

コンシューマ・システム論文

携帯端末による TV 画面の位置と姿勢の推定方法

川喜田 裕之^{1,2,a)} 中川 俊夫¹ 佐藤 誠²

受付日 2015年1月26日, 採録日 2015年5月21日

概要: 我々は、テレビの映像を画面外に拡張するサービス “Augmented TV” の研究を行っており、これまでに TV 映像に AR (Augmented Reality) の技術を適用したシステムを提案してきた。提案システムでは、携帯端末のカメラを通してテレビを見ることで、携帯端末のカメラから取り込んだ画像に 3DCG アニメーションをオーバーレイ表示することにより、テレビ画面内のキャラクターがテレビ画面の外に飛び出してくるように見える演出が可能である。このような演出を実現するためには、携帯端末で TV 画面の位置と姿勢を常時高精度に推定することが必要であり、我々はこれまでにカメラとジャイロセンサをハイブリッドに用いる推定方法を提案してきた。今回、方式の詳細について議論するとともに、市販のタブレットなどに実装して処理速度の評価を行い、方式の実用性を確認したので報告する。

キーワード: 拡張現実感, 位置姿勢推定, 画像処理, ジャイロセンサ, 処理速度

Estimation of TV Screen Position and Rotation Using Mobile Device

HIROYUKI KAWAKITA^{1,2,a)} TOSHIO NAKAGAWA¹ MAKOTO SATO²

Received: January 26, 2015, Accepted: May 21, 2015

Abstract: We have studied on a new TV service, named “augmented TV”, which is able to augment representation of TV programs beyond the TV screen. We have proposed a system in which animated 3DCG content interlocked with TV programs is overlaid on live video from the mobile device camera in the mobile device screen by augmented reality techniques. In the system, the representation of having a TV character coming out of the screen can be provided. To achieve such a representation, it is needed always accurately to estimate position and rotation of the TV in the mobile device. We have proposed an estimation method using the camera and the gyro sensor. This paper shows the details of our method and that it is effective by experiments using the demo content.

Keywords: augmented reality, 3DCG, augmented TV, position & rotation estimation, gyro sensor

1. はじめに

約 60 年前に TV 放送が開始されてからこれまでにカラー化, 高精細化, ワイド (アスペクト比 16:9) 化, 2 眼式の立体テレビ化などが進められてきた。これらの映像技術の進歩における主な目的の 1 つは、あたかも目の前にカメラでとらえた空間が広がっているかのような感覚としての「臨

場感」の向上である。我々は、さらに一步広げて、画面内のカメラでとらえた世界が画面手前の現実世界とつながっているかのように見せることをゴールとする “Augmented TV” を提唱してきた [1], [2], [3]。このゴールに対して我々は、一般の視聴者が Augmented TV のサービスを利用できるようにコンシューマデバイスを用いて構成を試みるアプローチを採用している。コンシューマデバイスは、近年センサ類の搭載など顕著に多機能化しており、そうした既存の機能を活用することで Augmented TV の可能性を探求する。

我々は、Augmented TV の一形態として、TV 映像に対して AR (Augmented Reality : 拡張現実感) の技術 ([4]

¹ NHK 放送技術研究所
NHK Science & Technology Research Laboratories,
Setagaya, Tokyo 157-8510, Japan

² 東京工業大学
Tokyo Institute of Technology, Yokohama, Kanagawa 226-
8503, Japan

^{a)} kawakita.h-dq@nhk.or.jp

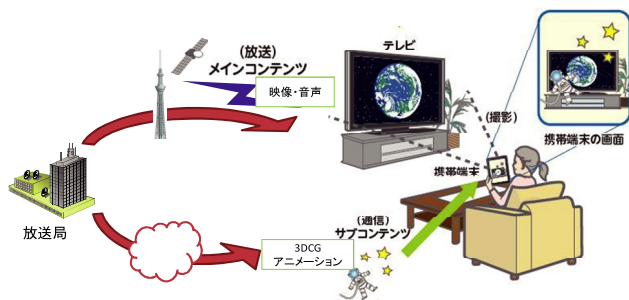


図 1 Augmented TV のサービスモデル
Fig. 1 Service model of augmented TV.

