

# [招待講演]デジカメ画像処理のしくみ

蚊野 浩<sup>†</sup>

<sup>†</sup> 京都産業大学 コンピュータ理工学部 〒603-8555 京都市北区上賀茂本山

E-mail: <sup>†</sup> kano@cse.kyoto-su.ac.jp

**あらまし** 画像処理アルゴリズムとそれを実装する半導体技術が劇的に進歩したことで、画像処理機能がデジカメにおける主要な差別化技術の一つになっている。デモザイクやノイズ除去、階調制御・高ダイナミックレンジ撮影というデジタル写真のための基本機能にとどまらず、顔検出と顔認識およびこれらに基づく付加機能、画像復元方式の手ぶれ補正、子供やペットなどの動被写体を逃さない被写体追跡技術が実現されており、これらの高度化、改良が進んでいる。本講演では、これらデジカメに実用化されている画像処理のしくみを解説し、さらなる技術開発を促進する。

**キーワード** 画像処理, デジタルカメラ, レンズ, 撮像, 画像処理エンジン

## [Invited Lecture] Image Processing Technology for Digital Camera

Hiroshi KANO<sup>†</sup>

<sup>†</sup> Faculty of Computer Science and Engineering, Kyoto Sangyo University

Motoyama, Kamigamo, Kita-ku, Kyoto, 603-8555 Japan

E-mail: <sup>†</sup> kano@cse.kyoto-su.ac.jp

**Abstract** Digital image processing has become a differentiating technology in a digital camera system, because of the recent progress in both of image processing algorithms and the semiconductor technology performing them. Not only basic functions such as demosaicing, noise reduction, tone management and high dynamic range imaging, but also face detection related functions, image stabilization based on image restoration and moving object tracking capabilities are developed and widely used. Engineers are still working to improve them and to create new image processing functions. In this lecture, I talk about the image processing technology used in the digital camera in order to accelerate the development.

**Keyword** Image Processing, Digital Camera, Lens, Imaging, Image Processing Engine

### 1.はじめに

デジカメ元年とされる 1995 年からの 15 年で、デジカメは長足の進歩を遂げた。当時、30 万画素の CCD で、数 MB のフラッシュメモリに 100 枚程度記録できたにすぎなかったが、現在では、1,000 万画素以上の画像を 10GB 以上のフラッシュメモリに数千枚も記録できるのが普通である。

デジカメ以前から放送用カメラや民生用ビデオカメラが存在し、それらに画像処理機能が実用化されていた。これらの機器は、映像をテレビ画面で観察することを目的としたものであり、また当時のデジタル技術の制約から、デジタル画像処理の適用は、階調特性の制御や色信号の補正、ノイズ除去・輪郭補正などの比較的狭い範囲に限られていた。

画像処理がデジカメの差別化技術として確立された背景は、デジカメ用システム LSI (本稿では「画像処理エンジン」と呼ぶ) を「XX エンジン」と固有名詞化する宣伝手法が定着したこと、および顔検出とい

うデジタル処理でなければ不可能な機能が実現されたことにある。顔検出の実用化は 2005 年であるが、その後、笑顔撮影や個人認証、美肌処理などの顔関連技術が実用化された。また、撮影シーンの自動判別、子供やペットなどの動被写体の追跡技術、超解像処理と呼ばれる適応的な鮮鋭感補正技術が実用化され、画像処理エンジンはより充実したものへ進歩した。

### 2.デジカメのしくみ

一眼レフデジカメの構成を図 1 に示す。一眼レフでないものは、ミラーレス一眼とコンパクト型であるが、図から反射鏡とペンタプリズムを取り去った構成になる。デジカメの主たる構成要素はレンズと撮像部、画像処理エンジンである。2 章ではこれらの基本的な機能について述べる。

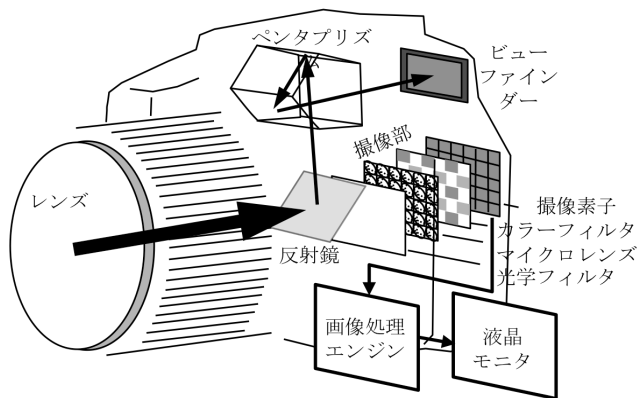


図 1. 一眼レフデジカメの構成

## 2.1. レンズ

レンズは被写体が発する光を集光し、撮像素子上に像を形成する。レンズの基本特性は口径、焦点距離、F 値である。F 値は焦点距離/有効口径で定義される数値である。有効口径が大きくなると光像は明るくなり、焦点距離が長くなると光像は暗くなる。F 値は光像の明るさと対応する。レンズの有効口径を小さくする部品を「絞り」と呼ぶが、F 値を絞り値とも呼ぶ。

