

コンピュータビジョンとパターン認識会議 CVPR'96 報告

池内 克史 (東大/CMU), 野村 由司彦 (名大),
中村 裕一 (筑波大/CMU), 志沢 雅彦 (NTT/東大),
宮島 耕治 (NTTデータ), セバスチャン ルージョ (電総研),
出口 光一郎 (東大)

1996年6月, 米国サンフランシスコで開催されたIEEEの「コンピュータビジョンとパターン認識会議 (CVPR'96)」について, 同会議のプログラム委員長と出席者有志により, その概要を報告する.

A Report on IEEE CVPR'96

Katsushi Ikeuchi (University of Tokyo/CMU),
Yoshihiko Nomura (University of Nagoya),
Yuichi Nakamura (University of Tsukuba),
Masahiko Shizawa (NTT/University of Tokyo),
Kouji Miyajima (NTT Data), Sebastien Rougeaux (ETL),
Koichiro Deguchi (University of Tokyo)

This report describes an overview of the 1996 IEEE Computer Vision and Pattern Recognition Conference (CVPR'96) held at San Francisco, USA in June. This report includes the message and summary by the Program Chair Prof.K.Ikeuchi, and comments on presentations by members of SIG-CVIM who attended the conference.

1 はじめに

1996年6月16日から22日、米国サンフランシスコで、IEEEの「コンピュータビジョンとパターン認識会議 (CVPR'96)」, および、関連ワークショップが開催された。本報告は、同会議について、池内克史プログラム委員長 (東大/CMU) による会議の沿革と概要と、出席者有志によるそれぞれの注目した発表に対するコメントを集めて、同会議を紹介したものである。日本からの出席者はそう多くなく、本報告ではすべてのセッションを網羅していない。会議の雰囲気、研究の動向を伝えようとした。各発表の詳細については、IEEEから出版されている、会議録 [1] を参照して頂きたい。

CVPR'96の沿革と概要 - プログラム委員長からの報告 -

今回で15回目になるCVPR (正式名称 IEEE Computer Society Conference on Computer Vision and Pattern Recognition) が、本年は サンフランシスコ の Hotel Nikko で 6月18日より20日まで開催された。この会議の前身は、Pattern Recognition 及び Image Processing (旧名称 PRIP) のアメリカの全国大会であったが、年を追うにつれ国際会議化していった。今回の実行委員長はカリフォルニア大学 Bhanu、プログラム委員長は ウィスコンシン大学の Dyer と私 (池内) がつとめた。

今年には551件の論文が集まった。(表1参照。) 投稿論文数は、92年 シャンペン (イリノイ州) 357件、93年 ニューヨーク 431件、94年 シアトル 470件と単調増加の傾向にある。(1995年は ICCV が北米で開かれたため CVPR は開催されなかった。) ただコンピュータビジョンの研究人口がこのように年々増加しているかどうかは定かでない。アメリカ全体の研究予算もそれほど増加していない。不思議な現象である。

全体として optical-flow や shape-from-shading といった純粋な理論研究よりも、それをを用いた face-tracking や real-time tracking system を設計するといった応用研究が増加して来た。これは、コンピュータビジョン分野の健全な進歩のためには歓迎すべき傾向だと思う。しかし、最優秀論文には Belhumeur & Kriegman, "what is the set of images of an object under all possible lighting conditions" とやはり理論研究の論文が選ばれた。理論と応用に対して別々に最優秀論文を選ぶべきであるとの意見もあったが残念ながら今回は見送られた。

査読は、査読者も著者を知らず、著者も査読者を知

らないという double-blind 方式を取った。コンピュータビジョンを8つの研究分野に分類し、査読はこの分野ごとに行なった。50名ほどのプログラム委員を6名づつほどの分野に分け、各分野毎に1名の主査、2名の幹事を割り当てた。各論文は分野内の3名のプログラム委員のもとに送られた。また、3名の査読者の内1名は必ず主査か幹事とした。論文の分野分類並びに査読委員の選定は、公平をきすため、筆者 (池内) が一人で行なった。500件余のアブストラクトを読み、1500名ほどの査読員を短期間で決定するのはなかなか骨が折れた。

