

A V I R G

会 報

Vol.34 No.4 (2000.12)

発行：視聴覚情報研究会(AVIRG)

代表幹事：伊藤 崇之

〒157-8510 世田谷区砧 1-10-11

日本放送協会放送技術研究所

TEL 03-5494-2361

FAX 03-5494-2371

. 11 月例会報告

「超高速映像検索と国際標準化：MPEG-7」

講演：山田 昭雄 氏（NEC）

報告：川田 亮一（KDD 研究所）

《概要と感想》

本講演では、来年標準化が予定されている MPEG-7、およびこの標準化活動において講演者山田氏が大きく寄与された画像の記述方法、さらにはその応用例について、デモンストレーションを交えながら、分かりやすい説明がなされた。

山田氏は、約15年前から画像の符号化を研究されており、その後、画像解析、さらに最近では MPEG-7にも関係する画像検索などを研究テーマとされているとのことである。本講演では、そのような氏の広い守備範囲を背景としつつ、次の3トピックについて話された。

- (1) MPEG-7規格
- (2) 検索技術について
- (3) アプリケーション

まず、MPEG-7についてである。この標準化の主体は、ISO IEC/JTC1/SC29/WG11である。MPEG-7の目的は、オーディオビジュアルコンテンツの表記方法にある。ビットストリームをデコードすることなく中身を知る方法といえる。しばしば誤解されるが、これまでのMPEG規格すなわちMPEG-1、MPEG-2、MPEG-4に代わるものではなく、インターネットを例に取れば、いわばブラウザとビットストリームの掛け橋となるべきものである。

MPEG-7の重要な機能の一つとして、検索がある。これには、次の3種類がある。

- (1) 信号レベルの検索。これは、画像の輝度値や色差値などを元にマッチングを取るもの。
- (2) 意味レベルでの検索。これは、キーワードなどにより検索するもの。
- (3) イベントレベルでの検索。これは、2よりも詳しく、画像の中身まで立ち入るものである。MPEG-7の機能には、この他、サーバーの制御や、コンテンツに付加価値を与えることなどがあげられる。

MPEG-7を使用した事業展開として考えられているものは、次のようなものがある。

- (1) デジタルテレビにおけるEPG（電子番組ガイド）のユーザごとのカスタマイズ
- (2) 放送局内でのデータベースマネジメント
特に、ポストプロダクション会社など外部とのやり取りの多いこの世界では、この種の標準化の意義は非常に大きい。
- (3) 家庭内におけるホームサーバ上の検索や、マニア向けのオーサリングツール
- (4) インターネットポータルサイトとしての応用。すなわち、異なるサイト間でコンテンツの記述方法を標準化することにより、検索工

エンジンの汎用性が広がる。

MPEG-7の標準化スケジュールとしては今後、2001年3月に標準化案が最終決定され、9月から実際に使用可能となる予定である。その3月のシンガポール会合では、デモなどのイベントも予定されている。

MPEG-7は、次の7要素からなる。基本はXMLである。

- (1)System
- (2)DDL
- (3)Visual
- (4)Audio
- (5)MDS
- (6)RSW(フリーのリファレンスソフトウェア)
- (7)Conformance

本日の話の中心は、Visualである。MPEG-7 Visualでは、画像を時間空間で分割する。すなわち、まずシーンに分割した後、regionに分割し、さらにそれをsub-regionに分割する。また、画像の記述方法の種類として、次の5種類が規定されている。

- (1)Color
- (2)Texture
- (3)Shape
- (4)Motion
- (5)Location

このうちMotionの応用例としては、道路監視映像からの動ベクトル検出により、交通流量の多い画像と少ない画像の自動分類法などがある。

山田氏のグループでは、家庭用のPersonal video archive systemを最終目標にしている。ここでは、いろいろなタイプの映像が記録されるし、また、いろいろな機能が要求される。そして、コンテンツの記述が必須であることが分かっている。すなわち、MPEG-7との関連性があり、このことが、氏がMPEG-7と関わることになったきっかけのことである。

MPEG-7の標準化活動において、NECからは、合計3セット10提案がなされた。これには、オーディオビジュアルプログラムのナビゲーション方法や、類似画像の検索方法などが含まれる。そこで、次に、山田氏らが提案し、標準にも採用されたカラーレイアウト記述子について、類似

