

## テレビ視聴環境を利用した顔認識のための変動学習

松尾 賢治 橋本 真幸 河井 恒

†株式会社 KDDI 研究所 〒356-8502 埼玉県ふじみ野市大原 2-1-15

E-mail: matsuo@kddilabs.jp

**あらまし** 変動に対する顔認識の適応力を高めるため、別途収集した同一人物の多数の顔画像から現在登録されていない変動を抽出して変動を学習させる取り組みがある。本論文では、これまで人間が目視により手動で行う必要があった変動学習用の顔画像の収集を、テレビ視聴環境上で自動化する方法を提案する。電子テレビ番組表 EPG を利用し、また動画の連続性を利用して、テレビ映像から変動学習用の顔画像を収集する点に特徴がある。連続ドラマ 6 話を対象とした実験の結果、F 値を評価尺度とする顔認識精度は変動学習無しの 0.59 に対して変動学習時は 0.79 を示し、0.20 ポイント上回る変動学習による顔認識精度の改善効果を確認した。

**キーワード** 顔認識, 変動学習, 電子テレビ番組表, デジタルビデオレコーダー, セットトップボックス

## Incremental Learning for Face Recognition on Digital Video Recorder

Kenji MATSUO Masayuki HASHIMOTO and Hisashi KAWAI

KDDI R&D Laboratories Inc., 2-1-15 Ohara, Fujimino-shi, Saitama 356-8502, Japan

E-mail: matsuo@kddilabs.jp

**Abstract** In this paper, authors propose a novel incremental learning method by which face templates are trained incrementally for television receiving and viewing systems, such as digital video recorders and set-top-boxes. The proposed method has two advanced components. First, cast information in EPG: Electronic Program Guide is used to make face identification reliable. Second, sequential relations between video frames are employed to enable accurate extraction even of fluctuated faces appearing on video. Face samples used in incremental training are automatically collected without human inspection. Experiments confirm that the proposed method can train face templates incrementally and improves the identification F-measure of cast performers from 0.59 to 0.79 through 6 episodes of a drama program.

**Keyword** face recognition, incremental learning, electronic program guide, digital video recorder, set-top-box

### 1. まえがき

顔認識はこれまでセキュリティの分野で盛んに研究開発されてきた。現在では、似ている有名人を判定する web エンタテイメントや、テレビ番組に登場する人物を検索するためのユーティリティ機能など、様々な場面で応用され始めている。本論文でも、新しい応用先の一つである、テレビ映像からの顔認識について議論する。これらのいずれの場面でも、検出された顔が誰であるかを認識するためには、あらかじめ認識させたい人物の顔と名前を関連づけて登録する必要がある。

同一人物でも、姿勢・照明・表情等の変動により顔の見た目が大きく変化する。一般的に、正面向きの顔から抽出された特徴量しか登録されていない人物は、正面を向いていない顔に対する認識精度が低下する傾向を示す。向きだけではなく、登録されている顔の特徴量とは別の変動が付加された顔を正しく認識できないといった問題があった。高精度な顔の識別を実現するためには、安定的な環境下で撮影された正面顔だけ

ではなく、姿勢・照明・表情等が異なる様々な顔画像から特徴量を抽出し、変動を含んだ特徴量を登録する必要があった。

登録されている顔の特徴量に効率良く様々な変動の要素を含めるための方法として、変動学習が挙げられる。変動学習は、別途収集した同一人物の多数の顔画像を学習に用い、学習用の顔画像から、現在登録されている顔の特徴量には含まれていない変動の要素を抽出して、付け加える方法である。変動学習により特徴量を強化する取り組みはこれまでもいくつかあった。例えば、Hall らは、追加された特徴量だけを使って、主成分分析を再度行うことなく、特徴量の分布する空間を張る固有ベクトルを補正する、IPCA 法を提案した[1]。また、Ozawa らは Incremental Principal Component Analysis: IPCA を顔認識に適用するための具体的な方法を提案し、実際に認識性能が改善されることを確認した[2]。さらに、Kasabov らは、数種類のクラスタに分類されて登録されている特徴量に対し、クラスター