選考会は、1月20日、21日にこれらの主査、幹事が Hotel Nikko に集まり行なった。選考作業も当然分野毎に行なった。主査、幹事が査読結果 (必要なら原論文も) を読み、それに応じて論文の順位を決定し、採録を決定した。これにより、73件の口頭発表、64件のポスター発表論文を選んだ。(表参照。)

口頭論文は15セッション/ダブルトラックで発表された。(筆者自身は反対であるが) ECCV との重複発表を禁じたため最終的なプログラムを作成してから、あるセッションの全部の論文が ECCV でも発表されていることがわかり、そのセッションを潰して、パネルにするという事故もあった。これなど、もう少し早めに各著者が ECCV での発表を CVPR 事務局に連絡してくれば他のポスター論文を口頭に回せたのにと残念に思った。ポスター 並びに デモセッションは非常に盛況であった。デモセッションはもう少し広い部屋を確保するべきであった。

日本からの採録が少数であるとの意見もあるが、これはもともと日本からの投稿件数が少ないからという見方もできる。ヨーロッパ (特にイギリス) からの採録率が低いのは ECCV がロンドンで開催された影響であろう。ただ、ドイツ、日本などの採録率が低いのはやはり英語、表現方法の問題もあるかもしれない。

(正直言ってアブストラクトを読んでいてイライラするような論文も時々見受けられた。) ただ、本質的には CVPR はアメリカの国内会議であり、アジア地域のコミュニティーを盛り立てると言う意味では、日本はむしろ同時期に開かれることになる ACCV への投稿、採録を気にした方が良いのではないかとも思う。

次回は、実行委員長に南カリフォルニア大学の Medioni と Nevatia、プログラム委員長にイリノイ大学の Ponce と コーネル大学の Huttenlocher との陣様で来年6月に開かれる。場所は以前に太平洋の真中 Hawaii でやったのだからと言う理由 (?) で大西洋の真中の US Virgin 諸島の St. Thomas である。余談であるが、ハリケーンに注意した方が良いかもしれない。もともと予定していたホテルがハリケーンで破

壊され急ぎよ別なホテルと契約をかわしたとの噂を聞いた。なお、詳しくは、
<http://marathon.csee.usf.edu/cvpr97.html>
 を参照のこと。

(以上、池内プログラム委員長)

	Submitted	Accepted
USA	292	97
Canada	22	5
Mexico	1	0
Japan	24	3
Singapore	10	1
Australia	8	1
China	7	0
HongKong	5	1
South Korea	2	0
Taiwan	3	0
France	54	10
United Kingdom	40	5
Germany	19	1
Israel	22	3
other	42	11
Total	551	137

2 講演発表

TM1:Segmentation and Low-Level

本セッションでは、領域分割に関して2件、エッジ抽出に関して1件、物体検出に関して2件の発表があった。領域分割、エッジ抽出のいずれにも共通の問題である Multi-scale の問題を扱っているものが多く見受けられた。

K.Vincken, M.Viergever(Univ.of Hospital Utrecht) からは、Multiscale linking model に基づく領域分割について述べている。これは、まず画像を異なるスケールで表現し、この異なるスケール毎の画像で領域分割を行う。次に各スケールで分割された領域をスケール間で対応する領域にリンク付けて各スケール毎に分割された領域を tree 構造で表現する。そして、tree 構造で表現された領域を適切にグループ化することによって領域分割するものである。実験では人間の頭部の MR 画像から脳の領域を抽出した結果を示している。J.Elder(NEC Research), S.Zucker(McGill Univ.) は、scale-adaptive なエッジ抽出方法につい

て述べている。これは輝度勾配の2階微分を求め、その極値の間隔に応じて画像をぼかすスケールを局所的に決定するものである。画像中に混在する高周波領域と低周波領域を適切なスケールでぼかすことにより、エッジを安定に抽出している。また、本手法で提案したスケールの決定方法を用いて、画像を Multi-scale で表現した画像符号化について論じている。

