(2) メタデータの数値化・階層化

3-1. アノテーションの規格化と容易化

アノテーションの問題の一つが、同じシーンであっても、付ける人によって、メタデータの内容が変わることや、同じ人でも、同じメタデータを付けることが出来るとは限らないことがある。このような不確定さは、検索の際に問題となり、アノテーションの困難な理由となっていたと考えられる。

そこで、ここで、付与するメタデータを、数値化することにより、同じシーンに付与されるメタデータの統一性を確保すると共に、必要十分なデータ付与が確保されること、データ付加の容易性を実現する事ができると考え、その技法を提案する。

3-2. 付加するデータの種類

従来、カット、シーンへのアノテーションは、自然言語によって行われていた。しかし、これには、様々な問題があった。

- どのような状況説明をメタデータとして付加するか明確な規定が無かった
- 同じシーンに付けるメタデータも、付加する人によって異なる。同一人物であっても、同じシーンに同じメタデータを付けるとは限らない。
- 同じシーンに対して、何度も同じメタデータを付けることは、非効率である。

これは、MPEG-7が、映像に対してどのようなメタデータを付与すべきかの明確な規格を持っていないためである。そこで、映像に付与するメタデータの種類と、内容を、定義しておくことが必要とされる。

3-3. 付与するデータの数値化

データの数値化に当たって、4階層、各7項目の行う事とした。

各階層は4ビットからなり、上位3ビットが項目分

類、下位1ビットが、続く階層がある(1)か、無い(0)かを示すものとする。

3ビットでは、8種類の項目を表す事が出来るが、そのうち、000(2)~110(2)までの7項目を、各種用語分類に割り当て、111(2)を拡張に用いる。

なお、1110(2)は分類不能、1111(2)は、将来の拡張のために予約するものとする。

状況描写に当たっての最上位階層は、大分類として、次のように分ける。この階層の記述は必須であり、全てのメタデータは、この8項目を含むものとする。

時間 000 場所 001 人物 010 対象 011
行為 100 感情 101 結果 110 拡張 111
その下に、3階層の分類を付加する。

最上位階層 - 第1階層 - 第2階層 - 第3階層
例)

場所 - 建物内 - 食堂 - 居酒屋
0011 0011 0101 1010

のように、16ビット(2バイト)で表される。

なお、さらに、言葉による記述の必要なものは、言葉により補足する。

例：居酒屋名、人物名、場所名など。

第1, 2, 3階層は、必須ではなく、存在しない場合もある。また、例外的にさらに、多くの階層を設ける場合もある。また、各階層は、必要に応じて1項目について、2つ以上設ける場合もある。その場合は、最上位階層から記述するため、同一の最上位階層が複数存在することもある。

* どの階層も、他、不明の項目を含むが、記述は省略してある。

| 最上位 | 第1階層 | 第2階層 |
|-----|------------------------------------|------------------------------------|
| 時間 | 实在現在、仮想現在、实在過去、仮想過去、未来、年代不明、未設定、不明 | |
| 場所 | 屋外 | 路上、公園・遊技施設、山・海、空中・海中・地中、 |
| | 建物内 | 住居・宿泊施設、会社・職場、学校、店舗・食堂、工場・研究室、司令室、 |
| | 乗物 | 自動車、鉄道、2輪車、飛行機、船、ロボット、 |
| | 宇宙 | 宇宙船内、惑星上、宇宙空間、宇宙基地、ロボット、 |
| 人物 | 人数、主人公、家族、仲間・同僚、敵、 | |
| 対象 | 人間 | 本人、家族、友人・知人、職場同僚、恋人、敵、 |
| | 持物 | 鞆、食品、ナイフ、銃、 |
| | 動物 | ペット、野生動物、実験動物 |
| | 機械 | 家庭内小物、電話、家電、コンピュータ、工場機械、 |
| 行為 | 立つ・座る、歩く・走る、話す、操作・操縦する、見る、考える | |
| 感情 | 喜び、悲しみ、怒り、疑惑、不快、恐怖 | |

fig.5 メタデータの階層の規格化の例

ここで、1階層の項目を7項目としたのは、人間が一度に認識できるもの数が7つ程度までという研究の結果をふまえ、処理画面上で、直感的に即時に分類出来る項目数として妥当と判断したからである。ビット数の制限を問題にしなければ、8ビットでも構わないはずであるが、あえてそうしなかったのは、8ビットのうち、7ビットを、同一階層の項目として使えば、最大、127種類の記述が出来るが、これだけ細かい分類をすると、一度に並べて表示したときに、適切な用語を見いだすのに時間がかかること。複数の選択肢の中で、どれが最も適切な用語かを判断するのが困難であり、結果として、アノテーションを行う際の表記のばらつきが生じやすくなることにある。

