「画像の認識・理解シンポジウム (MIRU'96)」1996 年 7 月

テンプレートマッチングによる動画像からの受講者の顔画像抽出

先山卓朗† 亀田能成† 美濃導彦† 池田克夫†
†京都大学大学院 工学研究科

Abstract: We describe a method of extracting front facial images in a classroom. For robust extraction, we use an image sequence because lecture attendants do not always look ahead. By taking a scene of a classroom whose desks and seats are fixed in location, using a fixed camera, we can determine one's facial position and its size in the image when he is seated. We use template matching using a general facial image as a template and the facial position in the image as a search area. We obtain an extraction rate of 70%.

1 はじめに

近年の計算機性能の向上や光ケーブルに代表される高速通信網の普及に伴い，音声や静止画像以外に高品質の動画像を実時間で遠隔地に送信することが可能となりつつある．ネットワーク接続された複数の大学間でインタラクティブに応答を返す遠隔講義システムの実験が行われており，環境が整備されるにつれて遠隔講義はさらに盛んになると考えられている．

遠隔講義においては，講師と受講者は視覚的にはカメラ映像を通じてのみ結ばれている．受講者は講師一人の様子が確認できれば十分であるが，講師側では受講者全員を把握する必要がある．しかし，カメラの設置される位置や視野の広さの問題から，遠隔講義室全体の情景を一つのカメラで撮影することは難しい．さらに，遠隔講義では受講者は複数の大学に散在することが多く，大学の数に応じてカメラの台数も複数になり得る．この場合，講師は複数のモニタを見て受講者を把握しなければならないが，これは非常に困難である．

そこで本研究では，遠隔講義室内の受講者の把握を計算機上で支援するシステムを構築する．その支援方法として，それぞれの受講者の正面顔を抽出するという手法を採用した．正面顔により，その受講者を識別でき，性別や年齢，その時の様子など非常に多くの情報も得られる．すなわち，正面顔は受講者を把握する際の重要な要素であると考えられる．

本研究で提案する手法では，講義室の情景が処理の対象となる．講義室内では，講師は前方で講義を行い，受講者は着席して講師の話を聞いていると考えられる．講義室のこうした環境を考慮して，講義室前方から室内の情景を撮影した画像を取り込み，受講者の正面顔画像を抽出する．

画像中から人物の顔を抽出する研究は，表情認識や顔を用いた個人識別のための前処理としても注目されており，多重精度のモザイク画像を利用して画像内での顔の位置や大きさを推定する手法 [1] や，一般顔をテンプレートとして集合写真などから顔の領域を抽出する手法 [2] などがある．これらの多くは，静止画像から，画像内の位置や大きさが不明な顔を抽出するものであるが，画像にはほぼ正面を向いた人物が存在することを条件としているもの [3] や，画像中の人物は一人であると仮定したもの [4] も少なくない．

しかし，講義室での受講状況を考えると，

- 各自の手元にテキストや資料がある場合，講師よりも手元の資料を眺めていることが多い．
- 板書をノートに書きとる場合，黒板を見る動作よりもノートに書き込む動作の方が時間がかかる．
- ある受講者が正面を向いていても，その前の受講者に隠れてしまい，前方からは顔を確認できない場合がある．

などの理由から，受講者が常に正面を向いていると期待することはできない．したがって，意図せずに撮影した静止画像から受講者全員の正面顔画像を抽出することは不可能である．ロバストに正面顔画像を抽出するには，動画像を利用して受講者が正面を向くまで観

Facial Image Extraction of Lecture Attendants in an Image Sequence using Template Matching
Takuro Sakiyama†, Yoshinari Kameda†, Michihiko Minoh†, Katsuo Ikeda†
†Kyoto University

察し続け，その受講者の正面顔が画像に現れたときに抽出するようにしなければならない．このとき，処理されずに読み飛ばされる動画像のフレーム数を減らすため，動画像を高速に処理する必要がある．