TM2:Object Recognition and Pose Recovery

姿勢推定に関しては、画像からエッジセグメントを抽出しこれらのエッジセグメントの方向から物体の姿勢を推定するものが1件、2画像間の対応点問題を MAP 推定によって解いたものが2件発表された。前者は物体の特徴を表すエッジセグメントが容易に抽出される対象についての実験結果が示され、後者は実画像でなくシミュレーションによって提案手法の有効性を述べている。しかしながら、これらの発表は Appearance based による姿勢推定問題との位置付けについて明確な記述がなく、対象をより一般的な環境で撮影した場合への拡張については疑問が残る。物体認識に関しては、Active sensing による3D物体認識に関して1件、Appearance based による物体認識に関して1件の発表があった。

J.Krumm(Sandia National Lab.) は、部分的に隠蔽がある状態での物体の姿勢推定方法を提案している。これは村瀬(NTT)らが提案したパラメトリック固有空間法を部分画像に適用し、その結果を統合することにより背景、隠蔽の変化の問題に対処している。物体の特徴を表わす30個程度の部分画像を物体の見え方毎にあらかじめ学習しておき、隠蔽が生じている物体の入力画像に対してこれらの部分画像を検出し、検出された位置、見え方の角度から物体を認識している。

(以上、宮島)

TN1: Human and Articulated Motion

人物動作計測に関する研究発表が4件、関節を持つ運動の解析に関する論文が1件あった。動作計測が現在 CV 研究の流行となっているのを反映して、聴衆も多く盛況であった。

従来から、関節を持つ剛体の集まりとして人体のモデルを記述する、モデルベースの認識が行なわれているが、このセッションで目だったのは、それに加えて、多眼(ステレオ視ではなく、多方向からの離れたカメラ群からの撮影)の利用と見え方モデルとのマッチングである。

Gavrila, Davis の研究では、ほぼ直交する4方向から撮影された画像を用いており、Kakadiaris,

Metaxasの研究では3方向からの画像を用いている。いずれの研究も、人体では避けることのできないオクルージョンの問題に対処するために多眼視が有効であることを主張している。方法は全く異なるが、Darrellらの研究でも、解像度の異なる2つのカメラを用いることを前提としており、Cui, Wengの研究では機械的に腕の動きをトラッキングする際に得られる多数の(例えば19枚)画像を用いることを提案していることも興味深い。

また、これらの研究に共通することとして、2次元の見え方を基にしてモデルと観測データとのマッチングを行なっていることがあげられる。では、見え方モデル(固有空間法)を複数の位置、照明条件に適用するために、固有空間法を拡張している。T.Darrellらも複数の見え方を用いる際の階層的なモデルの選択方法について提案している。また、では人体の3次元モデルを持ち、それを基に予想される2次元の見え方(予想されるシルエット)と実際の画像とのマッチングを行なっている。前者は古くからある chamfer matching の手法を用い、後者はエッジ間の距離を用いて、変形のための仮想的な力を求めている。いずれの研究でも、実験ではかなり良好な結果が得られているが、モデル(超二次曲面等)のシルエットと観測画像のエッジとのマッチングを行なっているため、衣服や背景が複雑な場合には課題が残っている。

RN2: Image Database and Document Analysis

画像データベース関係の発表2件、固有空間法に関する発表1件、文字認識関係1件の発表があった。同じ日の午後に Content-Based Retrieval と題するパネルセッションがあったことも(これについては簡単に後述する)、画像データベースに対する一般の関心の高さを物語っている。

しかし、上述のように画像データベース関係の発表は2件しかなく、結果として低調であった。Zabihらの研究は、映像のカット分割に関する手法を提案している。この発表では、エッジの出現/消滅を検出することにより、映像内の物体の動きがある場合でもフェードイン/アウト、ディゾルブの検出の精度が良くなることを報告している。また、Minka, Picardの研究では、画像データベースにおける組織化(グルーピング)の問題を扱っている。この研究の主張は、画像の類似性の定義が非常に多種多様となり、コンテキストに依存するため、画像特徴との直接の対応がとれないこと、そのため、ユーザの反応等を教師として利用して画像間の類似性を随時学習していくモデルの構築にある。

