

AVIRG

会報

Vol.33 No.6 (2000.4)

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

3月例会報告

「マンマシンインタフェース構築に向けた顔画像認識」

講演：福井 和広 氏((株)東芝研究開発センター マルチメディアラボラトリー)

報告：Simon CLIPPINGDALE (NHK)

《概要と感想》

顔画像認識の研究は近年、計算機技術の進展に伴って盛んになってきたが、日常生活へ応用されるようになるまでには多くの問題を解決する必要がある。福井氏は顔画像認識研究の前提と歴史の概要を述べられた後、問題を解決しようとするアプローチの一つを説明された。

顔画像認識の応用としてはセキュリティへの利用が挙げられる。顔認識処理によってユーザアイデンティティを確認できれば、ユーザに接触することなく、高いセキュリティレベルを実現できる。ユーザは比較的自由に動くことができ、繰り返し行ってもユーザの負担にならない。応用例として、将来のペトロボットのような家庭で使われるロボットが接触せずに主人を識別できたり、入場管理でIDカードより安全な認証が行なえる。また、コンピュータのスクリーンロック解除におけるユーザ確認の場合、連続的にユーザを確認できる技術があれば、一回し

か入力しないパスワードより安全でユーザにとって使いやすい。

従来の顔画像認識システムは、入力画像中の顔のありそうな領域を、予め用意しておいた参照画像辞書と比較する。最大類似度が閾値を超えた場合、それを与えた辞書画像の被写体人物が認識されたと判断する。しかし顔の向き、表情、照明条件、背景などの変化により、同一人物でも顔画像同士が全く違うことがある。画素レベルで比較すれば人物Aの顔画像A1と、同一人物の違う顔画像A2より、異なる人物Bの顔画像B1との類似度が高いことが多い。ここで1と2は、例えば顔の向きまたは照明条件の同異を意味する。時間に伴い、入力対象人物が頭を回転したり表情を出したりすると最大類似度が不安定になってしまい、誤認識も出てくることがある。福井氏が紹介されたシステムはこの不安定性を克服しようとする。

類似度の計算法はいくつかある。昔は、瞳な

ど顔特徴の位置だけを比較したが、条件（顔の向きなど）を制約しても識別可能な人数は1000人オーダーに制限されてしまう。最近のシステムのほとんどは画像の濃淡情報も利用する。1991年から有名になった“eigenface”（固有顔）法では、参照画像のサイズを正規化して全画素の濃淡値をN次元ベクトルとする。全ての参照画像から $M < N$ 次元の部分空間（いわゆる顔部分空間）を主成分分析という統計的手法により生成する。入力画像から顔のありそうな領域を正規化して顔部分空間との直交距離または角度を類似度とし、顔部分空間への射影先の位置により人物を認識する。

福井氏が紹介されたシステムは固有顔法に基づいているが、上述の不安定性を克服するため、入力フレーム一枚ではなく複数のフレームからなる部分空間を利用する。それに加え、一つの顔部分空間の代わりに、辞書に登録してある人物毎に、複数の参照画像からその人物の参照部分空間を生成する。入力部分空間とそれぞれの参照部分空間との正準角を類似度とし、固有値問題の解として求める。各参照部分空間はその対象人物が頭を動かしたり話したりしながら撮影された180フレームから生成され、入力部分空間は連続で撮影された30フレーム（1秒間）から生成される。類似度計算には、それぞれの部分空間の固有値の大きな6次元だけが使用される。このようにして求めた類似度を入力フレーム一枚だけを使った場合と比較すると、安定した類似度を得ることができると同時に、識別能力も向上するという。

さらに頑健性を増やすため、個人変化と照明変動の入力画像への影響を統計的に学習し、なるべく除去しようとされている。その方法は例えば、同一人物の異なる表情を示す画像の差分、または異なる照明条件を示す画像の差分より、差分空間を構築する。この差分空間に平行な成分を入力部分空間と各参照部分空間から除去す