など)を適用することにより、映像を画面外に拡張するシステムを提案している(図 1)。提案システムでは、TV 画面をスマートフォンやタブレットなどの携帯端末(以下、携帯端末)に内蔵されたカメラで映して見ることにより、カメラで取り込んだTV映像に対して3DCGアニメーションを重ね合わせることでTV画面内のキャラクターなどが画面外に飛び出してくるといった演出が可能である。このように視聴者が通常のTVでは予想できないような状況を演出することにより、驚きやキャラクターの実在感をともなうコンテンツを放送と通信が連携したマスメディアサービスとして提供することが期待できる。

提案システムでは、図 1 のようにTV画面に表示される映像は放送コンテンツで、3DCGアニメーションは通信経由であらかじめ携帯端末にダウンロードしてあることを想定している。キャラクターの実在感を演出するためには、キャラクターが画面の内と外を違和感なく出入りしているように見せる3DCG描画の時刻や位置が重要となる。携帯端末において、カメラで取り込んだTV映像と3DCGアニメーションとの時刻同期やTV画面の位置(併進成分)・姿勢(回転成分)を正しく推定する必要がある。

Augmented TVと同様のサービスを想定した研究の1つに、時刻を同期するために電子透かしを用いた研究[5]があるが、キャラクターが違和感なく出入りするように見えるほど同期精度が十分とはいえない。我々は、出入りのスムーズさを重要視し、より高精度に時刻同期する方式[1]を開発するとともに、カメラとジャイロセンサを用いたTV画面の位置・姿勢の推定方式[2],[3]などを開発してきた。今回は、これまでの研究成果をふまえ、位置・姿勢の推定方式について新しいアルゴリズムも加えて詳細に議論することでカメラとジャイロセンサをハイブリッドに用いることの有効性を明らかにし、実験などにより検証する。

2. TV画面の位置・姿勢の推定方式の要求条件

2.1 従来方式の制約

我々は、文献[1]においては、TV画面の位置・姿勢を推定するために、Augmented TVの視聴環境に制約を設けていた。TV画面を検出するための手がかりとするため



(a) マーカー式 (b) 白黒「枠」式

図 2 TV画面を検出するための手がかり
Fig. 2 Cue to detect TV screen.

にTV画面の四隅につねにマーカーを表示し、携帯端末のカメラでつねにマーカーをとらえて、マーカーの位置と向きを画像処理により検出して推定する方式[4]であった(図 2(a))。したがって、携帯端末のカメラでTV画面をとらえていない場合は推定することができない。

この制約は、TV画面のすぐ近辺でしか演出に使用できないことを意味している。たとえばTV画面から出てきたキャラクターが自由に部屋の中を大きく飛び回るといった演出は困難である。また、視聴者がつねに携帯端末のカメラでTV画面をとらえ続けなければいけないという条件を負担に感じる可能性もある。そこでこの制約を解消し、携帯端末のカメラでTV画面をとらえていない場合であってもTV画面の位置・姿勢を推定できる方式が望まれる。

また、TV画面にマーカーをつねに表示するというのも演出上の大きな制約である。文献[1]では実用的なマーカーのサイズについては議論したが、マーカーのデザインパターンの演出に対する影響については触れていない。マーカーは、その位置を一意に特定するという目的から固有のデザインパターンであるがゆえに、演出意図に反して目立ってしまう可能性がある。TV画面の検出の手がかりは、演出を阻害しない程度のシンプルなデザインが望ましい。

