

平成 22 年度 国立情報学研究所 市民講座 第 7 回
「マルチメディアと検索技術—キーボードを使わずに検索するには?—」
講師： 片山 紀生
(国立情報学研究所 コンテンツ科学研究系准教授)

◆ 講 義 ◆

本日は市民講座にお越しいただきましてありがとうございます。私はご紹介にあずかりました国立情報学研究所コンテンツ科学研究系准教授の片山です。

1 時間ほどお時間をいただき、研究動向のご紹介をさせていただこうと思っています。

・スライド 1 「マルチメディアと検索技術」

本日の副題に「キーボードを使わずに検索するには?」とありますが、これは私が付けたものではなく、この市民講座を企画しているグループから私に提案があったものです。

・スライド 2 「キーボードを使わないとは?」

これをご覧になって皆さんいろんなことを想像されたと思いますが、キーボードを使わずに検索するとは、どういうことなのでしょう。

私が最初にこの副題を見て想像したのは、脳波入力です。

最近では脳波を計測する技術が進んでおりますので、脳波を検知することによって、手を使わずに文字を入力するようなソフトウェアキーボードの仕組みが研究・製品化されています。

最初はそれかなと想像したのですが、キーボードを使わないものは、他にもいろいろあります。

最近では、Apple 社の iPad のように、指で触れて文字を打つなどしてコンピュータを操作することもできます。

それもキーボードを使わないといえ、使わないわけです。

他にも、音声で入力するということも想像されます。

しかし、実は今回のテーマは、こういう入力装置をどうするかという話ではありません。

・スライド 3 「マルチメディア情報と検索技術」

では、どういう話なのでしょう。

キーボードを使わずに検索するマルチメディア情報検索技術とは何かをまずご紹介したいと思います。

マルチメディアという言葉は新聞などにも掲載されている身近な言葉ですが、確認のために、言葉の意味を押さえておきたいと思います。

マルチメディアの「マルチ」は複数、「メディア」は情報媒体を表しています。

ですから、マルチメディア情報とは、複数の媒体から構成されている情報という意味になります。

一番分かりやすいのはビデオやテレビ放送で、これらは映像もあり、静止画もあり、音声も、テキストの情報もあります。

このように、いろいろな媒体を使って情報が伝えられるものがマルチメディア情報と呼ばれるものです。

このようなマルチメディア情報に、キーワードなどの言語情報が付与されていれば、インターネットの

Web ページなどと同様に、そのキーワードで検索することが十分に可能です。

すなわち、皆さんが Google や Yahoo!などの検索エンジンを使って Web の情報を検索するのと同じように、何か言葉を入れればそれに該当する画像や映像が出てきます。

そういった検索は、言語情報が付与されていれば可能なわけです。

実際、Google 社の画像検索では、キーワードを入力すれば Web 上の画像が検索結果として出てきます。また、静止画や映像の投稿サイトとして Flickr や YouTube などがありますが、これらもキーワードが付与されていれば、そのキーワードによって映像や画像を検索することが可能です。

しかし、こういう方法だと、結局、検索結果として返ってくるものがマルチメディア情報というだけで、実際にそこで行われているのはテキスト検索そのものです。

それに対して今日のテーマは、テキストだけでは扱えない検索、マルチメディア情報そのものを検索するという手法についてのご紹介です。

・スライド4「マルチメディア情報の内容による検索」

具体的に考えてみたいと思います。

例えば、テレビ放送ではいろいろな番組が放送されていますが、その番組を検索するという状況を想像してみてください。

いろいろな方法が考えられますが、例えば番組名で検索することも可能です。

知っている番組があれば、番組名そのものを入力してもいいですし、スポーツの番組を検索したければ、タイトルに「スポーツ」と入っている番組を検索することもできます。

出演者の名前で検索すれば、お気に入りのタレントの出るテレビ放送を検索することも可能です。

最近ですと各番組には解説文が付与されていますので、それを使って検索することもできます。

これらはいずれもテキストによる検索になります。

一方、例えば野球場の映像が映っているところを見つけたい、記者会見中の映像を見つけたい、特定の人や物が映っている映像を見つけたいという場合、テキストだけでは検索できません。