ングを再度行うことなく、新たに収集した顔画像の所属する最適なクラスタを判定して、クラスタの構成を更新する、Evolving Clustering Method: ECMを提案した[3]。また、Liuらは、動画画像からの顔認識にECMを適用するための具体的な方法を提案し、実際に認識性能が改善されることを確認した[4]。

上述の通り、従来から変動学習アルゴリズムは多数提案されてきた。しかしながら、これらの従来法は、認識精度に対する一定の効果を示したものの、実際に変動学習に使用する学習用画像を収集する具体的な場面および効率の良い手順については一切触れられていなかった。単純に考えると、人間が目視により逐一確認しながら、学習させたい人物の顔画像を手動で収集することになる。姿勢・照明・表情が様々に異なる顔画像を漏れなく収集するには、多大な手間がかかっていた。特にテレビ映像を対象とした顔認識を行う場合、テレビに出演する可能性がある人物が対象となる。日本タレント名鑑だけでも俳優、モデル、アーティスト、キャスター等の総計15,000人におよぶ人物が登録されており、その全員の学習用画像を収集するには膨大な時間を要し、現実的ではなかった。

このように、変動学習のための学習用画像を効率良く収集する方法が求められている。この要求を満たすことを目的とし、本論文では変動学習のための学習用画像を効率良く提供する方法を提案する。

## 2. テレビ視聴環境を利用した提案変動学習法

変動学習には次の2つの要求条件を満たした顔画像を用いる必要がある。

- ・顔画像に正しい人物名が付与されていること
- ・まだ登録されていない変動が付加されていること

前節でも述べた通り、変動学習用の顔画像を収集する作業には膨大な手間がかかっていた。これまでは人間が目視により誰の顔であるかを識別することで前者の要求を満たし、大量の顔画像を手動で収集することで後者の要求を満たしていたためである。そこで、顔認識技術を適用し、テレビ視聴環境を活用することによって、収集作業を自動化する方法を提案する。

テレビ映像には姿勢・照明・表情の変動により見た目に変化した顔が様々に含まれている。また、日々新しいテレビ映像が配信され、様々な人物が出演する。そのため、テレビ視聴環境は、変動学習用の顔画像を収集するための好適な環境である。学習用画像の収集と変動学習を再帰的に繰り返すことによって、最終的には、顔認識の精度を日々向上させる効果が期待できる。想定しているテレビ視聴環境を図1に示す。いわゆるハードディスクレコーダやセットトップボックスSTB等を使った、映像だけでなく文字情報を取り扱うテレビ視聴環境である。放送局からはケーブルまたは放送波を介して最新の番組映像が各家庭に送信されると共に、電子テレビ番組表EPGも毎日送信される。電子テレビ番組表には、番組の放送スケジュールの他に、ジャンル、番組内容、出演者等の詳細情報が番組単位で記載されている。図の例ではEPGの情報から、チャンネル6の19:00～19:57にMr. Challenge!というタイトルの番組が放送され、David CruiseとTom Bauerというテレビタレントが出演することが読み取れる。

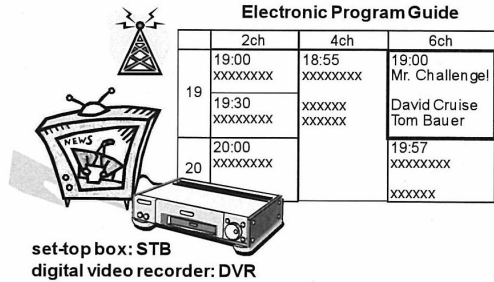


図1. テレビ視聴環境

このようなテレビ視聴環境において収集作業の自動化を実現するために、本提案法はEPGを活用した点に1つ目の特徴がある。例えば上述の例では、EPGに記載された情報からMr. Challenge!というタイトルの番組にはDavid CruiseとTom Bauerが出演していることがわかる。したがって、顔認識技術を適用することによって、同番組のテレビ映像からDavid CruiseとTom Bauerの顔画像を抽出できる可能性がある。ここで十分高い性能を有する顔認識技術を適用すれば、顔画像に正しい人物名を自動的に付与することができる。このことから、テレビ映像とそれに付随するEPGの出演者情報を利用することによって、変動学習用の顔画像に求められる1番目の要求を満たせると考えられる。