2.2 要求条件

推定方式は2.1節で述べた制約を低減するだけでなく、今後の技術開発の進展を想定した有用性の高い方式とすることが望ましい。近年、携帯端末に搭載されるカメラも“2K”から“4K”(水平画素数の概数を示す)へと高精細化してきており、フレームレートも30fpsから60fpsへと高速化が図られている。提案システムでは携帯端末を手で把持して動かすため、速い動きにも対応するには高フレームレートが望ましい。その一方で、AR分野などでは“ウェアラブル機器”といった小型で軽量ゆえに従来のPCなどと比べて計算資源が乏しい機器が用いられることも多い。以上から、高精細化・高フレームレートの映像に対して計算資源が乏しい機器でも対応できる処理方式が望まれている。

提案システムでは、携帯端末においてカメラ取込み画像の各フレームに対して3DCGのレンダリングを行うため、フレームレートに合わせてリアルタイムに推定処理を行う

必要がある。

また、マスメディアサービスとしての実用性を考慮すると、デバイスだけでなく照明やインテリアなどの視聴環境の多様性も考慮に入れる必要がある。今回、TV画面や携帯端末に対して機能や構成を変更せずにコンシューマデバイスをそのまま用いることを前提とする。

2.1節、本節での議論をまとめて、推定方式の要求条件を次のように定義する。

- 携帯端末のカメラがTV画面をとらえていないときも推定が可能であること
- TVに表示するTV画面の検出の手がかりは演出を阻害しないほどシンプルなデザインであること
- 計算資源の乏しい携帯端末でもリアルタイムに処理ができるように処理負荷が小さいこと
- TV画面以外の環境には影響を受けにくいこと

3. 提案する推定方式

3.1 アプローチ

携帯端末のカメラで一度認識したTV画面がフレームアウトした場合に、TV画面の位置・姿勢を推定するためには、携帯端末に搭載されたセンサにより自身の位置・姿勢を推定して求める方法と、画像処理によってTV画面以外の家具や壁の外見の特徴量を抽出して推定する方法 [6] がある。前章で述べたとおり、今回はTV画面以外の環境には影響を受けない方式を目指しているため、カメラが画面をとらえていない場合は前者の方法であるジャイロセンサを用いて推定する方式を考案した。

ジャイロセンサは、近年多くの携帯端末に搭載されてきたが、携帯端末の角速度を計るセンサであり、この値を積算すれば積算開始からの携帯端末の姿勢が分かる。Augmented TVでは主に座りながら携帯端末を動かして見ることが想定されるが、その場合などに携帯端末を回転させて利用する限りはジャイロセンサを用いてTV画面の位置と姿勢を推定できる。ただし、角速度を積算することから、ジャイロセンサのみを利用した場合は推定結果に測定誤差や計算誤差が蓄積するため、時々キャリブレーションを行わないと次第に推定が誤ってしまう。そこで、推定方式としては、ジャイロセンサによる推定結果をベースにしつつも、TV画面をカメラでとらえた場合には画像処理による推定を優先する方式を考案した。

さらに、ジャイロセンサによる推定でおおまかな位置が分かるため、画像処理によるTV画面の探索範囲を限定することができる。その結果、TV画面の検出の手がかりとして従来のマーカー方式 (図2(a)) のような目立つものにする必要がなくなり、シンプルなデザインにすることが可能である。今回、TV画面に図2(b)に示すような白黒の枠 (以下、枠) を表示させて認識することとした。2色用いることで、白と黒の境界線を認識すれば、表示する映像

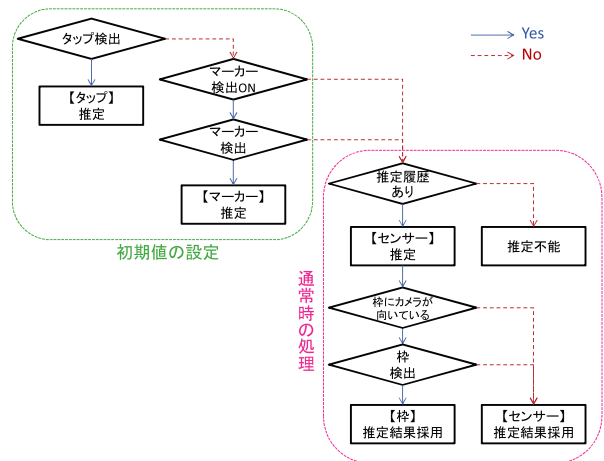


図3 全体フロー

Fig. 3 Flow chart of the estimation.