映像の中に実際に何が映っているかを計算機によって解析して見つける処理が必要になります。

こういった映像や画像の内容を解析して検索するというのが、本日のテーマです。

ここまでが前置きです。

・スライド5「あらまし」

本日の市民講座では、全体としてこのような流れでご説明したいと考えています。

第1に、今回の講座の背景として、マルチメディア情報が世の中に大量にあふれる時代になっていることをご説明したいと思います。

第2に、概論として、マルチメディア情報検索の概要について、マルチメディア情報の検索はどういう技術なのかという全体像を説明したいと思います。

第3に、主題として、そのマルチメディア情報の検索の中でも、テキスト検索ではなく、内容解析によって類似しているものを見つけるという、特徴量による類似検索という手法についての説明をします。

第4に、そのような類似検索の応用例として、テレビ放送からの類似ショットの検出という、応用事例についての説明をしたいと考えています。

・スライド6「マルチメディア情報があふれる時代」

最初に、本日の講義の背景である「マルチメディア情報があふれる時代」についてのご紹介です。

21世紀に入って、記憶装置、デジタル情報を蓄積するためのハードディスクドライブなどの記憶装置のコストが非常に下がり、大量のデジタル映像を蓄積できるようになりました。

その結果、現在では、記憶装置の容量を表す単位として、従来使われていたギガバイトの1,000,000倍、テラバイトの1,000倍であるペタバイトが使われるようになっていきます。

今、家庭用のハードディスクデバイスの容量が1テラバイトくらいですが、その1000倍規模の記憶装置を用意して、そこに何千時間、あるいは何万時間の大量のデジタル映像や静止画などを蓄積することが行われるようになってきたのです。

その結果、先ほどご紹介したように、Googleの画像検索ではWebページ上の画像を検索できるようになっていますし、写真投稿や動画投稿のサイトでは、一般の方々から投稿された大量の写真や動画が蓄積されるようになりました。

また、放送業界でも、皆さま名前を耳にされたことがあると思いますが、「NHKアーカイブス」という大規模アーカイブを蓄積するようになっていきます。

そのように、大量の映像をデジタル形式で蓄積して、それをオンデマンドでいつでも好きなときに呼び出して見ることができる時代になっているのです。

一方、商業ベースの試みと同時に、学術研究目的でテレビ放送の映像アーカイブを作る研究も、世界的にいろいろな形で行われています。

例えば、アメリカのカーネギーメロン大学がいち早くCNNの映像を使ってInformediaという研究プロジェクトを立ち上げ、テレビ放送のデジタルアーカイブの活用研究を行っています。