る、すなわち各部分空間を「制約部分空間」に射影する。これにより、20人分の辞書に対しての認識パフォーマンスは92%から99.6%まで著しく改善された。

講演の後半では、氏の顔画像認識処理の概要について説明された。詳細は参考文献に示されているが、福井氏の手法においても、元の固有顔法と同様に、認識の処理を行う前に、入力画像中の顔のありそうな領域を検出し、拡大、回転、平行移動により空間的に正規化し、ヒストグラムイコライゼーションなどにより濃淡ドメインに正規化するプリプロセスを行う必要がある。

多重のスケールにおいて、ウインドウで入力画像をスキャンし、「顔」と「顔でない」という二つの検出用部分空間に射影することにより顔のありそうな領域を指摘し、後の検出と正規化の処理をその領域中で行う。検出処理として、「分離度フィルター」というアルゴリズムを実行する。このフィルターは、狭い範囲の濃淡値の分布が二重の環形にどの程度近いのかを測るもので、瞳と鼻孔に対して大きく反応する。瞳と鼻孔の候補を再度部分空間で調べた後、その位置を使用して顔のありそうな領域を正規化する。この検出プロセスは単純なフィルター特徴を使用するのにもかかわらず、意外にうまく行くということは、おそらく顔と顔でない部分空間の効果を示すのであろう。

顔画像認識の研究者として、非常に興味深く聴講させていただいた。福井氏の研究ではまだ限定された人数の顔を対象とした実験をされているが、さらに多くの人物を扱った場合に、処理のパフォーマンスがどのようになるかは今後の課題であらう。

福井氏は講演の最後に、エンターテインメントとして、顔認識処理を使って入力人物の髪の毛や耳などをリアルタイムでアニメのものに置換えたビデオも見せてくださった。これは遊び

的な応用の一つである。また、実際に東芝のノートブックコンピュータにもこの顔認識処理が搭載されているそうである。

これらの話を聞いていると、顔画像認識が日常生活に応用されるのは時間の問題だという印象が残った。そして、それはおそらく遠い将来ではないであろう。

《参考文献》

- [1] O.Yamaguchi, K.Fukui, K.Maeda, "Face Recognition using Temporal Image Sequence", Proceedings of the third

International Conference on Automatic Face and Gesture Recognition, pp.318-323, April 1998.

- [2] 福井 和広, 山口 修, 鈴木 薫, 前田 賢一, “制約相互部分空間法を用いた環境変化にロバストな顔画像認識 - 照明変動を抑える制約部分空間の学習 - ”, 電子情報通信学会論文誌 (D-II), vol. J82-D-II, no.4, pp.613-620, April 1999.
- [3] 福井 和広, 山口 修, “形状抽出とパターン照合の組合せによる顔特徴点抽出”, 電子情報通信学会論文誌 (D-II), vol. J80-D-II, no.8, pp.2170-2177, August 1997.

「人間の聴覚的空間知覚特性」

講演： 大倉 典子 氏（芝浦工業大学工学部）

報告： 遠藤 徳和（トヨタ自動車東富士研究所）

《概要と感想》

報告者は人間の聴覚特性について、特に後天的に修得する空間把握特性は視覚特性と関連が強いとの思いから長年にわたって研究されている。

今回の報告では、まず今までの研究のまとめをわかりやすく紹介されたのち、空間把握特性の測定において、知覚された音像の位置が空間に固定されたスピーカークの視覚情報に誘導されてしまうこと、任意の位置に設定できる仮想イメージと仮想音源の組み合わせを使うことによってこの問題を解決できることを示された。これにより視覚ホロプタと同様聴覚ホロプタの存在を明らかにされ、その特徴の視覚と違いも指摘された。

報告の主題は、視覚での平行アレー・等距離アレーと同様の特性が聴覚でも認められることを同様の仮想音源を用いた心理物理実験で確認され、これをニューラルネットワークによる学

習で模擬することにより説明出来ることを報告された。

実験し易い視覚特性の研究結果と対比して聴覚特性を予想され、聴覚ゆえの実験設備の困難さを最新のバーチャル技術を応用して解決され、実験結果が聴覚と視覚で同様の傾向であることを示され、さらにその解析までニューラルネットワークモデルにより説明されたのは、素人の聴講者の私にとって非常にわかりやすく新鮮で興味のある内容でした。