画像検索への応用の観点から、説明がなされた。

このカラーレイアウト記述子の大きな目的は、信号レベルでのビデオコンテンツ管理である。超高速検索や、ビジュアライゼーションへの応用を見越している。デジタルアーカイブの検索方法としては、ビデオマッチング、静止画マッチング、スケッチマッチングがある。いずれの場合も、計算時間と検索パフォーマンスのトレードオフが成立する。山田氏らは、このトレードオフカーブを押し上げるため、1画像あたり8バイトしか要しないカラーレイアウト記述子を提案した。これにより、高速・高パフォーマンスが達成された。例えば、24時間の映像データから、平均0.1秒で検索が可能である。具体的方法としては、映像データのY,U,Vそれぞれにつき、8x8の縮小画像を作成する。次に、それぞれをDCTし、非線型量子化を行う。最終的には、合計12係数だけを取りだし、記述子とする。

MPEG-7の標準化の場において、本記述子を用いた検索方式は、独立グループ作成によるグラウンドトゥースデータ5500枚を使用したコンペティションにおいて見事最高の性能を示し、採用が決定した。

最後に応用システムの紹介があった。まず、インターネット映像ポータルでのデモでは、サンプル画像やラフスケッチの画像から、類似している静止画や動画がほぼ問題なく検索されることが示された。実システムの実装上は、動画の連続フレームなど、よく似たシーンに対する後処理が重要となるとのことである。

次にCM自動調査システムについて、興味深いデモンストレーションが行われた。

CM調査は、実際にCMが契約どおりテレビ局で放映されているかを調べるために行われている。現在は主に人間による人海戦術で行われているが、これを自動化する試みである。方法としては、映像による方法、音声による方法、ID信号付与による方法がある。しかし音声による方法は、同一音声で別映像というCMには対応できない。また、IDによる方法は、その付与作業が煩雑なのと、故意に違うIDを付与された場合に対応できない。

そこで、山田氏らは、映像による手法を検討

している。具体的には、前述の記述子を使用し、類似度をチェックし、その結果をwebやmailで知らせるシステムである。テストシステムを構築して実験した結果、極めて良好な特性が得られている。アナログテレビ特有のDCの変動などがあるため、記述子のうちDC成分に対する重みを小さくするなどの最適化は必要であるものの、十分実用に耐えられる性能となっている。この理由としては、一般に静止画に比べ動画は検索性能がよくなることや、CMは各社それぞれ独自性を出しているため、各CMはかなり異なっており混同の恐れが少ないことが上げられる。本システムにおいては、検索時に、そのとき放映しているのがCMか本番組かの区別はしていない。これで誤検出や検出漏れが0%というのは驚異的なパフォーマンスであると感じた。CM会社の調査依頼には、十分使用可能であるといえる。

山田氏によると、本システムの適用分野としては、この他、マーケティング調査、すなわちライバル会社が今どのようなCMをどこで放映し

ているかを調べるというものがある。この場合は、まずCMのデータベースを作成しなければならず、これを自動化で行う場合はまだ20%ほど本番組をCMと誤る場合があり、今後の課題とのことである。また、時間方向で長さや再生速度を変えた場合への対応も、今後の課題とのことである。

以上のように、本講演は、今非常にホットな話題である画像検索やMPEG-7について、実際に中心的活動をしておられる山田氏から直接お話を伺うことができ、極めて有意義であった。大成功を収めつつあるDCTを応用した記述子などは、符号化のバックグラウンドのある氏だからこそできた提案であると思われる。研究においては広い分野のカバレッジが独創につながるという格言を再認識できた。

本講演でも言及されていたが、今後、画像による監視などのアプリケーションが広がると予想され、その場合にはコンテンツレベルの記述子も重要になってくると考えられる。今後のこの分野の一層の発展を願う。

「分身生成のためのマルチモーダル表情合成」

講演：森島 繁生 氏（成蹊大学工学部）

報告：近間 正樹（奈良先端大学 知能情報処理学講座）

《概要と感想》

本講演では、人物の表情のコピーを計算機の中でリアルに表現する技術の最近の成果についてデモ映像を交え講演していただいた。

近年のバーチャルアイドルなどの登場によってCGによるリアルな顔CGは身近なものになってきたが、より微妙な表情、感情までも表現できるようなCGが望まれてきている。また、研究の対象としても、従来は基本6感情の分類などがあったが、最近では表情というノンバーバルな情報を伝えるためのモデル化という点で、ホットな領域になってきている。

しかし、現在の技術では合成した表情のCGと本物の表情とでは受ける印象が異なり、明らかに合成したCGが不自然に見えるという問題がある。これに対して、ハイスピードカメラによ

る撮影を行って、本物の笑いの表情と合成した表情の違いを比べてみたところ、眉の動きや口もとの動きの差はNTSCでの検出限界以下であることが分かった。氏のグループでは、この差が人間には知覚できず不自然さを感じるのではないかと仮説をたて、更に調査・実験を重ねているとの事である。

