

背景差分法に基づく高速背景除去方式に関する一考察

高村誠之[†] 八島由幸[†]

あらまし テレビ電話・遠隔会議システムなどにおいて、符号化の前段として映像シーケンスから高速に背景を除去する方式が望まれる。筆者らが提案した背景除去方式は、ソフトウェア実装にて実時間の10倍以上の速度を実現し、入力画像の自動コントラスト調整や自動焦点調整に対しても頑健であることを確認している。本稿では、本方式の基礎をなす背景差分法その他の構成要素について、工学的観点から考察する。

A Note on Fast Background Removal Based on Background Subtraction

Seishi Takamura[†] and Yoshiyuki Yashima[†]

Abstract: For the realization of TV-phones or teleconferencing codec systems, background image removal is of the fundamental user demands. In all-software implementation of such systems, as video coding part is usually the most time-consuming process, this process, i.e. background removal, must be processed as quickly as possible. We have developed 'super real-time' robust background removal algorithm that can process more than ten times faster than real-time. While achieving high speed, it was shown via computer simulation that the algorithm was robust enough with background image change caused by camera disturbances. In this paper, we investigate on the algorithm's component techniques, as well as background substitution technique as its core, from engineering point of view.

1 はじめに

テレビ電話システム、遠隔会議システムなどにおけるユーザの基本的要求の一つに「映像からの背景除去」がある。放送局のように青背景の前で撮影するといった方法はシステムの自由度の低下やユーザの設置負担の増加を招くため好ましくない。加えて、複数フレームをバッファリングし精度を高めるような試みも、遅延を増大させる原因となるため好ましくない。さらに、処理量の大きい画像符号化処理に最も多くの時間を割り当てるためには、実時間よりはるかに短い時間「超実時間」で背景除去処理が完了することが望ましい。

また、テレビ電話システム、遠隔会議システムにおいて、カメラ側の AF(Auto-Focus) 機能により

画像のボケ具合が位置・時間により変動したり、AGC(Automatic Gain Control), AOC(Automatic Offset Control) 機能により画像のコントラストが変動したりといった(画像処理側にとっての)擾乱がありえ、これらへの対処が同時に必要となる。

従来よりオブジェクト抽出・背景除去などに広く用いられる方法として「背景差分法」がある。これはあらかじめ取得しておいた背景画像と入力画像の差分に基づきマスクを生成し、オブジェクト抽出・背景除去を行うものである。本方式もこの方法を基本としている。

背景差分法に基づく従来の研究として、例えば、照明条件の変化に頑健な松山らの方法[1]は、収束までの繰返し処理を含むため超実時間処理には向かない。

実時間処理の提案として、背景を動的更新する高藤らの方法[2]や背景マスク誤抽出領域除去に距離変換を用いる橋本らの方法[3]があるが、両者とも画像処理専用ハードウェア上での実装であり、ソフト

[†] 日本電信電話株式会社 NTT サイバースペース研究所
〒239-0847 横須賀市光の丘 1-1 Y-517A
NTT Cyberspace Laboratories, NTT Corporation
Phone: 0468-59-2371, Fax: 0468-59-2829
E-mail: {takamura.seishi, yashima.yoshiyuki}@lab.ntt.co.jp

ウェアによる超実時間処理を指向したものではない。

これらの要求にかんがみ、筆者らは頑健で高速・低遅延を満たす背景除去技術を開発した [4]。本稿ではその要素技術について、理想的な処理および工学的な側面からの実装を対比させて考察する。

2 提案方式の要素技術

提案方式は背景差分法に基づいており、概略は以下のようである。

- 1) まず背景のみの画像 (背景画像) を入力・記憶する
- 2) 逐次入力される前景込み画像と、背景画像との差分の小さい部分を抽出・二値化し背景領域マスクを作成する
- 3) マスクの誤抽出領域を除去し、出力する

本方式では以下のような要素技術を用いている (処理順序の詳細は [4] を参照)。

- 1) 画像の縮小
- 2) 画像平均化
- 3) フレーム間差分検出
- 4) AGC/AOC 補償
- 5) AF 補償
- 6) 信号値雑音に起因する誤抽出領域除去
- 7) 前景・背景の色類似および背景の微小変化に起因する誤抽出領域除去