焦点距離は光像の大きさとも関係する数値である。焦点距離 20mm のレンズの像は 10mm の像の 2 倍である。しかし、前者の像を面積  $10 \times 10 \text{mm}$  で画素数  $1,000 \times 1,000$  画素の撮像素子でデジタル化したものは、後者を  $5 \times 5 \text{mm}$ ,  $1,000 \times 1,000$  画素の撮像素子でデジタル化した画像と同じになる。これでは不便であるので、カメラ仕様上の焦点距離は、撮像面が 35mm フィルムと同じ大きさに換算した数値を用いる。

デジカメ画像の画素数は大きくなってきた。どこまで大きくなるのか？という問題に対して、レンズ性能から決まる限界と撮像素子の解像度限界から答えることができる。レンズによる限界は MTF (Modulation Transform Function) から解析される。[1]によると、F 値 4.0 のレンズで波長  $546.1 \text{nm}$  の光を結像する場合、 $1.75 \mu\text{m}$  周期の正弦波の MTF は 20% 程度、 $1.0 \mu\text{m}$  周期で MTF がほぼゼロとのことである。一方、撮像素子の微細化の限界も  $1.0 \mu\text{m}$  付近に収束する傾向が見える。従って、民生品での限界は、 $1 \mu$  四方の画素を 35mm フィルムサイズに敷き詰めた  $35,000 \times 24,000$  画素程度になる。これは、とんでもなく大きな画像であるが、コンパクトデジカメの撮像素子は、 $1 \mu$  オーダーのピッチになっており、ほぼ限界である。

画像にどれだけのぼけが発生した時、人間はぼけを感じるであろうか？点光源を撮影したときに許容されるぼけの限界を許容錯乱円と呼ぶが、その直径として、フィルム写真では撮像面对角長の  $1/1000 \sim 1/1500$  が基準として使われていた。これをデジカメに当てはめると、1,000 万画素クラスの撮像素子に対して 3.5 画素、

2,000 万画素クラスの撮像素子に対して 5 画素程度が許容錯乱円の直径になる。これに従うと、デジカメではこの程度のピンぼけ、手ぶれが許容されることになる。

結像に許容錯乱円までのぼけが許されるので、図 2 に示すように、撮像面とレンズの距離に誤差が許容され、これを焦点深度と呼ぶ。これに伴って、被写体側でも正確にピントが合っていると見せる範囲が存在することになり、これを被写界深度と呼ぶ。被写界深度は、①焦点距離が短いほど深い、②口径が小さいほど深い、③物体までの距離が遠い、などの性質がある。同じ F 値のレンズであっても、レンズ口径が大きければ、被写界深度は浅くなる。これによって大口径レンズを用いると、主要被写体だけにピントが合い、その前後の被写体をぼかせる独特な写真撮影が可能になる。

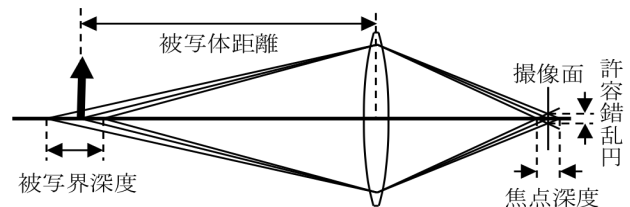


図 2. 焦点深度と被写界深度

## 2.2. 撮像素子

撮像素子はレンズが結像する光像を電気信号に変換する。通常、撮像素子の前部に赤外カットと高周波成分除去のために光学フィルタが配置される。

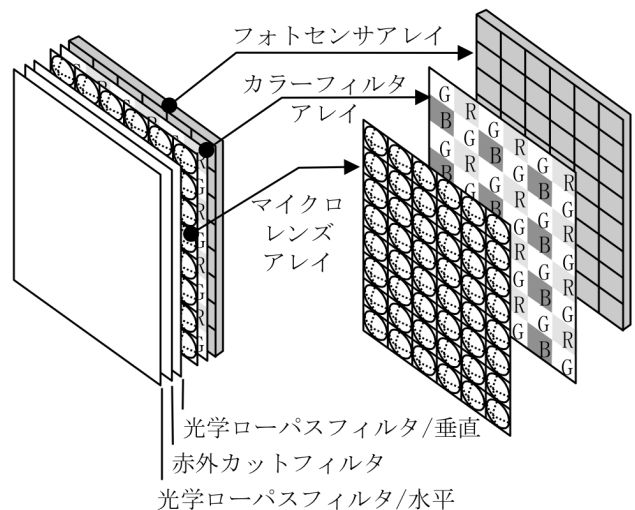


図 3. 単板撮像部の構成

カラー画像を取得する方式に単板式と三板式があり、本稿では単板式について記述する。単板式の撮像部は図 3 のような構成になる。フォトセンサアレイには CCD センサか CMOS センサが用いられる。カラーフィルタアレイは光像を 3 原色に色分解するためのカラーフィルタであり、RGB (Red Green Blue) 個別の微小フィルタ

