

Variable interference filter とスペクトル画像エッジを利用した 多波長パノラマ画像の生成

猪狩 壮文[†], Robby T. Tan[†], 池内 克史^{††}

[†]東京大学大学院情報理工学系研究科

^{††}東京大学生産技術研究所

本論文では、variable interference filter を使って多視点で撮影された画像を、エッジ検出を利用して位置合わせを行い統合することで、高精度の多波長パノラマ画像を作成する手法を提案する。Variable interference filter を用いると多波長パノラマ画像を簡単に得ることができるが、シーン上の各点のスペクトル分布を知るためには、多くの視点で得られた画像を位置合わせする必要があるが、異なった波長で撮影された物体同士の位置合わせを行うことは非常に難しい。そこで本論文ではエッジ検出を利用した画像同士での位置合わせにより、正確な位置合わせが実現できることを示す。

A model to Create High Precision Panoramic Multispectral Image: Image Mosaicing Using Edge Detection

Akifumi Ikari[†], Robby T. Tan[†], Katsushi Ikeuchi^{††}

[†]Graduate School of Information Science and Technology, University of Tokyo

^{††}Institute of Industrial Science, University of Tokyo

This report will explain a method of generating a high precision panoramic multispectral image by the method of determining the position relation using edge detection between images with which taken from many viewpoint using variable interference filter. By using variable interference filter, a panoramic image can be obtained easily. However, to obtain spectrum distribution, it is necessary to determine the position relation between images which taken with many viewpoint. And it is difficult to determine the same points, which has different location in the input images, precisely. So, this report shows a precise position relation determination are realizable by using edge detection.

1 まえがき

mosaicing(画像の統合)とは、複数枚の画像をつなぎ合わせて、新しく広視野な画像を得ることをいい、古くからリモートセンシングなどの分野において研究されてきた。カメラの解像度や撮影角等の制限により必要となるシーンの一部分しか一度に撮影できないときに、画像の一部が重なり合うようにしてシーンを複数回に分けて撮影して統合することによってそのシーンを得ることが出来る。これにより、画素数の大きな画像や、360°パノラマ画像等を得ることが出来る。そしてこの統合において一番重要な問題となるのが複数の画像間の位置関係をいかにして正確に決定するかということである。なぜなら、正確な位置決定を行わないと、精度の良い画像を得ることが出来ずにぼやけた画像になってしまうからである。そしてその位置決定の方法には大きく分けて2種類ある。1つ目は、2つの画像間の対応点を見つけて決定する方法、2つ目は、画像同士の差分を利用して決定する方法である。前者は高速であるがノイズに非常に弱く対応点を見つけ出すのが非常に難しく、後者は非常に計算時間がかかるが、簡単で、より正確に位置決定が行える。この論文では正確さが要求されるために後者を利用した。

次に多波長であるとは、多数の色情報を持っているということである。多数の色情報を持つことによってデータの量は増えてしまうが、その分様々な解析に利用することが出来る。これにより色推定、物体認識等が行える。

この論文では、variable interference filter を利用して撮影した画像をエッジ検

出して mosaicing することで、高精度な平面シーン及びパノラマの多波長画像を作成する。まず2章では今回利用する spatially varying filter について説明し、3章では mosaicing 全般について説明する。

2 Variable interference filter

本研究では variable interference filter を使用する。Variable interference filter とはフィルターに沿ってその透過する光の波長が線形的に変化するフィルターである。フィルターの端から端に向かって透過光が可視光内の700nmから400nmへと線形的に変化していくのである(図1)。よって、このフィルターをモノクロカメラに装着して撮影すると、画像の左側からそれぞれ赤、緑、青色の物体のみが画像に写るのである。では、あるシーン x - y を考える(図2の x - y 平面がシーン)。シーン上の各 pixel において方向の波長毎に輝度値を持っている。そしてこのシーンを Variable interference filter を装着したモノクロカメラで撮影すると、図2(左)のように斜め方向にシーンを撮影することができる。そして、図2(右)のように連続して複数枚の画像を撮影することで、シーンの多波長データを得ることが出来る。そしてこのデータを密に取ることによって、物体の反射スペクトル分布を得ることが出来るのである。

