

解像度の違いにロバストなまばたき検出と ベストショット画像の自動生成

尾崎 勇也[†] 今井 順一[†] 金子 正秀[†]

電気通信大学 大学院電気通信学研究科[†]

1 まえがき

人物写真(特に集合写真)においては、何人かが目を閉じた状態で撮影されてしまっていることを良く経験する。筆者らは、ビデオカメラを使用して0.5秒間の動画像を取込み、各フレームにおいてまばたきの状態を調べ、全員が目を開いた画像(ベストショットと呼ぶ)を選択する方法を開発した[1]。しかし、この方法では、標準TV方式のビデオカメラによる動画像入力となるため、画像の解像度を大きく出来ず、記念撮影等には向いていない。また、インタレース画像となっている。

まばたきの開閉判断については、例えば[2]の研究がある。しかし、一人の人物に対する一定距離からの撮影という条件が課せられており、様々な条件で撮影される一般の集合写真への適用には向いていない。これに対し本論文では、デジタルカメラの連写機能により連続画像として撮影した一般の集合写真を対象にする。また、まばたき検出の性能の向上を図る。

2 処理対象画像

表1: 文献[1]との処理対象画像の違い

| | 文献[1] | 本論文 |
|---------|-----------|-------------|
| 入力機器 | ビデオカメラ | デジタルカメラ |
| 画像サイズ | 720 × 480 | 2048 × 1536 |
| インタレース | 有 | 無 |
| フレームレート | 15フレーム/s | 3フレーム/s |
| 処理フレーム数 | 15 | 6 |

文献[1]との処理対象画像の違いを表1に示す。文献[1]に比べ、画素数が約9倍多く、一方、フレームの時間間隔は広がっている。人間のまばたきは、目が閉じ始めてから開ききるまで平均0.2~0.3秒かかり、約5秒間に1回である。本論文では、約0.3秒に1回撮影が行われる。そのため、1回まばたきがあっても目を閉じてしまうフレームは1枚か2枚だけであり、さらに1回まばたきをすると次は約5秒後になると考えられるので残りのフレームにおいてまばたきがある可能性は低いと考えられる。

Blink Detection Method Robust for Variations in Spatial Resolution and Automatic Generation of Best-shot Image.

Yuya OZAKI[†], Jun-ichi IMAI[†], and Masahide KANEKO[†]

Graduate School of Electro-Communications, The University of Electro-Communications[†]

1-5-1, Chofugaoka, Chofu-shi, Tokyo 182-8585, Japan
{ozaki, imai, kaneko}@radish.ee.uec.ac.jp

3 顔領域及び目領域の検出

まず、入力画像中から人の顔領域を検出する必要がある。顔検出については、様々な手法があるが、本研究では、オムロン社製のOKAOVisionによる顔領域検出機能を用いた。

3.1 初期フレームにおける目領域検出

検出した顔領域に対して目領域の検出を行う。初期フレームにおいて、図1に示す顔の寸法比を使って目領域を抜き出す。この顔の寸法比は経験的に決定したものである。

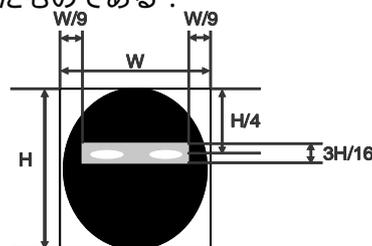


図1: 顔の寸法比に基づく目領域検出

3.2 2フレーム目以降の目領域検出

初期フレームで検出した目領域をテンプレートとしたテンプレートマッチングによって、6フレームすべてにおいて目領域の検出を行う。図2に3.1において検出した目領域を示す。この目領域の画像をテンプレートとして、他のフレームで検出した目領域を図3に示す。

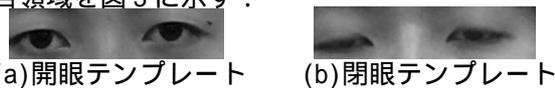


図2: 初期フレームの目領域

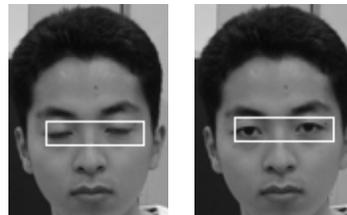


図3(a): テンプレートマッチング結果

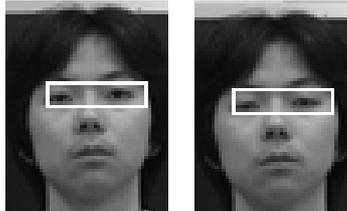


図3(b): テンプレートマッチング結果

図2(a)は開眼のテンプレートで、図2(b)は閉眼のテンプレートである。いずれの場合をテンプレ

ートにしても図 3 に示すように目の位置を正確に検出可能である。

4 まばたきの開閉判断

図 4：300 人分の平均顔の目領域

まばたきの開閉判断には、図 4 に示す 300 人分の平均顔の開眼目領域を使用する。図 4 の目領域のサイズは、人間が目で見てもまばたきの開閉が正確に判断できる限界として 30×6 画素とする。このように小さい画像で計算することで、大きな画像から小さい画像まで広い範囲の画像においてまばたき検出が可能となる。まばたきの開閉判断には、平均顔の目領域と入力画像との類似度を利用する。最も類似度の高い位置を調べるため、ある程度自由度を持たせて走査する必要があるため、入力画像は顔横幅が 35 画素となるように縮小する。3 によって検出した目領域の上端から顔検出時の縦幅の半分までの範囲の中で、式(1)で計算される類似度が最も高い位置を求める。