がモザイク状に配列される。マイクロレンズアレイは、各画素に入射する光束をフォトセルに集光する。フォトセンサアレイは配線層などフォトダイオード以外の層が積層されるため、実際の開口率はかなり低くなる。マイクロレンズによって光を集光することで、実質的な開口率を稼いでいる。光学フィルタは、光学ローパスフィルタと赤外カットフィルタで構成される。光学ローパスフィルタは水晶の複屈折性を利用したデバイスで、入射光の半分を水平および垂直に若干ずらせた光線を生じ、それを入射光に重畳させることで、光像の高周波成分を減衰させる。

撮像素子のダイナミックレンジは、一つのフォトダイオードが出力する最大電荷量と、ノイズとして発生する電荷量の比である。通常、 $20\log(\text{最大電荷量}/\text{ノイズ電荷量})$  で計算され、dB単位で記述される。コンパクトデジカメで用いられる撮像素子で50dB程度、レンズ交換式で用いられる撮像素子で70dB程度であるので、SNは300~3000倍程度である。

### 2.3. 画像処理エンジン

画像処理エンジンは撮像素子から出力される生画像に処理を加え、人間が観察するのに適した画像に変換する。これは図4に示すように組み込みCPUやDSP、顔検出などを行う専用演算回路、プログラムメモリ、データメモリ、画像メモリなどで構成されるシステムLSIである。システムLSIの高集積化・高機能化の恩恵によって、豊富な画像処理機能が付加される傾向が続いており、デジカメの主要な差別化要素の一つになっている。

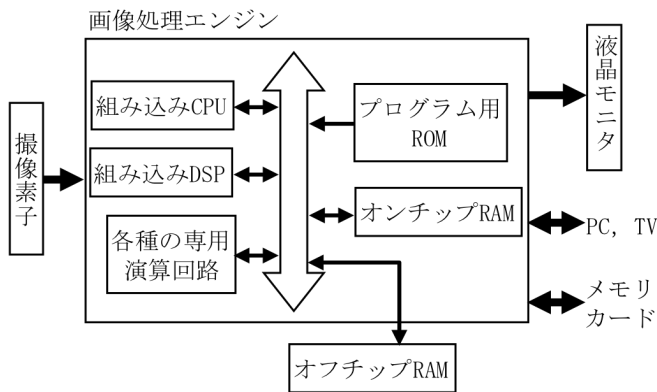


図4. 画像処理エンジンの構成

## 3. デジカメの主要機能

デジカメには、失敗なく美しい写真が撮影できるように多くの機能が搭載されている。ここでは、基本的なものを説明する。

### 3.1. オートフォーカス

オートフォーカスは被写体に自動的に焦点を合わせる機能である。これを実現するには被写体までの距離

情報が必要になる。主に用いられる位相差方式とコントラスト方式について説明する。

位相差方式は一眼レフカメラで標準的に用いられる方式である。図5に示すように、被写体からの光線のなかで撮影レンズの端部付近を通過する光線をセパレータレンズで分離する。撮像素子とは別に設けたAF用センサ上に2つの像ができ、その像の位置の差(位相差)から被写体までの距離を求める。距離を直接求めるため、オートフォーカスが高速に動作する。

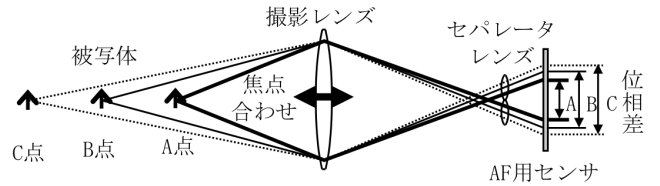


図5. 位相差方式オートフォーカス

コントラスト方式はコンパクトデジカメで標準的に用いられる方式である。図6において、B点に焦点を合わせると最もシャープな画像を撮影することができる。その前後のA点、C点に焦点を合わせると画像がぼける。カメラが焦点を合わせる位置を横軸にとり、撮影される画像のコントラストの評価値を縦軸にとると、図6下のようなグラフになる。この性質を利用してコントラストの評価値が最大になるように焦点位置を制御する。この方式は繰り返し計算が必要になるので、合焦に要する時間が長くなる傾向はあるが、合焦精度は高くなる。

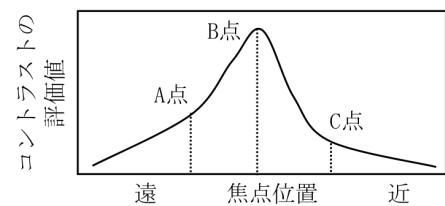
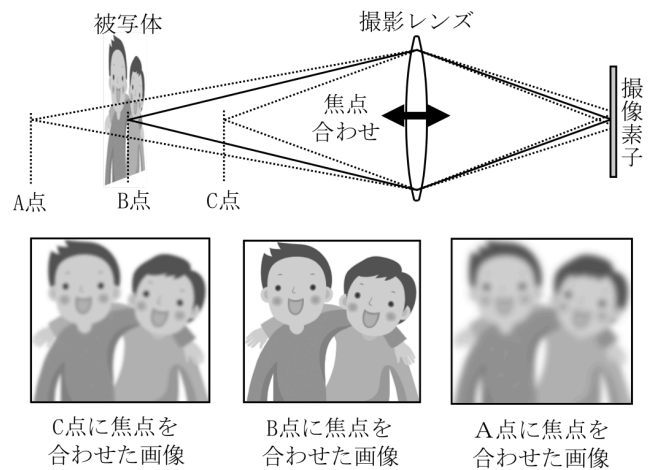