このような高速処理に適した方法として，本稿ではテンプレートマッチング法を利用する手法を提案する．テンプレートマッチング法は，処理が単純なため，高速に処理することができ，さらに，正面顔のように特定の形状のものの存在を調べるという処理内容にも適している．特に本研究では，テンプレートマッチング処理専用のハードウェアを利用することにより，リアルタイムで動画像のマッチング処理を行っている．

2　テンプレートマッチング法の利用

講義室内の受講者の正面顔を抽出するための手法として，本研究では，テンプレートマッチング法を利用する．

2.1　テンプレートマッチング法の有効性

本研究では，処理の対象を講義室内の情景に限定しているため，次のような講義室という環境の特徴に適する処理方法を採用した．

1. 講義室内の机や座席は床に固定されており，講義中，受講者は座席に座っている
2. 受講者は，少なくとも数回は正面を向く
3. 室内には複数の受講者が存在する

1. から，カメラパラメータを固定して講義室の情景を撮影することにより，画像内での受講者の顔の位置や大きさを限定することができる．ある限られた範囲から特定の形状のものを検出するという処理にはテンプレートマッチング法が有効である．2. は，受講者が正面を向く機会を逃さずに顔画像を抽出しなければならないことを意味する．このためには，講義室の情景を観察し続ける必要があるが，テンプレートマッチング法は処理の高速性から動画像処理にも適用できる．また，3. については，一般顔をテンプレートとして使用することにより，複数の受講者に対応する．

以上から，本研究にはテンプレートマッチング法が妥当であると考えられる．特に本研究では，動画像の取り込みとテンプレートマッチング処理専用のハードウェアであるトラッキングビジョン (富士通株式会社) を用いることにより，動画像に対して実時間で画像間の差分和 (相違度) を求めることが可能となった．

2.2　一般顔モデルの作成

講義室内の複数の受講者の顔に対応する方法としては，個人顔のデータベースを用意する手法と，一般顔を用意する手法が考えられる．しかし，個人顔のデータベースを利用しても，テンプレートマッチング法では表情の変化などには対応できないため，登録された顔データがその本人の顔と一致するとは限らない．検出処理も複雑となり高速化に向かないため，本研究では一般顔を用意する手法を採用する．

一般顔のモデルとしては，

- 画像内に複数の受講者が含まれるように講義室の情景を撮影するため，テンプレートに使用する画像は小さい．
- 濃淡画像を直接テンプレートとして扱うため，目・鼻・口などの各部品の位置関係とともにそれらの輝度値情報も重要である．

といったことを考慮し，次のような手法を採用した．

1. 一般顔モデル作成のための原画像として，実験環境と同じ条件で撮影した正面顔画像を用意し，顔の大きさが等しくなるように各画像を正規化する．
2. 顔の位置を合わせて全ての原画像の平均をとる．
3. 2. の影響でぼやけたエッジや輪郭を (1) 式によって強調する [5]．

   2. で作成した画像を F，F を 3×3 マスクで平滑化した画像を $\bar{F}$ とし，一般顔モデル画像を G とする．

$$
\begin{aligned}
\bar{F}(x, y) &= \frac{1}{9} \sum_{i=x-1}^{x+1} \sum_{j=y-1}^{y+1} F(i, j) \\
G(x, y) &= \frac{F(x, y) - \alpha \bar{F}(x, y)}{1 - \alpha} \qquad (1) \\
&\ (0 < \alpha < 1)
\end{aligned}
$$

2.3　テンプレートマッチングのパラメータ指定

トラッキングビジョンでマッチング処理を行うには，パラメータとしてテンプレート画像と探索範囲を指定する必要がある．本研究では，画像内での机や座席の位置は変化しないので，次のように指定した．

テンプレート画像

一般顔モデルをテンプレートとして使用する際，

顔全体をテンプレート画像に指定すると，髪型や輪郭の差異，及び顔回りの背景の影響を受けて相違度が大きくなる傾向がある．そこで，作成した一般顔モデルを画像の奥行きに合わせて縮小し，顔の輪郭内部のみをテンプレート画像として指定する (図 1 の右側)．