しかしながら、単純な顔認識技術では、そもそもその時点で特徴量として登録されていない未知の変動を含んだ顔画像を正しく認識できないため、テレビ映像から新しい変動の要素を含んだ顔画像を収集することはできない。そこで本提案法は、以下に概要を示す通り、動画画像の連続性を利用し、テレビ映像から未知の変動を含む顔画像を収集する点に2つ目の特徴がある。動画画像の連続性を考慮した変動学習用の顔画像の収集の概念図を図2に示す。オクルージョンやショットチェンジが発生しない限り、前のフレームに写っていた顔が次のフレームで突然消えることは無く、同じ位置もしくは近い位置に連続して写っていると考えられる。したがって、検出された顔のフレーム間での人物同一性を判定することで、比較的簡単に同一人物が連続して出現する顔画像群のまとまりを抽出できる。顔画像群のまとまり単位で誰であるかを識別することにより、未知の変動が付加されているため単純な顔認識技術では識別できなかった顔画像も、誰であるかを推定することが可能となる。フレーム単位で単純に顔認識を適用しただけでは識別できなかった顔画像にはまだ登録されていない未知の変動が付加されている。これらの顔画像を変動学習に用いることで、変動学習用の顔画像に求められる2番目の要求を満たせると考えられる。

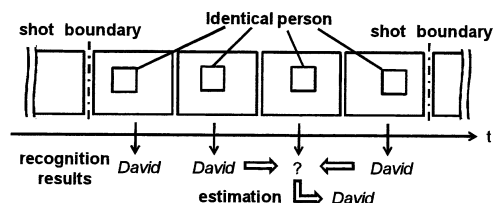


図2. テレビ映像からの同一人物

### 3. 提案法の詳細

#### 3.1. 距離と類似度による顔追跡

動画の場合、前後フレームの検出結果を利用して、突発的に生じる誤検出を訂正できると考えられる[5][6]。そこで図3に概要を示す通り、前のフレームにおいて既に検出された顔との連続性を保つことによって、次のフレームの顔を安定的に追跡抽出する方法を提案する。また、代表顔を抽出するためには、第一にショット内の同一人物の顔を途切れることなく連続的に抽出することが重要であり、提案顔検出法はこれを実現することを目的とした顔検出技術でもある。

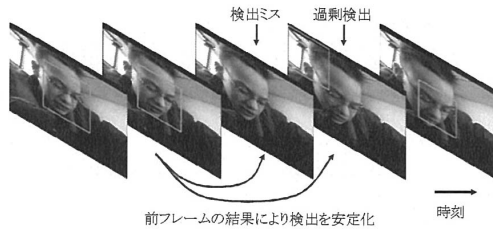


図3. テレビ映像からの顔検出の課題

提案顔検出法の処理手順を図4左に示す。まず「フレーム単位顔検出」で一フレームずつ顔検出を行う。この時点では検出された顔が正しい顔であるか誤検出であるかがまだ不明な状態とし、ここで検出された顔は候補として留める。後段の「距離による対応付け」および「類似度による対応付け」により、フレーム間での連続性から検出された顔の候補を正しい顔として確定または誤検出として棄却する。「距離による対応付け」は過剰な誤検出を解決し、また複数の顔の検出に対応させるための処理であり、「類似度による対応付け」は「フレーム単位顔検出」における汎用性の高い顔検出技術だけでは検出できなかった顔を見つけ出すための処理である。

図4右の検出例と共に各処理の詳細を説明する。説明簡便化のため、ショットチェンジを含まない映像が入力されるものと仮定する。例では図4(a)に示す通り、前フレームでは2名、現フレームでは3名の顔が出現した場合であり、図4(b)に示す通り、前フレームまでの処理で2つの顔  $i=1, 2$  が正しく検出されたと仮定して説明を進める。

##### ・フレーム読み込み・フレーム単位顔検出

画像を1フレーム読み込み、複数の顔に対応した顔検出技術により、候補となる顔を検出する。例では図4(b)に示す通り、現フレームにおいて3つの顔  $j=1\sim 3$  が候補として検出されたとする。