やTVのベゼル (外枠) のデザインや材質に依存しない枠線を確保することができる。

3.2 「推定」の幾何学的な意味

提案するアルゴリズムを説明する前に「TV画面の位置・姿勢を推定すること」の幾何学的な意味について述べる。TV画面の位置は、TV画面を基準とする3次元座標系 (以下、TV座標系) において固定値であり、TV座標系から携帯端末の画面の2次元座標系 (以下、スクリーン座標系) に変換する行列を求めることができれば、携帯端末の画面においてTV画面がどこにあるのかが分かる。この変換は、TV座標系から携帯端末のカメラを基準とする3次元座標系にビュー変換し、さらにスクリーン座標系に射影変換することに相当する。幾何学計算においては、それぞれの変換を行列で扱い、ビュー変換行列と射影変換行列と定義する。TV画面と携帯端末の相対的な位置・姿勢関係を示すのはビュー変換行列であり、推定することとはこのビュー変換行列を求めることに相当する。

ビュー変換行列を3次元座標系とスクリーン座標系との対応関係から求める問題は、Perspective-n-Point (PnP) 問題として知られ、解を特定するためには少なくとも4点の対応が必要となる。今回は枠の白黒の境界線で表される矩形の4頂点をこの対応点とし、実際にビュー変換行列を求めるためにOpenCV [7] の cv::solvePnp() 関数を用いた。

3.3 全体フロー

図3と表1に、提案する推定方式の全体フローを示す。カメラから入力されるフレームごとにこのフローが実行される。本方式では、ジャイロセンサによる推定のために必要なTV画面の初期位置・姿勢を得た後、ジャイロセンサによる推定を起点としてカメラからの画像を処理して枠を認識する。

表 1 推定方法

Table 1 Estimation method.

【タップ】	タップした位置から上下左右に探索を開始する【枠】推定
【マーカ-】	マーカ-検出による推定(従来方式)
【センサー】	直近の推定結果にジャイロセンサーの回転量を反映させた推定(3.5節)
【枠】	枠を画像処理により探索する推定(3.6節)

3.4 初期位置・姿勢の取得

ジャイロセンサによる推定を行うためには、TV 画面の位置・姿勢の初期値が必要である。初期値を得るために、タップ（携帯端末のタッチパネルを指で軽くたたく動作）による視聴者の能動的な指示による推定と文献 [1] と同様のマーカ-による自動認識による推定を用意した。

タップによる推定は、初期値の設定だけでなく、ジャイロセンサの積算値に大きな誤差が蓄積してしまったときなどに、画像処理によって枠が認識できない場合に視聴者が明示的に位置・姿勢のキャリブレーションを行う方法としても利用することができる。

マーカ-による推定は、放送局などのコンテンツ提供側の意志で表示切り替えができるため、コンテンツ開始時に強制的に初期設定することに利用できる。また、データ放送などを利用して TV リモコンによる操作でマーカ-の表示の ON/OFF を切り替えれば、タップによる推定と同様に、視聴者によるキャリブレーションに利用できる。

3.5 ジャイロセンサによる推定

ジャイロセンサによる推定の手順を示す。まず、ジャイロセンサの値を取得する。ここで、ジャイロセンサは直交する 3 軸のものを想定しており、独立した各軸回りの角速度が得られる。次に、各値に前回推定時刻からの差分時間を乗じることで、各軸の携帯端末の姿勢の変位角を求める。各変位角に対する回転変換行列を求め、任意の順序で前回の推定結果のビュー変換行列に乗じることで、今回の推定結果としてのビュー変換行列を求めることができる。