図6. コントラスト方式オートフォーカス

### 3.2. 測光機能

人間の視覚は、明るさと色の恒常性を備えているため、さまざまな環境光のもとでも、物体本来の明るさと色を正しく知覚することができる。それに対して、カメラは光のRGB成分の物理的な強度を記録するため、環境光の影響を強く受ける。人間の視覚を模擬するように光の絶対量を補正する機能を自動露出、色のバランスを補正する機能をオートオートホワイトバランスと呼ぶ。

### 3.3. シャッター機能

シャッターには物理的に光の開閉を行うメカニカルシャッターと、撮像素子内で電荷蓄積モードと廃棄モードを切り替える電子シャッターがある。

電子シャッターでも、CCDはグローバルシャッター、CMOSはローリングシャッターと大きな違いがある。グローバルシャッターは、その言葉の通り、全画素が同じタイミングでシャッター動作をする。ローリングシャッターは、撮像素子の画素ごとにタイミングをずらしながらシャッターを切る。その結果、撮像素子の上部では時間的に早いタイミングで情景を捉え、下部では時間的に遅いタイミングで情景を捉える。

図7にグローバルシャッターとローリングシャッターによる撮影画像の違いを示す。自動車は右から左に移動する場合、グローバルシャッターでは全画素が同じタイミングの情景を撮影するので、自動車の像に歪みを生じない。それに対してローリングシャッターでは、画像の上部と下部で時間的に異なった情景を撮影するので、図のように歪んだ画像を撮影してしまう。このような画像歪みをフォーカルプレーン歪みと呼ぶ。

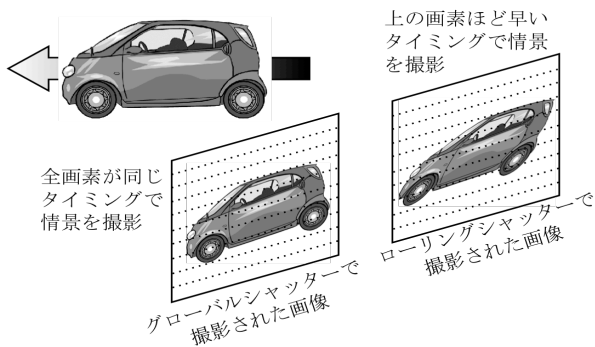


図7. グローバルシャッターとローリングシャッターによる画像の比較

電子シャッターは機構部品が不必要で、制御性に優れるが、いくつかの課題がある。CCDセンサは電荷転送中に撮像面に強い光を受けると、転送路への光の漏れこみが原因となって、スミアと呼ばれる筋状の白とびが発生する。メカシャッターを用いて電荷転送中の光を遮蔽することでスミアを避けることができる。CMOS

のローリングシャッターによるフォーカルプレーン歪みも大きな問題である。この問題を軽減するためにもメカシャッターが利用される。

### 3.4. 手ぶれ補正

撮影時の手ぶれを補正する方式には光学式と電子式がある。光学式はレンズから撮像部までの光学系を機械的に制御することで光像を安定化する。電子式は、ぶれて撮影されてしまった画像や映像を、画像処理によって補正する。電子式については4章で説明し、ここでは、光学式についてふれる。

光学式手ぶれ補正には、図8に示すように、レンズシフト方式とセンサシフト方式がある。レンズシフト方式は、撮影レンズの一部を上下左右に移動させることで、撮像素子に対する光像の位置を安定化させる。撮像素子シフト方式は、撮像素子そのものを上下左右に移動させることで、光像の位置を安定化させる。両方式とも効果的に手ぶれを抑えることができるが、センサシフト方式の特徴は、一眼レフデジカメに適用した場合、交換レンズの選択肢が広がることである。レンズシフト方式はコンパクト型とミラーレス一眼で、一般的に用いられている。

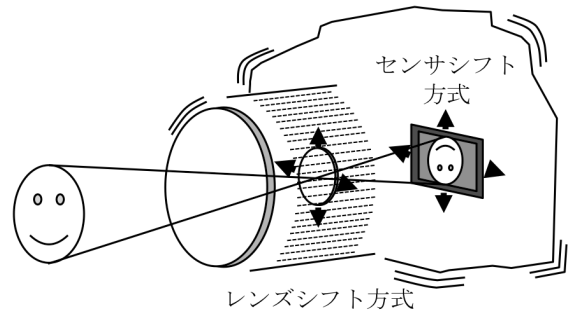


図8. 光学式手ぶれ補正

## 4. 基本的なデジカメ画像処理

画像処理エンジンで実行される処理には図9のようなものがある。生画像に対して一連の画像処理が、順次、実行される様子を画像処理パイプライン、画像処理の全体をデジカメ画像処理と呼ぶ。その中で基本的なものを4章で、高度なものを5章で説明する。

