

対応点の確度分布を考慮した多眼画像からの奥行き推定

今泉 浩幸、 蓼沼 眞、 片山 美和、 金次 保明

NHK 放送技術研究所
〒157-8510 東京都世田谷区砧 1-10-11

あらまし 自然な立体テレビシステムの開発を目指して、観察者の位置の移動に伴う任意視点画像を再現できる多視点立体画像システムの検討を行っている。被写体の画像情報、被写体までの距離情報などの3次元情報を、多眼カメラで撮影した複数の画像から正確に抽出することで、被写体の3次元情報を効率的に伝送し、受信側では任意視点画像を再生し表示する。本稿では、このシステムの実現を目指して、多眼カメラによる正確な距離情報取得のための対応点探索手法を検討する。多眼カメラでの対応点探索のために、対応度を新たに定義するとともに、探索を行うエピポーラ線上での対応度の分布を考慮することで、距離情報の取得精度が向上することを示す。

キーワード 奥行き推定、 対応度の分布、 多眼画像システム、 任意視点画像、 立体テレビ

Depth Estimation Algorithm for Multicocular Images using Distribution of Matching Cost on an Epipolar Line

Hiroyuki Imaizumi Makoto Tadenuma Miwa Katayama Yasuaki Kanatsugu

NHK Science and Technical Research Laboratories
1-10-11, Kinuta, Setagaya-ku, Tokyo 157-8510, JAPAN

Abstract This paper describes a robust depth estimation algorithm for multicocular image systems. Based on this algorithm, images can be synthesised for any intermediate viewpoint for smooth motion parallax in multiview systems. A block matching algorithm, using a new multicocular matching cost as well as epipolar constrains, provides robust depth estimation. Furthermore, validity of corresponding points, using a distribution of matching cost on an epipolar line, is introduced to improve the depth estimation algorithm.

Key words Depth estimation, distribution of matching cost, multicocular image system, intermediate viewpoint, stereoscopic television

1. はじめに

近年、バーチャルリアリティ技術の応用として、視点追従型の任意視点立体表示装置を使ったテレビプレゼンスシステム[1]や半球形の全天周ドームに多数のカメラを配置して3次元情報を取得し、再生時に視聴者が自由なカメラワークを選択できるシステム[2]なども開発され、これらの画像入力装置として多眼カメラが使用されている。また、家庭用ゲーム機の高機能化が進み、視点変換やテクスチャマッピングを高速に実現できるCGエンジンの開発も進められている。

放送の分野でも、日本では2000年にBSデジタル放送が開始されるなど、デジタル技術を利用した新しい放送サービスを提供できるインフラストラクチャが整備されてきている。

筆者らは将来のデジタル放送の新しいサービスとして、より自然な立体テレビ放送の実現の可能性を検討している。より自然な立体テレビシステムのひとつとして、運動視差を再現できる多視点立体画像システムが有望である。立体システムでは表示装置が実現の最も重要な要因であるが、実現性や評価の観点から、運動視差により立体感を再現するシステムであれば、必ずしも両眼融合による立体知覚機構を利用しなくても立体感を再現することができる。したがって、視点追従などの技術を併用することで、十分な解像度や表示面積を持った現行のHDTVモニタなどをそのまま利用することができる。また、将来、両眼立体視

のための表示装置が普及した場合には、容易に立体画像を再生できる。

多視点立体画像システムの画像入力装置として、多眼カメラを利用する場合、番組制作のための可搬性や経済性を考慮すると、最小限のカメラ台数が望ましい。また、多眼情報を効率的に圧縮符号化して伝送する際には、距離情報とカメラパラメータを使った視点変換を利用し、オクルージョン領域など視点変換だけでは再生できないデータのみを残差情報として伝送する方法が有望と考えられる。そこで、できるだけ少ないカメラ台数で正確な距離情報を取得し、視点変換後の残差情報を低減することがシステムの基本的な要求条件となる。