また、フランスで行われているInamediaプロジェクトは国策に近いものですが、フランスのテレビ放送をデジタル化して蓄積し、研究目的で活用しようという試みです。

日本でも国立情報学研究所が研究用の放送映像アーカイブとしてNII TV-RECSを作っています。

このような形で、計算機を使ったデジタル映像の活用技術の研究が盛んに行われているのです。

・スライド7「マルチメディア情報検索は情報学のフロンティアのひとつ」

そのようなマルチメディア情報についての検索技術は情報学のフロンティアのひとつとなっています。

では、どういう点がフロンティアなのでしょう。

今お話ししたように、何千時間分、何万時間分の大量のデジタル映像を蓄積することが可能になりましたし、それをオンデマンドで呼び出すこともできるようになりました。

ただし、大量に蓄積した内容の解析は、現在の高速プロセッサを用いても挑戦的な課題です。

蓄積だけなら、予算が多くあれば大量のハードディスクを購入でき、大量に蓄積できますが、いざ内容を解析しようとする、情報学分野での研究が求められるという状況です。

具体例を一つ紹介します。

国立情報学研究所のTV-RECS という研究用のアーカイブでは、東京地区の地上波のテレビ放送7チャンネルを昨年7月より蓄積しています。

どれぐらいの映像が蓄積されていくかといいますと、1日分だけでも168時間、1年を超えると累計で8万時間以上になります。

そのため、大容量のHDDを用意し、200TBのディスク装置を確保してそこに蓄積しています。

また、解析処理を行おうとすると、高速なプロセッサを用いても非常に時間がかかりますので、高密度サーバとして400コアの計算能力のものを使っています。

どのぐらいの処理性能があるかというと、家庭用の一般的PCではデュアルコア（2コア）なので、その数百台分の処理能力があるということです。

それだけの計算能力を使っても、簡単な処理を行わせるだけでも、1日分を処理するのに1時間、1か月分を処理しようとするとう丸1日処理し続けないと結果が出ません。

こういう状況なので、より複雑な、より詳細な解析を行わせようとするとう、何日も、あるいは何週間もかけて処理しないと結果を出せない状況です。

すなわち、蓄積は容易になりましたが、処理させるには非常に大きな計算能力が必要な状況なのです。

以上が現在の大量入手が可能になったマルチメディア情報の背景についての紹介でした。

・スライド8「あらし」

大量に蓄積されたデジタル情報があつたとき、それを活用するには、「見つける」という作業をして、自分に必要な情報に絞り込んで発見するという検索処理が必要になります。

では、検索をどうすればいいか。これが2番目の問題となります。

・スライド9「マルチメディア情報の検索」

このスライドには、マルチメディア情報の検索という応用において、具体的にどういう状況が考えられるかが列挙してあります。

マルチメディア情報には、映像や画像、音声など、いろいろなメディア情報が含まれます。

まず、画像の検索の具体例としては、報道写真の検索、博物館などが収集している写真や絵画の検索があります。

医療画像の検索も重要な分野で、レントゲン写真やMRI写真など、医療の現場で生み出される画像を効率的に検索することが重要な研究分野となっています。

身近なところではWeb画像の検索がありますし、航空写真や衛星写真の検索なども重要な応用分野の一つです。

一方、映像検索に関しては、テレビ放送の映像や一般家庭のホームビデオの検索があります。

また、防犯や犯罪捜査の視点では、監視カメラ映像も検索対象として重要な応用分野の一つです。

音声検索の場合、人が話している音そのものはテキストではありませんが、一度それを音声認識すれば、あるいは速記によって文字に起こせば、話している言葉はテキストとほとんど同じになります。

従って、音声言語の検索に関しては、テキスト検索とほとんど同じ性質を持っています。
一方、音楽はハミングや音色によって検索したいという要求もあり、重要な応用分野の一つとなります。

画像や映像、音声の検索手法には大きく分けて二つの方法があります。
一つがメタデータによる検索で、もう一つが内容解析による検索です。
それぞれについて、次のスライドで具体的に紹介します。

・スライド10「マルチメディア情報の検索」

まず、メタデータによる検索です。メタデータとは、「データのためのデータ」という意味です。
デジタル化されたマルチメディア情報として映像や音声があったとき、それを説明するために付与するデータがメタデータです。
属性情報という言い方をしてもいいかもしれません。
具体的には、映像や画像に付与されたキーワードや注釈のテキスト、ほかに、分類番号などの分類情報があれば、それもメタデータになります。
また、いつ撮影されたかという時刻に関する情報も重要な属性情報になります。
具体的に報道写真などであれば、その写真に何が写っているか、どういう内容の写真なのかといった具体的な説明文が得られる場合もあるので、そういった情報を使った検索もできます。
すなわち、映像や画像そのものの情報というよりは、それに付与された付随した情報を使って検索するというのがメタデータによる検索になります。
これが一つ目の検索方法です。

もう一つが、内容解析による検索で、英語では Content-Based Retrieval という名前で呼ばれています。
この内容解析による解析にも二つのアプローチがあります。
一つが特徴量による類似検索で、もう一つが内容解析による属性情報の解析です。
二つ目の内容解析による属性情報の解析は、メタデータによる検索に近いものです。
先にこちらを説明しますと、文字認識によって映像や写真の中に表れている文字を認識するものです。
例えば、どういう文字が表れているかを映像や写真の中から見つけたとします。
そうすると、その文字情報が人手で付けたメタデータと同じような役割を果たすわけです。
逆に音声認識を使いますと、今、私が話していることも機械的にテキスト形式にある程度の精度で変換することができます。
この場合、もともと音というメディアだったものがテキスト形式に変換されるわけです。
これと同様、計算機による内容解析によってテキスト形式の情報に変換し、それをメタデータと同様に検索するというのが、こちらのアプローチになります。
ですから、もともとは映像や画像の生のデータだったものをテキストや言語情報に変換して、それを使って検索するという、かなりメタデータによる検索に近い位置付けになります。