3.6 枠認識

3.5 節で求めたビュー変換行列を用いるとスクリーン座標系において TV 画面の位置を推定することができる。その位置を基点として画像処理により白黒の枠を発見し、枠の頂点を見つける枠認識のアルゴリズムについて述べる。

3.6.1 関連研究とアプローチ

枠を認識するためには枠の画像特徴により探索を行うが、ある線分上をヒューリスティックにたどりながら探す方法と、カメラで取り込んだ画像上をくまなく探す方法に大別され、一般的に計算量と認識率とのトレードオフの関係にある。特にその計算量は、1 辺のサイズを N とすれば、前者は $O(N)$ であり、後者は $O(N^2)$ となる。前者の

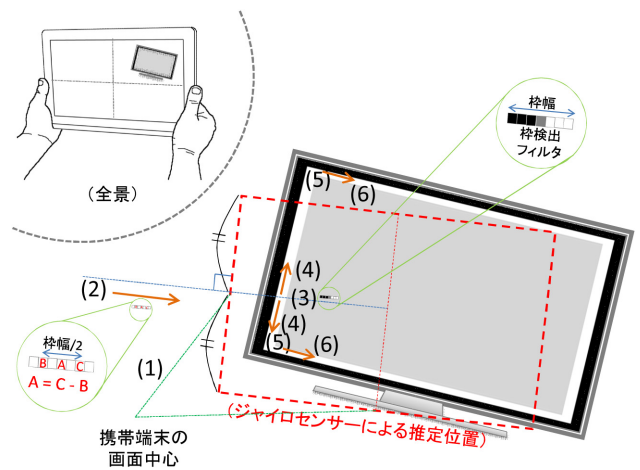


図 4 枠の探索経路

Fig. 4 Search path of TV frame.

方式の関連研究として、画像の縁を探索するアルゴリズム [5], [8], [9], [10] がある。文献 [8] には、後者の方式に分類されるハフ変換やパターンマッチングに比べて 1/100 以下の処理時間で探索できることが示されている。しかし、これらの関連研究は、すべて電子透かしの検出を目的としているためにカメラフレーム内に対象の画像が収まっていることやカメラフレームの中央に画像が配置されることなどを前提としており、今回のように枠の位置が不定のケースには適用できない。また、文献 [8] と [9] は静止画を対象としており、枠を設けない方式である。そのため、文献 [9] によれば画像領域の検出成功率はそれぞれ 61% と 93% であり、動画に適用して安定したビュー変換行列を得る程度には、画像と画像以外の区別が十分であるとはいえない。

以上をふまえ、今回は次の 3 つの点を考慮してヒューリスティックにたどりながら枠を探す方式を採用し、動画を対象とした文献 [5], [10] と同様に、枠を設ける方式とした。

- 計算資源の乏しい携帯端末でも処理時間が短いこと
- ジャイロセンサにより探索範囲を限定できること
- TV 画面に枠を表示することにより照明などの環境に依存しない検出の手がかりを利用できること

特に今回は高精度に推定するために、カメラで TV 画面を一部分とらえた場合であっても画像処理による推定を優先する方式を考案した。

3.6.2 枠認識アルゴリズム

図 4 に、画像処理により枠を認識する提案方式の探索経路の例を示す。探索アルゴリズムは以下のとおりである。

- (1) スクリーン座標系において、3.5 節で得られたビュー変換行列を利用して枠の推定位置（図では破線）と枠の幅（白枠と黒枠の合計）の推定値を求める。さらに、推定される枠の各辺の midpoint から携帯端末の画面の中心までの距離を求め、距離が最も短い辺を最初の探索辺とする。
- (2) 探索辺の midpoint を垂直に横切る直線上を、枠外から枠の