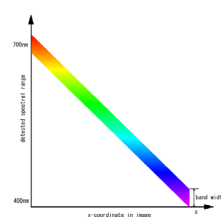


図1 フィルターの x 位置における透過光の波長

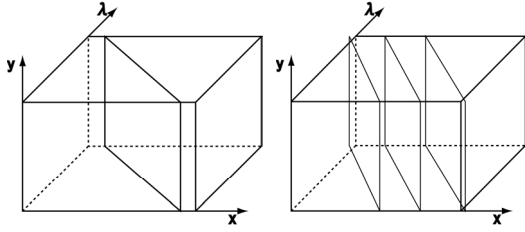


図 2 撮影シーンと撮影される画像(左)、シーンの連続撮影(右)

3 Mosaicing

3.1 カメラキャリブレーション

まず画像統合を行う前に、カメラキャリブレーションを行う。カメラのレンズの歪みやカメラの内部パラメータである焦点距離や画像中心等を求めて、撮影された画像をカメラモデルで撮影される画像に変換する。この論文では Zhang の手法[6]を使いカメラキャリブレーションを行い、画像の補正を行った。

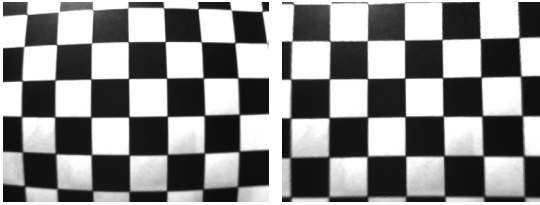


図 3 補正前(左)、補正後(右)

3.2 光源推定

Variable interference filter を装着して撮影された画像は光源の影響を強く受けてしまっている。例えば図 4 (左上)のように、蛍光灯の下で撮影するとそのスペクトルが画像の水平方向に出てしまう。そのため、位置合わせを正確に行うためにその影響をできるだけ取り除いておく必要が出てくる。そこで、光源のスペクトルを得る必要がある。光源のスペクトルを得る方法としては

直接光源を測定する方法も考えられるが、本論文では次の方法を使う。

まず撮影される $\hat{x}-\hat{y}$ のシーンを考えると、 k 番目に撮影された画像の座標 (x, y) での輝度は次のような式で表せる。

$$I_k(x, y) = L_{ill}(\lambda) l_{system}(\lambda) r(x, y, k, \lambda) \quad (1)$$

ここで L_{ill} は光源のスペクトル、 l_{system} は

カメラのシステムのスペクトル反応、 r は反射スペクトルである。ここで、

$$x \sim \lambda \quad (2)$$

であるので、次の式で表せる。

$$I_k(x, y) = L'_{ill}(x) l'_{system}(x) r'(x, y, k) \quad (3)$$

ここで、全ての画像に対して y 方向に x 軸に沿って和をとっていったものを考える。

$$M(x) \propto \sum_k \sum_y I_k(x, y) \quad (4)$$

また、次のように画像上の全ピクセルでの各波長での反射スペクトルの平均を考える。

$$r_{average}(\lambda) \approx \sum_{\tilde{x}} \sum_{\tilde{y}} r'(x, y, k) \quad (5)$$

ここで、撮影が十分な密度で行われたとすると、(3)(4)(5)式より次の式で表せる。

$$M(x) \propto L'_{ill}(x) l'_{system}(x) r_{average}(x) \quad (6)$$

ここで、 $r_{average}(\lambda) = Const$ すなわちシーンの

の平均したスペクトルが一定であると仮定

し、 l_{system} も一定であれば、 $M(x)$ を光源の

スペクトルと考えることが出来る。

そして、その値を利用して全ての撮影された画像に対して補正を行う。

$$I'_k(x, y) = \frac{I_k(x, y)}{M(x)} \quad (7)$$

以上によって、光源を推定してその影響を取り除くことが出来る。ただし、より正確な光源のスペクトルを得るためには、反射成分が既知なものを撮影して推定するか、直接光源を測定する必要がある。またこの時光源は、単一光源であることが要求される。

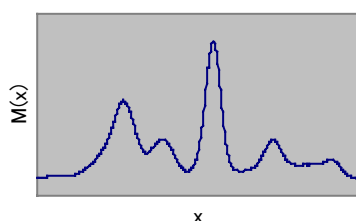


図 4 フィルターを装着して撮影された画像の例(左上)、それらの画像により推定される画像(下)、それを使って補正された画像(右上)