本稿では、この要求条件を実現するために、多眼カメラによる正確な距離情報取得のための対応点探索手法を検討する。従来、距離情報取得のための対応点探索手法では、2眼カメラを対象とした検討が多く、N眼カメラを用いる場合にも N_C 対のカメラ対に対してブロックマッチングによる差分2乗和を平均する[3,4]などの方法が一般的であった。今回、多眼カメラでの対応度を新たに定義するとともに、探索を行うエピソード線上での対応度の分布を考慮することで距離情報の取得精度が向上できることを示す。

本第1節につづいて、第2節では、想定している多視点立体画像符号化システムの要求条件を整理する。第3節で新しく定義した多眼カメラの

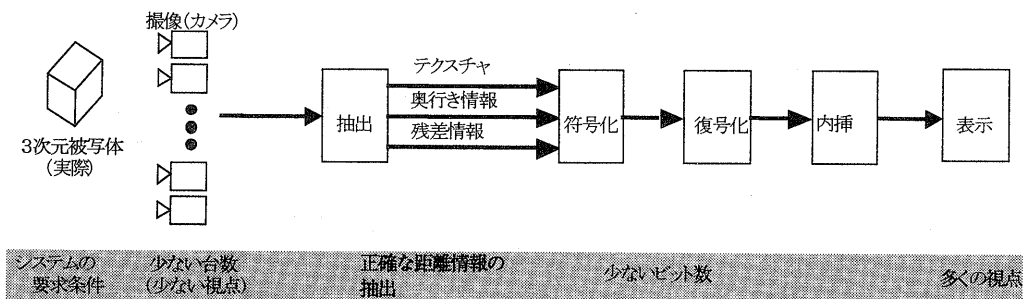


図1 多視点立体画像システムの全体構成

対応度について述べ、エビポーラ線上での対応度の分布を示す尖度を導入する。第4節では、シミュレーションの実験条件を設定し、第5節で実験結果と考察を述べる。

2. 多視点立体画像符号化システム

より自然な立体知覚を得るシステムとして、観察者の視点位置の移動に伴う任意視点画像を再現できる多視点立体画像システムの検討を行っている。想定しているシステムの全体構成を図1に示す。

考慮すべきシステムの要求条件として、

撮像、抽出部：

- できるだけ少ない台数のカメラで、再生に必要な被写体の3次元情報（画像情報と距離情報）を取得する。

符号化部：

- 撮像した多眼画像の画像情報と距離情報を効率的に符号化する。

再生部：

- 復号した3次元情報から視点内挿を行い、視点移動に対応した任意視点画像を再生する。

などが挙げられる。

これらの要求条件を実現するために、

- 1) 高精度に距離情報を抽出するための対応点探索アルゴリズムの検討
- 2) 最小限のカメラ台数で効率よく3次元情報を取得するための最適なカメラ配置の検討
- 3) 距離情報とカメラパラメータ（位置、光軸の向き、画角など）を使って視点変換を行い、視差補償予測により効率的に多眼画像を圧縮符号化する手法の検討
- 4) 視点内挿により任意視点画像を補完して再生する手法の検討

などを行っていく必要がある。なかでも、1)の検討課題の解決が、2),3),4)として挙げた課題の解決に与える影響が大きく、システムを設計する上で最も基本的な要因の一つと考えられる。

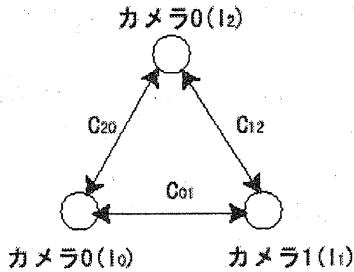
3. 距離情報取得のための多眼画像の対応点探索

3.1 多眼カメラでの対応点探索のための対応度の定義

従来、距離情報取得のための対応点探索として、2眼カメラ画像を用いて、ブロック相関が最大となる点や差分の絶対値または2乗和が最小となる点を対応点と決定する手法が一般的に用いられている。また、多眼カメラ画像の場合には、 N 眼カメラ画像から2台のカメラを選択する組み合わせ ${}_N C_2$ 対の相互相関値や差分絶対値和、2乗和などの重み付き平均を対応度の定義とする手法[3,4]が用いられてきた。また、5眼カメラを、中心とその上下左右に配置し、中心カメラを基準カメラとして、他の4つのカメラとの差分2乗和のうち小さい値2つの和を対応度とする手法が提案されている[5]。