また、オクルージョンに対応可能な固有空間法の拡張を Leonardis, Bischof らが提案している。この研究では、識別のために、固有ベクトルとなるパターン及び未知パターンのすべての部分を用いる必要がないことを利用して、一部が隠蔽されているパターンへの固有空間法の適用を提案している。その方法として、ここでは未知パターンからランダムに選び、それらを用いて固有空間中の位置を求めている。実験では、固有ベクトル数の3倍~5倍の点をランダムに選んで複数の仮説を生成し、Minimum Description Length principle を用いて、採用する仮説を選択している。カテゴリ(モデル)の数、種類によって、選ぶべき点の数、位置などが変わってくるのが予想されるが、その基準が明らかになれば、実用的な手法になるだろう。

(以上、中村)

TA1: Contour Reconstruction

J.Shah は、curve evolution、分割、そして異方性拡散の問題を共通に解釈できる枠組みを提案した。彼は、まず、形状復元に用いられてきた curve evolution の枠組みにヒントを得て、セグメンテーション汎関数(エッジ強度関数、および smoothed image 関数の汎関数)を導入した後で、その汎関数から勾配降下方程式を求め、曲線の駆動に関わる停止項、曲率項、水平移動項を導いた。さらに、物体境界の不連続やY接点が表現でき、領域分割に有効であることを示した。

M.Heath, S.Sarker, T.Sanocki and K.Boyer は、心理学の分野における実験手法と統計的解析手法を応用し、4つのエッジ検出器、すなわち Canny エッジ検出器、Nalwa-B inford エッジ検出器、Sarkar-Boyer エッジ検出器、ヒステリシスしきい値処理付 Sobel エッジ検出器を比較した。要因として、エッジ検出器以外にも、画像、パラメータ設定(パラメータを固定する、またはパラメータを画像毎に適応させる)を取り上げ、それらの組としての要因(相関の考慮)も含めて、ANOVA なる分散分析によりどの要因が支配的であるかを比較した。結果の要点は、(1) 画像により最適なエッジ検出器は異なり、どの画像にも最適なエッジ検出器はない。強いて優劣を挙げれば、最良が7、最悪が1という評価でいうと、固定パラメータでは(Canny と Nalwa-Binford が4.4)、(Sarkar-Boyer と Sobel が3近辺)であり、適応的パラメータ設定では(Canny が4.8)、(Nalwa-Binford と Sarkar-Boyer が4強)、(Sobel が3.5)であった。(2) Nalwa-Binford エッジ検出器は、最適パラメータが画像によって変わらない。(3) 適応的

パラメータ設定では、エッジ検出器による差がほとんど現われない。ただし、わずかに8枚のサンプル画像のほとんどが人工物であることに象徴されるように、表題通り初期の研究成果という印象を受けた。

S. Casadei and S.K.Mitter は、Canny オペレータで検出されたエッジをグリーディに結合するやり方では曲線の特異点近傍で失敗しやすいことを指摘した後で、高精度にエッジ輪郭を再構成できる階層的手法を提案した。階層構造は、下層から上層へ、明度、エッジ点、エッジ線分、正規曲線、結合曲線からなる5つ層である。最上の結合曲線層で線分を結合する仮説を生成し、下位階層で明度の一様性、線分端点の近接性、線分の直線性などの規範に基づいてその仮説を検証するもので、トップダウン的な線分結合アルゴリズムである。

D.Geiger, K.Kumaran and L.Parida は、拡散という物理現象を模擬して図と地を分離する視覚の組織化手法を提案し、Kanizsa 正方形の主観的輪郭の問題に適用した。主観的輪郭の生成は、一般には、エッジ端点を伸張するか、または領域ベースで行われている。彼らは、まず、コーナー点、線分端点、そしてT接点とY接点に着目し、各点、各仮説毎に、その点の回りに遮蔽の仮説を立て、周囲にその仮説を拡散させる。この仮説の拡散は、遮蔽面である可能性を表す測度の変化として表され、拡散場をなす。次に、拡散場が等しいという点で最適な仮説の組を選択する。拡散という現象を用いた結果、曲線までもが生成された点が興味深い。