なお PC(Pentium-4 1.4 GHz) 上で ANSI C 言語 (gcc) により本方式を実装した結果、CIF サイズの RGB 画像 (352x288, 24bit/pel) について、一連の処理に要する時間は画像 1 枚あたり 3 ~ 5 ms であった。想定される後段の画像符号化フレームレートは毎秒 10 ~ 30 (33 ~ 100 ms/フレーム) 程度であるため、実時間の 10 倍以上の速度である。

2.1 画像の縮小

映像入力部分で一律に 2x2 ブロック毎の画素値平均によりサイズを縦横それぞれ 1/2 に縮小する。縮小時に折り返し歪みを発生させないためには、sinc 関数の畳み込み・サブサンプリングによる縮小が必要であるが、一般用途カメラからそれほど広帯域な画像信号は入力されないことを仮定しこのような処理とした。

本処理により処理対象画素数が削減され高速化が

期待できる一方、空間方向の若干の頑健さ、すなわち、カメラのサブピクセル単位の微小なブレもある程度まで吸収できるという利点も生じる。

2.2 平均化による背景画像の作成

入力信号に重畳している雑音の主たる要因である、熱雑音に起因する雑音は、フレーム間での相関は無視できる。ここで、撮像装置から連続して得られる背景映像信号 (静止) の、ある画素位置における信号値 (雑音を含む) を k_i (添字 i はフレーム番号) とする。重畳雑音は電力 σ^2 、平均 0 とする。 $k_i = k + n_i$ と考えられる。ここで k はこの位置における真の画素値、 n_i は重畳雑音である。

着目画素位置における N フレームの平均値 $\bar{k}$ を求めると、

$$\bar{k} = \sum_{i=1}^N k_i / N = \left(kN + \sum_{i=1}^N n_i \right) / N = k + \sum_{i=1}^N n_i / N$$

$\bar{k}$ の真値からのずれを n' とすると、

$$n' = \bar{k} - k = \sum_{i=1}^N n_i / N$$

となる。 n' の電力は、 n_i 間の相関が無視できることから σ^2 / N となる。したがって N を大きくとればとるほど $\bar{k}$ は雑音の低減した、真値により近い信号となっていく。

本方式では $N = 16$ としている。この場合 8bit 信号の PSNR にして約 12 dB の画質改善となる。

前景込み画像については、このような一律な平均化はできないため本処理は行わない。しかしながら以下に、前景込み画像での雑音低減ができなくとも、背景画像の雑音が低減していれば、フレーム間差分の雑音が低減することを示す。

前景込み画像の、ある位置の画素値を l_n とする。この画素においては真の画素値 l に電力 σ^2 の n_l の雑音が重畳していると考えられ、 $l_n = l + n_l$ と書くことができる。背景画像の、同一位置の画素値を $k + n_k$ とする。ここで k は真の画素値、 n_k は重畳している雑音である。 n_k の電力を σ_k^2 とする。背景画像の雑音を低減していなければ $\sigma_k^2 = \sigma^2$ であり、上述のような平均処理により低減していれば $\sigma_k^2 = \sigma^2 / N$ である。

両者の差分 (フレーム間差分) を求めると

$$l_n - (k + n_k) = (l + n_l) - (k + n_k) = (l - k) + (n_l - n_k)$$

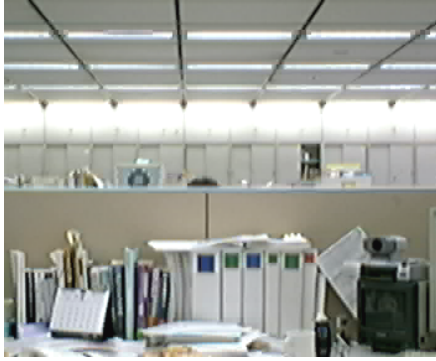


図 1: 平均背景画像



図 2: 入力画像の例

となり、真の差分値 $(l-k)$ に雑音項 $(n_l - n_k)$ が重畳している形となる。 n_l と n_k の相関は無視できるので、該雑音 $(n_l - n_k)$ の電力は $(\sigma^2 + \sigma_k^2)$ となる。したがって σ_k が小さいほどフレーム間差分の雑音は低減されることになる。