階層を、3階層に限定したのは、一つのもの事を記述するのに、3階層あれば十分であり、それ以上の階層に分類すると、煩雑となることや、さらなる細かな分類が必要な状態の場合は、言語による記述を併記した方が、より適切であると判断したからである。(※5)

この3階層を全て使えば、1つの大分類ごとに、 $7 \times 7 \times 7 = 343$ 種類の記述が出来るが、実際には、各種のばらつきから、100種類以下の物事の記述しか出来ないと思われる。しかし、それでも日常の事象の記述には十分であると考えられる。

さらに、分類の際だけでなく、検索に当たっても、3階層以上の階層化は、むしろ不便であると考えられるからである。3階層であれば、目的の分類に到達しなかったときに、上位に戻っての再検索が容易であると思われる。

3-4. 言語によるメタデータ付与と秘匿度の設定

登場人物や場所などの固有名詞等は、あらかじめ規定の数値化することは困難であるので、言語でメタデータを付与する。あるいは、下記の方法で、拡張設定を用いて、メタデータ化する。いずれにしても、既成の分類だけで不十分な場合は、言語による補足説明が必要である。また、感情の変化など、数値化の困難な物も言語による補足説明が必要である。また、「このときのA氏の右手に注目」のような、解説者的立場からのメタデータの付与も必要とされる場合もある。ただし、このような解説的メタデータには、いわゆる「ネタばらし」となる記述が含まれる場合もあるので、例えば、推理ドラマで「犯人A氏は、このとき、毒を入れた」のような記述に関しては、検索者が必要としているかどうかの判断も必要となる。従って、これらの記述には、「秘匿度レベル」の設定のようなパラメ

ータを付加することが必要と思われる。

例えば、「全てのデータを表示：0000」「犯人に関する情報は表示しない：0100」「解説者Bの解説は表示しない：1100」というようなメタデータ秘匿レベルの設定も必要となるとと思われる。

3-5. 個別定義データの必要性

映像により、登場する人物や、場所、特殊な設定など、一般的なメタデータとして、数値化出来ないものがある。こうしたメタデータに関しては、たびたび登場するものであれば、数値化するのが望ましい。そこで、そのための、個別定義データを任意の時点で追加できるように仕組みが必要である。連続ドラマの場合、初回にほとんどの定義が可能であるが、回を追うごとに、変化が生じる場合もあり、リアルタイムで放映されている番組などでは、あらかじめ、全ての定義を行っておくことは出来ない。

そこで、任意の時点で、新しいメタデータの数値化定義を行えるようにしておく。こうしたデータとして考えられるのは、登場人物、場所、固有の物品名のほか、特有の概念も含まれる。例えば、アニメ「エヴァンゲリオン」に於ける「人類補完計画」などという用語は、物語の根幹をなす用語で、多数登場するものであるにもかかわらず、他のアニメでは使用されない特殊な概念であり、その意味するところが、物語の最後に至るも、謎のまま残されている。このような用語も、あとから定義出来なければならぬ。

また、同じくアニメでは、「ガンダム」のように、

拡張 - 場所 - スペースコロニー - サイド3

拡張 - モビルスーツ - ザク

拡張 - 科学設定 - ミノフスキー粒子

といった、通常用いられない架空の設定を多数必要とする。

また、映画、ドラマなどで、登場する人物、物なども、毎回言語で記述するよりも、今後たびたび登場すると予測された時点で、拡張設定に入れておいたほうが良い場合もある。

こうした定義は、別の形式で、いつでも行う事が出来るようになっていく必要がある。また、拡張設定は、作業中に行ったときに、それ以前の部分にも影響を及ぼすことがあるために、入力インターフェースとしては、拡張設定時に、それ以前の、言語による設定を数値に置き換えるような作業を行う必要がある。これは、処理時間等を検討して、即時に行うか、後で行うかを選択できるようになっていることが望ましい。