探索範囲パラメータ

講義室内を撮影した画像において，その画像内の各座席位置に対し，その席に受講者が座ったときの顔の位置を実験的に調べ，顔が出現する近傍領域をその座席位置に対する探索範囲とする (図 1)．すなわち，それぞれの座席位置と画像上での探索範囲は一対一に対応している．

なお 3.2.2 節の III. で述べるように，この対応関係から，抽出した顔画像も座席位置ごとに管理する．

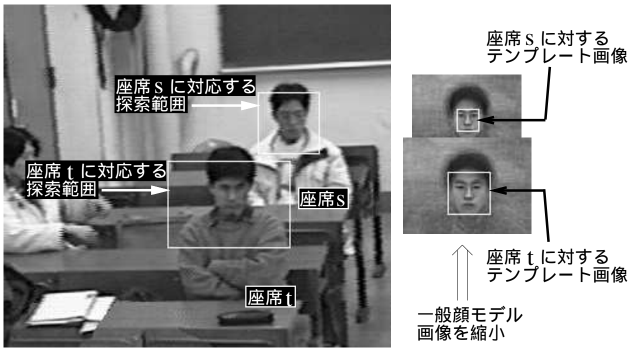

図 1: テンプレート画像と探索範囲の指定

3　顔画像の抽出・除去処理

3.1　処理の概要

本手法では，講義室の情景を撮影した動画像を観察し続け，受講者の正面顔が含まれるフレームが発見されれば，その画像に対して静止画処理を行い，顔画像を抽出する．この処理では，動画像のフレーム間の繋がりは考慮されていない．

この手法では，画像上で探索範囲に指定している部分を受講者が通過する際に，体の一部を抽出してしまう場合がある．このような誤って抽出された画像 (誤抽出画像) は，フレーム間の長期的な変化を調べることによって除去することができる．

そこで，本研究では，動画像の個々のフレームに対する顔画像の抽出処理 (3.2 節) と，フレームの時間的変化を調べる誤抽出画像の除去処理 (3.3 節) の二つの処理を用意する．

3.2　顔画像の抽出処理

濃淡画像に対してテンプレートマッチング処理を行うため，テンプレート画像に類似した領域の他に，木目調の机の表面や単一色の洋服など平均的にテンプレート画像に近い輝度値を持つ領域 (平坦領域) においても相違度が小さくなるという現象がある (図 2)．このため，マッチング処理結果がそのまま顔画像であるとは言えない．

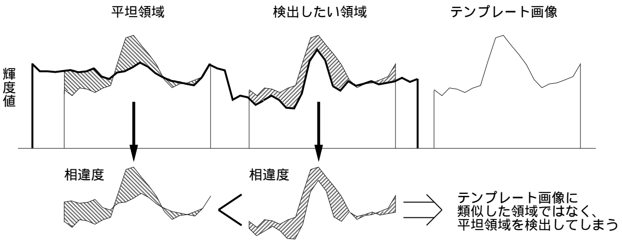

図 2: 平坦領域 (一次元の場合)

そこで，動画像から顔らしい領域 (顔候補領域) を検出する処理と，顔候補領域に本当に正面顔が含まれるかを判定する処理を分離する．ハードウエアマッチングを行う検出処理に比べて判定処理は速度が遅いので，顔候補領域が検出された場合のみ判定処理を行うことで処理を高速化できる．

3.2.1　顔候補領域の検出

顔候補領域の検出にはテンプレートマッチング法を用いる．2 章で述べたように，カメラパラメータを固定して撮影することで，画像内で受講者の顔が存在する位置やその大きさを限定する．

講義室内を前方から撮影した画像内の各座席位置に対して，その席に受講者が座った時の顔の位置を実験的に調べ，その近傍をテンプレートの探索範囲として指定する．テンプレート画像には，一般顔モデルを各座席位置の奥行きに合わせて縮小したものを用意し，その顔の輪郭内部のみを使用する．これを一般顔テンプレートと呼ぶ．さらに，平坦領域のうち時間的に変化しないもの (机の表面など) を本処理段階で除去するため，探索範囲に対応する背景画像を予め用意しておき，これもテンプレート画像として使用する．この背景画像には主に机が写っているので，これを机テンプレートと呼ぶ．