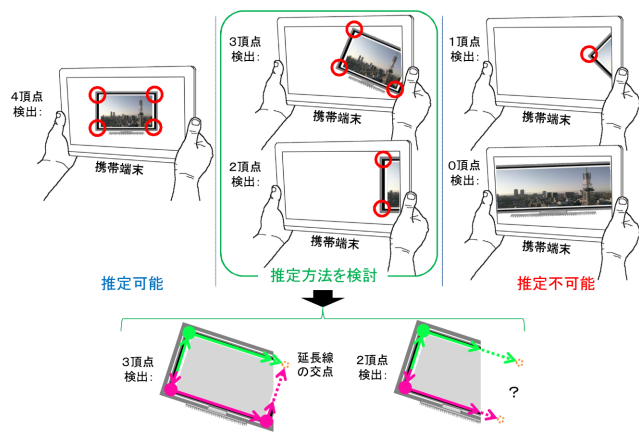


図 5 枠を部分的にとらえた場合の分類

Fig. 5 Classification of captures of TV frame.

中心に向けて辺長の長さ分だけたどる。このとき、枠の幅の半分だけ進行方向に離れた画素との輝度値の差を記録し、閾値を超えたものを枠候補として記録する。枠候補がなければ (7) へ。

- (3) (2) の候補について、枠の幅のサイズの「枠検出フィルタ」により閾値を超えたものを枠候補として選抜する。枠候補がなければ (7) へ。
- (4) (3) の地点から、(3) と同じフィルタを用いて黒白の境界線を両側とも探索する。探索の初めに境界線が伸びている方向を 8 方位のうちから隣接する 2 候補を決定し、その方向にのみ探索を行う。
- (5) (4) の探索が途切れたときに (3) の地点から一定長 (枠の幅の定数倍) にわたって探索ができていれば辺として有効とする。その地点から直角に、(4) と同様に次の辺が想定される向きに探索を行う。
- (6) 片側の探索ごとに、探索が途切れるか、3 つの辺を探索した時点で探索を終了する。(3) の地点以外の辺と辺の接続点を頂点とする。2 頂点以上検出できていれば終了し (頂点数が 2 や 3 の場合の推定方法は 3.5.3 項で述べる)、そうでなければ (4) に戻り、次の枠候補にての探索を行う。次の枠候補がなければ (7) へ。
- (7) (1) で次に距離が短い辺を探索辺として (2) へ。探索辺がなければ枠認識できないことを出力して終了する。

3.6.3 TV 画面の一部分をとらえた場合の推定

3.6.2 項のアルゴリズムでは探索で 4 頂点検出できれば TV 画面の位置・姿勢が推定できるが、カメラが枠の一部分をとらえたことにより検出した頂点数が 2 や 3 の場合であっても、とらえた部分的な枠を活かした推定を行う方式を考える。今回は、図 5 のように頂点検出数で分類し、3 頂点あるいは隣接する 2 頂点を検出した場合の推定方法を検討した。

3.6.3.1 3 頂点を検出した場合の推定

3 頂点のみを検出した場合は、単純にスクリーン座標系において、3.6.2 項のアルゴリズムでの探索が途切れた 2

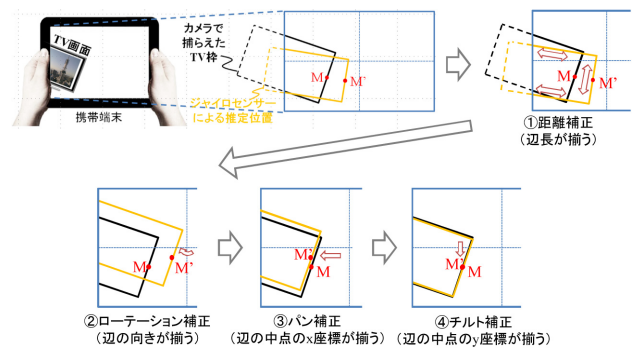


図 6 2 頂点検出時の補正

Fig. 6 Rotation correction at detection of 2 vertices.