何分専門的知識もなく初めての聴講で報告の栄を賜りましたので、間違った理解を多々していると思いますが、以下に発表の概要をまとめます。

発表はプログラム通り

1. 聴空間におけるホロプタの距離依存性
2. 仮想環境を利用した距離に関する音源定位
3. 聴空間におけるアレイの距離依存性
4. 聴空間におけるホロプタやアレイを説明す

る数学モデル

5. 聴覚情報を利用した人工現実感システムの順で話された。

1. 聴空間におけるホロボタの距離依存性

視覚では、目の前で奥行き方向（正中面）に垂直な平行線を何本か書かせると1m強の位置でまっすぐな横線となるが、これより手前で凸に湾曲・遠くでは凹に湾曲な線となってしまう。これをホルムヘルツのホロボタと称して一般的に知られている。

聴覚でも同様な現象となることを大倉氏は心理物理実験によって確認された。但し、聴覚でのまっすぐな線となる位置は1.5～2.5mの間にあり多少遠方になる事も明らかにされた。

実験は全てスピーカを用い、真正面奥行き方向で左右に置かれたスピーカと同じ距離の3つ目のスピーカ位置を答えるというもので、恒常法による。音はランダムノイズと音声で検討されたが、最終的には先入観の入らないランダムノイズのみにした。

ホロボタを説明するため、視覚に対して前田氏は両眼視差・輻輳角が視空間を学習する際に空間制限のあることに着目されてモデルを作成され現実をシミュレートされている。大倉氏は聴覚に対しても両眼視差・輻輳角の代わりに音の強さと時間差をパラメータとしてシミュレートし、現実の特性を説明できることを報告された。

議論として、近い距離ではスピーカの指向性が問題となり左右見開き角のあるスピーカと真正面にあるスピーカを等距離に知覚出来るのか（知覚しても音質の違いが問題になるのでは？）との意見が出され、仮想音源でこれらの要因をうち消した実験の必要性を感じる伏線となっていた。

2. 仮想環境を利用した距離に関する音源定位
視覚ではホロボタの他に、平行アレー・等距

離アレーといわれる特性のあることが知られている。これは、正中面から左右等距離な点を結ぶ2本の直線は手前が狭くなり遠ざかる程開いていき（等距離アレー）、正中面に平行な直線を2本奥行き方向に書かせる（平行アレー）と等距離アレー以上に手前の狭いものになる。

この特性を聴覚で確認するために、スピーカを奥行き方向にたくさん配置しても手前のスピーカが奥のスピーカの音波を妨害してしまい、正常な実験が出来ないことが仮想音源を作るきっかけとなった。

距離に関しては音量（強度）の変化で、方向に関しては左右の音量差と位相（時間）差でシミュレートし、ヘッドフォンで提示する。都合、真正面単一音では左右差がないのでモノラル音となる。音の提示と同期して、スピーカの仮想画像をシースルー型ヘッドマウントディスプレイ（HMD）を使って表示し、現実空間に重畳する事で視覚・聴覚の現実感を向上している。確認として、真正面直線上での距離知覚実験を行い、1～5mの実空間で行った過去の実験結果と同じ傾向が得られた。仮想音源により10m以上遠くにも仮想スピーカを置くことが出来るようになった。

議論は、ヘッドフォンも各種特性があり選定に注意が必要、スピーカ位置による頭部伝達関数（HRTF）の違いを考慮していないが大丈夫か、やはり仮想音源は現実を十分シミュレート出来ないのでは（でも実実験は大きな無響室が必要/スピーカの干渉は？）等の意見が出てかなり盛り上がった討論となりました。また、聴覚実験は常に基準位置の音と比較したものであるが絶対的な距離感を計測する必要は無いのか、との意見に対し1対比較が一般的だとの事で、工学的見地からは多少物足りない感じを受けました。

3. 聴空間におけるアレイの距離依存性

上記仮想音源を用いて、漸く聴覚の平行ア

レー・等距離アレーの実験が行えるようになった。4√2 m先の基準左右2対のスピーカに対し、手前の平行または等距離の2対のスピーカを答えることで平行または等距離アレーを調べた結果、両アレーとも直線にならず平行アレーの方がより手前で狭くなるとの視覚と同様の傾向が明らかになった。