各探索範囲 i ごとに，一般顔テンプレートに対するしきい値変数 T^i_{face} を用意し，初期値として一定値 $FaceDist$ を与える．机テンプレートに対してはしきい値 $DeskDist$ を設定する．なお，3.2.2 節の IV. で

述べるように，しきい値変数 T_{face}^{i} は一旦顔画像が抽出されると次第に小さくなる．

すべての探索範囲において，一般顔テンプレート，及び机テンプレートを用いて並列にマッチング処理を行う．探索範囲 i における最小相違度を，それぞれ Dst_{face}^{i}, Dst_{desk}^{i} とし，以下の手順によって顔候補領域を検出する．

1. $Dst_{desk}^{i} < DeskDist$: その探索範囲は背景成分が占めており，受講者の顔は存在しない．
2. $Dst_{face}^{i} > T_{face}^{i}$: その探索範囲には受講者の顔は存在しない．
3. $Dst_{desk}^{i} \leq Dst_{face}^{i}$: 受講者の顔が存在するかも知れないが，背景成分である可能性の方が大きい．
4. 以上のいずれにもあてはまらない場合 : 最小相違度が Dst_{face}^{i} となった探索範囲 i 内の相違度最小領域には顔が存在する可能性が高いとみなし，顔候補領域として検出する．

すべての探索範囲について検出処理を行い，一枚の動画フレームについて一ヶ所でも顔候補領域が検出されれば顔候補領域の判定処理へ進む．

3.2.2 顔候補領域の判定

以下の処理によって，検出された顔候補領域のうち，平坦領域からなるものを除去し，正面顔を含む領域のみを顔画像として保存する．

I. 頭髪領域による判定

顔候補領域の検出処理では輝度変化の少ない顔の輪郭内部をテンプレートとしているため，誤って平坦領域を検出してしまうことが多い．そこで，顔の輪郭内部とは輝度値の異なる頭髪領域を検出することにより，平坦領域を除去する．

頭髪領域の検出にもテンプレートマッチング法を使用することが考えられるが，髪型は人によって差が大きく，固定テンプレートでは対応不可能なのでハードウェアによる高速化が難しい．

そこで本手法では，顔候補領域の中央上部に取った矩形について，輝度値が髪しきい値 $HairDist$ より小さい画素の数を調べる．その数が一定値以上なら頭髪領域が存在するとみなす．少なければその顔候補領域は平坦領域の一部であると判定し，その領域は除去する．

II. 水平方向輝度均一性による判定

顔候補領域に実際に受講者の顔が含まれる時，肌色領域である顔の内部と背景成分である輪郭の外側とでは輝度値に差があるものと考えられる．一方，平坦領域からなる顔候補領域は，ほぼ一様の輝度値を持つ．

そこで，以下の処理によって水平方向の輝度均一性による判定を行う．

1. 256 階調の輝度値を p 個の輝度領域に分割し，その輝度領域を $r_0, r_1, \ldots, r_{p-1}$ と呼ぶ．
2. 顔候補領域の中央付近の水平方向 m 行のうちの第 k 行について，輝度分布を調べ，輝度値が t 番目の輝度領域 r_t に含まれる画素の個数を L_{kt} とする．さらに，L_{kt} の t に関する最大値を L_k とする $(L_k = \max_{t=0}^{p-1} L_{kt})$．
3. 水平方向の m 行に関する L_k の最大値を L とする $(L = \max_{k=0}^{m-1} L_k)$ ．
4. L がしきい値 $LevelMax$ 以下ならば，その顔候補領域は平坦領域ではないと判定する．