つの線分を延長し、それらの交点を残りの頂点位置とする (図 5 の左下)。以降は 4 頂点検出と同様な推定が可能である。

3.6.3.2 2 頂点を検出した場合の推定

隣接する 2 頂点のみを検出した場合は、3 頂点の場合のように画像処理のみによる方法で残りの頂点位置を推定することは原理的に不可能である。そこで、検出した 2 頂点の位置を活用してジャイロセンサの推定結果を補正する方法を検討した。

ジャイロセンサのみによる推定を続けているとその結果には、次の誤差が発生する。

- A) 携帯端末は回転させるだけで併進運動をともしない場合に、ジャイロセンサによる推定の過程で累積していく携帯端末の姿勢の誤差
- B) 携帯端末が併進運動をともしなう場合に、運動した分の位置の誤差と運動により TV 画面との方位関係が変化する分の姿勢の誤差

上記 A の誤差を補正するためには、最終的な推定結果がカメラでとらえた部分的な枠に合うように、携帯端末の姿勢推定を補正することで対応が可能である。

図 6 に、スクリーン座標系における提案する補正方法を示す。まず、視聴者が座っているなど、位置を動くことなく携帯端末を回転させるケースを想定し、誤差 A に対する補正について説明する。上記のように、携帯端末の姿勢を補正するために、カメラがとらえた部分的な枠のうち、2 頂点間の辺の向きと中点 (M) 位置を合わせるように、携帯端末の姿勢推定を②ローテーション、③パン、④チルトの順で補正する。実際には、補正の角度を算出して回転行列を作り、②～④の順にビュー変換行列に乗じる。これにより、A の誤差については正確に補正することができる。

続いて、B の誤差については正確に推定を行うのは不可能であるが、運動方向を限定し、A の補正との関係性も考慮しながら推定結果がカメラでとらえた部分的な枠に合うように併進成分をある程度補正することが可能である。

B の誤差を補正が必要となりそうな場面を想定すると、2 頂点のみの検出を連続的に継続しながら併進運動を行う

ことになる。そのような動きは、たとえばカメラでTV画面の左辺をとらえながら、その場所に表示された3DCGをもっとよく見るために前後に移動するか、または別の角度から見るためにTV画面を中心として左右に回り込んで見るように移動するケースが想定される。前者の場合は、TV画面と携帯端末が離れている場合はTV画面と携帯端末を結んだ方向の移動と近似できる。図6では、カメラがとらえた2頂点の辺長に合うように、Aの誤差の補正の前にTV画面と携帯端末間の①距離補正を行う。実際には、辺長の比率をもとにビュー変換行列の併進成分を定数倍する。なお、AとBの誤差の補正により、2つの頂点の座標は一致する。また、後者の場合は、自ずと携帯端末の姿勢を回転させることになり、Aの補正により回り込んだ見え方になり、また距離もBの補正により近い値になるため前述の想定したケースに補正される。

4. 実装評価

提案する推定方式を検証するために、市販のTVとタブレットを用いて実装評価を行った。表2に、評価に用いた機器の仕様などの実験環境を示す。

4.1 処理時間

3.6節で示した枠認識アルゴリズムは、その計算量が推定全体の処理に対して支配的であるため、各フレームにおける枠認識の処理時間を計測した。図7に、タブレットを左右に回転させたときの処理時間の推移例のグラフを示す。実際の利用環境を想定し、CGのレンダリングを含めてAugmented TVを実現するすべての機能を実行したうえで枠認識処理の前後でシステム時刻を出力し、その差分を処理時間として測定した。そのため、今回の処理時間は当該処理のCPU使用時間の上限に相当する。また、処理時間が計測単位である1ms前後と短かったため、より正確に計測するために対象処理を10回ループさせて後処理で1回あたりの平均処理時間を求めた。

表2 機器の仕様などの実験環境

Table 2 Specification and parameters of implementation.