$$R(x, y) = \frac{\sum_{x', y'} [T(x', y') \cdot I(x+x', y+y')]^2}{\sqrt{\sum_{x', y'} T(x', y')^2 \cdot \sum_{x', y'} I(x+x', y+y')^2}} \quad (1)$$

R は類似度、 T は平均顔の画素値、 I は対象画像の画素値、 (x, y) は入力画像の参照座標値、 (x', y') は平均目領域画像の参照座標値である。正規化相関を行っているため明るさに影響されずロバーストに検出が可能となる。図 5 にまばたきの開閉判断の結果を示す。



図 5(a) 閉眼を含む画像 図 5(b) 全員開眼の画像

5 開眼画像合成

多人数になればなるほど 6 フレームの画像の中に全員がまばたきをしていない画像がない可能性が高くなる。その場合は、開眼の人物が最も多い画像において、閉眼の人物だけ他のフレームから開眼の画像を探してその開眼画像を閉眼の部分に合成する。

5.1 合成する目領域の抜き出し

合成する際は、片目ずつ合成するため、目領域の検出には閉眼の片目ずつの平均の目領域を使用し、最も類似度の高い座標値とその検出した画像を記憶する。テンプレートとして用いる閉眼の片目ずつの画像を図 6 に示す。図 6 は各々 10×4 画素の極小のサイズである。対象の画像も、まばたきの開閉判断の際に使用した顔の横幅を 35 画素に縮小した画像を使用する。

図 6：平均顔から作成した閉眼の目領域

目単独の領域が図 6 のように小さくなると、はっきりとした目ではなくなってしまふ。そのため、開眼の画像よりも白目のないまつ毛だけの閉眼の画像のほうが目の候補として一致し易くなると考えられる。この方法により、入力画像から正確に目領域を抜き出すことが可能になる。目領域を抜き出して合成する際、各目の横幅の間隔と縦幅の間隔の距離を固定して考える。目の間隔を固定して合成しないと、万が一目の間隔がずれた場合に違和感のある画像となってしまうためである。

5.2 目領域の合成

目領域合成の位置については、5.1 で抜き出した目領域の座標を基本とする。5.1 では類似度の最も高い位置を検出しているため、各フレームにおいてすべて完全に同じ位置を検出するのではなく、微妙なズレが生じる可能性が考えられる。そこで、検出した座標だけではなく、周辺を走査して最も適合する位置を探索する。位置の評価計算方法を式(2)に示す。

$$R(x, y) = \sum_{i, j} \sum_{\text{隣り合う周囲の } x', y'} \left[(T(x', y') - I(x+x', y+y'))^2 + (T(x', y') - I(x+x'+i, y+y'+j))^2 \right] \quad (2)$$

i, j は、合成する開眼画像の周囲画素と隣合う入力画像の画素への座標値のズレ、その他の変数は式(1)と同じである。基本座標位置から周辺±5 画素の範囲を走査し、式(2)によってそれぞれの位置で評価値を計算する。最も評価値の小さかった座標へ開眼画像を合成する。合成した結果を図 7, 8 に示す。合成したことがほとんど分からないぐらい正確に合成できている。

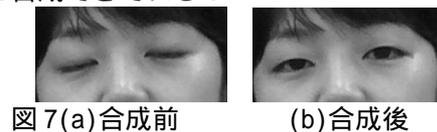


図 7(a) 合成前 (b) 合成後

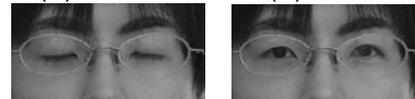


図 8(a) 合成前 (b) 合成後

6 むすび

提案手法により、様々な条件下で撮影された 14 種のサンプルにおいて 1 例(逆光の場合)を除き正しくまばたきの開閉判断まで行うことが出来た。開眼画像の合成については、めがねを掛けている場合においても正確に合成出来た。今後の課題として、入力画像の枚数に関する検討が挙げられる。

参考文献

- [1] 入江, 他: "まばたき検出に基づくベストショットの選択及び加工", 2005 信学総大, D-12-89, 2005.
- [2] 助川, 他: "目つぶり排除機能をもつ顔撮影システム", 信学論 D-, Vol. J84-D-, No.6, pp.1053-1060, 2001.