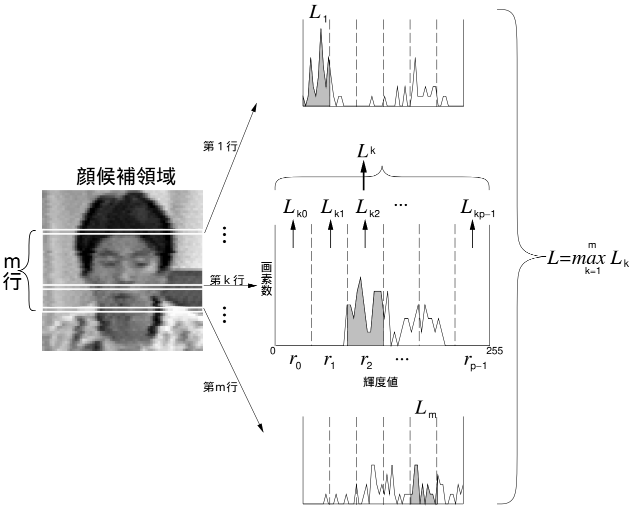

図 3: 水平方向輝度均一性の判定

平坦領域の場合，輝度値の変化が少ないので，2. における L_k が大きくなる．逆に，正面顔が含まれていれば，輪郭の外側に肌色と同系色の領域がない限り，この値はしきい値 $LevelMax$ 以上にはならないと考えられる．

III. 類似顔の判定

3.2.1 節で述べたように，座席位置ごとに探索範囲が指定されているので，探索範囲 i で抽出された顔画像

は，対応する座席位置 i に座っている受講者の正面顔であると考えられる．

しかし，講義室内を撮影するカメラの角度によっては，前後隣接する座席位置に対する探索範囲が重なってしまい，複数の座席位置で同一受講者の顔画像が抽出されることがある．そのため，隣接する座席位置に既に抽出された顔画像(既抽出顔画像)が存在する場合，その画像との類似性を調べておく必要がある．ここでは，次の二種類の判定のどちらかが成立すればその顔候補領域を類似顔と判定する．

- 隣接する座席位置での既抽出顔画像をテンプレートとし，顔候補領域を探索範囲としてマッチング処理する．その結果，最小相違度がしきい値以下の場合，類似顔とする．
- 隣接する座席位置で既抽出顔画像が検出されたときの画像内での座標位置と，顔候補領域の検出座標位置が，近接している場合，類似顔とする．

類似顔と判定された場合，複数の座席位置に同一受講者の顔画像が保存されることになるので，それらは類似顔であるという情報を残しておく．

IV. 顔画像の保存及び顔候補領域の検出しきい値の更新

以上のすべての判定処理において除去されなかった探索範囲 i の顔候補領域は，座席位置 i に座る受講者の顔画像として保存する．

一旦顔画像が抽出されれば，同一座席位置の受講者について何度も抽出を繰り返す必要はない．既抽出顔画像よりさらに最小相違度が小さい顔候補領域が検出されたときのみ判定・保存処理をするようにすれば，顔候補領域の検出 (3.2.1 節) のみで抽出処理が終わる可能性が増え，高速化につながる．そこで，顔画像を保存する際，その領域を顔候補領域として検出したときの一般顔テンプレートとの最小相違度 Dst^i_{face} を，その探索範囲 i に対する新たなしきい値として T^i_{face} を更新する．

3.3 誤抽出画像の除去処理

顔画像の抽出処理 (3.2 節) において, 講義室内を移動する受講者が画像上で探索範囲に指定している部分を通過する際に，体の一部が誤って抽出されることがある．このような，一時的に存在する物体に対する誤抽出画像は，探索範囲のフレーム間の時間的変化を調べれば誤りであると判断することができる．そこで，一定期間ごとに以下の処理を行う．

i. 既抽出顔画像を持つ各座席位置 i について，顔候補領域の検出処理 (3.2.1 節) における一般顔テンプレートによるマッチング処理を行う．相違度最小領域を，Dst^i_{face} としきい値 T^i_{face} との大小関係に関わらず，すべて顔候補領域とする．

ii. 顔候補領域の判定処理 (3.2.2 節) における処理 I. と II. を行う．

iii. ii. の判定処理中に，正面顔を含まないとして顔候補領域が除去された座席位置については，その動画像フレームにおいて受講者が存在しないと判定する．

iv. 一定期間以上受講者が存在しない状態が続けば，その座席位置における既抽出顔画像は誤抽出であるとして取り消し，しきい値 T^i_{face} を $FaceDist$ に再初期化する．