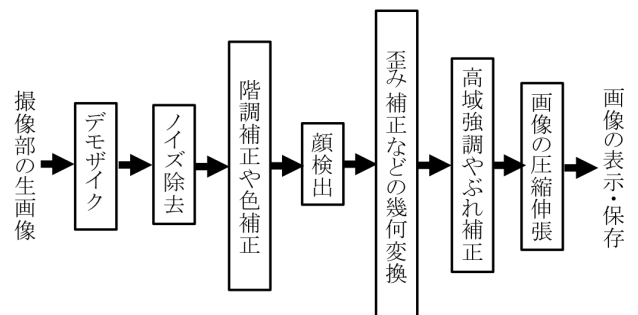


図9. デジカメ画像処理の画像処理パイプライン

#### 4.1.デモザイク

単板撮像素子で撮影される生画像である CFA (Color Filter Array) 画像は、各画素が RGB いずれか一つの値を保持する不完全なカラー画像である。これを RGB フル解像度の画像に変換する処理をデモザイクと呼ぶ。デモザイクは、要するに画像の補間であるから、バイリニアやバイキュービックなどの基本的な補間手法を用いることもできる。しかし、これだけでは細かい絵柄やエッジ部分に、本来存在しなかった色が偽色として発生して好ましくない。これを軽減するための手法が実装されている。代表的なものは、局所的に色信号の割合が一定であると見なす、色相関補間方式である。

図 10 に色相関補間方式によるデモザイク処理を模式的に示す。図中、左側の R, G, B の小画像はベイヤー配列の CFA 画像を各色に分解したものである。まず、一般的な補間処理によって G 成分だけをフル解像度の画像に変換する (図右上)。また、これを平滑化 (LPF 処理) した中間画像を生成する ( $G_{lpf}$ )。R と B 成分についても補間処理と平滑化によって中間画像を生成する ( $R_{lpf}$ ,  $B_{lpf}$ )。R と B の最終画像は、局所的に R/G, および B/G が一定であるという仮定から、図中の式によって計算する。

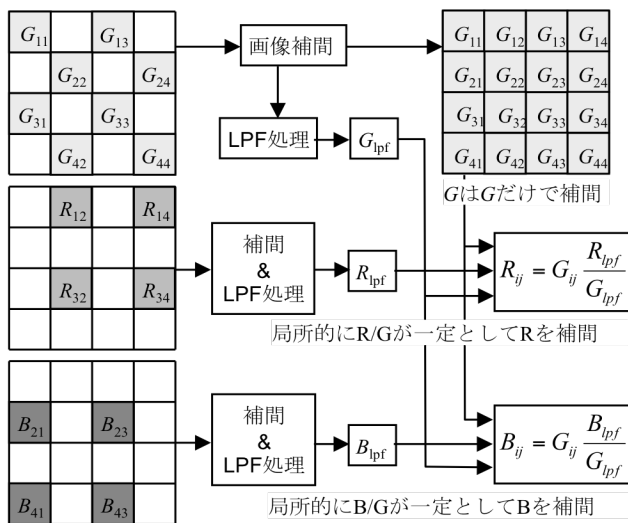


図 10. 色相関補間方式によるデモザイク

#### 4.2.ノイズ除去

画像に表れるノイズは撮像素子の特性によるものや撮像素子表面に付着したごみ、レンズ光学系に起因するものがあるが、ここでは素子特性が原因となるものを考える。その主たるものは、フォトダイオードの暗電流による固定パターンのインパルスノイズ、CMOS ではフォトダイオードに近接した増幅器の  $1/f$  ノイズやランダムテレグラフノイズ、画素をリセットするとき発生するリセットノイズ、および電荷を転

送するときに発生するさまざまな回路的なノイズである[2]。素子特性に起因するノイズは、半導体プロセスの改良と回路的な工夫によって年々、改善されている。しかし、画素数の増加と撮像素子の小型化も並行して進んでおり、フォトダイオードのセルサイズは縮小されている。従って画素あたりのノイズは必ずしも改善されておらず、画像処理によるノイズ除去への期待は大きい。

古典的なノイズ除去手法は平滑化フィルタやメディアフィルタである。本稿では代表的なエッジ保存型平滑化フィルタであるバイラテラルフィルタ (BF) を、1次元信号を用いて説明する。図 11 左上はノイズを含むステップ信号波形である。この波形上の、ステップ直前の注目点に対して BF 処理を施す場合を説明する。まず、注目点からの距離に応じたガウス分布重みと、注目点との値の差に応じたガウス分布重みからなる、2つの重み係数を設定する (図左下)。これを掛け合わせることで、注目点からの距離の差と値の差の両方を考慮した重み係数を決定する (図右下)。この最終的な重み係数で近傍画素を平滑化する。全ての原信号に BF を適用してノイズを除去した信号波形を右上に示す。ガウス型平滑化フィルタと比較して、優れたエッジ保存性があることがわかる。

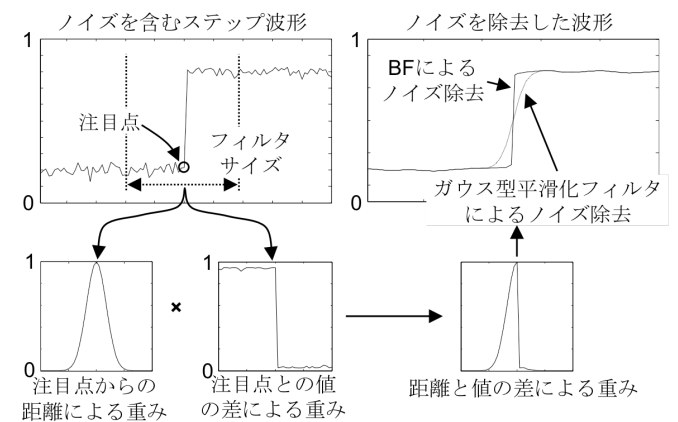


図 11. バイラテラルフィルタによるノイズ除去

#### 4.3.光学的劣化と幾何学的歪みの補正

カメラにはレンズや撮像部に由来するさまざまな特性や劣化、歪みが存在する。球面収差、コマ収差、非点収差、像面収差、歪曲収差はザイデルの5収差と呼ばれるが、これらは光が1点に集光しない現象や、像が幾何学的に歪む現象である。また、光の色によって屈折率が異なるため、軸上色収差と倍率色収差と呼ばれる色のにじみが発生する。画像の周辺は光量が減少するため暗くなる。この現象はシェーディングと呼ばれる。これらの劣化・歪みの中で歪曲収差、倍率色収差、シェーディングが画像処理によって補正される。

#### 4.4.階調と色の補正

撮像素子は、基本的に、被写体の輝度に比例した電荷を出力する。一方、代表的な表示装置であった CRT ディスプレイの画面輝度は入力電圧の約 2.2 乗に比例し、この特性をガンマ特性と呼ぶ。カメラと CRT ディスプレイで構成される入出力システムで忠実な階調再現を行うには、カメラ信号に 1/2.2 乗の特性を持たせることが好都合であり、これを逆ガンマ補正と呼ぶ。液晶など最近のディスプレイ技術は、必ずしも CRT と同じ表示特性ではないが、ディスプレイ装置としてはガンマ特性を持たせ、カメラが出力する画像信号は逆ガンマ特性を有する。

明・暗が混在しダイナミックレンジが広いシーンに対して適切な露光量と階調特性を決定することは重要である。デジカメの自動露光機能では、入力画像を 10×10 程度に分割した領域ごとの平均画素値を求め、その全体的な分布からこれらを決定する。さらに、あるレベル以上の明部の信号を圧縮して階調再現するニー補正が行われる。ニー補正のように、シーンの性質に応じて適応的に階調特性を制御することをトーンマネジメントと呼ぶ。より高度なトーンマネジメント手法は、画像のヒストグラムを解析し、画素値が密集する階調近傍を伸長し、画素値が疎になる階調近傍を圧縮する手法などである。

現実シーンの輝度は図 12 のようになる。従って、撮影されるシーンのダイナミックレンジが 100dB（対数スケールで 5 桁）以上になることもある。このようなシーンに対して、コンパクトデジカメのダイナミックレンジは 50dB、高性能な一眼レフデジカメでも 70dB 程度であるので、カメラのダイナミックレンジが不足する。これを補うために、同じシーンに対して露光量を変化させて複数枚の画像を取得し、それらを合成することでハイダイナミックレンジ（HDR, High Dynamic Range）画像を取得する手法が用いられる。

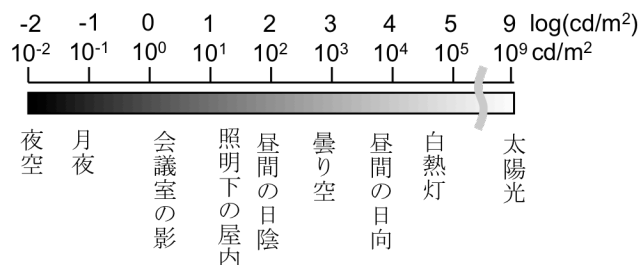


図 12. 現実シーンの輝度

画像合成による HDR 画像の取得は、図 13 左と中央に示すように、露光オーバーとアンダーで複数枚の LDR (Low Dynamic Range) 画像を取得し、それらを正確に位置合わせした後、HDR 画像に合成する。HDR

画像を直接表示できるデバイスは少ないため、通常、トーンマッピングと呼ばれるダイナミックレンジの圧縮が行われる。トーンマッピングにもさまざまな手法があり、その一例による結果画像を図 13 右に示す。



図 13. HDR 画像の取得（北九州大・奥田研提供）

#### 4.5.電子式手ぶれ補正

電子式手ぶれ補正は、ぶれて撮影されてしまった画像を、画像処理によって補正する技術である。動画に対する補正手法と静止画に対する補正手法がある。

高い倍率で撮影した動画や歩きながら撮影した動画は、手ぶれが生じて見難くなる。これを補正する電子式動画手ぶれ補正は、図 14 に示すように、連続するフレーム間の動きを計算し、フレームごとに画像を位置合わせした後、画像の表示枠を再設定する処理である。フレーム間の動きをジャイロセンサで計測することも可能であるが、画像処理による位置合わせを用いることで、これが不要になる。

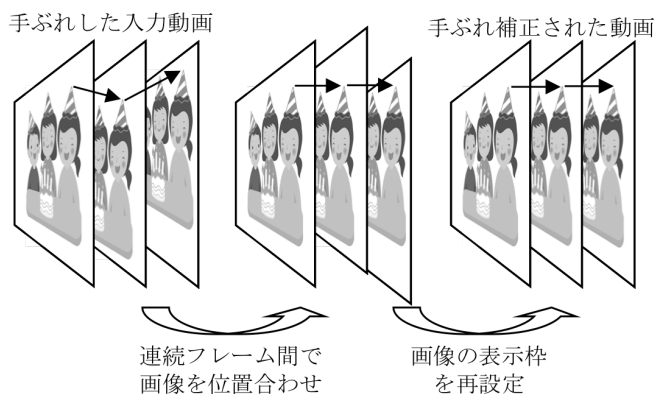


図 14. 電子式動画手ぶれ補正

電子式の静止画手ぶれ補正技術は、レンズや撮像素子を物理的に移動させることなく、静止画に生じる手ぶれを補正する技術である。いくつかの考え方があり、代表的なものを表 1 に示す。加算合成式は高速連射が可能な CMOS センサが必要であるが、処理が比較的単純なため、確実な補正性能を期待できる。画像復元式は、手ぶれの原因である PSF (Point Spread Function, 点広がり関数) を推定した後、ウィナーフィルタなどで画像復元する。演算に畳み込みを含むため、ノイズを増幅する傾向がある。長短露光 2 画像合成式は加算

合成式に近い方法であり、露光時間の異なる2枚の画像を取得し、平坦部は長露光画像を用い、エッジ部は短露光画像を用いるように適応的に画像合成する。

表 1. 代表的な電子式静止画手ぶれ補正

| 方式        | 処理の内容   |
|-----------|---|
| 単純加算合成式   | 通常の1/4程度の短い露光時間で複数枚の画像を連射撮影し、それらを位置合わせて加算合成する。    |
| 画像復元式     | 手ぶれが生じた1枚の画像を、推定したPSFを用いて画像復元する。                  |
| 長短露光画像合成式 | 手ぶれが生じやすい通常露光画像と、手ぶれが生じにくい短露光画像を連射し、それらを適応的に合成する。 |

図 15 に画像復元式静止画手ぶれ補正の処理内容を例示する[3]。画像復元式では PSF の推定が重要な課題である。図 15 の場合、手ぶれ画像と、これとは別に取得した手ぶれの少ない短露光画像の2枚の画像を用いて PSF を推定する。推定された PSF と手ぶれ画像から、ウィナーフィルタなどの画像復元フィルタを用いることで手ぶれが補正された画像を生成する。

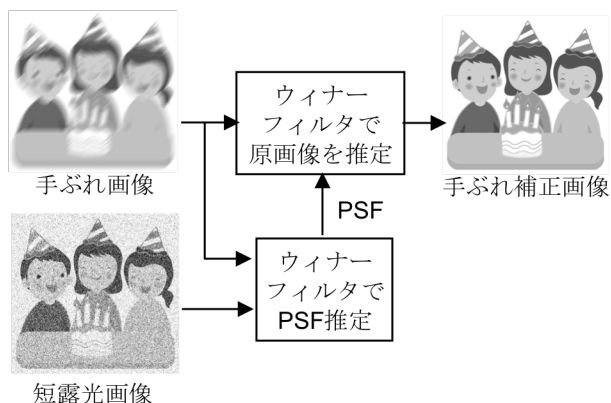


図 15. 画像復元式手ぶれ補正の処理の例

## 5. 高度なデジカメ画像処理

最近のデジカメに実用化されている画像処理機能は非常に高度であり、ますます多種多様になっている。少し数え上げただけでも、顔検出関係、被写体追跡、パノラマ画像合成、シーン分類、超解像、各種のオートフィルタ、3D 画像の取得をあげることができる。本稿では、これらの中で代表的なものである顔画像処理と被写体の追跡を説明する。

### 5.1. 顔画像処理

画像中の正面顔を検出する技術は、Viola[4]らの研究をきっかけに急速に発展した。その手法は、Haar-like 特徴と呼ばれる一種のエッジ検出オペレー

タを用い、顔中の目鼻口など大小さまざまな部位に対応した多数の Haar-like 特徴を、AdaBoost と呼ばれる学習アルゴリズムで、効率的に選択する。さらに、計算を高速に実行するために積分画像を用いるのであるとか、階層的に計算を進めるなどの工夫がなされていた。現在の顔検出技術はアルゴリズム的により高度である。また、正面顔だけでなく、首振り回転や顔を傾けることによる変形、表情変形、人種の差にも頑健で、広く実用化されている。

検出された被写体の顔情報に基づいて、以下のような機能が実現されている[5]。

- ①顔最適撮影：顔オートフォーカスや顔オートアイリス、特定個人優先撮影など。
- ②顔最適タイミング撮影：笑顔撮影、目つむり防止など。
- ③顔最適補正：美肌補正、美白補正、赤目補正、小顔補正など。
- ④顔画像の検索：顔画像による分類、検索など。

### 5.2. 被写体の追跡

画像処理による被写体の追跡とは、指示した対象が、動画の中でどのように移動したかを推定する問題である。簡単な被写体追跡は、テンプレートマッチングや色追跡などの方法で実現されていた。しかし、従来手法は、被写体の移動や姿勢変化による見えの変動に弱く、利用場面が限られていた。それに対して、最近の手法は、コンピュータビジョン技術の成果を取り入れることにより、動作が頑健になっている。本稿では、領域ベースの手法と特徴点ベースの手法について簡単に説明する。

領域ベースの手法は対象に対応した矩形小領域を追跡する。従来手法であるテンプレートマッチングは、対象の変形に非常に弱く、計算量が大きいという問題があった。これに対して、最近の手法はテンプレートマッチングのように画素値を直接の手がかりとするのではなく、カラーヒストグラムを用いることが多い。カラーヒストグラムは、要するに、色信号の度数分布であるが、これによって変形に強くなる。領域間のカラーヒストグラムのマッチングにはヒストグラムインタセクションと Bhattacharyya 係数[6]が用いられる。

領域ベースの手法で、特に重要になるのは、探索手法である、テンプレートマッチングのように全探索が必要なものは計算量が大いだが、それでも、動画像全体として見た場合に、最適解が発見できるとは限らない。領域特徴としてカラーヒストグラムを用い、全探索を効率化する手法にアクティブ探索[7]がある。アクティブ探索によって 100 倍程度高速化される。アクティブ探索は探索範囲内をもれなく探索する。それに対

して、局所的な演算によって移動位置を決定するアルゴリズムとして Mean Shift[8]がある。これはアクティブ探索と比較しても、さらに高速である。

アクティブ探索と Mean Shift は、2フレーム間で移動後の位置を一意に決定する。このため、長い動画像の間に、追跡結果が徐々に正しい結果から離れる場合があり、動画像全体として判断した場合に最適な追跡が行えない場合がある。それに対して、パーティクルフィルタ[9]は各フレームでの追跡を確率的に実行する。追跡結果として可能性がある複数の位置を同時に追跡するので、一時的には可能性が低くなった位置でも、追跡計算を継続し、最終的に正しい結果を得る。

領域ベースの手法は、基本的に、領域の大きさが変化しないものとして追跡を実行する。領域の大小変化を考慮することも可能であるが、探索空間が急激に増大するので、あまり適さない。このような場合、特徴点ベースの手法が有効である。

特徴点の追跡手法として有名なものに KLT 法があるが、この手法は被写体の変形に十分に追従できないとされる[6]。これに代わって、最近用いられる特徴点は SIFT (Scale Invariant Feature Transform)である。SIFT 特徴は、この特徴の抽出アルゴリズムの中に、大きさと回転を不変化するメカニズムが組み込まれている。これによって、被写体の大きさや姿勢の変化にも追従できる性能が、大きく改善される

## 6.まとめ

電子式カメラはテレビジョン技術の中で発展したものであるが、デジカメ時代になって、その枠組みが外れ、結果的に大きく発展した。本稿では、現在のデジカメを高度化するための画像処理技術の一部を説明したが、デジカメ画像処理の全体はもっと膨大なものであり、技術開発の余地が大きい。

デジタル写真技術に対する新しいアプローチとして、Computational Photography と呼ばれる技術が提案されている。これは、デジカメ画像処理をさらに一歩踏み出して、レンズ・撮像部の設計変更までも考慮したデジタル写真技術である。このように、デジカメ画像処理は従来技術が発展しながら新しい技術が導入され、まだまだ進歩する。

## 文 献

- [1] 竹村裕夫, “CCD・CMOS カメラ技術入門,” pp.82, コロナ社, 東京, 2008.
- [2] 高橋秀, “CMOS イメージセンサの低ノイズ化動向,” 映像情報メディア学会誌 vol.62, No.3, pp.303-306, March 2008
- [3] H. Hatanaka, S. Fukumoto, H. Kano and H. Murata, “An Image Stabilization Technology for Digital Still Camera Based on Blind Deconvolution,” IEEE ICCE, Tech. Paper 4.3-1, 2009
- [4] P. Viola and M. J. Jones, “Robust real-time face detection,” Int. J. Comput. Vision, vol. 57, No. 2, pp. 137-154, 2004.
- [5] 山下隆義, 井尻善久, 木下航一, 川出雅人, “シャッターチャンスを逃がさない 顔の検出・表情の認識技術,” 映像情報メディア学会誌, vol.62, no.5, pp.708-713, May 2008.
- [6] 藤吉弘亘, “ねらった被写体を逃がさない 物体追跡技術,” 映像情報メディア学会誌, vol.62, no.6, pp.849-855, June 2008.
- [7] 村瀬洋, V.V. Vinod, “局所色情報を用いた高速物体探索: アクティブ探索法,” 電子情報通信学会論文誌, D-II, vol.J81-D-2, No.9, pp.2035-2042, Sept. 1998.
- [8] D. Comaniciu, V. Ramesh, and P. Meer, “Real-time tracking of non-rigid objects using mean shift,” IEEE Conf. on Computer Vision and Pattern Recognition (CVPR’00), vol. 2, pp. 142-149 vol.2, 2000.
- [9] P. Brasnett, L. Mihaylova, D. Bull, and N., Canagarajah, “Sequential monte carlo tracking by fusing multiple cues in video sequences,” Image and Vision Computing, vol. 25, No. 8, pp. 1217-1227, 2007.
- [10] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” Int. J. Comput. Vision, vol. 60, No. 2, pp. 91-110, 2004.