一方、特徴量による類似検索は、言語情報を使うアプローチではなく、映像や画像そのものの本質的な特徴を抽出して、その特徴同士が似ているかどうかを比較することで、似ているものを探そうというも

のです。

この方法が今回の講座の一番のテーマで、後の説明で具体的な方法についてご紹介していきます。

特徴量としてどんなものが使われているかをここにリストアップしていますが、色、形状など、写っているものの特徴を見つけていきます。

例えば音楽なら、高い音か低い音か、音色といった周波数特性を使って検索が行われます。

・スライド11「メタデータ vs 内容解析」

メタデータによる検索と内容解析による検索の二つのアプローチを比較すると、このようなリストを作ることができます。

かなり好対照のアプローチになります。

メタデータは基本的にテキストベース、言語情報を使った検索になりますので、Web のサーチエンジンの検索など、テキスト検索と同様の手法を使うことができます。

ですから、内容解析に比べると比較的处理コストが小さいというメリットがありますし、検索精度も高く、文字の照合や単語の照合で検索結果を得られるので、実用性の高い検索手法になります。

従って、その映像・画像にキーワードなどを付与できるのであれば、それを使うことで基本的な検索はできます。

一方、内容解析の方は、処理コストが大きいという問題点があります。

先ほど、大量に蓄積されているテレビ放送を処理しようとするとなん日かかかるという話をしました。

映像や画像の内容を細かく分析するのは、計算機がどれだけ速くなっても、コストとしては非常に大きくなるという不利な点があります。

それでも内容解析をすることによって、メタデータの付与だけでは見つけられないようなもの、映像や画像の本質的な特徴にかかわるような内容を生かした検索が実現できるのが一つのメリットです。

また、メタデータをすべて人手で付けるとなると、非常に時間もかかりますし、人的コストが非常に大きくなります。

特に大量に画像や映像を蓄積できる時代になっていますので、これらすべてを人手で付けるのは効率的に限界があります。

そこで、計算機にそれに代わる処理をさせたいのですが、メタデータの自動付与をさせるという目的のためには、内容解析の技術が必要になります。

従って、即戦力になるのはメタデータによる解析ですが、それを補う形で内容解析による手法もこれから計算機精度が上がるとともに応用範囲が広がることが期待されています。

ここまでは漠然とした話だったので、内容をつかみにくかったかもしれませんが、マルチメディア情報を検索するというテーマについて、全体像を見渡すような形で概要をご紹介させていただきました。

・スライド12「あらまし」

続いては、今回のメインテーマである、特徴量による類似検索がどのようなもので、どのような方法を使って行うかということについて説明します。

・スライド13「特徴量による類似検索」

最初から文字だけのスライドで恐縮ですが、特徴量による類似検索の流れを簡単に図で表すと、このようになります。

原理としては非常に簡単です。

まず、映像や音声をデジタル化します。

もちろんアナログでは計算機は処理できないので、アナログからデジタルに変換して、デジタルデータとして記憶します。

そして、デジタル化された映像や音声から、形状や色など、内容を表すような特徴量の抽出を行います。特徴量を抽出した結果、特徴量同士を照合することで、ある映像や画像とある映像や画像が似ているか、似ていないのかを特徴量同士の比較で照合します。