##### ・距離による対応付け

前フレームで既に検出された顔と現処理中のフレームで検出された顔候補との距離を総当りで測定し、フレーム間で同じ位置に存在する顔との対応を調べる。例では図4(c)の  $i=1$  と  $j=1$  が同一人物の顔として対応付けられる。また、本処理により対応が付き残った現フレームの顔候補は、過剰な誤検出領域または新たな人物の出現である可能性が高いと判断し、次のフレームでの判定に委ねる。例では同図の顔候補  $j=2, 3$  が

これに相当する。

##### ・類似度による対応付け

距離による処理で対応が付きなかつた前フレームにおける顔は、「フレーム単位顔検出」だけでは現フレームで顔が検出できなかった可能性が高い。そのため、図4(d)に示す通り、前フレームで検出された顔  $i=2$  をテンプレートとして用い、現フレームから類似度がしきい値を上回る領域を同一人物の顔として検出する。

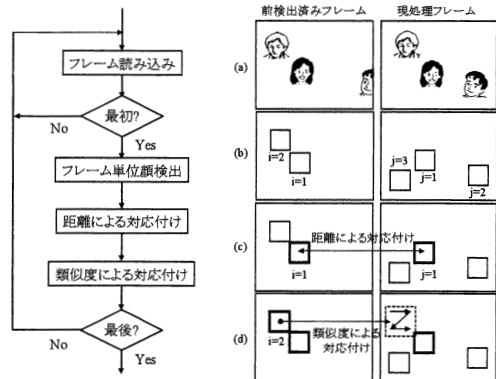


図4. 顔追跡の流れ図と処理結果の一例

ここで、提案顔追跡が必要とする3つのパラメータの詳細について以下に述べる。

##### i). 距離による対応付けにおける判定しきい値

フレーム間で人物が画面の端から端まで急激に移動することは無いため、顔として想定される範囲の移動距離をカバーする値にしきい値を設定する。例えば、ドラマ映像は様々な動きを含むが、30fpsの映像に対しては、前フレームで検出された顔領域の一边の20%にしきい値を設定した場合に、予備実験から経験的に良好な結果が得られた。ただし、画面内の人物の移動距離はフレームレートに依存するため、フレームレートによりしきい値の調整を必要がある。

##### ii). 類似度による対応付けにおける探索領域の制限

テンプレートによる類似領域の探索領域は、現フレームの画像全体ではなく、テンプレートとして使用する顔の位置する近傍に限定する。これにより誤検出を減らし、かつ、計算量を減らすことができるようになる。テンプレートとして使用する顔領域の一边の±20%の近傍領域に探索範囲を制限した場合に、予備実験から経験的に良好な結果が得られた。ただし、距離による対応付けにおける判定しきい値と同様に、フレームレートを加味し、顔として想定される範囲の移動距離をカバーする値に探索領域を設定する必要がある。

##### iii). 類似度による対応付けにおける適用回数制限

予備実験により特に動きが無く静止している映像に対しては、雑音等の影響により突発的に出現する顔の類似パターンが誤って検出されることを確認した。このとき、類似度による対応付け処理により検出され続け、誤検出が伝播され続ける恐れがある。したがって、類似度による対応付けに適用回数の制限を設け、10回以上連続しないようにした。

### 3.2. 全自動変動学習アルゴリズム

前節で述べたテレビ視聴環境において、変動学習を行うための手順を詳細に説明する。図5に提案法のフローチャートを示した。

- ・顔検出

テレビ映像の中に出現する顔をフレーム単位で検出する。検出された顔は矩形で特定される。

- ・顔追跡

フレーム間で検出された顔の同一性を判定し、顔を追跡する。同一人物の顔が連続して出現する区間の情報が得られる。

- ・特徴量変換

検出された顔の切り出し位置と大きさをフレーム単位で正規化し、特徴量に変換する。本論文では縦  $H$  画素横  $W$  画素に正規化された顔画像の輝度信号を Local Binary Pattern[7]に変換した。

- ・尤度算出

特徴量 DB に登録されている特徴量との尤度をフレーム単位で算出する。

- ・判定

尤度算出によりフレーム単位で算出された尤度と、顔追跡により得られた同一人物の顔が連続して出現する区間の情報に基づき、検出された顔が誰であるかを識別する。変動学習用の顔画像を決定すると同時に、顔の識別結果は出演者検索のためのメタ情報となる。