L.R.Williams and D.W.Jacobs は、確率的な充填場を局所的並列的に計算することにより、例えばKanizsa 三角形の主観的輪郭を説明できる手法を提案した。これは一種のエッジ端点伸張方式である。2つの線分の端点を吐出源と吸込源とし、ランダムウォークする粒子の流出の軌跡により生成される場と流入の軌跡により生成される場の積で表される新しい場が、主観的輪郭となるという着想である。ただし、粒子の移動方向は確率的に定められ、その分布は現在の方位を中心とする正規分布でモデル化している。その際、粒子の崩壊も考慮しており、例えば局所的な輝度勾配で変調する異方性崩壊を用いることにより、輪郭抽出もできることを示している。主観的輪郭の問題において、曲線を生成する能力、遮蔽の多層構造の復元性という点では、上記のD. Geigerら方法の方が優れているように思われる。

RN1:Texture and Shading

A.J.Stewart and M.S.Langer は、利用する画像が1枚、対象の物体の反射率が一定との条件のもとで、

拡散光源下で影と相互反射を考慮した陰影からの形状復元方法を提案した。ポイントは、物体の底面は物体自身の影響や、物体を支える床の暗い部分からの2次反射光しか受けられないことが多く、一層暗くなりがちであるというような観察に基づき、“物体の表面上の点の輝度とその点に2次反射の影響を与える周囲の点の輝度は等しい”とした相互反射モデルにある。このモデルは、大胆な仮定に基づいているが、相互反射の問題に対する一つの手掛かりになるであろう。

L.Wang and G.Healey は、照明変化と幾何学的変換(回転とスケール)を受けたカラー画像のテクスチャの識別方法を提案した。ポイントは、(1)2次元自己相関関数行列の各成分毎にとったZernike モーメント行列は、照明変化に関しては線形変換を受け、幾何学的変換に関しては線形座標変換を受けるだけであり、容易に求められる、(2)Zernike モーメントは絶対値が回転不変であり、かつその行列が大きくテクスチャ特徴の表現能力が高い。ただし、スケールは、簡便のため、2次元自己相関関数行列の原点中心の円周上の積分値を用いて求めている。照明に関する線形変換行列は事前に較正しておくのか、扱い方が示されていない。

W.Y.Ma and B.S.Manjunath は、テクスチャの検索を目的として、ガウス型ウェーブレットとして知られているGabor フィルタ(複素三角関数で変調されたガウス関数)を用いたテクスチャ特徴の抽出方法を提案している。ただし、テクスチャ特徴の抽出という点では、上記のD.Geigerらと同じ問題である。Gabor フィルタは、方向とスケールが調節できるエッジ・線分検出器といえる。テクスチャ特徴は、画像の一定範囲内におけるGabor フィルタ出力の平均値と分散を用いている。この特徴量は、6方向、4スケールでの個数が48個で、上記D.Geigerらの自己相関関数に基づく特徴量に比べると、かなり少なくなりそうであるが、波長分離性が良く、画像の局所的なばらつき表現能力が高いように思われる。

M.Watanabe and S.K.Nayer は、焦点外れからの奥行き復元に用いことのできる最少のオペレータの組を提案した。条件は、ぼけが非点光束のない錯乱円でモデルされる、合焦点の前後の2枚の入力画像を用いる、レンズ開口径、錯乱円半径、焦点距離、撮像面位置が既知、などである。基本的には2枚の画像のフーリエ変換の比をとると原画像の影響が消え、ぼけ関数のフーリエ変換のみが残り、それから錯乱円の直径、さらには合焦点位置が求められるという原理を次のように発展させた。第1には、フーリエ変換する2枚の画像として、2枚の画像の差画像と和画像を用いる点である。そうすることにより、焦点外れや空間周波数に対して、フーリエ変換の比が線形を少し補正した単