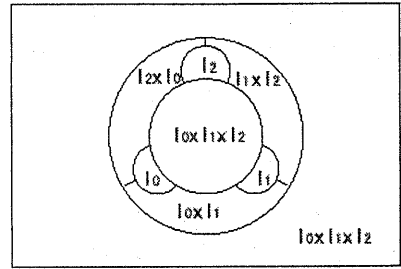
しかしながら、従来の手法では、2眼カメラの場合には、半遮蔽領域が発生する割合が多く、また、エビポーラ線と同じ方向に周期的なパターンがある場合、対応点を正しく特定できない問題があった。多眼カメラの場合にも、オクルージョンなどにより、一部のカメラ対でしか対応点が特定できない場合に、平均を取ることで、正しい対応点の対応度が低くなってしまい、誤って対応点を決定してしまう。5眼カメラの手法では、ある程度オクルージョンの影響を低減できると考えられるが、4眼のうち2眼を利用するという制約を N 眼の場合に、どのように一般化するかという問題がある。これらの問題点を解決するために、以下のような手法を提案する。

多眼としての最小眼数である3眼カメラを例にオクルージョンの発生をモデル化する。3眼カメラの配置を図2のように仮定し、それぞれのカメラ0、カメラ1、カメラ2で撮像される画像を I_0, I_1, I_2 とする。 I_0 と I_1 のブロック相互相関値を C_{01} 、 I_1 と I_2 のブロック相互相関値を C_{12} 、 I_2 と I_0 のブロック相互相関値を C_{20} と表す。また、図3に示すように、実際の被写体がどのカメラで撮影されているかによって、領域を分類することができる。



I_j : カメラ j で撮像された画像
 C_{jk} : カメラ j とカメラ k の対応点での相互相関係数

図2 N眼カメラモデル (N=3の場合)



$I_0 \times I_1 \times I_2$: カメラ0, 1, 2で撮像される領域
 $I_j \times I_k$: カメラ j, k のみで撮像される領域
 I_j : カメラ j のみで撮像される領域

図3 オクルージョン領域の発生モデル

領域 $I_0 \times I_1 \times I_2$ は3台のカメラすべてで撮像される領域である。領域 $I_0 \times I_1$ はカメラ0およびカメラ1によってのみ撮像され、カメラ2からはオクルージョンとなっている領域を示す。同様に、領域 $I_1 \times I_2$ はカメラ0のオクルージョン領域、 $I_2 \times I_0$ はカメラ1のオクルージョン領域である。領域 I_0, I_1, I_2 はそれぞれカメラ0、カメラ1、カメラ2のみによって撮像される領域である。

領域 $I_0 \times I_1 \times I_2$ では、対応点での相互相関値 C_{01}, C_{12}, C_{21} が同時にピーク値をとるはずであり、このピークを特定するための対応度の定義として、

$$C_{01} \times C_{12} \times C_{20} \quad (1)$$

を用いることで感度良く対応点を特定できると予想される。

次に領域 $I_0 \times I_1$ では、カメラ2に対するオクルージョン領域であることから、対応度を定義する際に、ブロック相互相関値を I_2 を使って算出する C_{12} および C_{20} は雑音成分になってしまう。したがって、領域 $I_0 \times I_1$ では対応度の定義として、

$$C_{01} \quad (2)$$

を利用すると良いと考えられる。同様に、領域 $I_1 \times I_2$ での対応度の定義は C_{12} 、領域 $I_2 \times I_0$ では、 C_{20} によって鋭敏に対応点を決定できると考えられる。探索している対応点が、領域 $I_0 \times I_1, I_1 \times I_2, I_2 \times I_0$ のいずれかにある場合には、この3つのうちどの領域に対応点があるのか特定できなくても、3つの相関値の最大値、すなわち、

$$\max(C_{01}, C_{12}, C_{20}) \quad (3)$$

を取ることで、対応点の存在する領域に最も適した対応度の定義を用いていることになる。ここで、 $C_{01} \times C_{12} \times C_{20}$ と $\max(C_{01}, C_{12}, C_{20})$ を重み係数 C_w で規格化し、和をとった

$$C_{01} \times C_{12} \times C_{20} / C_w^3 + \max(C_{01}, C_{12}, C_{20}) / C_w \quad (4)$$