- ・変動学習

収集された変動学習用の顔画像  $N$  枚と特徴量 DB に現在登録されている  $M$  個の特徴量から、クラスタリングにより未知の変動を抽出し、 $K$  個の特徴量を特徴量 DB に再登録する。具体的には、まず  $N+M$  個の特徴量に対し、尤度の高いもの同士を二分木により縮退させて  $K$  個のクラスタを作成する[8]。各クラスタを代表する特徴量を1つずつ、計  $K$  個の特徴量を抽出し、変動学習後の特徴量として登録する。

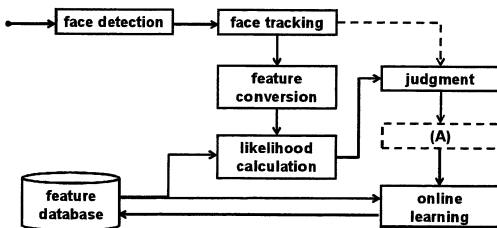


図5. 提案法のフローチャート

### 3.3. 半自動変動学習

ここで、変動学習による効果を予備実験により確認したところ、LBP を特徴量とする顔の識別性能には依然として改良の余地があり、誤識別を生じる場合が見受けられた。この場合、ここまでで述べた変動学習用の顔画像を完全に自動で収集する方法では、誤って識別された特徴量が登録され、変動学習を繰り返すたびに誤りが伝播する恐れがあった。そこで、誤りの伝播を防ぐため、図5の(A)に示す位置に、人間の目視確認を行う手順を追加し、完全自動ではなく半自動学習法を新たに考案した。変動学習後に生成された新しい  $K$  個の特徴量に付随する顔画像を提示し、人間が目視に

よる正誤確認を行う。この結果、候補の中で正しいと選択された顔画像だけが特徴量として登録される。考案した半自動学習法は適用する顔認識技術の性能が人間とほぼ同じレベルを達した場合の完全自動学習法に相当すると考えられる。

### 3.4. フィードバック型半自動変動学習

さらに、半自動変動学習に関する予備実験を繰り返す中で、登録されている特徴量の中でも、特に誤識別を起し易い特徴量が存在することが判明した。このような特徴量は変動学習を進めても、常に誤識別を起こす可能性が高いため、次の手順で排除することを試みる。すなわち、人間が目視で行う正誤判定する際に、誤っていると判定された顔画像を誤識別させた特徴量を特定し、特徴量 DB から排除するようにした。

ここまでで述べた全自動、半自動、フィードバック型半自動学習時の特徴量の選択の様子を図6に示す。図では、登録されている  $M=3$  個の特徴量を使って、テレビ映像からの  $N=6$  枚の顔画像が認識される。クラスタリングにより  $K=6$  個の特徴量が抽出され、全自動変動学習法ではこれらが次段以降の特徴量として特徴量 DB に登録される。その後、人間の目視により誤識別した顔を  $K_n=1$  枚排除し、その誤識別が生じる理由となった  $K_f=1$  個の特徴量を削除する。最終的に半自動変動学習法では  $K-K_n=5$  個の特徴量が、またフィードバック型半自動学習法では  $K-K_n-K_f=4$  個の特徴量が選択され、次段以降の顔認識に使用する特徴量として登録される。

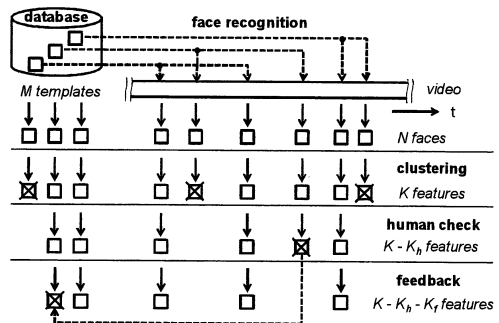


図6. 変動学習時の特徴量選択の様子

## 4. 実験

### 4.1. 顔追跡の効果

フレーム単位の顔検出技術として OpenCV[9]の Haar-like 特徴と Adaboost 識別器を用いた顔検出法を使用した。テスト映像として、ドラマ、バラエティ、音楽番組を含む、地上波アナログテレビ映像の MPEG ファイル 6 本を用いた。各番組は幅 320 画素、高さ 240 画素で約 800 フレーム前後の映像である。OpenCV のみを適用する顔検出法を従来法とし、比較により提案法の有効性を評価する。