3.3 エッジ検出

次に位置合わせを正確に行うためにエッジ検出をかける。エッジ検出をした画像を利用する理由は次の2つである。

1. 各画像での各 pixel の輝度値は、画像上の各位置(各波長)での物体の反射スペクトルを表すため、同じ物体でも画像上の位置によって輝度が大きく変わってくる。そのため輝度値の差分は利用が難しい。
2. 一般の mosaicing は画像同士の重なり合った部分の加重平均を取って新しい画像を作るが、この論文での

mosaicing はそれとは異なり、重なり合った部分のデータはスペクトル分布として得られるためにより正確な位置合わせが必要であり、輝度の差分よりエッジを利用した差分の方が正確である。

また、本論分では、エッジ検出を行う前に平滑化を行った。その理由は、位置合わせには光源補正がされた画像を使用するが、各 pixel の輝度値の誤差 $G(x)$ は、 g を撮影誤差とすると、

$$G(x) = \frac{g}{M(x)} \quad (8)$$

となっている。すなわち、元の画像で暗かった部分が補正後の画像では誤差が大きくなってしまっているからである。よってそのノイズを取り除くために平滑化が必要となってくる。そして平滑化にはガウス関数を利用した。平滑化の範囲を蛍光灯の下で撮影された画像では $\sigma = 3$ とした。またエッジ検出にはソベルオペレータを使用した。

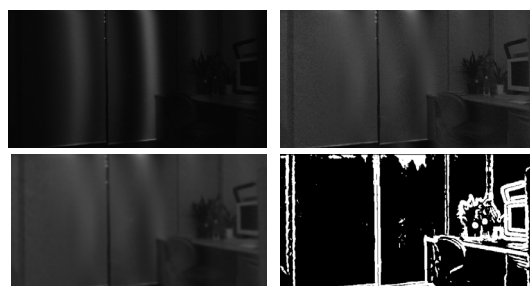


図 5 光源補正前(左上)、光源補正後(右上)、平滑化後(左下)、エッジ検出後(右下)

3.4 位置決定

統合を行うシーンには視差が無いものを扱う。ここで、視差の無いシーンには大きく分けて2種類がある。

1. 平面シーンを撮影したもの。(ただし本

論分では、平面シーンに対してカメラ垂直に向けて平行移動して撮影を行った。))

2. カメラを、その主点を中心に回転して撮影したもの。(ただし本論分ではカメラをパン回転のみさせて撮影を行った。)

まず平面シーンを撮影した画像の位置決定だが、位置関係は次の式で表すことが出来る。

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} \sim Mx = \begin{bmatrix} m_0 & m_1 & m_2 \\ m_3 & m_4 & m_5 \\ m_6 & m_7 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (9)$$

ここで、カメラを平面に向かって垂直にし、水平に移動しながら撮影を行えば、

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} \sim Mx = \begin{bmatrix} 1 & 0 & m_2 \\ 0 & 1 & m_5 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (10)$$

と表せる。そして、

$$Error(i, j) = \frac{\sum_{overlap} |E_1(x, y) - E_2(x+i, y+j)|}{\sum 1} \quad (11)$$

が最小になる (i, j) が画像間の位置関係を表すとする。ただし E はエッジ検出された画像。

次にパン回転で撮影した画像の位置決定だが、撮影された画像を円柱に投影(図 6 右)して、で表せる。

$$Error(\theta', n') = \frac{\sum_{overlap \theta, n} |E_1(M_1(\theta)+u, M_2(\theta, n)+v) - E_2(M_1(\theta-\theta')+u, M_2(\theta-\theta', n+n')+v)|}{\sum 1} \quad (12)$$

$$M_1(\theta) = \frac{f}{k} \tan \theta$$

$$M_2(\theta, n) = \frac{1}{|\cos \theta|} n$$

ここで、 (u, v) は画像中心、 f は焦点距離、 k は画素の大きさである。そして同様にエ

ッジ画像の差分が最小となるものを求める。ただし、スペクトル分布を得る以上、カメラはある程度一定の動きをする必要があり、そのため全数探索する範囲を絞り込んだ。

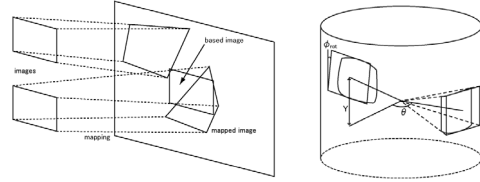


図 6 平面シーンの投影(左)、360° パノラマシーンの投影(右)