これが特徴量による類似検索です。

・スライド14「色特徴の例:カラーヒストグラム」

具体的な例をご紹介します。

これからご説明する話は、特徴量の一例としてしばしば使われるカラーヒストグラムという、割と単純な特徴量を使った例です。

これは、ある画像に黒や白や赤などいろいろな色が出てくるわけですが、それぞれの色がどれくらいの割合で含まれているかという出現割合を特徴量として数値化したものです。

例えば、スペースシャトルが着陸する画像を、まず右上、左上、右下、左下の四つの領域に分割します。こうして分割した中で、左上の4分の1の領域に注目してみてください。

この中に黒がどれくらい含まれているか、ねずみ色、白、赤がどれくらいかを計算機で数え、その結果を数値化し、特徴量として保存します。

このグラフは正確に元の絵を計算したものではなく、あくまでもイメージ、挿絵としてご覧いただきたいと思います。

本当に青がこれだけあるわけではありません。

このようにして、元の静止画にどれくらいの色が含まれているかという割合を数えると、四つの領域それぞれの色の分布を、数値として数え上げることができます。

ここでは黒、ねずみ色、白、赤、黄色、緑、水色、青、ピンクの9色をそれぞれ取り出しているので、全部で $9 \times 4 = 36$ の数値が得られます。

この色の分布を特徴量として使うのです。

ここで色が占有する割合を取ったグラフがカラーヒストグラムというもので、計算機で比較的簡単に求めることができるので、高速に使える特徴量の一つとしてよく使われます。

・スライド15「画像・映像から特徴量を抽出」

こういう形でそれぞれの写真について、どういう色の分布であるかを計算しておきます。

この図では元になる映像としてオープンリールのイラストを使っていますから映像っぽいですが、映像でも写真でも構いません。このようにして元の映像や写真のフレーム画像などから、カラーヒストグラ

ムを計算しておきます。

すると、それぞれの画像や映像に対応した特徴量がベクトルとして一つの空間の中に生成されます。

先ほど、カラーヒストグラムは36の数値で表されると言いましたが、そうした複数の数値で表されるものは数学上ではベクトルになります。

もともと映像や画像だったものが、特徴量の抽出によって複数のベクトルで表現されるわけです。

そのベクトル同士が近いかどうかです。

簡単に言えば、ベクトルの先に丸印が付いていますが、その印が近いかどうかで「似ている・似ていない」を判定します。

ある程度遠いと全然似ていない、近ければ似ているということです。

これらの特徴量によって見つけるというのが、特徴量に基づく類似検索です。

・スライド16「特徴量の照合による類似検索」

このようにもともと蓄積している画像や映像から特徴量を抽出しておくのですが、それを検索するときはどうするのでしょうか。

特徴量を検索するとき、検索したい人、すなわちユーザーは、キーボードをたたいて手で特徴量を入力するわけにはいきません。

例えば赤が10%、青が30%と、そんなことをいちいち入力していたのではとても使い物になりません。類似検索のときの入力方法として代表的なものが、この四つです。

一番簡単なのは、見本の映像をリスト表示して、探したいのはどれですか、似ているものを選んでくださいというものです。

例えばスペースシャトルの着陸シーンが検索したいのだったら、ちょうどありますから選べばいいですし、ぴったりのものがなくても、割と近いものがあればそれを選ぶというように、見本となる画像から選ぶというメニュー形式の検索です。

