

## 論文の内容の要旨

論文題目 TV局における映像メタデータ付与技術ならびに  
映像検索の効率化に関する研究

氏 名 伊 藤 学

一般的に、映像メタデータ付与（生成）技術は、大きく分けて 2 つの方法がある。1 つは、PC などを用いる“自動付与”と、人による“手動付与”である。両者ともに、意味的に価値のあるメタデータ生成を目標としているが、生成にいたる時間や中身などで大きな差が生じている。現在行われている多くの研究は、自動付与が中心となっており、機械的な情報をいかに意味的情報に変換するかについて報告されているが、満足な性能を得ている研究はほとんどない。そこで本研究では、この逆のアプローチ、すなわち、手動付与を前提に、付与時間を高速化する方法に取り組んだ。例として、“記述に 10 秒かかるメタデータを 100 個記述”を挙げ、その高速化する 3 つの方法論について説明する。方法論 1 は、“100 個より価値ある 10 個のメタデータ記述で、100 個と同等の検索効率を得る”で、方法論 2 は、“10 秒かかる記述を 1 秒で行う”で、そして方法論 3 は、“10 個記述すると、のこり 90 個は自動で付与される”である。

また、手動付与によるメタデータおよびその生成に対し、5 つの要求条件を定めた。1 つ目は、“記述が高速化できること”である。2 つ目は、“揺らぎがないこと”。3 つ目は、“独立性に優れていること”。4 つ目は、“分解能に優れていること”。5 つ目は、“実現容易性があること”である。前記した 3 つの課題により生成されるメタデータは、これらすべての要求条件を満たす手法である。

手動記述メタデータは、人が対象となる映像を見て、その特徴を記述することになる。そのため、時間や労力など多くのコストがかかりこれまであまり研究されていない。しかしながら、上記の要求条件を満たすことにより、自動付与により生成されるメタデータよりも、価値のあるものが生成可能となる。

本論文では、上記 3 つの方法論を可能とするアプリケーションとして、TV 局が保有する

2種類の映像である番組と素材を取り上げ、検索効率促進を目指す。

方法論1のアプリケーションとして、課題1では、視聴者による番組検索に検索対象となるメタデータ付与手法として、“過去の放送済映像に対しメタデータ付与の箇所を自動推定し、メタデータ付与を高速化する手法”と題し取り組んだ。ここでは、従来研究の成果を踏まえ、番組の中でも特にその構造化が困難な物語番組より、映画を取り上げる。過去に放送された映画を流通させようとしたとき、メタデータ記述しなければならないが、その映画を見たことのない記述者は、容易に取得できる、タイトル、キャスト、あらまし、などしか記述できない。検索者がタイトルやキャストを思い出せない場合、記憶に残っているシーン内容で検索することになる。現在は、記憶に残っているシーンがどこであるか分からないため、映画の全シーンの内容を記述するしかない。しかしながら、そのためには映画全体の実時間視聴が必要となるだけでなく、記述コストが膨大となり現実的でない。そこで、その工程を高速化するための方式が必要である。

本課題の解決策は、視聴者の記憶に残っているシーンを、映像ストリーム全体の中から見つけ出すことである。ここでは、映画の映像ストリームにおいて、視聴者が特に記憶に残りやすいシーンを感情シーンと定義し、その部分だけを視聴しメタデータを詳しく記述することで、インデキシング者の記述作業の効率化を可能とした。

提案手法では、映画を構成するさまざまな要素の中より、感情を強く受ける視覚的な要素が、ショットサイズの大きさであることを発見し、感情シーンの判別式を構築した。感情シーンの判別式は、解析実験のサンプルに用いた5つのアカデミー賞受賞映画を用い、それぞれのシーン単位で、ショットサイズや顔の向きなど、4つの量的データを取得し、その平均と分散を分析して得ることができた。また、評価実験により、感情シーンの判定式が有効であること、感情シーンが視聴者の記憶に残りやすいシーンであることを、主観実験及び客観実験により証明された。また、総合実験としての評価に至っては、メタデータの記述コストを約1/10にしながらも、約80%の映画検索ヒット率を得た。