を3眼カメラでの対応度と定義できる。

領域 $I_0 \times I_1 \times I_2$ と領域 $I_0 \times I_1 \cup I_1 \times I_2 \cup I_2 \times I_0$ は排他的であり、(4)式のように和をとれば、領域 $I_0 \times I_1 \times I_2$ では第1項が支配的となり、領域 $I_0 \times I_1 \cup I_1 \times I_2 \cup I_2 \times I_0$ では第2項が支配的となる。また、(4)式の $1/C_w^3, 1/C_w$ は第1項と第2項の重み付けに相当するが、このように定義しておくと、

$$C_{01} = C_{12} = C_{20} = C_w \quad (5)$$

のときに、第1項と第2項の重み付けした対応度が等しくなる。

(4)式で定義したオクルージョン領域を考慮した多眼カメラでの対応度は、容易にN眼の場合に拡張できる。

3.2 尖度による対応点の確度指標

(4)式に示される対応点の対応度は、輝度の平坦な部分や画像のテクスチャが周期的なパターンを有している場合などは、対応度は高いにもかかわらず正しい対応点でない場合があり得る。

そこで、対応点の確度を表す指標として尖度の導入を検討した。対応点 p の尖度 $K(p)$ は対応点探索を行うエピソード線上の対応度の分布 ρ_i から、

$$K(p) = \frac{\sum_{i=d_{\min}}^{d_{\max}} \rho_i \sum_{i=d_{\min}}^{d_{\max}} \rho_i (i-p)^4}{\left(\sum_{i=d_{\min}}^{d_{\max}} \rho_i (i-p)^2 \right)^2} \quad (6)$$

d_{\min}, d_{\max} : 対応点の探索範囲、
 ρ_i : エピソード線上の点 i での対応度

(6)式のように定義した。

尖度とは相関関数のピークの集中度を表す指標であり、探索範囲内で同程度の対応度を示す対応点の候補が広い範囲に分布しているとき低い値となる。すなわち、対応点探索の手がかりとなる顕著な特徴が無い領域、たとえば、輝度の分布が平坦な領域や、テクスチャが周期的なパターンの繰り返しを含む場合などに、低い値を取る。

一方、ステップ状やインパルス状に対応度に変化する点では、対応度が低くても尖度の値は高くなる場合がある。対応度が低い場合には、1台のカメラでしか撮像されていないオクルージョン領域、または、探索窓内に複数の距離の被写体が含まれていることなどが考えられ、対応点の確度は低く示されることが妥当である。そこで、(4)式と(6)式の積で対応度の確度を定義することとした。

4. シミュレーション実験条件

4.1 カメラ配置と評価画像

実験に用いたカメラ配置を図4に示す。カメラの台数は多眼カメラでの最小の台数である3台で実験を行った。3眼カメラでは3眼を水平に並べるより、中央のカメラを上方に配置する方が、距離の推定精度が高くなるとの検討結果[6]に基づき、右、左、上の三角形のカメラ配置とした。

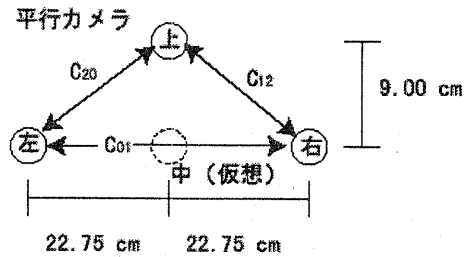


図4 実験でのカメラ配置

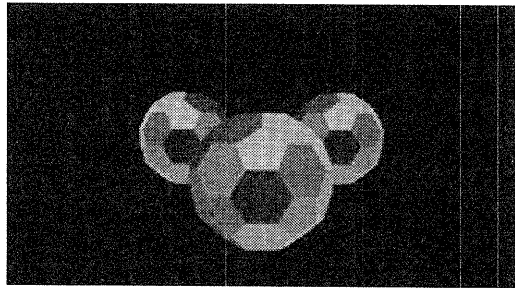


図5 評価画像 (CGにより作成)