これはしばしば行われます。

また、見本からではなく、もっと自由に選びたい場合は、スケッチする方法もあります。

簡単な例では、赤い丸と緑の三角を描いて、山の景色を選ばせるというものです。

そのようにスケッチをするという選択もあります。

また、お仕着せの見本では嫌だ、自分で入力したいという場合は、自分で撮影したものを検索したい対象として計算機に入力することもできます。

音であれば、実際に自分でハミングするなり、歌うなりして入力して、自分が録音した音と似たものをデータベースから探すことも行われます。

このように、特徴量の照合によって類似検索するときには、キーボードは使わずに、見本からの選択や撮影や録音したものをキーワードの代わりに与えることになります。

そうすると計算機で特徴量を計算し、特徴量の比較で検索結果を返してくるわけです。

本日の副題の「キーボードを使わずに検索するには」というのは、この部分を指しています。

例えば、ある検索キーとして、このベクトルが選ばれたとします。

これに一番近いのはどれだと計算機が計算し、近いものを「これだよ」と検索結果として返してきます。

これが、特徴量の照合による類似検索です。

原理はこういうものですが、実際にどんな結果が返ってくるのでしょうか。

・スライド17「検索例」

これはNASA が公開している映像や写真を対象にした検索システムの例です。

最初に、どういう映像がデータベースにあるか一覧表示しておきます。

その中から、検索するユーザーは、自分が検索したいものに近いのを選びます。

この例では、一番左上にあるスペースシャトルが着陸しているシーンを選んだとします。

そうすると、カラーヒストグラムを計算して、カラーヒストグラム同士が近いものを計算機は探します。

・スライド18「検索例」

すると、色の分布が似通ったものが検索結果として表れます。

・スライド19「検索例」

今度は右下の違う例で説明します。

スペースシャトルが打ち上げられるときの写真を検索対象として選んだとします。

・スライド20「検索例」

この画像のカラーヒストグラムの色の分布が似通ったものだけを検索結果として返してきます。

すなわち、特徴量によって似ているかどうかを判定し、似ているものを返してくるわけです。

このように、特徴量を介して検索が行われるわけです。

これが特徴量による類似検索という手法です。

今の具体例をご覧くださいと、問題なく動いているように見えると思います。

実は既にある程度の精度では動いているのです。

しかし、この手法は情報学の分野ではまだ発展途上にあり、研究開発が進められている真っ最中で、技術が日進月歩で進んでいる状況にあります。

・スライド21「類似検索における研究課題」

では、研究者はこの手法においてどういうところを頑張っているのか、何を解決するために苦労しているかを、このスライドでご紹介したいと思います。

この流れは、先ほどのスライドでお見せした類似検索の基本的な流れになります。

研究者が頑張っている一つは、どういう特徴量を使うかで、これが大きなテーマになっています。

もし検索したい対象が完全に同一な映像や写真なら、数字と数字を比較したり、文字と文字を比較したりするのと同じで、簡単です。

アルファベットの A と A は同じです。100 と 100 は同じです。

しかし、類似検索の場合には完全に同一なものではなくて、似ているのだけれども違うものを見つけることが必要になります。

世の中、たくさんの画像や写真を集めたり蓄積したりできるようになっていますが、全く同一なものは、そう滅多にはありません。

元の写真をいろいろな形で複製したのなら別ですが、先ほどのスペースシャトルの画像のように、いろいろな写真が撮られていて、全く同一ではないけれども状況としては同じという写真・映像があふれています。

そういう状況で、このスライドにあるような「軽微な差異」、多少、回転していたり、縮尺が違ったり、歪みがあったりしても、「漏れや誤り」、すなわち検索漏れや誤って返ってくることが少ない特徴量を作るにはどうしたらいいか。

画像や映像のどういう特徴を使い、それをどう数値化すればいいのか。

そこが非常に大きな研究課題となっています。

もう一つの研究課題は非常に単純な話で、スピードです。

映像や音声を現在の高速なプロセッサを使って処理しようとしても、大量な計算処理能力が必要です。ですので、非常に高速、あるいは効率的、コストが少ないアルゴリズム、計算機の処理の手順を考えていく必要があります。

それを考えるのが、もう一つの情報学分野での大きな研究テーマです。

すなわち、誤りを少なくし、効率的・高速に処理することが、特徴量による類似検索において、情報学分野で研究されている課題になります。

・スライド 22 「特徴量の例」

一つ目の研究課題である「どのような特徴量を使えばいいか」を考えるために、カラーヒストグラム以外に具体的にどのような特徴量があるのかを、このスライドでご紹介したいと思います。

