

# 対象追跡映像からの3次元形状復元のための 多視点カメラキャリブレーション

松村 和機, 延原 章平, 松山 隆司

京都大学大学院情報学研究科知能情報学専攻

**概要** 広範囲を移動する対象の精細な3次元形状を得るには, 広範囲かつズームアップ撮影という相反する2つの条件を満たさなければならない. これらは広範囲を撮影するにはカメラの視野角を広角に, ズームアップ撮影を行うには視野角を狭角にする必要があるため相反する. このような要件に対して, 本研究ではアクティブマルチカメラを用いた対象の追跡撮影を提案する. 従来は, このようなアクティブカメラ群をすべて視点固定とみなした手法が提案されていたが, これには追跡可能範囲が狭く, 機械的な読み取り誤差があるという問題点が存在する. これらの問題点は, カメラの運動モデルを先に定義している, という点に起因する. そこで本研究では, 撮影された映像そのものからカメラの位置及び姿勢を直接推定するアプローチを提案し, これと従来法との比較検討を行う.

## A multi-camera calibration for reconstructing 3D shape by images of tracking the object

Kazuki Matsumura, Shohei Nobuhara and Takashi Matsuyama

Department of Intelligence Science and Technology,  
Graduate School of Informatics, Kyoto Univ.

**abstract** To reconstruct an accurate 3D shape of the object moving in a wide area, we need to accomplish conflicting objectives – wide area, but close-up multi-viewpoint observation of the object.

To solve this problem, we propose a framework which tracks the object with active multiple cameras, and focus on the camera calibration of them. While some papers proposed calibration methods which assume the cameras to be active and fixed-viewpoint, this assumption on their motion model will break down or be valid only for a narrow pan/tilt angles if the object is relatively close to cameras. Moreover, the fixed-viewpoint active camera model gives its calibration data (rotation and translation) from not only this model but also its motion parameters. However, it is difficult to sense accurate motion parameters from pan/tilt unit in action. In order to overcome these difficulties, we propose a direct calibration of active cameras using observed images themselves.

## 1 動的キャリブレーション

### 1.1 背景

人の色・形状・動きなどの3次元復元は, ある撮影空間中で人が動作を行い, これを様々な方向から撮影し, 撮影された映像を処理することで行うことができる. このとき, 対象をより広い範囲で撮影すること, そしてより精細に対象の形状を復元することはコンピュータビジョンの分野において重要な課題である.

これらが実現することで, 例えば舞などの複雑で繊細な対象の形状や所作が, 能舞台のような広い範囲の空間で忠実に復元することが可能となる.

撮影可能な空間を広く取るためには

1. カメラの向きは固定で, 設置台数を増やす
2. アクティブカメラを用いて対象を追跡撮影する

の2つの手法が考えられる. しかし, カメラの設置やコスト, そしてキャリブレーションの観点からカメラ台数はより少ない方が望ましい. このため, 2

の手法を用いることを考える。

また、対象の撮影精度を向上させるためにはズームアップ撮影を行う必要がある。しかし、ズームアップ撮影を行うと、これに伴ってカメラの視野が狭くならざるを得ない。このため、限られた台数のカメラで、広い空間内を移動する対象を高精細に復元するためには、対象もしくはその一部を追跡撮影する必要がある。

そこで本研究では、複数のアクティブカメラで対象を追跡撮影しながら形状復元を行うことを考える。形状復元を行うためにはカメラのキャリブレーションが必要不可欠である。そのため本論文では、対象を追跡撮影した複数のアクティブカメラのキャリブレーションを行うこととする。このキャリブレーションの目的は全時間における全カメラの内部・外部、両パラメータを推定することであるが、簡単化のために内部パラメータは既知であるとし、外部パラメータを推定することとする。

複数のアクティブカメラのキャリブレーションを行う手法として次の2つが考えられる。

1. カメラの実際の運動機構をモデル化し、撮影時の制御量からカメラパラメータを推定する。 [6, 11]
2. 撮影画像を用いて、撮影時のカメラパラメータを推定する

このうち前者は、カメラモデルのもつパラメータをあらかじめ推定しておくことによってキャリブレーションが可能である。例えば、投影中心が固定で回転運動のみを行うようなカメラモデルならば、投影中心の位置と撮影時のパン・チルトに対応したパラメータを測定しておくことでカメラの位置及び姿勢を推定することができる。したがって、キャリブレーションは撮影前に一度行えばよい。このようなキャリブレーション手法を静的キャリブレーションと呼ぶ。しかし、一般に実際のカメラ運動を完全にモデル化することは困難であるため、ある測定に対してモデルを合わせるとその測定状態に対して最適化され、カメラの位置・姿勢が変動すると精度が劣化する [1, 10]。

これに対して後者は、撮影画像を用いて撮影時のカメラパラメータを推定する。また、カメラが自由に空間中を移動するため、撮影時のカメラ位置は不明であるから、毎フレームごとにキャリブレーション

を行う必要がある。しかし、キャリブレーションがカメラの運動モデルなどに影響されないため、どのようなカメラの位置・姿勢であっても均一な精度を得ることができる。このようなキャリブレーション手法を動的キャリブレーションと呼ぶ。

本研究では広い空間中を移動する対象を追跡するため、大きなカメラ運動を伴い撮影する。このため動的キャリブレーションを用いる。

従来の動的キャリブレーションの例として因子分解法が挙げられる [9]。この方法はカメラ間の相対位置、すなわち回転及び並進を求めることができるが、カメラの投影モデルに透視投影モデルを用いることができない [7]。この制約により、対象が十分遠くに存在して弱透視投影とみなすことができるか、対象が実空間中の1平面上に存在する、といった環境でなければ用いることができない。このため、対象をズームアップ撮影するという用途には適さない。

また、別の例としてセルフキャリブレーションが挙げられる [3, 8]。この手法は、空間中を自由に動く1台のカメラで撮影した映像を用いてそのカメラの動作のキャリブレーションを行うものである。よって、セルフキャリブレーションでは比較的視差の小さい連続フレーム間でキャリブレーションを行うが、本研究では空間的に異なる位置に置かれた視差の大きいカメラ間でキャリブレーションを行う必要がある。

## 1.2 提案手法

アクティブカメラを用いた広範囲・高精細撮影という問題に対するカメラキャリブレーションの手法として、以下の方法を提案する。

複数フレームの対象追跡映像を用いて

- Step 1. まず各フレームでの多視点カメラ群のキャリブレーションを行い、フレーム間で独立な座標系におけるカメラの相対位置を求める。
- Step 2. フレーム毎のカメラ位置と、世界座標系との間のキャリブレーションを行う。

という手順でキャリブレーションを行う。

この両者の特徴として、前者は空間的なカメラキャリブレーションであり、

- 視差が大きい

- 全ての特徴点を用いることができる
- それぞれのカメラの組合せで異なった特徴点の集合となる
- 共通視野が少なく対応点が少ない

という特徴を持つ。

一方後者は時間的なカメラキャリブレーションであり、

- 視差が小さい
- 運動しない特徴点しか用いることができない
- 各時刻における特徴点集合がほぼ同じ
- 共通視野が多いため対応点が多い

という特徴が挙げられる。本研究では、前段の手順である空間的なカメラキャリブレーションに焦点を当てることとする。

空間的なカメラキャリブレーションは、

1. カメラ群の中から特定のペアに関してキャリブレーションを行う
2. 複数の 1. を統合する

ことによって実現できる。一方、時間的なカメラキャリブレーションについてもある一台のカメラに注目して各フレームにおけるカメラ姿勢について同様のことを行うことで実現できる。したがって、本研究では2台のカメラのキャリブレーションに焦点を当てる。この精度を従来の視点固定型カメラのキャリブレーションと比較することにより本手法の有用性を検証する。

## 2 キャリブレーションの手法

本研究では、内部パラメータは既知として外部パラメータ（回転・並進ベクトル）を求め、この精度を評価する。外部パラメータは以下の手順で求める。

1. 対応点群を、重心が0、分散が1となるように正規化する。
2. 内部パラメータが既知であるため、対応点から8点アルゴリズムを用いて Essential 行列を求める。

- (a) 対応点の同次座標表示  $m = (u \ v \ 1)^T$  と  $m' = (u' \ v' \ 1)^T$  を用いて、Essential 行列は次のように表せる。

$$m^T E m' = 0 \quad (1)$$

ここで、 $E$  行列の要素を1次元ベクトルに並べると、上記の方程式を

$$m^T e = 0 \quad (2)$$

に変形できる。ここで  $n$  個の対応点  $m_1 \cdots m_n$  が得られているとして、この各々について、式 (2) を縦に積み重ねて

$$A_n e = 0 \quad (3)$$

が得られる。ここで、

$$A_n = \begin{pmatrix} m_1^T \\ \vdots \\ m_n^T \end{pmatrix} \quad (4)$$

である。この方程式を  $e$  について解くことによって Essential 行列が定まる。これは  $A_n$  の特異値分解を用いると、最小固有値に対応した固有ベクトルとして求めることができる。

- (b) 求められた Essential 行列は一般に rank3 であるため、再度特異値分解し、rank2 であり Essential 行列の性質を満たすような特異値に修正する。
3. Essential 行列から Hartley のアルゴリズム [2] を用いて外部パラメータを求める。

- (a) Essential 行列を分解し、 $E = R \begin{bmatrix} T \\ \end{bmatrix}_x$  ( $R$ : 回転行列,  $\begin{bmatrix} T \\ \end{bmatrix}_x$ : 方向ベクトルの歪対称行列) を満たすような  $R, T$  の組を見つける。

- (b) 得られた  $R, T$  の組のうち、特徴点は両カメラの前面に存在するという条件から、物理的に正しい  $R, T$  の組を得る。

## 3 モデルとアルゴリズム

視点固定型モデルとは回転中心と投影中心が一致すると考えるモデルであるが、本来はこれらは別々

に存在する．本章では，新たに回転中心と投影中心が異なるカメラモデルを3.1節で定義し，このカメラモデルにおいてキャリブレーションを行うアルゴリズムを提示する．

### 3.1 カメラモデル

静的キャリブレーションでは撮影前にキャリブレーションを行い，そこからのパン・チルトの変化を加えて撮影時のカメラパラメータを計算する．これに対し，動的キャリブレーションでは撮影時の画像を用いてキャリブレーションを行う．このため，撮影にカメラ運動によるモーションブラーが影響を与えない限りにおいて，各フレームではカメラは固定されているとみなすことができ，カメラの動作を考慮に入れる必要がない．以上より次のようなカメラモデルを定義する．

- カメラは雲台に固定されており，雲台上で回転する．回転によって投影中心が微動する．つまり，各フレームにおいてはカメラは固定とみなすことができ，かつ，フレーム間においてもカメラは微動ししない．
- カメラはパン・チルトに回転が可能．ズームは固定とする．

### 3.2 アルゴリズム

#### 1. 撮影

- (a) 数メートル四方の撮影空間を用意し，周囲に撮影用のアクティブカメラを配置する．
- (b) 配置したアクティブカメラを視点固定カメラとみなして，静的キャリブレーションを行う．
- (c) 撮影空間内に撮影対象を配置し，撮影を行う．

この結果，撮影画像とともに，撮影時の雲台のパン・チルトが得られる．

#### 2. 特徴点検出

- (a) 撮影で得られた画像を，カメラモデルをピンホールカメラモデルとして考えることができるように補正を行う．

- (b) 得られたピンホールカメラモデル上の画像において特徴点を検出する．

#### 3. 特徴点の対応付け

- (a) 2節で得られた特徴点のうち，対応付けが可能であるものを視点固定カメラキャリブレーションの結果を参考にして選択する．
- (b) 空間中の同じ点を写していると考えられる，両画像中の2特徴点を対応付ける．
- (c) 可能な限り2.を繰り返す．

#### 4. カメラ対のキャリブレーション

3節で得られた対応点集合を元に，2章の原理を用いてキャリブレーションを行う．

## 4 評価実験

4.1節で本実験の目的を述べた後，4.2節で実験に対する評価基準を定義する．この評価基準の元で4.3節の実験を行う．

### 4.1 目的

本研究で提案するキャリブレーションの精度とその特性を検証するため，共通の評価基準の元で他のキャリブレーション方法と比較する．比較対象としては視点固定カメラの静的キャリブレーションを用いた [12]．

### 4.2 評価基準

対応点によって与えられるそれぞれのカメラの投影中心と各対応点を結んだ2本の直線の距離を評価基準として用いる．本研究で用いているキャリブレーションは撮影画像の対応点座標を用いており，この対応点から引かれる2直線は一般にねじれの位置にあり，この線間の距離が小さければ，空間中の真の特徴点を正確に表していると考えられる(図1)．よって，これを本研究におけるキャリブレーションの精度評価基準とする．

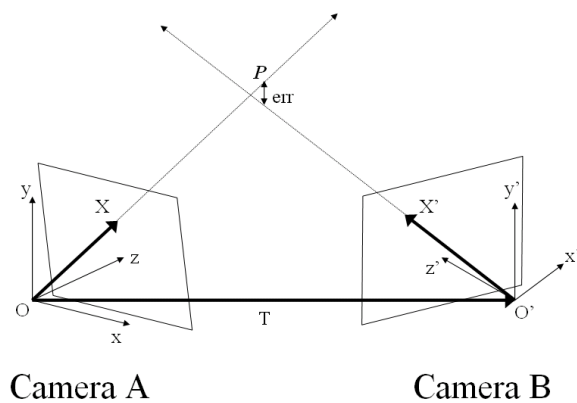


図 1: 評価基準．直線間の距離，つまり、中心の  $err$  の部分が評価基準となる．小さいほど評価値はよい．

### 4.3 実画像による評価

実際に撮影した画像を用いて精度を比較評価する．本研究ではピンホールカメラを仮定しているが，現実のカメラはより複雑な光学系で撮影を行う．このため，ピンホールカメラで写した画像として扱うことができるように初めに画像の補正を行う [5]．

アクティブカメラを用いているため，カメラは回転運動を行う．したがって，カメラの位置・姿勢は視点固定カメラキャリブレーションを行った位置・姿勢とは異なる．また，運動中に撮影を行うため，雲台から読み取られるパン・チルトの値が不正確となる．

これら 2 つの誤差要因に注目して精度比較を行う．4.3.1, 4.3.2 節で前者を，4.3.3, 4.3.4 節で後者を扱う．

#### 4.3.1 実験手順：カメラの回転運動

複数のカメラ位置・姿勢で撮影を行い，各カメラ状態においてそれぞれキャリブレーションの精度評価をする．カメラは部屋の中心より 3m ほど離れた床に，間隔を 1.5m ほど離して 2 台設置した．床からの高さは 80cm 程度であった．撮影は片側のカメラをパン方向に回転させ， $10^\circ$  刻みで行った．

そして，撮影した画像の組み合わせに対して動的キャリブレーションとその精度評価を行い，各画像のペアごとに反復した．

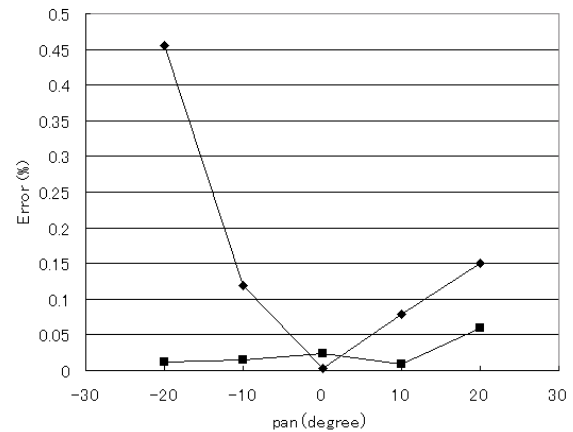


図 2: カメラの回転運動に対するキャリブレーション精度．横軸が静的キャリブレーション時のカメラ姿勢からのパン方向の回転角，縦軸が評価値． $\diamond$  : 静的キャリブレーション， $\square$  : 本手法

#### 4.3.2 評価：カメラの回転運動

図 2 は，事前のキャリブレーションを行ったカメラの位置・姿勢に対して，片側のカメラの位置・姿勢を固定し，もう片側のカメラを  $\pm 20^\circ$  の範囲でパン方向に首を振って撮影した画像に対してキャリブレーションの評価を行った結果である．ここでパン角の  $0^\circ$  とは，静的キャリブレーションを行ったときの向きを意味する．また，評価に用いた対応点は対象上もしくは背景上のものを用いている．図中の評価値は，各カメラ対から見た対応点の評価値の平均値となっている．

これより，視点固定カメラキャリブレーションではキャリブレーション時の位置・姿勢からずれるごとに誤差が増加するのに対し，動的キャリブレーションではほぼ一定の誤差範囲でカメラパラメータが求まっていることがわかる．ただしこの精度は，静的キャリブレーションにおいて，パン・チルトの読み取りが正確に行われている場合のものである．

しかし一方で，事前のキャリブレーションにおける評価値 ( $0^\circ$  のときの評価値) と比較すると，一桁ほど精度が悪くなっている．これは

1. 対応点からキャリブレーションを行う際に，SVD を用いて最適化を行っている．これは対

象とする線形連立方程式が定義された空間における最小2乗誤差解を与えるが、物理的に意味のある最適化となっていない [4]

2. キャリブレーションに用いている対応点の数が少ない
3. 特徴点抽出を手作業で行っているため、抽出精度が保てない
4. 計算機における数値計算上の誤差の扱いが最適化されていない

といった原因が考えられる。これらに対して、それぞれ

1. 現在のキャリブレーションを初期値として、投影誤差が最小となるように最適化をする
2. 特徴点検出・対応付けを自動化することで用いる対応点の数を増やす
3. 特徴点検出を自動化し、一定の抽出精度を保持する
4. 誤差の扱いを含むような実装をする

といったことが考えられる。

#### 4.3.3 実験手順：読み取り誤差

静的キャリブレーションでは、撮影時のカメラパラメータを求めるために、キャリブレーション時のカメラ姿勢からの差分パラメータ（ここではパン・チルト）を用いる。しかし一般に、雲台から得られたこれらの差分パラメータは誤差を含んでいる。

この誤差の影響を調べるために、4.3.1節で得られたデータを用いて、静的キャリブレーションに用いるカメラのパン・チルトの読み取りに誤差が含まれたと考えてシミュレーションを行う。具体的には、カメラの実際のパン・チルトに対して一定量の誤差が含まれていると考え、静的キャリブレーションの外部パラメータを求める際にこれらの誤差を重畳する。誤差に対する評価値は、様々な方向に誤差を加えて得られた結果を重畳したものの平均を取ることとし、

- 静的キャリブレーション時の向きにおける誤差の影響

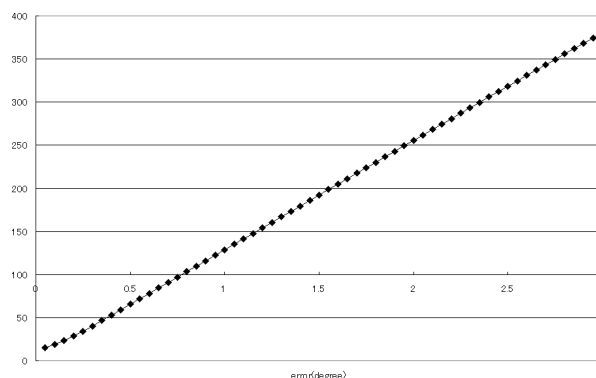


図 3: 静的キャリブレーション時のカメラ方向における誤差の影響。横軸が誤差の角度、縦軸が静的キャリブレーションの評価値 / 本手法の評価値。

- カメラの向きを変えた場合の評価値の変化

について調査する。実験では、片側のカメラの向きは固定し、もう片側のカメラに対して運動させたり誤差を乗せたりといった操作を行う。

#### 4.3.4 評価：読み取り誤差

図 3 は、1 対のカメラを静的キャリブレーションでキャリブレーションした向きに向けた状態で撮影した画像を用いて実験したシミュレーションの結果である。

実験では、本手法の評価値に対して、得られたパン・チルトに誤差が乗った状態での静的キャリブレーションの評価値をそれぞれの誤差の値に対して求め、これを比較する。

誤差の拡大に従い静的キャリブレーションの精度は悪化し、平均して約 0.75 °以上の誤差を含むような場合には本手法の方が有効であることが示された。

図 4 は、4.3.2 節の結果に、誤差を含んだ場合のものを加えたグラフである。静的キャリブレーションにおいては雲台からのパン・チルトの読み取りに誤差を含むと精度が劣化する。そのため、誤差が大きくなるとグラフは上にシフトし、一定以上の誤差（本実験の場合は約 0.8 °）を含むような場合に対しては従来法に比べて本手法の方が有効であることが示された。

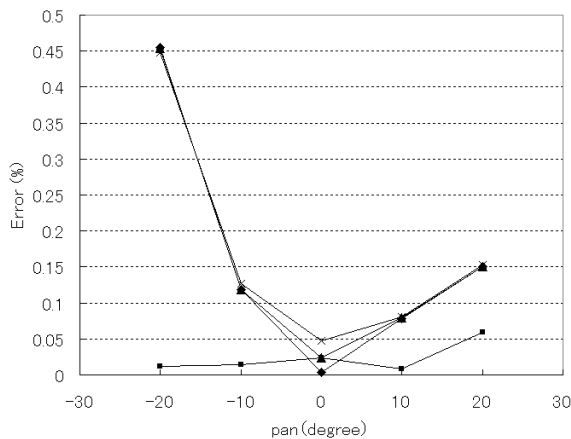


図 4: カメラの回転運動に対するキャリブレーション精度．横軸が静的キャリブレーション時のカメラ姿勢からのパン方向の回転角，縦軸が評価値．以下，静的キャリブレーションを SC と表記．□：本手法，○：SC（誤差なし），●：SC（誤差 0.8°），×：SC（誤差 1.6°）

## 5 結論

本研究では，広範囲高精細な 3 次元復元をするために，対象追跡映像を用いたカメラの動的キャリブレーションを

Step 1. まず各フレームでの多視点カメラ群のキャリブレーションを行い，フレーム間で独立な座標系におけるカメラの相対位置を求める．

Step 2. フレーム毎のカメラ位置と，世界座標系との間のキャリブレーションを行う．

というようなステップに分解し，前段のステップのうち最も基本となる 2 台のカメラのキャリブレーションを行った．そしてこの精度を視点固定カメラキャリブレーションのものと比較した結果，本研究で提案した動的キャリブレーションがある範囲の均一な精度でキャリブレーションを行うことができること，一定範囲以上のカメラ位置・姿勢の変化に対しては精度が劣化する従来法と比較して安定した精度を得ることができること，そして撮影時の Pan/Tilt が読み取り誤差を含むような場合には従来法は非常に弱く，これに影響されない本手法が有効であることを示した．

未解決の問題としては

- 3 台以上のカメラのキャリブレーション
- 精度の向上
- 対応点抽出の自動化

等が考えられ，より頑健で精度の高いキャリブレーションの実現が望まれる．

## 参考文献

- [1] Lourdes Agapito, E. Hayman and I. Reid, Self-Calibration of Rotating and Zooming Cameras, International Journal of Computer Vision 45(2) (2001) 107-127
- [2] Richard I. Hartley, An Investigation of the Essential Matrix (1993)
- [3] Richard I. Hartley, Self-Calibration of Stationary Cameras, International Journal of Computer Vision, Vol22 Issue1 (1997) 5-23
- [4] 三島等 金谷健一, 基礎行列の最適計算とその信頼性評価, 情報処理学会研究報告 99-CVIM-118-10 (1999) 67-74
- [5] 松山隆司 久野義徳 井宮淳, コンピュータビジョン 技術評論と将来展望, 新技術コミュニケーション (1998) 10 37-53 83
- [6] T. Matsuyama, S. Hiura, T. Wada, K. Murase and A. Yoshioka, Dynamic Memory: Architecture for Real Time Integration of Visual Perception, Camera Action, and Network Communication, Proc. of Computer Vision and Pattern Recognition Conference (2000) 728-735
- [7] C. Poelman and T. Kanade, A Paraperspective Factorization method for Shape and Motion Recovery, Computer Science Department (1993)
- [8] M. Pollefeys, R. Koch and L. Van Gool, Self-Calibration and Metric Reconstruction in spite of Varying and Unknown Internal Camera Parameters, International Journal of Computer Vision, 32(1) (1999) 7-25

- [9] C. Tomasi and T. Kanade , Shape and motion from image streams under orthography: A factorization method , Int. J. of Computer Vision , Vol.9 No.2 ( 1992 ) 137-189
- [10] 和田俊和 , 浮田宗伯 , 松山隆司 , 視点固定型パン・チルト・ズームカメラとその応用 , 電子情報通信学会論文誌 , Vol.J81-D-II No.6 ( 1998 ) 1182-1193
- [11] Norimichi Ukita and Takashi Matsuyama , Real-time Cooperative Multi-target Tracking by Communicating Active Vision Agents , Computer Vision and Image Understanding , Vol.97 No.2 ( 2005 ) 137-179
- [12] J. kondou, X. Wu and T. Matsuyama, Calibration of Partially-Fixed Viewpoint Active Camera, CVIM2, Vol.44 No.SIG3 ( 2003 )