

I-36 類似顔による誤検出抑制を用いた人物検索システム TV cast search system with false detection suppression using similar faces

斯波 尚志† 井上 晃† 田島 譲二†
Hisashi Shiba Akira Inoue Johji Tajima

1. はじめに

近年、映像データが個人レベルで容易に大量に蓄積される環境が整ってきた。今後は効率的な検索システムがより強く要求されることになるであろう。膨大な映像データに対する検索キーとしては“顔”が最も効果的だと考えられる。そこで筆者らは市販の顔認識ライブラリを用いた人物検索システムを作成した[1]。これは CMU の Name-It [2][3] や Virage 社の VideoLogger [4]等従来の映像アノテーションシステムとは異なり、画像のみを用いた人物検索システムである。横顔のみのシーンを検出できないといった制限はあるものの、未学習データに対して誤検出率約 10%、検出漏れ 20~30% を達成した。

更なる性能向上を目指し、検索対象かどうかを判別するしきい値の最適値を探索したところ、ある値を境に急激に誤検出率が増加する傾向が見られた。実際の運用を考慮すると、しきい値に対する過敏性を軽減する必要がある。誤検出された画像は、特定の人物に偏在していることが分かった。これらは検索システムにとって類似した顔であると言える。この類似した顔をあらかじめ登録しておくことにより、誤検出を抑制できると期待できる。ここでは類似顔による誤検出抑制方法及び、検索システムの性能向上について説明する。

2. 人物検索システム

作成した人物検索システムの画面構成は図 1 のようになっている。入力した画像を表示する部分、顔の検出・照合を示す部分、検索対象を検出したフレームを表示する部分からなる。検出・照合した顔は図 1 のように四角い枠で囲まれる。枠の色で本人かそれ以外かを示す。

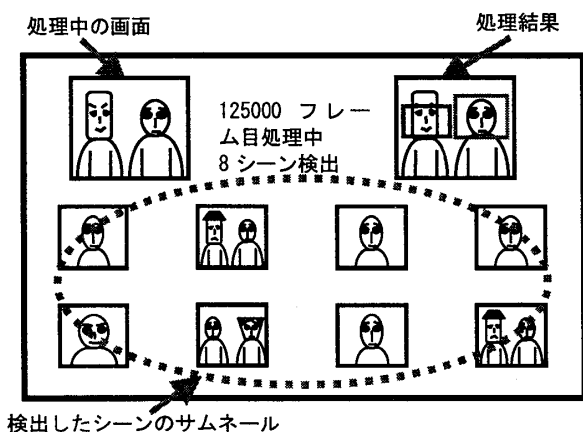


図 1. 人物検索システム

この人物検索システムは、動画像を入力し、6 フレーム毎に顔検出・照合を行い、検索対象人物が含まれているフレームを検出する。顔検出の際に肌色や動きを用いて探索領域を制限することにより、最も処理量の大きい検出部分の処理時間を 1/3 程度にまで抑えている。カラーヒストグラムを用いたカット検出によりシーンを区別し、シーン毎に検出結果を出力する。シーン内に 1 フレームでも検索対象人物が含まれているならば、そのシーンに検索対象人物が含まれていると見なし、代表的なフレームをサムネイル画像として表示・ファイル出力する。サムネイル画像については市販の映像キャプチャソフトで閲覧し、サムネイル画像が見れるシーンへジャンプすることができる。

3. 類似顔による誤検出抑制

まず、学習用映像データから 15 フレーム前後を選択して、検索対象人物の辞書を作成する。次にこの辞書を用いて、学習データに対して、検索システムを適用する。検索の結果、誤検出された人物を類似顔として記録する。類似顔の人物から、登場頻度と類似度を考慮して 4~10 人を選択する。これも一人当たり 15 フレーム前後を登録する。検索対象人物、類似顔人物ともに、選別の煩雑さを軽減するため、番組の開始から 20~30 分以内のフレームを集約的に選択する。なお、ここで誤検出された人物は、検索対象人物と似ていないように見えるものが多く、検索実験を行わずあらかじめ目視だけで類似した人物を選択することは好ましくないと考えられる。

検索時、検出した人物について、検索対象の辞書データに対する信頼度が、類似顔人物の辞書データに対する信頼度よりも大きくかつ、あらかじめ設定したしきい値より大きい場合に、そのフレーム内に検索対象人物が存在するとみなす。

4. 実験

3つのニュース番組(番組 A、B、C)を実験対象とした[1]。ニュース番組各々に対し放送日の異なる映像データを 2 つずつ選び、一方を学習画像選択用とした。番組の長さはそれぞれ 30 分、1 時間 15 分、2 時間である。番組 B からは、2 人の人物を検索した。各々 α 、 β とする。他の 2 つの番組については 1 人だけ検索対象とした。番組 A の検索対象を γ 、番組 C の検索対象を δ とする。各映像データにつき、検索対象の登場シーンの数は 20~40 であった。

検索対象本人かどうかを示す指標である信頼度(最大値 100)に対ししきい値を設定し、60~90 の範囲でしきい値毎の性能の推移を調べた。検索性能については、正解率(=検出したシーンの内対象人物が含まれていたシーンの数/システムが対象人物だと見なして検出したシーンの数)と検出率(=実際に検出できたシーンの数/番組全