この二つの画像は、テレビのニュース番組から取ってきたものです。

これが同一か、似ているかどうかを比較したいと思います。

人間は簡単です。鳩山前総理が映っている、同じ人物だとすぐに分かります。

しかし、この二つは全く同一な映像ではありません。

というのは、実は放送局が違うのです。

ですので、映像の中に加えられている文字に違いがあるというのが 1 点です。

もう 1 点、よく見ていただけると分かりますが、背景の位置が微妙にずれています。

恐らくカメラの向きが若干違うわけです。

このようなわずかな差があるときに、それには影響されずに、これは鳩山前総理が映っているのだから

同じである、少なくとも似ている、全く別物ではないということは、人間には簡単に分かるのですが、それをコンピュータに判断させるのは実は非常に難しいのです。

コンピュータは人工知能であると言うと格好よく聞こえますが、原理的には単なる計算機で、計算することしかできない機械です。

数字の処理は高速にできるのですが、人間の視覚のように、似通っているということを具体的に処理することは実は難しいのです。

そこで、特徴量の設計が非常に重要になります。

下にずらずらと画像が出てきていますが、それぞれ横の行が特徴量の種類になります。

一番上の行に表れている画像は輝度の特徴を表しており、画像の明るさだけを取り出したものです。

一方、2段目と3段目は「線成分」と書いておきましたが、2段目は水平方向に直線っぽいものを、計算機による画像処理で取り出したものです。

3段目は縦方向の線成分を取り出したもので、このような画像になります。

それぞれ元の画像に対して、明るさだけを取る、水平方向の線だけを取る、垂直方向の線だけを取るといった画像の処理を行うと、このように加工された画像が得られるのです。

これである程度特徴量を抽出したことにはなるのですが、今度はこれら同士を比較するわけです。

比較するときに、同じ行の二つの画像をいきなり比較することはしません。

なぜかという、ぴったり一致するとは限らないからです。

先ほどお話ししたように、カメラの向きが微妙に違っていたりしますので、多少のズレがあります。

そういったズレを許容するために、DCT という手法を使って、スライドにあるような画像を作るのです。

何かこう、ぼやけていますね。

ぼやけていて全然駄目ではないかと思われるかもしれませんが、これはわざとぼかしているのです。

どういうことかという、高周波成分という細かい変動を捨てているわけです。

細かい変動を捨てて、主要な、より大きな変化だけを残すと、こういう画像になります。

こうした画像同士を比較すると、計算機でも「似ているな」と分かります。

突き合わせてみると、確かに差が少ないことが計算機にも分かります。

従って、細かい違いは捨てて、いかにして画像や映像が持っている本質的に重要な特徴に絞って比較できるかが、特徴量による類似検索の検索精度を高めるための基本的な原理・原則になります。

今回の例では輝度成分や特徴量を使っていますが、いろいろな特徴量が提案されていて、それらを組み合わせることも行われており、どうすれば誤りや漏れの少ない検索が実現できるかが日々研究されています。

以上が、特徴量による類似検索の概略の説明です。

・スライド23「あらまし」

残りの10分弱の時間を使って、発展的な応用事例として、テレビ放送からの類似ショットの検出につい

て説明します。

・スライド24「テレビ放送からの類似ショットの検出」

現在、テレビで放送された番組を大量に蓄積することがいろいろな形で行われています。

NII の放送アーカイブもそうですが、各国でテレビ放送を使った学術研究用の実験が行われています。市販の製品でも、ハードディスクレコーダーの大容量化により、複数チャンネルを同時に録画する製品が出ています。

テレビ放送を大量にデジタル記録することが一般的になっているのです。

例えばニュース番組でも、現在の技術を使えば、1年分のニュース番組を蓄積しておくことができます。しかし、1年間いろいろな形で報道された映像がコンピュータの HDD の中に入っているとして、それを活用したいという場合、1年分あったとすると、それを見るのに1年かかったりします。全部見ることはできないので、何とかして絞り込まないと活用し切れません。

例えば、ここにニュース番組から抜き出したビデオフレームが並んでいます。

ニュース番組では、もちろんニュース内容に直結した映像もありますが、ニュースキャスターや現場のレポーターの映像も使われています。

天気予報ももちろんあるでしょう。

また、ニュースの内容と直結しているとは言えませんが、臨場感、雰囲気伝えるための挿絵のような役割をしている映像もあります。

例えば新型インフルエンザの話だと、マスクをした子どもたちが映っています。

これは、この子どもたちがどうこうではなく、雰囲気伝えるための映像です。

ニュース映像の中には、重要な映像とそうではない映像があるわけです。

1年分もためると、全部見るのは大変なので、重要なものだけに絞る必要性が出てきます。

では、どうするかということです。

現在の計算機の性能があれば、1チャンネルだけでなく、例えば東京地区だと教育テレビを除いて六つのチャンネルがあり、同じようにニュースが流れて、ほとんど同じ映像が使われています。