総合的な有効性を客観値により比較評価する。フレーム単位で顔の存在領域を目視により判別し、その正解領域をどの程度精度良く検出できているかを測定した。結果を Recall と Precision で表1に示す。共に値が

1に近い程良好な検出法と言える。全ての映像に関し、提案法は従来法を上回り、平均で Recall は約 0.146 ポイント、Precision は約 0.003 ポイント改善された。Precision の改善から「距離による対応付け」により顔ではない領域が誤って検出される誤検出を排除できることが確認できる。また Recall の改善から「類似度による対応付け」において最新の顔を反映させたテンプレートを用いることにより、「フレーム単位顔検出」における汎用性の高い顔検出技術だけでは検出できなかった顔の検出が可能となることが確認できる。以上から、テレビ映像から安定的な顔追跡が達成できたことを確認することができた。

表 1. 顔追跡の精度

	OpenCV			提案法(括弧内改善度)		
	Recall	Precision	F-Measure	Recall	Precision	F-Measure
1 ドラマ	0.509	0.996	0.674	0.623 (0.114)	0.997 (0.001)	0.767 (0.093)
2 バラエティ	0.644	0.999	0.783	0.728 (0.083)	1.000 (0.001)	0.842 (0.059)
3 バラエティ	0.624	1.000	0.768	0.809 (0.185)	1.000 (0)	0.894 (0.126)
4 バラエティ	0.547	1.000	0.708	0.785 (0.238)	1.000 (0)	0.880 (0.172)
5 音楽	0.528	0.987	0.688	0.711 (0.183)	1.000 (0.013)	0.831 (0.143)
6 バラエティ	0.759	0.999	0.863	0.834 (0.075)	1.000 (0.001)	0.910 (0.047)
平均	0.602	0.997	0.747	0.748 (0.146)	0.999 (0.003)	0.854 (0.107)

#### 4.2. 変動学習の効果

提案した変動学習法の有効性を確認するために、計算機上でのシミュレーションを行った。下記の a~d の 4 方式のそれぞれについて認識精度を Recall, Precision で測定して、比較評価を行う。

- a. 変動学習無し (初期特徴量を使い続けて顔識別)
- b. 全自動学習 (全自動の変動学習を適用)
- c. 半自動学習 (目視確認を取り入れた変動学習)
- d. フィードバック (フィードバック型半自動学習)

使用したテスト画像は、日本の放送で毎朝 15 分ずつ放送されているテレビドラマから、連続して放映された 6 話分のテレビ映像を使用した。EPG に記載された出演者は全 3 名であり、ここでは便宜上、女性 A, 女性 B, 男性 C とした。Web から収集した画像を初期登録特徴量した。顔の正規化サイズは  $W=32, H=40$  とした。

女性 A に関して、測定した Recall, Precision の値から F 値を算出し、その推移で図 7 に示した。参考までに測定した Recall および Precision の詳細を表 2 に示す。F 値は値が大きいくほど認識精度が優れ、変動学習の効果も高いことを意味する。図 7 より、a~d の 4 つの方式共に、1 話目の顔認識には Web から収集した同一の登録特徴量を用いたため認識精度に違いは無い。2 話目以降、変動学習無し a は他の変動学習を行う方式 b~d と比較して全体的に低い認識精度を示す。変動学習を行う方式 b~d では、1 話目に対する認識結果から正しく人物が識別された顔画像が収集され、新しい変動要素が学集され、2 話目では良好な認識精度を示すことが確認できる。一方、全自動学習 b は、しかしながら、2 話目に対する認識結果から収集した顔画像に誤識別が発生して、3 話目の認識精度が低下し、それ以降も誤りが伝播し認識精度が大幅に低下している。これに対し半自動学習 c は 2 話目で発生した誤識別結果を目視により排除することができるため、3 話目以降も高い値を示すことがわかる。