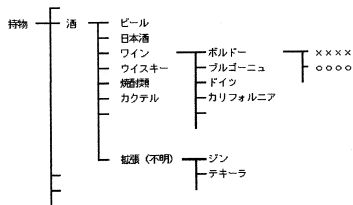


fig. 6 メタデータの階層化の例

なお、全ての設定は、一度設定した後での変更は困難であるが、不可能ではない。より適切な分類法に変更する場合には、全てのメタデータをチェックして変換しなければならないが、実用上は、それほど長時間を要する作業ではないはずである。(現在のPCの性能からすれば、数万行の文書の検索置換も、数秒とみならず可能である。)

4. メタデータ入力ツールの開発

こうしたアノテーションツールは、IBM、リコーなどが出しているが、上記の仕様に合わせた物とするため、現在独自に開発中である。上記の仕様を満たすためには、大きく2つの機能が要求される。

1) カット・シーン分割の機能と、それを補正する機能。シーンのクラスタ化の機能。適切な閾値の設定機能と、映像の種類別の適切なパラメータを、保存、読出、修正する機能。

2) 階層化メタデータの構築と、それを映像を見ながら、入力する機能。頻出する場面に対するメタデータを記録し、再度類似映像が表示された際に、それを再利用することができる機能。言語によるメタデータの付与、拡張メタデータの入力機能を備える。

5. 検索ツールの開発

上記のようにして付けられたメタデータを元に、映像の中から、検索したい部分を探して表示する機能を持った映像ブラウザの開発が、同時に行われなければならない。

映像の検索に当たっては、自然言語による検索と、前記の数値化・階層化されたメタデータによる検索の両方の機能を備え、類似度を判定して、該当する映像のサムネイルを表示し、指定の映像を再生する機能を要する。必要に応じて、指定箇所を出力する機能を備える。類似度の判定アルゴリズムに関しては、今後研究を要する。

6. 制作時のメタデータの付与

こうしたメタデータの付与は、本来映像を制作する時点で行われるのが望ましい。しかし、通常のテレビ放送などでは、映像制作に忙しく、メタデータを付与する余裕は無いものと思われるが、今後のデジタル放送や、DVD等の普及にともない、こうした付加価値を付けた映像が、商業的利益の増大をもたらすようになれば、制作時のメタデータの付与が一般的になるものと思われる。制作者、特に監督や脚本家がそのような作業を行う時間的余裕は、少ないと思われるが、少なくとも、専属スタッフを擁するだけのメリットをもたらすことが出来るようになれば、より一層、MP E G-7 が普及することになるであろう。

7. まとめ

MPEG-7の実用化のために、上記のようなシーン分割法と、メタデータの規格化について、現在研究を進めている。これにより、メタデータの入力が容易となり、入力者による表現の揺らぎも少なく押さえることが出来ると同時に、検索に当たっても、実用的な検索が可能となることから、MPEG-7を普及させ、新旧の映像に対して、メタデータを付与することによる付加価値が増大し、商用レベルに乗るものとするのが可能であると考えられる。これにより、今までに蓄積された膨大な映像資料の検索が容易となり、今後作成される映像にもメタデータを付与した物が増えていくものと期待される。

文 献

- 1 浅利純、金子隼之: "MPEG-7による映像のデータベース化", 稚内北星学園大学2002年度総合研究論文
- 2 蔵知晃、石川健太、山根康裕: "MPEG-7による映像データベース化と検索エンジン" 稚内北星学園大学2003年度総合研究論文 <http://www.wakohok.ac.jp/~fujiki/>
- 3 藤木 文彦: "MPEG-7のメタデータ付与に関する規格拡張の研究", 映像情報メディア学会2006年度冬期大会
- 4 日本規格協会 ISO/IEC 15938(1-8)
- 5 伊藤・比企・小暮・石橋・長谷川 "MPEG-7記述を用いた映像検索プロファイルの検討" 山形映像アーカイブセンター情報メディア学会誌 2003年 Vol. 57-7

連絡先

〒097-0015 北海道稚内市若葉台1丁目 稚内北星学園大学

E-mail: fujiki@wakohok.ac.jp

ffujiki@mail.goo.ne.jp

URL: <http://www.wakohok.ac.jp/~fujiki/>

<http://homepage3.nifty.com/ffujiki/>