これを人手で見つけようとする大変ですが、特徴量を用いた類似性の照合技術を使うと、コンピュータにより自動で抽出することができます。

そうすると、1年分の映像があっても、各チャンネルにどれだけ共通した映像があるかが機械的に処理できるのです。

例えば「お父さんのためのワイドショー講座」など、テレビ番組の中で同じようなことをやっているではないかと思うかもしれませんが、あれは基本的には内容単位、話題単位です。

それぞれのニュース、話題ごとに時間を集計して、この話題が何分何秒放送されたと番組内で伝えているわけです。

しかし私が言っているのは、話題単位ではなく、放送に表れた映像そのものが同じかどうかを比較しようという話なのです。

これを人間がしようとすると大変です。

どんな映像が表れたか覚えているのは大変なので、そこはコンピュータが機械的にする方が効率的です。ここでは、放送局をまたがって表れている同一のショットを見つけることを計算機でさせるわけです。ショットとは、一つの映像を一つのカメラで撮影した範囲のことです。

しかし、民放の場合は、CM も放送局をまたいで放送されているので、ニュース映像だけではなく、CM も同じものが放送されているとして検出されます。

そうだとすると、CMは15秒単位で流されるという規則があるので、そういう情報を使うと、CMは除去してニュース映像だけ取り出すことができます。

・スライド25「出現チャンネル数での重み付けによるニュースショットのカレンダー表示」

こういうことを行くと、あるニュース映像が幾つのチャンネルで放送されたかが機械的に計算できます。その情報を生かすと、こういう表示が可能となります。

これは、一番上の行は9月1日になっていますが、9月2日、9月3日と、日ごとに、四つ、五つ、六つのチャンネルに共通して表れたショットだけ選んで表示しているのです。

六つのチャンネルで表れたものが一番大きく、五つのチャンネルで表れたものが少し小さく、そして四つのチャンネルで表れたものが一番小さく表示されています。

こういう表示をすると、非常に絞り込まれた映像だけでそれぞれの日の特徴を表すことができます。

ここでは四つ以上のチャンネルに表れたものだけ表示していますが、この裏には三つだけ、二つだけ、あるいは特定の放送局でしか表れなかったものがあるわけです。

そういう中で、厳選された映像だけを絞り込むことができるということです。

このようなカレンダー表示ができると、例えば1年間という長期間ニュース映像を蓄積したときに、1年の間にどんな映像が放送局によって流されていたのかを把握するのに効果があります。

こういう形で類似検索の仕組みを使うことで、効率的な閲覧表示が可能になるということです。

・スライド26「まとめ」

そろそろ終わりの時間に近づいてきましたので、まとめに入ります。

本日の市民講座では、背景、概論、主題、応用と四つの段階に分けて説明しました。

この分野の今後の期待される展開について最後に説明して終わりたいと思います。

今後、研究者がこのテーマについて研究を進めていくことで、内容解析法についての精度が上がっていくことがひとつ期待できます。

また、いろいろな映像に適用できるように汎用性が広がり、実用性が高まることを目指して、今後研究が進められていくと思われます。

この分野の今後の発展にぜひご期待いただければと思います。

・スライド27「関連書籍」

最後にもう1枚スライドがあります。

関連書籍をご紹介します。

国立情報学研究所の方でシリーズ化している情報研シリーズの12番目に、『石頭なコンピュータの眼を鍛える—コーパスで人間の視覚にどこまで迫れるか—』という本があります。

この本は、国立情報学研究所の佐藤真一教授によって書かれたもので、コンピュータによる映像理解技術の最近10年の研究動向を紹介する内容になっています。

本日の話の内容に関心を持たれた方であれば、恐らくこの本も面白く読んでいただけるのではないかと思います。

私からのご説明はこれで終わらせていただきます。

ご清聴ありがとうございました。