純な関係になってしまう。よって、事前に、合焦点位置とフーリエ変換の比との関係を較正しておけば、実際のフーリエ変換から逆に合焦点位置が求められる。第2には、周波数領域の関係を逆フーリエ変換することにより、実空間でのわずか7×7画素の大きさのフィルタ出力のみを用いて合焦点位置が求めている点である。

(以上、野村)

WM1: Active and Real-Time Vision

The first presentation of this session was devoted to 3D scene reconstruction. The method presented by *E. Marchand and F. Chaumette* (IRISA-INRIA) uses controlled camera motion for precise primitive extraction (focus on points of interest, alignment the edges of the images along the row and column of the CCD, etc.) and global scene retrieval (using the motion of the camera to suppress occlusion and obtain a complete model of the scene). Adopting the continuous "structure from controlled motion" approach, this technique looks accurate and stable but is limited to predefined primitives such as points, lines or cylinders.

The next talk by *H. Rotstein and E. Rivlin* (Technion) was more dedicated to the "servoing" of active vision systems and did not include any results on image processing. Using target motion models, they discussed many problems such as the determination of the optimal size of the foveal window or the opportunity whether one or two controllers (fixation and saccade) are suitable for such systems. The control theory were explained in detail, but no experimental results were presented.

Berthouze, Rougeaux, Chavand and Kuniyoshi (ETL) then presented some calibration techniques for an active vision head with foveated wide-angle lenses nicknamed ESCHeR. The methods presented, taking advantage of the camera motion to model the space-variant resolution of the lenses, can be achieved online without calibration setup. A video showing demonstration of target tracking and visual servoing was presented as a validation. However only very simple image processing (bright light targets in black backgrounds) was used for the experiments.

That was all for 'active vision' part of this session, the rest of the talk dealing with passive cameras.

K. Toyama and G. Hager (Yale University) intro-

duced their hierarchical architecture for robust visual tracking. The principle is to use several trackers (homogeneous regions, edges, corners, SSD, etc.) and intelligent selectors to switch from one to another depending on their efficiency during the tracking process. This coarse to fine approach, or "incremental focus of attention", enhance drastically the overall performance the system in critical conditions when target recovery is often requested. The system is also capable of determining object position and rotation in the finest tracking stage (correlation), but only with a reasonably slow motion of the target.

T. Kanade, A. Yoshida, K. Oda, H. Kano and M. Tanaka finally presented a spectacular real-time stereo machine for video-rate depth mapping. This systems uses 5 cameras and some basic SSD matching techniques implemented on a fast dedicated hardware (a C40 DSP array) to recover the depth of a scene at a resolution of 256x240 at frame rate. They then demonstrated how to use this system for merging virtual and real images on the basis of depth information. They also showed a video of their virtual studio, a set of more than 50 cameras equally arranged on a semi-sphere and capable of reconstructing off-line a whole 3D scene (size of a human being) using similar correlation techniques.

RM2: Systems and Integration

One of the most impressive result presented at this session was the implementation of *P. Belhumeur's* theory (see the best paper award in session WN2) by *G. Hager* in an application for tracking object under varying illumination conditions. The method looks robust and uses standard correlation techniques with no additional computational cost, but requires an "illumination basis" (usually 5 images under different lighting condition) of the tracked object and thus have a complex initialization process.

Also noticeable in this session was the work of *S. Bull and B. Sridhar* (NASA) for detecting obstacles on a runway to assist the landing of airplanes. The method uses on-board cameras to detect image flow disparities between expected and computed flow fields (using a planar surface model for the runway). Experimental results obtained on real flight were presented, yet the image processing was done

off-line due to expensive computational costs.

Another promising work for road surveillance was presented by *T. Tan, G. Sullivan and K. Baker* (University of Reading). Their method, using ground-plane assumption and different car models, is capable of retrieving position, model and 3D orientation of cars using snapshots of highway's traffic.