図 1 に平均背景画像の例、図 2 に最初の入力画像をそれぞれ示す。

2.3 背景差分の測度

前景と背景のフレーム間差分を求めるにあたり、「差の測度」が必要となる。一般に背景信号と前景信号の間には何ら相関はないため、両者は単にスペクトルが異なるとしか仮定できない。従って、差の測度として(輝度色差のような)人間視覚特性を用いることには意味がない。マルチバンド映像および照明光のスペクトルを用い「分光反射率画像」を求め、この映像に対してフレーム間差分を求めるのが理想的であるが、一般のカメらは3バンド入力であるため、3次元信号値間のユークリッド距離(L2 ノルム測度)が、この仮定のもとで選択できる差の測度として適切かつ現実的である。



L1 ノルム測度



L2 ノルム測度

図 3: 背景差分測度による背景分離の違い

そこで、前景込み画像と背景画像の同じ空間画素位置における RGB 値をそれぞれ (r_f, g_f, b_f) , (r_b, g_b, b_b) としたとき

$$\sqrt{(r_f - r_b)^2 + (g_f - g_b)^2 + (b_f - b_b)^2}$$

が閾値 TH より小さい画素位置を抽出し、その集合を「背景領域マスク」とする。TH の値は実験から定め、30 とした。

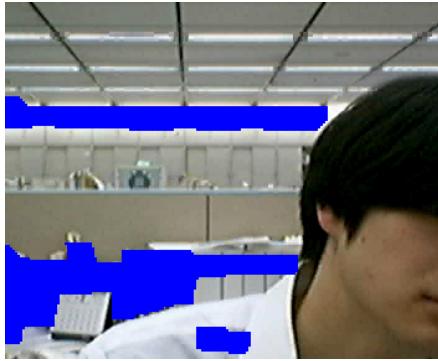
同時に L1 ノルム測度

$$|r_f - r_b| + |g_f - g_b| + |b_f - b_b|$$

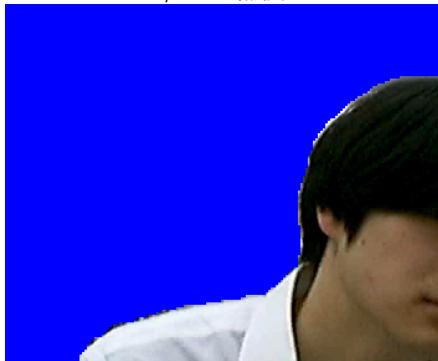
も比較した。両測度ともに閾値を調整した結果を図 3 に示す。L1 ノルムでは背景を前景と誤抽出する領域、ならびに前景を背景と誤抽出する領域がともに増えていることがわかり、背景分離能力は L2 ノルムが優れていることがわかる。

2.4 AGC/AOC 補償

AGC/AOC 機能によるコントラスト変化を補償するために、前景込み画像を f 、背景画像を b とした



AGC/AOC 補償なし



AGC/AOC 補償あり

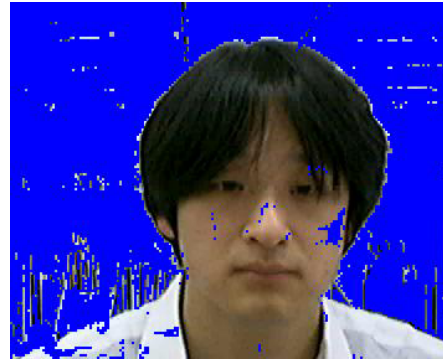
図 4: AGC/AOC 補償の有無によるマスク画像最終結果の違い

とき、 f を 1 次関係式 $gb + o$ により最小 2 乗近似する。なお前景オブジェクトの信号値をこの計算から除外するため、一旦背景領域マスクを仮に求め、そこでの画素値を用いる。 g がゲイン、 o がオフセットの自動制御をそれぞれ補償する。得られた係数により、背景画面全体を $f := gb + o$ と上書きし、カメラの自動制御を補償する。

図 4 上下に、同一シーケンス内で、頭を大きく振って画面内の髪領域の面積が変動しているため AGC/AOC 機能が動作している部分における処理画像を示す。AGC/AOC 補償を行わないと、同図上に示すように、二段階の誤抽出領域除去後であっても、かなりの誤抽出が見られる。これに対し、AGC/AOC 補償を行うことにより、同図下のように頑健な処理が実現できていることがわかる。