項目	値	
OS	Windows8	
CPU	Intel Core i3 1.8GHz (2コア)	
システムメモリ	4 GB	
イメージセンサ解像度	1920 x 1080 px	
カメラ解像度設定	640 x 480 px	
カメラフレームレート	30 fps	
カメラ水平画角	50.4度 (外付レンズ有)	
ジャイロセンサー	3軸	
外形	295 x 191 x 11 mm	
重量	955 g	
有効画面	886 x 498 mm (40型)	
TV	白黒の枠幅	上下辺 白:13 mm 黒:12 mm
		左右辺 白:18 mm 黒:28 mm
TV-タブレット間距離	約130 cm	
コンパイラ	Visual C++ 2012	

提案方式の枠認識の優位性を検証するために、比較実験として、同様の実験環境において直線検出手法であるハフ変換 (OpenCV の cv::HoughLinesP()) を行った。図8に処理結果を示す。ハフ変換の処理時間は、前処理であるエッジ抽出 (OpenCV の cv::Canny()) の6.8msを含めると30.2ms (10回の平均値) であった。ハフ変換の処理時間は対象画像やパラメータにも依存するが、枠の線分をすべて検出するためには同程度の時間が必要である。また、ハフ変換の場合はこの後に検出された線分を評価して枠を認識する処理が必要であり、さらに処理時間がかかる。

表2に示したように今回の環境はフレームレートが30fpsであるため、1フレームの画像処理は33ms以下に抑える必要がある。提案するアルゴリズムは1~2ms程度とまったく問題ない値であり、ハフ変換と比べてもより実用的であることが分かった。

4.2 枠認識

数分間の試作コンテンツを用いていくつかの視聴環境にて枠認識のテストを行ったところ、ほとんど途切れることなく正しい頂点が検出されることを確認できた。特に2頂点検出できている場合は、ジャイロセンサのみで推定を行うよりも少なくとも2頂点間の辺付近は高精度に位置・姿勢が推定できた。ただし、まれにTV画面以外の照明などあるいはTV画面内の映像におけるコントラストの強い部

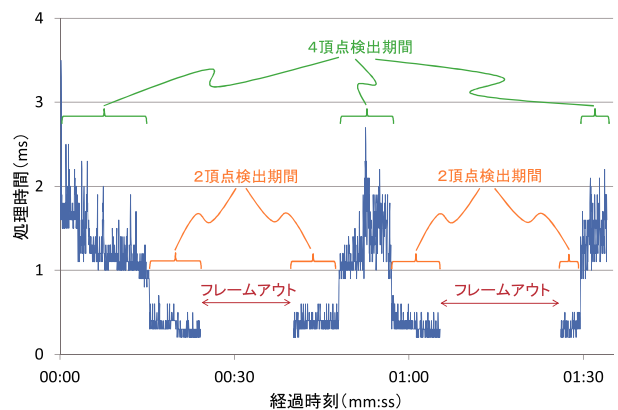


図7 枠認識の処理時間

Fig. 7 Processing time of the frame recognition.

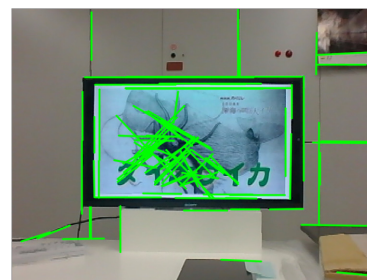


図8 ハフ変換の実行結果例

Fig. 8 Example of the results of Hough transformation.

分を枠として誤認識することがあった。この問題については、5.2節にて考察し、改善を図る。

また、実装を通じてジャイロセンサのみによる推定は、強く振った場合に特にその回転方向に誤差が蓄積されるという特徴が見られた。そこで今回、携帯端末を小刻みに強く振ることにより意図的にジャイロセンサに誤差を与えて、枠認識の可否を調べる実験を行った。その結果、ジャイロセンサのみによる推定においては大きく推定位置が外れてしまい、再びTV画面をフレームインさせても枠を再認識しない場合があったが、画像処理による頂点検出数が2~4である場合には、いくら強く振っても枠を認識できなくなることはなかった。

5. 考察

5.1 処理時間

提案する枠認識アルