

焦点ボケ量に基づく人物抽出方式とオブジェクト符号化への適用

5G-05

高田 巡

平池 龍一

NEC インターネットシステム研究所

1. はじめに

IMT2000 がサービスインを迎え、MPEG4 によるテレビ電話機能を備えた新端末「ビジュアルホン」に注目が集まっている。MPEG4 には、映像を任意形状のオブジェクトに分離して符号化するオブジェクト符号化が規定されている。これによって映像から人物領域を抽出して符号化することが可能となり、伝送効率の向上、画質改善が期待できる。

人物領域の抽出には、従来から肌色成分抽出、背景差分、フレーム間差分などが多く用いられている。しかし、これらの手法では、携帯テレビ電話の利用シーン(任意の対象人物、任意の照明条件、移動中や街中など背景の変化が激しい条件)に対応することは困難である。

一方、携帯テレビ電話の映像には「至近距離の撮影対象」と「遠距離の背景」から構成され、焦点は撮影対象に固定されているという特徴がある。この特徴を利用し、合焦領域を抽出することで、注目対象である人物領域を効率的に抽出できる。合焦領域の抽出は、従来にもコンテンツ製作支援のための人物抽出[1]などが提案されているが、携帯テレビ電話に適したものは見られない。

以上の観点から、我々は、焦点ボケ量に基づく携帯テレビ電話向けオブジェクト抽出方式を開発した[2]。また、抽出結果に基づいて擬似的なオブジェクト符号化を行い、符号量の測定を行った。

本稿では、携帯テレビ電話での人物抽出に適したレンズ条件、画像処理の詳細、符号量について述べる。

2. レンズに求められる条件

提案方式では、焦点ボケ量に基づいてオブジェクトと背景の分離を行う。抽出精度を向上させるためには、オブジェクト(人物)では合焦し、背景ではボケが大きくなるレンズを選ぶ必要がある。

映像のボケ量は、①撮像面(CCD)のサイズ、②レンズの焦点距離、③絞りの3つのパラメータから求めることができる[2]。ここでは、携帯テレビ電話の利用シーンおよび後段の画像処理を考慮し、人物の顔が画面に収まること、画像解像度は 320×240 、カメラから 200cm 以上離れた背景で 3 画素以上のボケが生じることを条件として、焦点距離 f および絞り F を求め、以下のレンズパラメータを得た。

$$f: 13\text{mm 以下}, F: 1.5 \text{ 以下}$$

3. 人物抽出処理

オブジェクト抽出の概要を図1に示す。

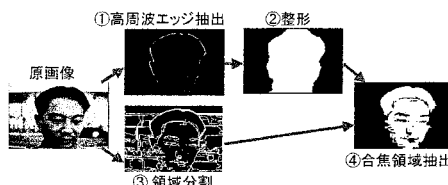


図1 処理の流れ

本方式では、動画像の各フレームにおいて、エッジに基づく合焦領域の推定(図1①②)と色相に基づく領域分割(図1③)を行い、それらの結果を統合して、合焦しているオブジェクトの領域を推定(図1④)する。

3.1 高周波エッジ検出

合焦しているオブジェクトでは、ボケたオブジェクトに比べてエッジ部分の輝度変化が急峻になり、輝度変化の急峻さは周波数領域における高周波成分となって顕れる。以下、このエッジを高周波エッジと呼ぶ。本方式では、合焦しているオブジェクトの輪郭を得るため、周波数解析処理によって高周波エッジを抽出する(図2)。

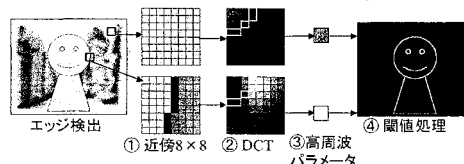


図2 高周波エッジ抽出

- ① 画像全体に 1 次微分フィルタを施し、エッジ強度が一定以上の各画素 (x, y) について近傍 8×8 画素の領域を取り出す
- ② 8×8 画素の領域に離散コサイン変換(DCT)を適用し、 d_{uv} ($0 \leq u < 8, 0 \leq v < 8$) を得る
- ③ 次式から高周波パラメータ $E(x, y)$ を求める

$$E(x, y) = \frac{d_{10}(d_{30} + d_{31}) + d_{01}(d_{03} + d_{13})}{(d_{10} + d_{01})} + d_{22}$$

高周波パラメータは縦／横それぞれの高周波成分について、低周波成分(縦／横方向)の比率で重み付けを行っている。すなわち、縦方向の低周波成分が強い領域では縦方向の高周波成分を大きく評価し、横方向の低周波成分が強い領域では横方向の高周波成分を大きく評価するように式を導出した。

- ④ 閾値処理により $E(x, y) > E_{th}$ となる点 (x, y) を高周波エッジとして検出する

3.2 輪郭整形

高周波エッジとして得られたオブジェクトの輪郭を整形し、輪郭に囲まれたオブジェクト内部の領域を得る(図1②)。3.1の処理で得た高周波エッジは途切れが生じているため、フラクタル法[3]を利用した輪郭整形処理によって、途切れた輪郭線を補間する。

3.3 色相に基づく領域分割

色相成分に基づく領域成長法によって、原画像を一般的な色相値を持つ領域ごとに分割する。

3.4 合焦領域の推定

3.2で抽出した領域と3.3の領域分割の結果から、合焦領域を推定する。具体的には、分割された各部分領域(図1③)が高周波エッジ内部(図1②)に含まれているかどうかを判定し、高周波エッジ内部に含まれている領域を合焦領域として抽出する(図1④)。なお、面積の小さい領域はノイズとして除去する。こうして得られた合焦領域を元に再度フラクタル法による整形を行い、抽出オブジェクトとして出力する。

4. 実験

実撮影データ(CCD:2/3 インチ, レンズ:f12mm, F1.4, 画像解像度:360×240, 1527 フレーム)を利用してオブジェクト抽出の実験を行った(図3)。正しく抽出されたフレーム(図3A)は 707 (46%), 髪などオブジェクトの一部が欠けたフレーム(図3B)が 540 (35%), オブジェクトが大きく欠けるなど抽出に失敗したフレーム(図3C)が 292 (19%)であった。処理時間は 160msec / frame(Pentium III 1GHz)であった。抽出失敗の原因としては以下の点が考えられる。

1. オブジェクトが過分割され、ノイズとして除去されてしまう(図3B)
2. 輝度変化が少ない領域で高周波エッジが十分抽出されない(図3C)

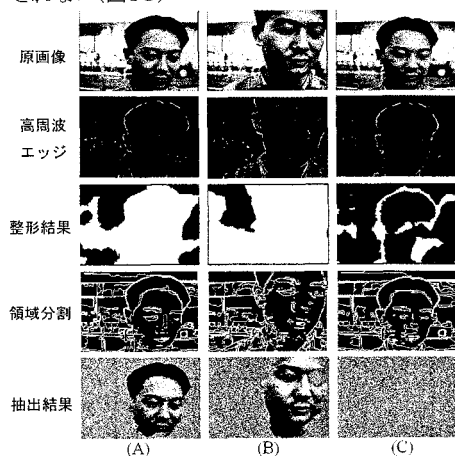


図3 抽出結果

抽出結果に基づいて、背景領域のみ品質を低下させるフィルタを施して符号化し、オブジェクト符号化を擬似的に実現した。使用したフィルタは以下の通りである(図4)。

- A) 背景領域に5×5平滑化フィルタを施して出力する
- B) 人物領域を30フレーム/秒のまま、背景領域を3フレーム/秒として出力する。具体的には、人物領域抽出後、「現フレームの人物領域」または「前フレームの人物領域」に含まれる領域のみ、出力映像を更新する。また10フレーム間隔で画面全体を更新する。

出力映像は、品質設定に関するパラメータを同一にしてMPEG4 符号化した。原画像をそのまま符号化したものと比較して、ファイルサイズがAで35%, Bで42%減少した。

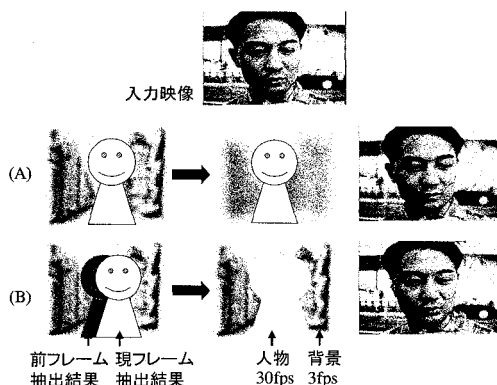


図4 オブジェクト符号化

5. おわりに

本稿では、焦点ボケ量に基づいて合焦オブジェクトを抽出する方式について述べた。高周波エッジ検出と領域分割の結果を統合することで、被写体の顔領域が抽出可能であることが確認できた。また、擬似的なオブジェクト符号化により、符号量の低減を確認した。

現状ではDCTによる周波数解析は輝度情報のみであるが、色情報なども含めて周波数解析をより高精度化する必要がある。また領域分割についても、動き情報などを併用して過分割を抑える必要がある。

参考文献

- [1] 鎌田, 杉浦, “焦点ボケ量推定を用いた単一画像からの3次元空間における領域抽出方法” 信学技報, IE99-126(2000)
- [2] 高田, 里田, 平池, “エッジの周波数特性を用いた携帯動画通信向け人物抽出方式” PCSJ2001, pp.69 (2001)
- [3] 井田, 堀, 三本杉, “自己相似画像を用いたリアルタイムオブジェクト抽出” 第6回画像センシングシンポジウム, 2000.6.14, pp.161-166