${}_3C_2 = 3$ 対のカメラ対で求めた相互相関値を図に示すように C_{01} 、 C_{12} 、 C_{20} と表すこととする。

仮想カメラを図4の中央カメラの位置に置き、中央カメラ画像上の1点が投影される右、左、上カメラ画像上のエピソード線上で、対応度を求め、対応度が最大となる点を対応点と決定する。

評価画像は図5に示すCGにより作成した画像を使用した。CGを利用した理由は、距離の真値が既知であり、方式による距離推定精度の違いを比較できることである。なお、このCG画像は、立体テレビチャート[7]の統計的な性質を反映して作成したもので、輝度勾配の分布、輝度勾配の方向性、視差(距離値)の分布などが考慮されている[8]。また、被写体は約2m~3.4mの範囲に存在している。カメラの垂直画角は18.92[deg]で、画面高の3倍の距離からの観察に相当する。画像の縦横比は9:16であり、垂直×水平画素数は1080×1920である。

4.2 多眼画像の対応点探索手法

(4)式で表される対応度の定義の中で、距離の推定精度が最も高くなるように重み係数を決める

$$\text{距離の平均推定誤差} [\%] = \frac{\sum_{\text{画素数}} \left| \frac{\text{距離の推定値} [mm] - \text{距離の真値} [mm]}{\text{距離の真値} [mm]} \right|}{\text{画素数}} \times 100 \quad (7)$$

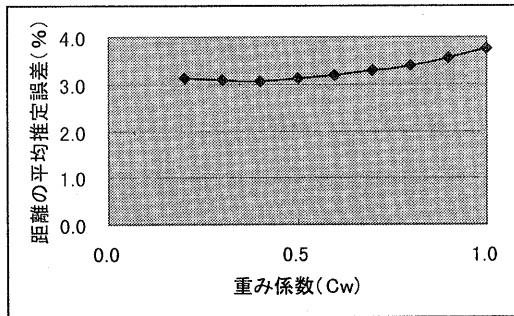


図6 (4)式の C_w と距離の平均推定誤差

ため C_w の値を変えて(7)式に定義される距離の平均推定誤差を求めた結果を図6に示す。この結果に基づき $C_w = 0.4$ に設定した。

対応点探索の評価窓のサイズは 15×15 画素、探索精度は、左カメラ画像、および、右カメラ画像上で1画素精度となるようにした。

表1に示す対応度の定義について対応点探索を行い、距離の平均推定誤差を比較する。方式1, 2, 3は、従来の2眼カメラでの対応度の定義である。方式4は、従来の多眼カメラでの対応度の定義として用いられている方式で、複数のカメラ対での相関値を単純平均する方式である。方式5が(4)式を用いて、今回新たに定義した対応度を利用する方式であり、方式6では、式(4)と式(6)の積で定義した対応点の確度を対応度と定義した場合である。

5. 実験結果と考察

各方式について、(7)式により求めた被写体までの距離の平均推定誤差を表1に示す。実験から次の結果が確認された。

- 1対の対応度(2眼)を用いる場合に比較して、3対(3眼)の場合に、距離の推定誤差の顕著な改善が認められた。
- 方式6では、尖度を考慮することで、距離の推定誤差に改善が認められた。

図7は走査線1本分(第540ライン)の画像について、縦軸に距離をとって、対応度の分布を示した。高い輝度値は高い対応度を示す。対応点探索は、縦軸方向に探索して輝度値の最も高い点を対応点と決めることに相当する。

従来の2眼カメラ(方式1)では、真の距離以外の点に、対応度(輝度値)の高い点が多く存在し、誤って対応点を決定する可能性が高い。従来の対応度の定義を用いた3眼カメラ(方式4)では、方式1に比較して、対応度の高い点が真の距離位置に絞り込まれてきている様子が認められる。さらに、新しく対応度を定義した提案方式(方式5)では、より真の距離値に近づいてきている。表1に示した平均推定誤差の数値からは、方式4と方式5は、ほぼ同程度の性能と評価されるが、図7の対応度の分布を比較すると、方式5による改善効果が期待できる。方式6は尖度を考

表1 対応度の定義と距離の推定誤差

対応点探索方式		対応度の定義	距離の平均推定誤差 [%]
方式1	1対(2眼)	左・右カメラの相互相関値 C_{01} (図4)	7.41
方式2		左・上カメラの相互相関値 C_{20} (図4)	8.36
方式3		右・上カメラの相互相関値 C_{12} (図4)	8.19
方式4	3対(3眼)	従来の多眼の対応度 $(C_{01} + C_{20} + C_{12}) / 3$	3.08
方式5		提案手法(4)式 ($C_w = 0.4$)	3.05
方式6		尖度を考慮した提案手法(4)式×(6)式	2.35

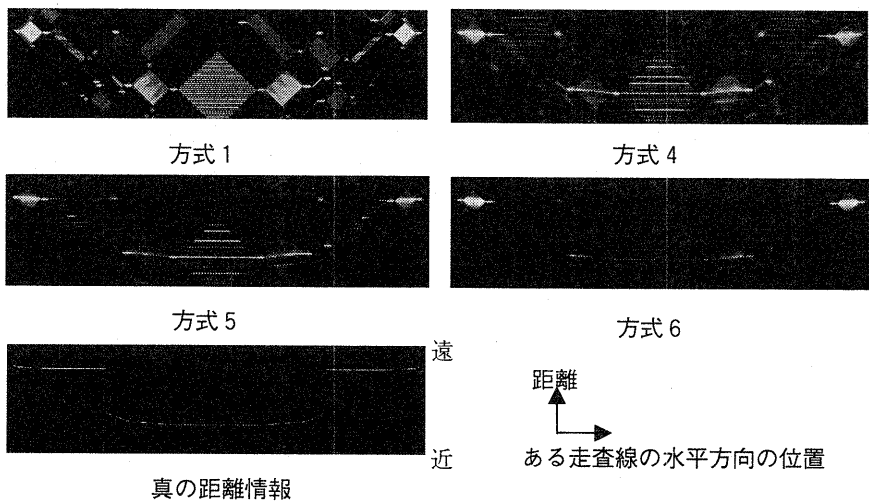


図7 対応度の分布：輝度が高いほど対応度が高いことを示す。

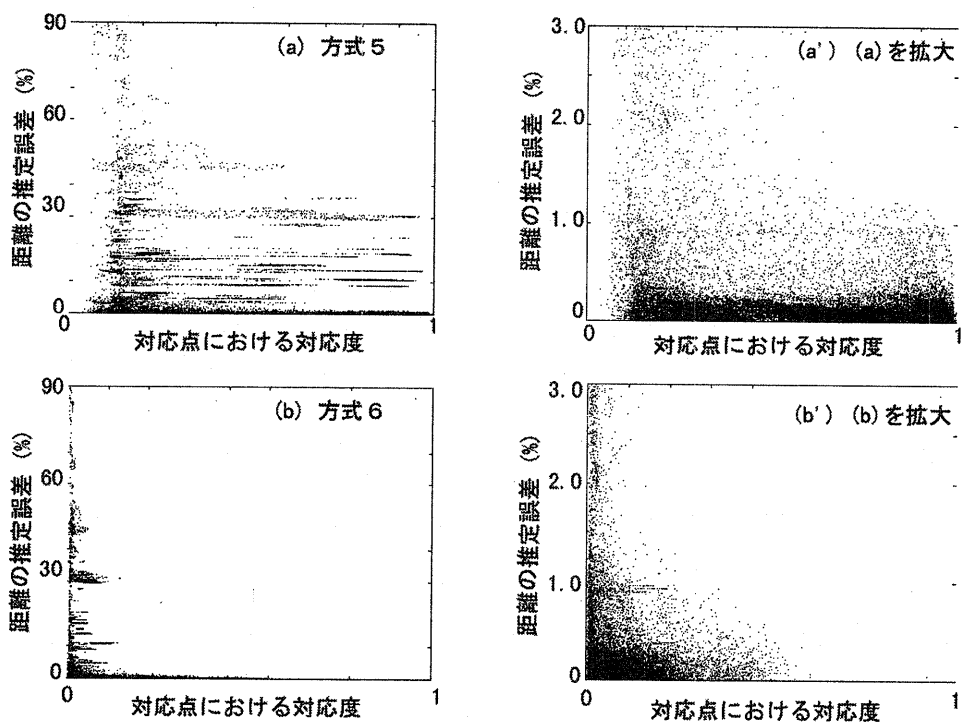


図8 対応度と距離推定誤差の関係

慮した対応点の確度を対応度と定義した場合であり、改善が認められる。表1の数値でも方式5に比較して距離推定精度の向上が認められる。

尖度の導入により距離推定精度が向上したことから、誤った対応点に比較して真の対応点では、対応度のピークが鋭くなる傾向があると予想されるが、理論的な証明は今後の課題である。

次に、方式6で尖度を考慮して定義した対応度の確度としての妥当性を検討する。図8は横軸に対応度、縦軸に距離の推定誤差をとって対応点における対応度と距離推定誤差の関係を調べた散布図である。(a)は方式5で用いた相互相関値を基本とする対応度の定義、(b)は方式6で用いた尖度と相関値の積によって定義する場合である。平均推定誤差に近い誤差の分布を調べるため、縦軸の表示範囲を3%付近までとして、(a)、(b)の垂直方向を拡大したグラフを(a')、(b')に示す。

相関値だけを用いた(a)では、対応度が高い場合にも大きな推定誤差が発生している場合が多いことが認められる。一方、相関値と尖度を考慮した(b)では、対応度が高くなるにしたがって、確実に大きな推定誤差が出現しなくなっており、対応点の確度を十分に反映する指標となりうることが理解される。この確度は視点変換の際のオクルージョン検出などにも利用できる。

6. まとめ

より自然な立体テレビシステムの実現を目指して検討している多視点立体画像システムに要求される条件と実現に向けての検討課題を整理し、最も重要な課題のひとつである多眼カメラによる対応点探索での正確な距離情報の取得方法を検討した。

オクルージョンの発生モデルに基づいて、多眼カメラでの対応点探索のための対応度を定義し、これによって距離情報の取得精度の向上を図った。また、2眼から3眼への多眼化による推定精度の改善を確認した。

さらに、対応度のエピソード線上での分布に着目し、対応度のピークの尖度を対応度の定義として考慮することで、距離推定精度の向上を図った。さらに、尖度と相関係数の積は、対応点の確度を表す有効な指標となることを示した。

今後は、3眼カメラでのカメラ配置の最適化を検討するとともに、4眼以上場合についてもカメラ台数と距離情報の推定精度の関係を明らかにし、カメラ台数やカメラ配置の最適化を図る。ま

た、CGではなく実写画像を使ってシステムの要求条件を満足する手法を検討するとともに、符号化方式、表示のための再生方式などの検討をあわせて行い、3次元情報の取得から表示のための再構成までのトータルシステムの観点から、各要素技術の研究を進めていきたいと考えている。

[参考文献]

- [1] EU Projects ACTS 092 PANORAMA (Package for New Imaging and Applications)
- [2] P.J.Narayanan, P.W.Rander and T.Kanade, "Constructing Virtual Worlds Using Dense Stereo", ICCV'98, India, Jan. 1998
- [3] T. Kanade 他, "Development of a Video-Rate Stereo Machine", IROS'95, Pittsburgh PA, Aug. 1995
- [4] N. Grammalidis and M.G.Strintzis, "Disparity and Occlusion Estimation in Multicocular Systems and Their Coding for the Communication of Multiview Image Sequences", IEEE Trans. on Circuits and Systems for Video Technology, vol8, No.3, June 1998
- [5] 朴、井上、「映像合成のための多眼カメラを用いた奥行き抽出」、信学技報 PRMU96-133, Jan. 1997
- [6] 蓼沼、今泉、金次、「画像内オブジェクトの統計量解析に基づいた多眼立体カメラの最適配置法の検討」、IMPS'99, Sep. 1999
- [7] 立体ハイビジョン標準静止画像、ITU-R 234/11
- [8] 蓼沼、金次、「立体ハイビジョン標準静止画像におけるオブジェクト輪郭の統計的性質」、映像メ学会全大, Aug. 1999