方法論2のアプリケーションとして、課題2では、“風景などの素材映像に対し、ゆらぎが生じずかつ高速付与可能な構図メタデータ記述手法”と題し、素材映像検索に効果的なメタデータを構築した。素材映像の検索は、多くの番組制作者を対象としているため、専門家ならでわのリテラシーを用いたメタデータ付与が必要となる。番組制作の専門家が使う素材であるが故、素材映像の内容は、番組内容に合う被写体（画）の構図が重要な要素である。しかしながら、現在のメタデータには、画の構図に関するものが必要であるにもかかわらず存在していない。そこで、構図メタデータ付与が可能な方式の構築に取り組んだ。番組制作の専門家であれば、必要とする画の構図は頭の中にイメージされている。そのイメージされた構図をいかに表現するかが重要となる。ここでは、画の見え方の共通性に着目し、専門家の経験値より共通的に認識される画の特徴として、被写体までの距離や、カメラの高さ及び仰角といった3つの特徴の組合せを用いた、揺らぎの少ない構図メタデータ記述方式を構築した。

構図メタデータ記述方式の記述内容は、立場の違い（クエリー創出側とインデキシング側）に影響されない、揺らぎの少ない特徴表現である。さらに、構築した構図メタデータ記述方式は、共通認識、独立性、検索適合率の3つの実験の結果、それぞれの実験において本手法の優位性を確認し、有効性が検証されている。本メタデータの記述負荷は、単純キーワード（例えば、“建物”や“富士山”など）の記述負荷の約1/10である。キーワード検索に本手法を組み合わせることにより、単純メタデータ検索と比較して、検索効率（絞込み率）を約10倍向上させることができる。

方法論2のアプリケーションとして、課題3は、課題2同様に素材映像の検索に有効なメタデータ記述手法として“大量の素材映像に対し、検索者の利便性と検索の演算量削減を目的とした上位概念メタデータ記述手法”と題し取り組んだ。素材映像を活用するユーザーは、検索者とメタデータ記述者が同局の場合だけではなく、局外の制作者も使用する。局外の場合、検索者は素材に映っている被写体の固有名詞を知らないことが多い。そのため、メタデータに被写体の固有名詞しか存在しない場合、検索にヒットしない。しかしながら、これらを改善するために上位概念などの名詞句を膨大に記述すると、検索対象とする名詞句が増えてしまい、データベースの検索負荷が増加する。そこで、上位概念のメタデータ付与において、検索時にはその演算（検索）時間を著しく短縮できる方式が必要である。単純に、固有名詞と上位概念を記述することは簡単である。しかしながら、汎用性を考えると、一般性のあるメタデータスキーマを用い、その特性を生かすべきである。そこで、本課題において構築すべきメタデータへの条件は、固有名詞とその複数の上位概念が階層記述できること、メタデータのデータ量は最小であること、一般性のあるメタデータスキーマを用いその体系を壊さないこととなる。

本手法では、木構造シソーラス辞書の日本語語彙大系と、マルチメディアコンテンツのメタデータ標準であるMPEG-7を採用し、その記述体系を構築した。候補となるタグとして、Part.5 MDS (multimedia Description Schemes)より、それぞれのスキーマに則り上位概念（非固有名詞）が記述でき、データベース負荷が最小かつメタデータのデータ量も最小な記述体系が<Semantic>の”ObjectType”であることを突き止めた。構築した記述体系は、1つのキーワードを付与することで、そのn倍の上位概念を自動記述できるだけでなく、並列記述されたキーワードに比べ、データベースの検索負荷を約1/nにする。（n:名詞句概念の階層数）

番組映像の中でも、映画に関する研究は、それがあらゆる自然画を用いていることや、決められた画面構造が少ないなど、構造化しにくいコンテンツである。そのため、関連する研究は極わずかである。しかしながら、日本における興行収入は史上最高を記録し続け、Webや販売店によるパッケージ販売やレンタルビデオなどのビデオソフト市場は、2004年に5,543億円（前年比105.1%）を記録し、今後も増加する傾向にある。一方、無数の映画が存在する流通市場では、巨額の資金を使い制作された映画にもかかわらず、膨大な映画コンテンツの中で、日の目を見ないまま埋もれてしまう現状もある。ゆえに、本編を見る

過程として、視聴者に閲覧・購買意欲を促す予告編や要約映像の生成は今後重要になる。本手法で判定される感情シーンは、その映画の記憶に残るシーンとなるため利用価値が高い。映画を対象とした研究は、いま行うべき重要な分野である。また、素材映像は、約1時間の番組を制作しようとした場合、その約100倍以上必要といわれている。現在、世界中で数百チャンネルの番組が放送されていることを考えると、素材映像は爆発的に増加している。現状では、これらの映像は他の番組制作に用いるなどの二次利用はほとんどされていない。増え続ける素材を有効利用するためにも、素材映像検索に関する研究は、今後ますます重要と考える。