最終的に 6 話目までの出演者 3 名の F 値の平均を比

較すると、変動学習無し a が 0.59, 全自動学習 b が 0.42, 半自動学習 c が 0.74, フィードバック型の半自動学習 d が 0.79 を示した。半自動学習 c とフィードバック型の半自動学習 d では目視による一手間が必要となるが、それぞれ変動学習無し a から 0.15, 0.20 ポイントも高い、良好な認識精度を達成することが確認できる。全自動学習 b の F 値は変動学習無し a の場合を下回ったが、使用する顔認識技術が人間と同レベルの高精度な性能を達成するようになれば、フィードバック型の半自動学習 d と同程度の認識性能を示すことができる。したがって、実際の場面では、最初の何日目まではフィードバック型の半自動学習 d で変動を学習させて、認識性能を向上させてから、全自動学習 b に切り替える、という運用法が有効であると考えられる。

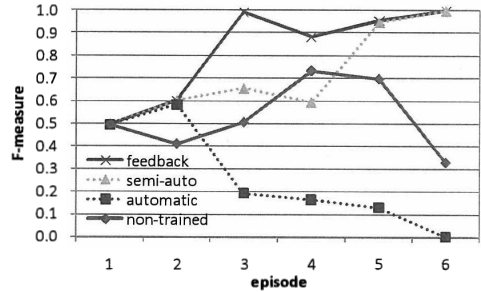


図 7. 各方式の示す F 値の推移

表 2. 認識精度と変動学習の効果

		episode	1	2	3	4	5	6	Average F-measure	
woman A	recall	non-trained	0.34	0.26	0.39	0.58	0.54	0.20	0.20	non-trained
		automatic	0.34	<b>0.41</b>	0.11	0.09	0.07	0.00	0.00	0.53
		semi-auto	0.34	<b>0.43</b>	<b>0.49</b>	<b>0.42</b>	<b>0.90</b>	<b>1.00</b>	<b>1.00</b>	automatic
		feedback	0.34	<b>0.43</b>	<b>0.99</b>	<b>0.82</b>	<b>0.99</b>	<b>1.00</b>	<b>1.00</b>	0.26
	precision	non-trained	0.92	0.93	0.73	1.00	0.99	1.00	1.00	semi-auto
		automatic	0.92	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	<b>0.72</b>
		semi-auto	0.92	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	feedback
		feedback	0.92	<b>1.00</b>	<b>0.99</b>	<b>0.96</b>	<b>0.92</b>	<b>1.00</b>	<b>1.00</b>	<b>0.82</b>
	recall	non-trained	1.00	0.98	1.00	0.93	0.99	1.00	1.00	non-trained
		automatic	1.00	0.86	1.00	0.93	0.84	1.00	1.00	0.46
		semi-auto	1.00	0.86	1.00	0.93	0.95	1.00	1.00	automatic
		feedback	1.00	0.86	0.98	0.36	0.85	1.00	1.00	<b>0.56</b>
precision	non-trained	0.26	0.45	0.41	0.14	0.65	0.06	0.06	semi-auto	
	automatic	0.26	<b>0.71</b>	<b>0.77</b>	<b>0.11</b>	<b>0.67</b>	<b>0.23</b>	<b>0.23</b>	<b>0.87</b>	
	semi-auto	0.26	<b>0.89</b>	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	feedback	
	feedback	0.26	<b>0.89</b>	<b>1.00</b>	<b>0.61</b>	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	<b>0.81</b>	
recall	non-trained	0.92	0.75	0.70	0.98	1.00	1.00	1.00	non-trained	
	automatic	0.92	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	0.79	
	semi-auto	0.92	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	automatic	
	feedback	0.92	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	<b>0.87</b>	
precision	non-trained	0.92	0.85	0.99	0.65	1.00	0.28	0.28	semi-auto	
	automatic	0.92	0.34	0.41	0.20	0.12	0.04	0.04	0.64	
	semi-auto	0.92	0.29	0.51	0.17	0.40	<b>0.96</b>	<b>0.96</b>	feedback	
	feedback	0.92	0.29	0.99	0.40	0.67	<b>0.96</b>	<b>0.96</b>	0.78	

Bold values show better performance than non-trained method.