(by Sebastien Rougeaux)

WA1: Motion Estimation and Structure from Motion

複数運動を扱った研究が2件, egomotionのアルゴリズムを実験的に比較した研究が1件, 滑らかな3次元物体のシルエットの運動からの3次元復元が1件, 多フレームのStructure from Motionに関する研究が1件, 合計5件の発表があった。

複数運動の推定に関しては, Toronto大のS.X.Ju, M.J.Black, A.D.JepsonとMITのY.Weiss, E.H. Adelsonの発表があった。特に目新しい手法の導入はなかったが, 運動数の決定法が重要な課題であり, 発表や議論でもその点に論点が集まっていた。この周辺の研究は, 動画像符号化の将来規格になる予定のMPEG4に関連している。複数面の重なり合いと面の不連続を同時に扱うことを含め, 複雑な運動や面構造を復元する方法は, 今後も基本技術としてさらに研究されるべきである。

Vijayakumar(Yale Univ.)らは, 滑らかな表面を持つ動物体のシルエットの単眼像から物体表面の3次元幾何構造情報とカメラの運動を復元するアルゴリズムを提案した。従来の方法では, カメラの運動は既知としていた。bi-tangent, 変曲点などの射影不変な特徴を用いて3次元運動の拘束方程式を解いた。

WN2: Shape Representation

照明条件と画像集合の次元に関する考察が1件, 3枚の濃淡画像から3次元物体の復元を行なったものが1件, 鏡映対称性, 点対称性に関する不変量の研究が1件, 可変形モデルに関する研究が1件, 複雑な表面復元に関する研究が1件の合計5件の発表があった。

P.N.BelhumeurとD.J.Kriegman(Yale Univ.)は, 照明条件を任意に変化させて得られる画像の集合を数学的に特徴付けた。主要な結果は次の通り。(1)任意数の無限遠方からの光源に照らされた凸Lambert面の n 画素画像の集合は, n 次元画像空間の中で凸錐面(これをillumination coneと名付けている)をなす。このillumination coneの次元は, 画像面上に観測されるにおける互いに異なる法線ベクトルの数に等し

い。(2)(1)の条件のもとで, 特定の物体のillumination coneは, 特異な条件でないように選ばれた3枚の画像から構成できる。(3)任意の照明条件で任意の表面反射率を持つ任意形状の物体の n 画素画像の集合も凸錐面をなす。この結果は, 照明条件の変化に弱いとされるコンピュータビジョンのアルゴリズムを改良するための指針となる。

L.V.Gool,T.Moons,M.Proesmans(Katholieke Univ., Leuven, Belgium)は, 線対称あるいは点対称の関係にある平面曲線が透視投影のもとで持つ射影不変量を導いた。3次元空間内の平面上で対称な関係にある曲線特徴を画像から検出するグルーピングに有用である。

RA1: Feature Extraction and Detection

スケールを自動化したエッジやリッジ検出の研究が1件, パラメトリック固有空間法を応用した特徴抽出法の研究が1件, 関係グラフとグラフスペクトルを用いた画像特徴のグルーピング法の研究が1件, 画像特徴分布のランダム性に関する研究が1件, Lie群論を用いた近似的にシフト不変なテンプレートの構成法が1件, の合計5件の発表があった。

S.K.Nayer, S.Baker, H.Murase (Columbia Univ, NTT基礎研)は, パラメトリック固有空間法をstep edge, line, cornerの検出に応用した。例えば, cornerの検出では, 2辺の角度とスケールをパラメータとする3次元多様体との距離が十分近いという条件をテストする。この3次元多様体を求めるため, 数万の例を生成している。この固有空間法の難点は, 精度を上げるために例をこのように非常に沢山必要とすることである。

(以上, 志沢)

3 パネル討論

RA2: Content-Based Retrieval

画像検索にたいするCVの技術の利用について, 現状と今後の課題に関する討議が行なわれた。そこであげられた主な意見としては: 現状では, 本当の意味でのContent-Based Retrievalができていない。限られたコンテキストではなく, 広く一般に通用するVision Algorithmが必要。テキスト情報などのマルチモーダルな情報処理が必要不可欠。正解が明らかではないために, まず正解についてコンセンサスのとれたテストベッドを作ることも必要。ユーザーインターフェースを含めて考えなければ, retrievalの問題は語れない。等

