

Webオークション出品時における疑似3D画像の生成

西 紗記子^{1,a)} 東野 進一¹ 坂本 竜基¹

概要: Webオークションでの商品説明における画像は、落札者に対する重要な情報源である。出品者は商品の実態がよく伝わるように、デジタルカメラを用いて商品を様々な角度から撮影するが、各画像の見えはあまりに異なるため、すべてを連続して閲覧しても商品を立体的にとらえることは困難である。一方、商品をターンテーブルに乗せ、カメラを固定したうえで、少しずつ回転させながら多数の画像を撮影し、それらを連続表示すれば立体感がでるが、ターンテーブルの準備や撮影の手間がかかるため、現在のところ大衆化されるまでには至っていない。本稿では、初心者によって商品を周囲から簡易的に撮影された画像群から、あたかもターンテーブルで回転させながら撮影したかのような画像群へと変換する手法を提案する。

Pseudo-3D Image for Online Auction Services

SAKIKO NISHI^{1,a)} SHINICHI HIGASHINO¹ RYUUKI SAKAMOTO¹

Abstract: For online actions, potential bidders obtain the item information through the web. Images as a part of the descriptions have significant effects on decision-making by reducing the concern of the bidders. Therefore sellers prefer to take pictures from different perspectives using digital camera to appeal the actual condition of the item. However, these pictures can hardly provide the 3D sensation experience as in a real-life scene. This can be achieved by placing the item on the turntable, and taking pictures while the turntable rotates slightly, and displaying the multiple pictures continually. However since the preparation of the turntable and the time-cost for shooting, it has not yet been popularized currently. In this paper, we propose a method for novice users to convert multiple pictures that are simply taking surrounding the item and provide the same experience as if using a turntable.

1. はじめに

Web上に流通する媒体はテキストから始まり音、画像、動画へとリッチになりつづけてきた。近年は単なる動画ではなく、HMDで没入感をもって閲覧できるよう両眼立体視を模した映像や、全天球パノラマ画像・映像も普及しつつある。これらは究極の超広角画像・映像であり、風景等の全体的な景観の把握に適している。例えば、ホテルや不動産の内部からみた部屋全体の見えや、景勝地やイベントにおける眺望や情景をこのようなパノラミックな画像・映像で撮影・提供することは、実際にその場に行き見ることができたであろう風景により近く有益であろう。

一方で、Webにおける通信販売やオークションにおい

て、商品をビジュアル情報によって効果的に表すには、このような画像・映像は適していない。なぜならば注目したい被写体は全天球に対してごく一部しかなく、その他の多くの部分が無駄になるからである。また、広範囲に撮影するといっても視点は一点であるため、被写体を立体的に閲覧することはできない。被写体を立体的に表現する最も単純な方法は、ビデオカメラで被写体の周囲を回りながら映像を撮影することである。しかし、高価なドリーやスタビライザーなどの業務用機器を導入しない限り、映像に手振れや振動がおこるため商品紹介のための媒体としては適さない。商品をターンテーブルに乗せ、カメラをターンテーブルにむけて固定したうえで、少しずつ回転させながら多数の画像を撮影すれば振動や手振れがなく撮影可能であるが、ターンテーブル調達や撮影のコストから現在のところ大衆化されるまでには至っていない。

これに対して、被写体の周りから簡易的に撮影した振動

¹ ヤフー株式会社
Yahoo Japan Corp. Mid9-7-1 Akasaka, Minato-ku
^{a)} snishi@yahoo-corp.jp

が含まれる映像を何等かの方法で滑らかに補正することで、商品を立体的に見せるためのサービスも導入されはじめている^{*1*2}。特に、被写体の周囲から撮影した画像群の中心にある共通する三次元位置に固定したかのような画像へ変換し、それらを連続的に見せる手法（以下、疑似 3D 画像と呼ぶ）は、画像間の補間や被写体の 3 次元モデルの推定といった不確実な要素を含まないため、見えに対して頑健である。しかし、自然特徴点に基づいて補正をするとすべてのシーンで確実かつ正確に補正をおこなうことは不可能であるため、ビジュアルマーカによる校正が適切である。本稿では、Web オークションなど、撮影者のスキルや事前の統制が難しい状況における不確実な自然特徴点ではなく明示的なビジュアルマーカによって補正する手法を提案する。