分身コミュニケーションを実現するためには、微妙な表情なども漏れなく伝える必要があり、そのためのモデルや表出のための適切なデバイスはどうかを考えなくてはならない。しかし、すべてを忠実にコピーする必要はなく、例えば頭髪などは近似でも十分である。リアルタイム性を保ちながら、近似でも良い部分はどこまでか、どこを忠実にコピーするか見極めが

重要である。

次に、実際に顔をモデル化する手順を説明された。

まず、Cyberware 社のカラー 3 次元レーザーレンジファインダを使って得られたデータをもとに、本人のワイヤーフレームモデルを作成する。このワイヤーフレームを制御して動かす必要があるため、その整合を取るためのツールを作成した。ワイヤーフレームを作成する際には、口元など表情表出に重要な場所はメッシュを細かく割り当てるなどの工夫をしたとのことである。しかしこのレンジファインダでは 1 回のスキャンに十数秒と時間がかかり、また装置も高価であるという問題点があることから、正面や側面などの複数視点から観測した顔画像をもとに三次元モデルを構築する手法も試みていて、視点を多くする事で精度を向上させている。

表情のコントロールの方法には Paul Ekman の FACS (Facial Action Coding System) に基づいており、左の唇をあげるなど 44 個の AU (Action Unit) の組合せで制御している。しかし、この動きをマーカをつけて追跡しようとしても十分な精度で追うことができず、また AU を使うことが顔合成において正しい適切な方法かは疑問であり、今後よりよい制御単位を提案していく必要がある。現在のところは、AU を使った制御方法で第三者の顔を変形させている。

ここで、このシステムをサイバースペース上の対話システムに応用したデモの紹介があった。このデモでは音声と表情のパラメータの情報をやりとりしており、現状では 4 人がネットワークを介して同時に対話できるという。

次に、口形の制御の方法について説明された。口形状と音声は密接に関係しており、6 つの基本形状 (/a/i/u/e/o/ と 閉口) からなっている。また、基本形状の補間を行うことで、母音の口形を使って子音の口形も表現可能である。

制御点としては、唇の左右両端、上唇、下唇の上・下端、および顎の 7 つを使用しており、これらの移動量をニューラルネットワークを使って学習している。このとき入力は例えば 'あ' の音 (スペクトルパターン) なら出力は 'あ' の口の形でマッピングを行っている。音量の情報は

まだ使っていないが、将来は口の大きさの制御を行おうと考えているとのこと。

このあと、SIGGRAPH で発表したプロトタイプシステムをいくつか紹介していただいた。最初は 95 年のもので、リップシンクを使った対話システムのデモであった。会場に来た人の正面画像を取り込んで、コンピュータ上にアバターを作成し、表情制御や入力音声に基づいて口の動きを与えたアバター同士での会話を行うものである。

次は 98 年のインタラクティブ=シネマシステムで、これまでは口や表情だけしか動かなかったものを、背景も動かしてやってみようということで、映画の主人公の顔の 1 シーンを置き換えて楽しむシステムである。残念ながら著作権の問題で reject されてしまい、あまり知られていないシステムだそうだ。

99 年には HYPER-MASK という顔のトラッキングとプロジェクションを組み合わせたシステムで、人物にお面をかぶせ、そのお面に顔画像を投影するというものである。プロジェクションができることで今回はお面 (マスク) に投影したが、例えばボールに投影することで、インタラクティブなプレイグラウンドにも応用の可能性を示した。

最近の研究として、ATR と共同で研究している VIDEO-Translaton (ビデオ翻訳) の紹介があった。このシステムでは翻訳による音声の吹替えと、それに合わせた口形の合成を行っている。これによって、言語間の音声の長さの違いも吸収している。合成は口形のみで表情はオリジナルの映像をつかっているため、自然な合成顔が得られる。合成音声は ATR の CHATR を使用しているとのこと。また日本語から英語へ、英語から日本語へのどちらも適用可能である。

さらに話はよりリアルな表情認識の方法について、解剖学的モデルを用いた制御方法を導入する点に進んだ。1 台のカメラで 3 次元情報を取り出したい。そのためには解剖学的な顔面モデルを持つ必要があるということで、Keith Waters が 93 年に提案した顔の筋肉モデルを導入した。これは表情筋をバネでモデル化したものである。

森島氏はこのモデルを拡張して、皮膚表面、筋肉ノード、骨格ノードの3階層モデルを構築し、これを元にポリゴンを形成し運動方程式を解いて表情を生成している。パラメータとしては、表情筋の強度のみを指定してやるだけでよく、AUをパラメータとするものに比べても精度がよくなる。

また、骨格情報が入ることによって、より自然な表情が可能になった。しかし、このパラメータを決定するのはたいへんなので、実写画像からオブティカルフローを計算して皮膚の動きを取り出し、そこから筋肉パラメータを推定している。この方法では喜びや怒りなどのカテゴリズを行わないので、中間的な微妙な表情をうまく表現できている。しかし、欠点として、推定にはニューラルネットワークを使うので、学習データに依存してしまう。

次に、EMG(筋電位)波形を使った表情測定法の紹介をされた。EMGは筋肉に電極を刺して筋電位をはかるもので、これによって表情筋の動きを測定しようというものである。しかし、この方法では顎の動き情報を取ることはできないので、顎の動きを制御する咀嚼筋と表情筋、それらと表情の関係の解明はまだまだこれからという段階である。

また、Watersの表情筋モデルを再構築する試みも行われており、実際に解剖を何体が行っている。それによると表情筋は個人によってかなり異なっており、筋肉のつきかた、発達、退化の仕方はさまざまである。これが、表情の個人差になるのではないかと推定しておられる。

最後に髪の毛のモデル化の話を紹介した。このモデルは、髪の毛1本で7つの制御点を持ち、

Bスプライン近似を行った房モデルを仮定している。レンダリングは円筒形状にはりつけ、また髪の毛の生え方は3種類のモデルを用意した。これらの運動方程式を解き、レンダリングを行ったところ本人よりもボリュームのある表現になった。つまり、本物の髪は痛んでいているのだ。すると用途としては、CG合成で綺麗な髪を生やしたり髪型を変えたりすることが考えられる。また、モデリングによって風になびく髪を表現することもできるようになった。課題としては、流れ場の導入、レイヤーの導入、衝突判定などがある。現在は風は顔をつき抜けているため、多少不自然になってしまうとのことである。

余談として、自然な表情の撮影風景を納めたビデオが紹介された。心理学で使われている嫌悪や喜びを示す映像を部屋の中に被験者一人だけに見せて、その様子を撮影している。自然な嫌悪や喜びを表させるのは結構簡単だが、怒りの表情などはなかなかとれず苦労しているとのこと。また、表情が硬い人もいて、そういう人には、フェイシャルエクササイズをしてもらい、そうすると普段は使わない筋肉が活性化して表情が豊かになるそうである。

その他の課題として、目はテクスチャを張り付けるだけでは死んだ目になってしまうので注意が必要。目はコミュニケーションに重要な役割を果たしているようだ。また、顔の皺を考慮したモデル作りなどをあげて講演を締めくくられた。CGの合成というエンターテインメント的要素だけでなく、より深く人間を知ると言う意味でも、表情の研究は重要かつ興味深いテーマであると感じた。

. 1月例会予定

2001年1月の例会は、

日時：1月18日(木)14時~17時

場所：東京大学工学部 6号館 2F 63号講義室
で開催します。

テーマは、

『初期視覚におけるテクスチャ情報処理』

で、以下の2件の講演を予定しています。

奮ってご参加ください。

「V4野における陰影テクスチャーの

特徴抽出」

講演者：花澤 明俊 氏

(生理学研究所高次神経性調節研究部門)

物体表面のテクスチャーは、その物体を識別したり、表面の材質、摩擦などの属性を判断する上で有用な情報源である。テクスチャー特徴の抽出にマカクザルV4野が関与するか否かを調べた。コンピュータ画面上に、輝度勾配(陰影)によって粒子状の細かい凹凸があるように見えるテクスチャー刺激を提示し、要素の密度、大きさおよび陰影の方向を変化させ、神経細胞の応答を調べた。多くの細胞がこれらの刺激属性に選択性を示した。それらの一部は空間周波数選択性では説明のできない振る舞いをした。まず、要素の陰影の方向に対し1方向性の選択性を示す細胞があった。上または下方向の陰影に選択的な細胞が多く、陰影からの立体知覚にみられる偏りと対応していた。また、粒子状のテクスチャーのみに応答し、正弦波格子やパー刺激に応答しない細胞があった。これらの結果から、V4野が、陰影から復元される3次元構造を含む、テクスチャーの特徴抽出に関与していることが示唆された。

《参考文献》

- [1] Ramachandran, V. S. Perception of shape from shading. *Nature* 331, 163-166(1988).
- [2] Lehky, S.R. & Sejnowski, T.J. Network model of shape-from-shading: neural function arises from both receptive and projective fields. *Nature* 333,452-454 (1988)
- [3] Gallant, J. L., Braun, J. & Van Essen, D. C. Selectivity for polar, hyperbolic, and Cartesian gratings in macaque visual cortex. *Science* 259,100-103 (1993).
- [4] Monkey V4 neurons integrating the direction of local luminance gradients in shaded 3-D textures. Hanazawa, A., Komatsu, H. Society for Neuroscience 29th Annual Meeting Abstracts 24, 370.3, (1999).
- [5] Extraction of texture features in macaque area V4. Hanazawa, A., Komatsu, H. Society for Neuroscience 30th Annual Meeting Abstracts 25, 357.18, (2000).