2.5 AF 補償

光学系によるボケは、光波の重ね合わせの原理から、光量ゲインが 1 の線形フィルタにより表現でき



AF 補償なし



AF 補償あり

図 5: 誤抽出領域除去前のマスク画像 (黒ベタ部分)

る。フィルタとして、ガウス関数や円筒関数などが一般的であるが、ここでは理想的な帯域制限を施す (位置不変な) ローパスフィルタと仮定する。より強い帯域制限のかかった画像領域に対し、より弱い帯域制限フィルタを施しても画像は不変であるので、ボケによる空間位置ずれを無視すれば、AF 機能により背景画像と入力画像で合焦部分が異なっても、両者に適切な (最もボケている部分の帯域制限とカットオフ周波数が合致した) 同一のローパスフィルタを施すことで、ボケの度合を画面全体で一致させることができると思われる。

本方式では保存した背景および前景込み画像両方に対し、下記係数のローパスフィルタを施す。

$$\frac{1}{16} \begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{pmatrix}$$

図 5 上下に AF 補償を行わない場合と行った場合の修正マスク画像を示す。空間周波数の高い部分 (特に背景の本棚領域) での差分が抑制されていることがわかる。

2.6 マスクの誤抽出領域除去

ここまでの処理で得られた背景領域マスクには、誤抽出領域が残存している。これには以下のような原因が考えられる:

- 1) (特に前景込み画像の) 信号値に重畳している雑音のため
 - 2) 背景が微小に移動・変動するため
 - 3) 前景と背景の信号値が近い値のため
- 1) に起因する誤抽出領域は比較的微細であり、2),3) に起因する誤抽出領域は一般に比較的大きい。

仮に理想的な条件下で完全な背景差分が得られれば、1) の誤抽出は発生しないが、2),3) は背景差分法が本質的に持つ問題であり、常に完全に回避することはできない。

竹島ら [5] は 4x4 画素単位で背景差分をとることで、1) の誤抽出を低減し、別処理により輪郭を抽出することで結果的に 2),3) の誤抽出を低減している。

本方式が対象としている画像の前景は通常人物像である。従って、分離される前景領域は微小なものではなく、ある程度以上の大きさを持つ、連続した領域であることが、ほとんどの場合期待される。そこで「正しい前景領域は比較的大きな連続領域である」という先験的知識に基づき、以下の二段階によりこれらの誤抽出領域を除いている。

2.6.1 第 1 段階: 微小誤抽出領域の除去

まず第一段階として、信号値に重畳した雑音に起因する、比較的微細な誤抽出領域の除去を行う。差分二値化前に高速アルゴリズムの存在するメディアンフィルタ [6] を施すことなども考えられるが、ここでは処理の単純な、二値化後画像に対するモルフォロジフィルタリングを行った。

すなわち、領域を境界において 1 画素広げる「膨張処理」および 1 画素狭める「収縮処理」の組合せにより誤抽出領域を除く。まず

膨張, 膨張, 収縮, 収縮

により背景マスクの微小な穴を埋める。続いて

収縮, 膨張

により前景中で背景と誤認識した部分を除く。

この処理により、小さな穴や孤立領域が除かれると同時に、マスク境界の微小な凹凸が平滑化される。

この処理においては、画素を順に走査し、それが 0 か 1 かの判定が必要なので、着目画素が境界領

域かどうかといった判定は不要である。

2.6.2 第 2 段階: 比較的大きな誤抽出領域の除去

前述の第 1 段階においては、微小な誤抽出領域は効率的に除去できるが、縦横ともに 4 画素を超える比較的大きな誤抽出領域は除き切れない。仮にそのような領域をも第 1 段階の機構で除くためには、膨張・収縮処理の回数を増やすことになるが、過度にマスク境界が平滑化されたり、処理時間が増大してしまうといった問題が生じる。メディアンフィルタを用いる場合もそのカーネルサイズを大きくとらざるをえず、処理速度の低下を招く。