(以上, 中村)

4 ポスター発表

TP2:Pattern Recognition

本ポスターセッションでは、領域分割に関して1件、物体抽出に関して2件、Shape recoveryに関して1件、特徴抽出に関して1件、パターン認識一般に関して3件、画像照合に関して1件の発表があった。

H.Shekarforoush, J.Zerubia(INRIA)は、多重位相相関に基づく画像間の位置合わせ手法を提案している。これは Fourier shift theorem に基づいた考えで2枚の画像間の移動量を推定するものである。J.Peng, B.Bhanu(UC Riverside)は、強化学習により領域分割のパラメータを推定する手法を提案し、これを光源の状態、カメラの位置を変化させながら撮影した一連の画像の領域分割に適用した。S.Baker, S.Nayer(Columbia Univ.)は *Pattern Rejection* という考え方を提案した。これは *rejector* を認識候補の中で認識するパターンと大きく異なるパターンを除去するアルゴリズムと位置付け、高次元の特徴ベクトルで表現されるパターンをデータベース中のパターンとマッチングする際に、マッチング候補を絞りこむことの効果について述べている。この他、Sonar Range Image からリアルタイムでの物体抽出、RBF network に基づくパターン認識などの発表があった。

Pattern Rejection という新しい発想の発表が2件あった。今後は、物体が認識されるためのパターン、及び、*reject* されるためのパターンとの識別に関する議論への発展が期待される。

TP3:Applications

本ポスターセッションでは、顔画像認識に関して2件、医用画像に関して2件、文字・文書認識に関して4件、画像検索に関して1件、物体追跡に関して1件の発表がなされたが、CVのアプリケーションとしての目新しい発表はなかった。

R.Uhr, N.Lobo(Univ.of Central Florida)らは、警察が発表する容疑者の似顔絵を基に顔画像データベースとの照合方法について提案した。これはかなり写実性が高い似顔絵画像から目、鼻、口を切り出し、スケッチによる shading がかかっている顔の肌の面の shading を滑らかにして、再度、目、鼻、口を埋め込むことにより“疑似顔画像”を作成し、これを顔画像データベースと照合するものである。照合方法は Turk & Pentland による“固有顔”を基に行っている。医用画像解析に関する発表として、S.Kumar(Cognex co.), D.Goldgof(Univ.of South Florida)は、画像間の対応をとらずに非剛体の動きを推定する手法を提案した。G.Funka-Lea, A.Gupta(Siemens)は、アフィン

変換を変形した Affine Plus Anisotropic Radial Scaling を大域的な動きモデル、Active Spring Mesh を局所的な動きモデルとし、これらの Hybrid モデルを基にして非剛体の動きを推定する手法を提案し、心臓の動きを抽出した結果を示している。B.Moghaddam, A.Pentland(MIT)らは、画像平面 $X-Y$ における画素 (x, y) における輝度値 $I(x, y)$ とした XYI 空間に点 $(x, y, I(x, y))$ によって構成される surface を作成し、これを特徴量とした顔画像認識を提案している。この他、文書中の表の解析方法、文書画像からのテキスト抽出、手書き英文の認識など、画像検索における複雑な形状の類似性の発表があった。

(以上 宮島)

5 おわりに

IEEEの「コンピュータビジョンとパターン認識会議(CVPR'96)」について、プログラム委員長による会議の沿革と概要と、出席者有志による発表へのコメントによって同会議を紹介した。

最初に述べたように、本報告ではすべてのセッション・発表を網羅していない。池内プログラム委員長と各出席者から頂いた原稿を、なるべく会議の雰囲気伝えるべく、出口が編集したものである。この分野の研究の動向を感じて頂ければ幸いである。

参考文献

- [1] Proceedings 1996 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, IEEE Computer Society Press, ISBN 0-8186-7258-7, 1996.