2. 先行研究

多視点から撮影された画像を 3 次元位置のある一点で各カメラの光軸が重なった状態で撮影されたかのような画像へ変換するための射影変換行列の導出は、各カメラが強校正、つまり位置姿勢を含むカメラパラメータが既知である前提である。カメラパラメータは自然特徴点から SfM で推定することが可能だが、安定した計算のためには撮影される 3 次元空間内に特徴点が遍在している必要がある。特徴点が見込めない場合や計算量を削減したい場合は、ARToolKit[1] をはじめとするビジュアルマーカを利用すれば強校正可能であるが、推定されるカメラパラメータの精度に問題がある。例えば、ARToolKit は正方形のビジュアルマーカを用い、その正方形の 4 辺からなる平行線の組からマーカとカメラとの回転角 (R) と平行移動量 (T) を推定するが、1 辺の長さが短い場合、 R の精度は高くない。本研究では、Web オークションでマーカと商品を同時に撮影し、なおかつ商品の周囲からカメラで撮影することを想定している。よって、マーカは比較的小さく、遮蔽にそなえて複数個配置する必要があるため、平行線に基づいたマーカは適していない。

平行線以外のビジュアルマーカとしては、複数個の円の見えに基づいてカメラパラメータを推定する方法がある。これは、ユークリッド空間における真円は楕円としてピンホールカメラモデルの画像平面に投影されるため、その楕円のパラメータから R を推定するアプローチである。Rune Tag[2] は小さな円を円状に並べて、その並びからタグの ID、円のパラメータから R と T を求めるビジュアルマーカであり、遮蔽に強いという特徴があるものの、構成要素が小さな円であるため、カメラが離れたり、解像度が低いカメラであると小さな円同士を分離できなくなったり、円としてフィッティングできなくなるという問題が

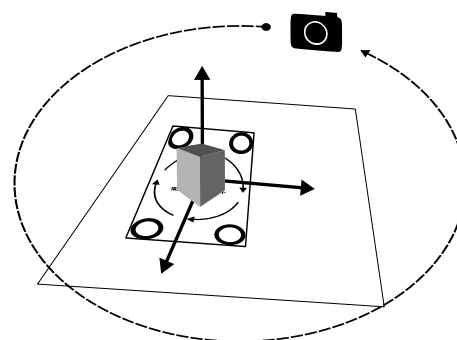


図 1 シートに商品をのせて撮影するイメージ

Fig. 1 The concept of capturing an item on the sheet.

ある。これに対して、本研究におけるビジュアルマーカは、比較的大きな円を要素としているため、このような問題は起こりにくい。

3. 実装方法

3.1 複数の円に基づく強校正

まず、出品者には一般的な家庭用プリンタで A4 用紙に四隅に大きさの異なる円が描かれたシートを印刷してもらい、シートの中心に商品を置いてもらう。その後、一般的なデジタルカメラやスマートフォンによって横方向に少しずつ角度を変えて撮影してってもらい、その画像群が入力されることを想定する（図 1）。よって、商品の体積がある一定以上であれば入力画像における四隅の円は、ほとんどの画像において 1 つないし 2 つの円は商品によって遮蔽されて見えない。

画像から検出された各楕円は、その楕円パラメータから法線が推定できる。この法線は 1 つの円につき 2 本算出されるので 2 つ以上の円の解を組み合わせると、その円を含む平面の法線を推定することができ、これをワールド座標系における Y 軸とするならば適当な円心同士を結ぶベクトルを仮の X 軸として R を求め、円心同士の距離を 1 とすると仮の T を推定できる。この仮の RT をもちいて、検出された円の本来の大小関係をもとめれば、どの円が楕円として検出できたのかを決定することができる。シートにおける円の位置は既知であるから、検出された円のワールド座標系における位置から、正しい X 軸と T を推定することができ [3]、これにより強校正が完了する。

3.2 3 次元的に光軸をそろえる射影変換

入力された画像群は商品の方向にむけて撮影されている想定ではあるが、厳密には光軸は上下左右にばらついており、入力画像をそのまま連続的に閲覧しても不連続性が際立ち、商品の説明画像としてふさわしいクオリティであるとは言い難い。そこで、各画像が撮影された際のカメラの光軸が 3 次元空間上の、ある 1 点を通っていたかのような画像へと変換する射影変換行列を導出する。