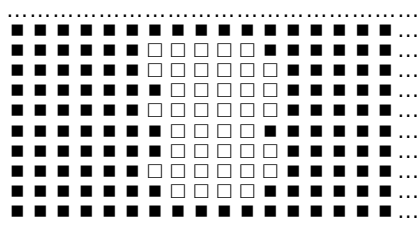
そこで以下に述べる、ラン長に着目した処理により、比較的大きな誤抽出領域を除く。まずマスクを水平に走査し、連続する同一値の画素 (ラン) の長さが閾値 r_{TH} 未満であれば、そのランを両脇の値で埋める処理を行う。

さらに、垂直方向にも同様の処理を行う。この際、走査方向は下から上とし、人物領域は画面下端から始まっていることが多いことに基づき、画面下端から続く前景領域 (非マスク領域) は、ランが短い場合であっても保存する。

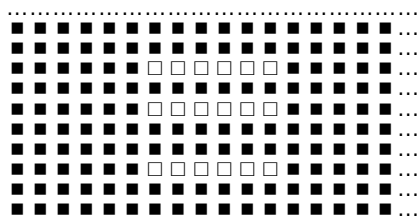
図 6 に具体例を示す。フレーム間差分および二値化閾値処理の結果、同図 (a) のような細長い領域が得られたとする。周囲は「■」で埋めつくされているとする。水平ランの長さを用いた誤抽出領域除去により、同図 (b) のような処理結果が得られる。時おりラン長がラン長閾値を超える部分がある場合、このように 1 ラインのみの水平ランが出現することがある。ここへ、さらに垂直ランの長さを用いた誤抽出領域除去を行うことにより、同図 (c) のように、残存していた誤抽出領域が除かれた結果が得られる。

本処理により、領域境界の凹凸をほぼ残しつつ、比較的大きな誤抽出領域を除去することができる。なお r_{TH} の値は実験から定め、CIF サイズ画像で 7 とした。

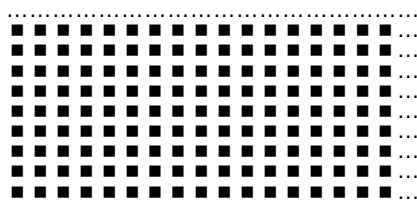
図 7 上下はそれぞれ修正マスク画像から誤抽出領域を除いた画像 (第 1 階および第 2 段階) である。第 1 段階で除ききれなかった誤抽出領域が第 2 段階により効果的に除かれていることがわかる。



(a) 比較的大きな誤抽出領域の例



(b) 水平方向に除去処理を施した結果



(c) 垂直方向に除去処理を施した結果

図 6: ランに基づく領域除去処理 ($r_{TH}=5$ の場合)



第 1 段階誤抽出領域除去後



第 2 段階誤抽出領域除去後 最終結果

図 7: 誤抽出領域除去後のマスク画像 (黒ベタ部分)

3 おわりに

筆者らが提案した背景差分法に基づく超実時間背景抽出方式の要素技術を考察した。本方式が固定的に用いている各種閾値の動的決定方式や、より高速な処理あるいは高品質な誤抽出領域除去アルゴリズムが望まれる一方、本方式が用いている背景差分法の課題(背景と同じような明るさ・色を持った前景部分は除去されてしまう、照明自体が変化する場合は人物の影が背景に含まれるような場合、背景画像が得られない、等の問題)において、処理速度を落とさず誤認識を減らす試みも今後の検討課題である。

参考文献

- [1] 松山 他: “照明変化に頑健な背景差分”, 信学論 D-II, vol. J84, no. 10, pp. 2201–2211, 2001.
- [2] 高藤 他: “空間微分および差分処理を用いた車両抽出法”, 信学論 D-II, vol. J80, no. 11, pp. 2976–2985, 1997.

- [3] 橋本他: “遠隔会議システムのための背景除去処理”, 平 11 信学ソサイエティ大, A-16-15, 1999.
- [4] 高村 他: “映像シーケンスからの超実時間背景除去方式”, 映メ誌, vol. 57, no. 7, pp. 873–877, July 2003
- [5] 竹島 他: “参加者の切り出し画像を実時間で共有するチャットシステム”, 2002 信学総大, SD-3-10, pp.391–392 Mar. 2002
- [6] 浜村 他: “ 3×3 メディアンフィルタの高速アルゴリズム”, FIT 2003, LI-9, Sep. 2002