4 実験

本稿で提案する手法に基づいて実際にシステムを構築し，実験を行った．

実験対象とする講義室として，遠隔講義の実験にも使われる本学科内の第一講義室を利用した．室内前方の中央上部に設置されたカメラで，通常の講義が行われている情景を撮影した (図 4) ．

図 4: 動画像の 1 フレーム

図 4 において，2.3 節のパラメータ指定方法に基づいて探索範囲を指定した座席位置を図 5 に示す．図 5 は講義室の座席配置を上から見たものであり，網目の

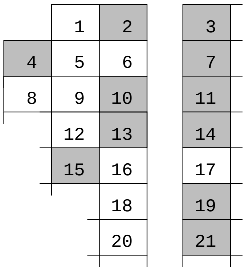

図 5: 処理を行った座席位置

図 6: 処理結果

かかった座席が，図 4 において受講者が座っている座席を表わしている．

3.2.2 節の III. で述べたように，抽出した顔画像はそれぞれ座席位置ごとに管理されている．したがって，処理結果を出力する際も図 5 のように座席位置に対応する一覧形式で出力する (図 6) ．この時点で，既抽出顔画像が類似顔 (3.2.2 節の III.) と判定されている受講者の顔画像は，一般顔テンプレートとの相違度が最小であったものを残して除去される．

図 6 において，座席位置 10 の受講者が座席位置 6 に表示されているのは，処理終了時に類似顔と判定されており，双方の最小相違度から座席位置 10 の画像が取り消されたためである．また，座席位置 19 の受講者は，処理を行った期間の前半に顔画像が抽出されたが，後半は一度も顔を上げていないため，誤抽出画像の除去処理において存在しない受講者として既抽出顔画像を除去されている．

図 7: うまくいかなかった画像

図 8: 処理結果

図 7, 8 は，顔画像の抽出処理がうまくいかなかった例である．この講義では，配布した資料について説明しており，手元の資料に集中して顔を上げない受講者が多かったためである．他に，講義中に OHP を使用した場合，照明条件の変化のために顔画像の抽出ができないことがあった．

こうした実験を 10 回の講義に対して行った結果を表 1 に示す．表の上側の項目は，左から順に，結果として出力されたもの，処理中に一度は抽出されたがその後除去されたもの，それらの合計を表わしている．実験を行った画像中には合計 81 人の受講者がおり，そのうちの約 7 割にあたる 57 人の顔画像を抽出することが可能であった．

表 1: 処理結果のまとめ

	抽出	除去	合計
正面顔	57	8	65/81
誤抽出	4	5	9

5 おわりに

本稿では，講義室内の受講者の顔画像を抽出する方法として，動画像に対してテンプレートマッチング法を使用する手法について提案した．その手法に基づいてテンプレートマッチング処理にハードウェアを利用して実験を行ったところ，約 7 割の受講者の顔画像を得ることができた．

参考文献

[1] 小杉信: “個人識別のための多重ピラミッドを用いたシーン中の顔の探索・位置決め,” 信学論, Vol.J77-D-II, No.4, pp.672–681, 1994.

[2] 桝永慎哉, 長尾智晴: “遺伝的アルゴリズムを用いた静止画像中の人物の顔領域の抽出,” 信学技報, PRU-160, pp.13–18, 1995.

[3] 坂本静生, 田島譲二: “画像からの顔の検出,” 信学秋期全大, D-211, p.214, 1995.

[4] 佐々木努, 赤松茂, 末永康仁: “顔画像認識のための色情報を用いた顔の位置合わせ法,” 信学技報, IE91-2, pp.9-15, 1991.

[5] D E. Knuth: “Difital Halftones by Dot Diffusion,” *ACM Trans. on Graphics*, Vol. 6, No. 4, pp.245–273, 1987.