† NEC マルチメディア研究所

体における対象人物の登場シーンの総数) を指標として評価した。

辞書作成に用いなかった映像データについて、誤検出抑制を行わない場合の検出率・正解率の推移を図2に示す。しきい値が90の時、検出率が最低(グラフ左上)で、しきい値が小さくなるに従って正解率がほぼ一定のまま、検出率が増加する(グラフ右側に進む)。そしてある点を境に急激に正解率が低下する

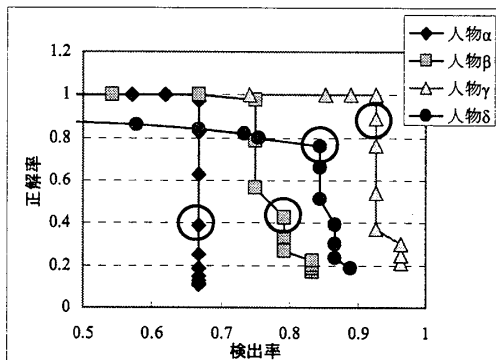


図2. 未学習データに対する検出率・正解率の推移

人物 δ ではしきい値72(グラフ上マルで囲んである点)が最適だと思われるが、このとき人物 α や β では正解率が0.4前後となってしまふ。検索対象毎に異なったしきい値設定を行うとすると、各々最適なしきい値を探索する必要が生じるため、煩雑になる。また、検索対象が同じでも映像データが異なると最適なしきい値も異なると考えられる。従って、どのような場合でも同一のしきい値が設定できることが好ましい。

類似顔による誤検出抑制を行った場合を図3(a)(b)に示す。しきい値72でどの人物の検索においても正解率0.75以上と、誤検出抑制を行わない場合に比較して大幅に向上している。また、しきい値の変化に対する正解率の変動が小さく且つ、この実験のしきい値の範囲で常に0.75以上となっている。なお、ここでは示さなかったが、学習に用いた映像データに対しても同様の傾向を確認した。

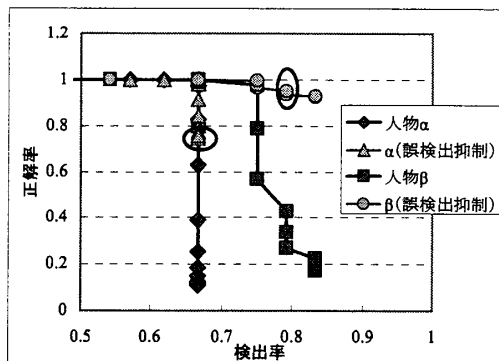


図3(a) 誤検出抑制を行った場合の検出率・正解率の変化(人物 α ・ β)

5. 考察

類似顔登録によって排除されるのは、その当人だけでなく、彼ら/彼女らに類似した人物も排除される。そのため、全ての登場人物を登録する必要は無い。しかしながら、どの映像データに対しても高い検出率・正解率を維持するには登録する類似顔を多くする必要があると考えられる。一方、検索対象本人であるにも関わらず、類似顔との類似性が高く排除される場合もあり、これが検出率向上の妨げとなっている。類似顔の登録数を更に増やすと、更に検索対象が排除される可能性がある。また、顔照合の処理時間は登録した顔の数に依存しており、登録数を大きくすれば、検索時間が大幅に増大すると予測される。従って類似顔登録にあたっては最適な数を慎重に見極める必要がある。

類似顔登録によりしきい値に対する過敏性が軽減され、より実用的になったと思われる。また、誤検出に対して目視による確認等が必要なアプリケーションの場合、ユーザの負担を軽減するには、正解率が高いほうが好ましい。従って、膨大な映像データを扱う場合は、類似顔による誤検出抑制が有効であろう。

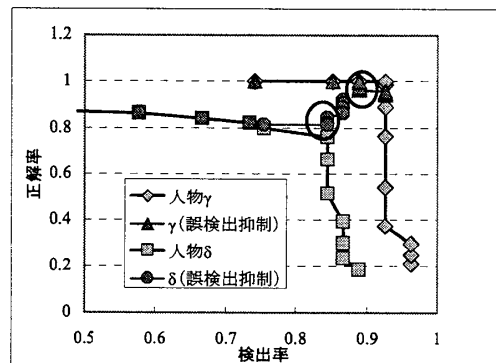


図3(b) 誤検出抑制を行った場合の検出率・正解率の変化(人物 γ ・ δ)

6. おわりに

人物検索システムにおいて、類似顔によって誤検出を抑制する方法を提案した。常に高い正解率を維持し、幅広いしきい値範囲で安定した性能を維持するため、検索システムの可用性が高まった。今後はバラエティ番組等多様なTV番組に適用して性能検証を行う予定である。また、実用化のために、煩雑な作業である検索対象及び類似顔の登録手順を簡略化するツールを作成する予定である。

参考文献

- [1] 斯波、井上、田島、"顔認識に基づくTV番組人物検索システム"、第8回画像センシングシンポジウム、2002
- [2] S. Satoh, Y. Nakamura, T. Kanade, "Name-It: Naming and Detecting Faces in News Videos", IEEE Multimedia, vol. 6, January-March, 1999.
- [3] S. Satoh, Y. Nakamura, T. Kanade, "Name-It: Association of face and name in video", Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 368-373, 1997.
- [4] http://www.virage.com/files/products/VL_DS_lores.pdf