「人間の初期視覚系における

テクスチャ画像の処理」

講演者：本吉 勇 氏

(NTTコミュニケーション科学基礎研究所)

入力画像からエッジを検出し、異なる領域に分割することは、最も基本的な視覚処理の一つである。人間の初期視覚系には、強度や波長に限らず、運動、形状、両眼視差など画像に含まれる様々な特徴の変化に基づいて、エッジを検出したり領域を分割する仕組みが備わっている。特に、細かな形状の違いに基づいてエッジを検出したり、領域を分割するはたらきは、テクスチャ分凝と呼ばれる。そのメカニズムは、かつてはシンボリックなトークンの比較とされていたが、現在では、画像に含まれる方位・空間周波数成分に選択的な線形フィルタ群の出力に対する、二次のフィルタリングと考えられている。これを受けて、我々はテクスチャ分凝機構の諸特性をシステム解析の観点から再検討し、その実像に迫ろうとしている。今回の講演では、以上の研究の流れをレビューするとともに、テクスチャ分凝機構が極めて優れた時間分解能をもつことを明らかにした、我々の最近のデータを報告する。

《参考文献》

- [1] Julesz, B. (1981). Textons: the elements of texture perception, and their interactions. *Nature*, 290 91-97.
- [2] Landy, M. S. & Bergen, J. R. (1991). Texture segregation and orientation gradient. *Vision Research*, 31 679-691.
- [3] Kingdom, F. A., Keeble, D. & Moulden, B. (1995). Sensitivity to orientation modulation in micropattern-based textures. *Vision Research*, 35 79-91.
- [4] Motoyoshi, I. & Nishida, S. (2000). Temporal-frequency characteristics of two subsystems involved in orientation-based texture segregation. *IOVS*, 41 S222.

．電子化検討の中間報告

2000年5月の総会決議に基づいて、AVIRG活動の一部電子化について検討を進め、前号の会報でもご報告いたしましたように、

avirg.org

のドメイン名を取得しました。さらに以下の3点について、検討状況を中間報告させていただきます。

(1) ホームページ開設

AVIRGのホームページを開設致しました。URLは、

<http://www.avirg.org>

です。内容は、AVIRGの紹介、会報、例会のお知らせ、関連イベント情報など、会員の皆様にとって有益な情報を数多く掲載できるよう努力して参ります。ご意見、ご要望等がございましたら

kanji@avirg.org

までお寄せください。

(2) メーリングリスト

会員の皆様への情報提供手段として、また意見交換の場として、AVIRG会員のメーリングリスト

kaiin@avirg.org

を開設いたしました。アドレスの登録、削除は、本文なしで以下のアドレスにメールをお送りください。

アドレスの登録：

To: kaiin-request@avirg.org

Subject: subscribe

アドレスの削除：

To: kaiin-request@avirg.org

Subject: unsubscribe

なお現在、2000年4月の電子化アンケートに記入して頂いたアドレスが登録しております。登録不要の方はご面倒ですが削除をお願いします。また、アンケートにアドレスを記入されなかった方あるいはアンケートを提出頂けなかった方で登録を希望される方は、上記の手順で登録をお願いします。このメーリングリストに関連情報などを投稿されたい方は、

kaiin@avirg.org

までお送りください。

(3) 会則改正について

AVIRG活動の一部電子化に伴い、会員の位置づけ、会の組織や役員の役割などについて、鋭意検討を進めております。早々に改正案をまとめ、会報ならびにホームページに掲載の上、会員の皆様のご意見をお伺いしたいと考えています。今しばらくお待ちください。

～ 会員登録情報の変更のお願い～

AVIRG会員の御所属、会報送付先など登録情報に変更がありましたら、お手数ですが以下のいずれかにご連絡ください。

(財)日本学会事務センター 会員業務係 電子メール kanji@avirg.org (AVIRG幹事宛)

(注) 会員の確認のために、御氏名とともに、必ず会員番号を明記して下さい。

会員番号および学会事務センターの連絡先は会報郵送時の封筒に印刷されています。