さらに、より正確な画像統合のために、 f を実験で求めた焦点距離として、

$$f' = \frac{2\pi + \theta}{2\pi} f \quad (13)$$

により、より正確な焦点距離 f' を得て、それを実際の焦点距離とした。ここで、 θ は 360° の統合を行ったときの一周した同一画像間での角度の差分である。

3.5 画像の作成

以上により x-y 平面での各 pixel におけるスペクトルデータを得ることが出来る。ここで、そのデータを視覚化し、利用しやすいようにするため RGB 画像に変換する。RGB 画像は、これはスペクトル分布とカメラ、目等の反応素子の積を積分することで得ることが出来る。

$$I_{r,g,b}(x, y) = \int_{\lambda} D_{r,g,b}(\lambda) I(x, y, \lambda) d\lambda \quad (14)$$

ここで $D_{r,g,b}$ は波長 λ での赤、青、緑の素子の反応量、 (x, y) はシーンの座標、 I は (x, y) での各波長の輝度、 $I_{r,g,b}$ は RGB 画

像である。

4 実験

デジタルモノクロカメラの SONY XC55 (non interlaced) に variable interference filter を装着してマクベスチャートを置いた平面シーンをカメラがシーンに対して垂直にかつ水平に移動するようにしながら連続撮影した。各 pixel が 29 回ずつ撮影され、その画像の統合を行った(図 7)。

また統合によって得られた (x, y, λ) 空間での図 9 のようにして得られる各波長での輝度値、すなわち各波長でのシーンの反射スペクトルを図 10 に示す。

次に、エッジ検出を使用した場合と使用しない場合での位置合わせを行った画像を示す(図 11)。

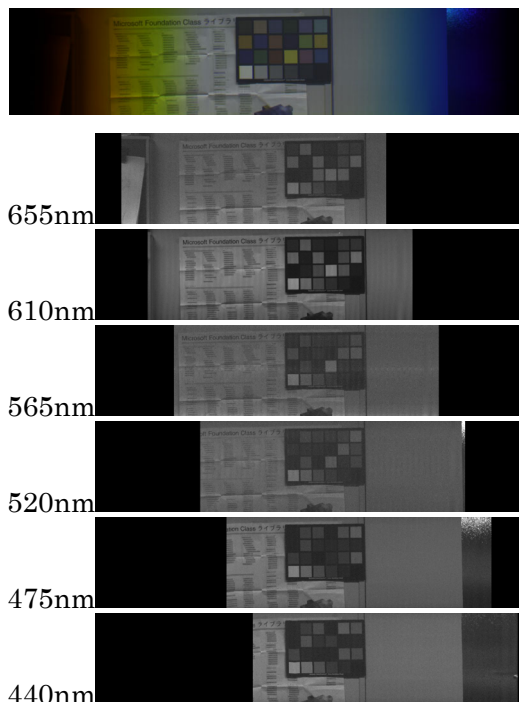


図 10 統合されたシーンの各波長での反射スペクトル



図 7 統合された画像

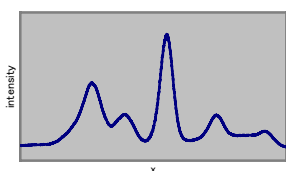


図 8 画像群から推定された光源

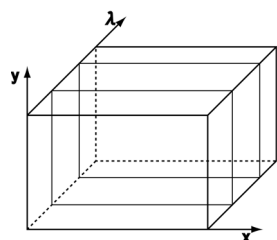


図 9 シーンの各波長での切断

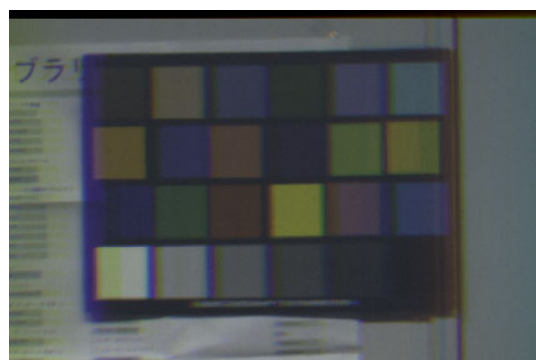


図 11 エッジ検出を用いる場合(上)、用いない場合(下)