議論として、アレーやホロボタが視覚と聴覚で一致するのは学習段階で聴覚空間が視空間と必ず関係しており視空間の影響が強く出るとの傾向が一致するのではないかと、との意見と聴覚空間だけでも学習出来視空間の影響を排除できるとの意見があり、聴覚しかできない後方で映像定位実験をしてみても（一般に難しくて手をつける人が少ないそうですが）との突飛な意見で盛り上がりました。

4．聴空間におけるホロボタやアレイを説明する数学モデル

上記3までで発表途中の活発な議論によって残り時間が少なくなり、空間知覚の数学モデルシミュレーションに関しては奔った発表となってしまったが、聴覚ホロボタの解析と同様のスカラ加算モデルを用いて説明出来ることが報告された。具体的には、平行・等距離アレーとも正中面平行直線とならないこと、平行アレーが等距離アレーより正中面側に振れることが種々のシミュレーションで実実験と同傾向となった。

5．聴覚情報を利用した人工現実感システム

実際の映像定位アプリケーションとして、電子総合研究所やNTT研究所のVR(バーチャルリアリティー)研究の例の紹介があったが、これも時間の都合で詳細は省略された。それぞれホームページで紹介されているので参照されたい。

全体の感想として、非常に活発で自由な議論がなされ、参加者の熱気が講義室全体に溢れていた。初めての私も素人の的外れな質問をたくさんしてしまいましたが、大倉教授や参加者の皆さんの分かりやすい応答とそれに続いて専門的な議論に発展する場面もあり、あっという間に予定時間が過ぎてしまいました。

報告のより詳しい研究内容を知るためには以下の参考文献を参照ください。

《参考文献》

- [1] 大倉典子, 館日章, “距離に関する音源定位と聴空間におけるホロボタ”, 計測自動制御学会論文集, 30-11, 1287/1292 (1994)
- [2] 大倉典子, 前田太郎, 館日章, “聴覚ホロボタを説明する空間位置知覚モデル”, 計測自動制御学会論文集, 34-10, 1472/1477 (1998)
- [3] 大倉典子, 柳田康幸, 前田太郎, 館日章, “仮想環境における聴覚アレイの測定とその数学モデル”, 電子情報通信学会論文誌, D-11-10, 2438-2446 (1998)

・特別講演・平成12年度通常総会のご案内

5月25日(木) 15時より特別講演を, 17時より平成12年度通常総会を, 東京大学工学部6号館 2F 63号講義室で開催致します。議題は次の通りです。

1. 平成11年度事業報告
2. 平成11年度収支決算報告
3. AVIRG会報等の電子化について
4. 平成12年度事業計画(案)および収支予算(案)
5. その他

なお, 特別講演としては以下を予定しておりますので, 奮ってご参加ください。

「1ms超並列ビジョンチップとその応用」

講演者: 石川 正俊 氏(東京大学)

ビデオレートの限界を破る1msのサイクルタイムを有する画像処理を実現する超並列・超高速ビジョンチップについて, そのアーキテクチャ並びにVLSI技術を用いた実装に付いて述べ, 具体的なビジョンチップの開発例を示す。また, このよう

な高速画像処理チップの応用例として, ロボット制御, 特にビジュアルフィードバック, ヒューマンインターフェイス, バーチャルリアリティ等への応用等を述べる。特に, ビジョンも含めて1msのフィードバックレートを有する感覚運動統合システムについて, 統合原理, システムアーキテクチャ, 具体的な動作として高速把握行動の実現について述べる。これらにより, 画像処理並びに応用分野が根底から変わることをビデオによる実例を交えて述べる。

《参考文献》

- 1) 石川正俊, “並列処理を用いた知能化センシング”, 計測と制御, Vol. 36, No. 9, pp. 648-654 (1997)
- 2) 石川正俊, “超並列・超高速視覚情報処理システム - 汎用ビジョンチップと階層型光電子ビジョンシステム -”, 応用物理, Vol. 67, No. 1, pp. 33-38 (1998)
- 3) 石川正俊, “センサ情報処理技術で変わるロボットの世界”, システム/制御/情報, Vol. 42, No. 4, pp. 210 -216 (1998)
- 4) 石川研ホームページ,
<http://www.k2.t.u-tokyo.ac.jp/>

～ 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: avirg-member@vision.STRL.nhk.or.jp (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。