また、図 7 から、フィードバック型半自動学習 d は 3 話目と 4 話目で半自動学習 c を上回る特性を示している。このことから、収集された変動用画像の候補の中で誤認識された顔画像を排除するフィードバックを加えたことにより、誤検出を生じやすい登録特徴量を排除できたことが確認できる。実際にフィードバック型半自動学習 d で行われるフィードバック処理により排除された登録特徴量数  $K_f$  の推移を表 3 に示す。女性 A は 4 話目と 5 話目で 2 ずつ、女性 B も 2 話目と 4 話目で 1 枚ずつ排除されているが、男性 C はフィードバックで排除される顔画像数がある他の 2 人に対して多い。

図7の結果と合わせてみると、女性Aおよび女性Bではフィードバック型半自動学習はF値の改善効果を示すが、男性Cは効果が見られないため、フィードバックが過剰に作用している可能性がある。過剰なフィードバックを抑制する方法は今後の課題である。

表3. 登録特徴量の数の推移

	woman A						woman B						man C					
Episode	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6
$K$	11	18	26	54	76	88	17	18	19	21	31	32	17	39	26	35	26	25
$K_b$	1	0	0	2	3	0	14	2	0	1	0	0	3	16	1	7	1	0
$K_f$	0	0	0	2	2	0	0	1	0	1	0	0	0	3	1	5	1	0

## 5. まとめ

本論文では、変動学習用の顔画像の収集を、テレビ視聴環境上で自動化する方法を提案した。電子テレビ番組表 EPG を利用し、また動画の連続性を利用して、テレビ映像から変動学習用の顔画像を収集する。全自動連続ドラマ 6 話を対象とした実験の結果、F 値を評価尺度とする顔認識精度は変動学習無しの 0.59 に対して変動学習時は 0.79 を示し、0.20 ポイント上回る変動学習による顔認識精度の改善効果を確認した。変動学習がまだ十分進んでいない初期段階では、収集された顔画像の候補を目視により評価する半自動学習が有効であることを確認した。また、TV 映像からの顔検出において突発的に生じる誤検出の問題を解消するために、前のフレームにおいて既に検出された顔との連続性を保つことによって、次のフレームの顔を安定的に追跡抽出する顔検出法を提案した。実験により Recall と Precision 特性を測定し、検出精度の改善効果を確認した。

提案した変動学習を行い、テレビ番組に出演者情報が高精度に付与されることにより、テレビ映像の中から所望の人物が出演する場면을効率的に検索できるようになる。

## Acknowledgements

Sound and Vision video is copyrighted. The Sound and Vision video used in this work is provided solely for research purposes through the TREC Video Information Retrieval Evaluation Project Collection.

## 文献

- [1] Peter Hall and Ralph Martin, "Incremental Eigenanalysis for Classification," Proc. of British Machine Vision Conference, vol. 1, pp. 286-295, 1998.
- [2] Seiichi Ozawa, Soon Lee Toh, and Shigeo Abe, "Incremental learning for online face recognition," Proc. of International conference on Neural Networks, Vol. 5, pp. 3174-3179, 2005.
- [3] Nikola Kasabov, Evolving Connectionist Systems: Methods and Applications in Bioinformatics, Brain Study and Intelligent Machines, Springer-Verlag, 2002.
- [4] Liang Liu, Yunhong Wang and Tieniu Tan, "Online Appearance Model Learning for Video-Based Face Recognition," Proc. of CVPR, pp. 1-7, 2007.
- [5] 奥村, 大綱, 平田, "番組の出演者一覧生成のための顔画像クラスタリング手法", FIT2006, pp.29-30, Sept. 2006

- [6] 西崎, 尾形, 中村, 大田, "会話シーンを対象とした自動撮影・編集システム", 信学論, Vol.J89-D, No.7, pp.1557-1567, 2006
- [7] Shu Liao, Albert C. S. Chung, "Face Recognition by Using Elongated Local Binary Patterns with Average Maximum Distance Gradient Magnitude," Proc. of ACCV, pp.672-679, 2007.
- [8] Anil Jain, Umut Uludag and Arun Ross, "Biometric Template Selection: A Case Study in Fingerprints," Proc. of AVBPA, pp. 335-342, 2003.
- [9] <http://sourceforge.net/projects/opencvlibrary/>