また、画像統合により作成された RGB 画像と SONY DXC-9000 を使用して撮影された画像とを比較する(図 12)。

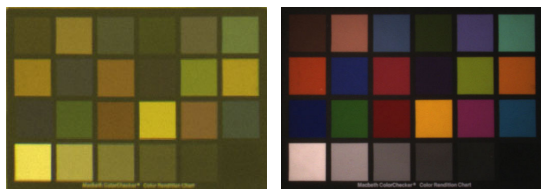


図 12 合成された画像(左)とデジタルカメラで撮影された画像(右)

次にカメラをパン回転させて、パノラマシーンを撮影する。同様に SONY XC55(non interlaced)にフィルターを装着しシーンを撮影した。カメラを 5° ずつパン回転させながら 144 枚の画像の撮影を行い、各 pixel とも 29 回ずつ撮影された。全数探索範囲をカメラの回転の 5° にまで狭めても 144 枚の統合に Pentium4 1.4Ghz で 1 時間かかった。

5 考察

実験の図 11 から、エッジ検出を利用したほうが位置合わせがより正確に行えていることが分かる。これは、波長により反射スペクトルが異なるために、単純な差分法で行うと正確な位置での評価関数が大きくなっていて、少しずれた位置との差がなくなってしまうためである。次に、実際に撮影を行って統合を行った画像が、直接 RGB カメラで撮影した画像と大きくずれている点についてだが、これはフィルターの透過光を正確に求められていないこと、

カメラ素子の反応を正確に求められていないこと、光源を正確に求められていないこと等によると思われる。よってフィルター及びカメラ素子の反応のデータを正確に求めておく必要がある。また光源については、直接光源を測定したり、反射スペクトルが既知の物体を利用する等が考えられる。次に計算時間だが、360°パノラマ画像の位置合わせに 1 時間かかってしまった。まだまだプログラム等に改良の余地があり早くなると思われる。また 360°は無理でも一定の角度分をリアルタイムに処理できるようにすれば物体の認識などの研究範囲が大きくなると思う。

むすび

この研究では variable interference filter を使って撮影された画像を、エッジ検出を利用して位置合わせを行い統合することで、より正確な多波長パノラマ画像の生成を行った。これからの研究としては、カメラの素子の反応を正確に調べてより厳密に多波長画像を得て、そしてそれによって得られた物体のスペクトルを利用して、光源推定等の分野の研究を行っていきたい。

参考文献

- [1]Richard Szeliski and Hung-Yeung Shum. *Creating Full View Panoramic Image Mosaics and Environment Maps*. Microsoft Research. SIGGRAPH'97, pp.251-258, 1997.



図 13 統合された 360°パノラマ画像

- [2]Yuu Nakatani. *Image Mosaicing Using Active Camera*. The Institute of Image Electronics Engineers of Japan, vol.29, no.5, 2000.
- [3]Yalin Xiong and Ken Turkowski. *Registration, Calibration and Blending in Creating High Quality Panoramas*. Fourth IEEE Workshop on Applications of Computer Vision, Princeton, New Jersey, October 19-21, 1998, pp. 69-74.
- [4]Naoki CHIBA, Hiroshi KANO, Michihiko MINOH, and Masashi YASUDA: *Feature-Based Image Mosaicing*. The Institute of Electronics, Information and Communication Engineers. D- , Vol.82-D- , pp.1581-1589, 1999.
- [5]Yoav Y.Schechner and Shree K.Nayar: Generalized Mosaicing: *Wide Field of View Multispectral Imaging*. IEEE Trans. on Pattern Analysis and Machine Intelligence 2001.
- [6]Zhengyou Zhang: *A Flexible New Technique for Camera Calibration*. Microsoft Research Technical Report MSR-TR-98-71. December 2, 1998.
- [7]Graham D. Finlayson and Steven D HorDley: *Color constancy at a pixel*. JOSA-A Vol. 18 No. 2, February 2001, pp. 253-264.
- [8]JIAN HO, BRIAN V. FUNT, MARK S. DREW: *Separating a Color Signal into Illumination and Surface Components: Theory and Applications*. IEEE Transactions on Pattern Analysis and Machine Intelligence. VOL. 12, NO. 10, OCTOBER 1990.
- [9]Shoji Tominaga: *Multichannel vision system for estimating surface and illumination functions*. Vol. 13, No. 11/November 1996/J. Opt. Soc. Am. A