*1 Fyuse <https://fyu.se/>

*2 疑似 3D 画像 <http://apps.yahoo-labs.jp/p3d/>

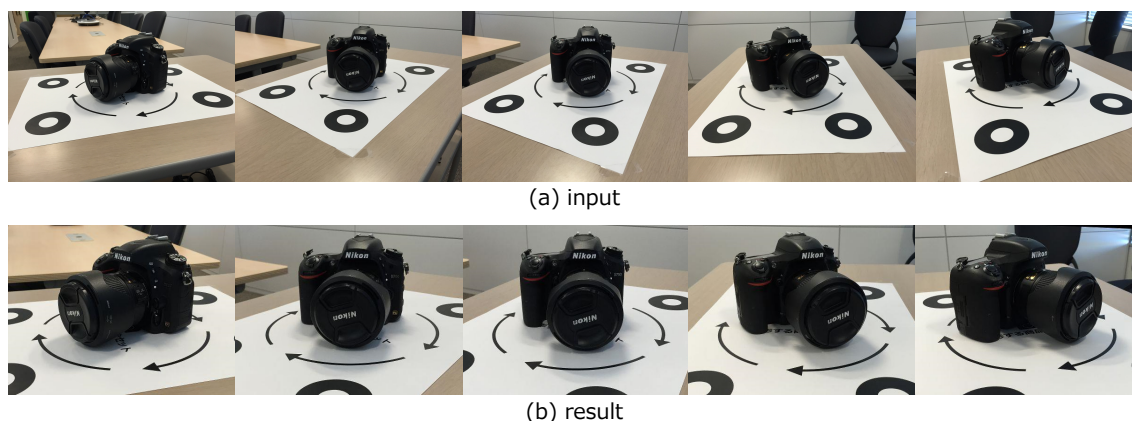


図 2 入力した 25 枚の画像の一部と変換結果
Fig. 2 A part of input images and the result.

このようなカメラ k における射影変換 $\mathbf{H}_k$ は、各カメラの $\mathbf{R}$ および光軸を通過させたい 3 次元点から以下の式で求めることができる。

$$\mathbf{H}_k = \mathbf{A}_k \mathbf{R}'_k \mathbf{R}_k^{-1} \mathbf{A}_k^{-1} \quad (1)$$

ここで、 $\mathbf{R}_k^{-1}$ はカメラ k における元の $\mathbf{R}$ であり、 $\mathbf{R}'_k$ は新しい回転、 $\mathbf{A}_k$ は内部パラメータである。新しい回転 $\mathbf{R}'_k$ は、カメラ位置とワールド座標系における回転させたい中心から求まる理想的な光軸と、元の光軸から求めることができる [4]。回転させたい中心は任意に設定でき、シートの中心でもよいし、変換後の余剰領域が減るように最適化してもよい [5]。

4. 実験

印刷したシートの上に商品にあたるデジタル一眼レフカメラを置き、スマートフォンのカメラを用いて全周方向から 25 枚撮影した画像のうち、45 度程度の部分を抜き出したものを図 2(a) に示す。奥に存在するはずのマーカーが商品で遮蔽されているが、射影変換した結果である図 2(b) では正確に強校正されて変換されていることがわかる。

5. おわりに

本稿では Web オークションにおいて商品をターンテーブルに置いて周囲から閲覧しているかのような疑似 3D 画像を生成する手法を提案した。疑似 3D 画像の生成にはカメラの強校正が必要であるが、オークションのような様々な撮影状況においては自然特徴点よりもビジュアルマーカーに基づいた校正のほうが頑健である。今後は、この頑健さを実験により証明していきたい。

参考文献

[1] Kato, H. and Billinghurst, M.: Marker tracking and hmd calibration for a video-based augmented reality conferencing system, *Proceedings. 2nd IEEE and ACM International Workshop on Augmented Reality(IWAR'99)*, pp.

85–94 (1999).
[2] Bergamasco, F., Albarelli, A., Rodola, E. and Torsello, A.: RUNE-Tag: A high accuracy fiducial marker with strong occlusion resilience, *CVPR2011*, pp. 113–120 (2011).
[3] Chen, D., Sakamoto, R., Chen, Q. and Wu, H.: Extrinsic camera parameters estimation from arbitrary coplanar circles, *SIGGRAPH Asia 2014 Posters*, ACM, p. 28 (2014).
[4] 富山仁博, 宮川 勲, 岩館祐一: 多視点ハイビジョン映像生成システムの試作: 全日本体操選手権での中継番組利用, 電子情報通信学会技術研究報告. PRMU, パターン認識・メディア理解, Vol. 106, No. 429, pp. 43–48 (2006).
[5] 坂本竜基, 陳 鼎: 解像度を維持する Bullet Time 用射影変換行列の導出および評価, 情報処理学会論文誌, Vol. 56, No. 5, pp. 1388–1397 (2015).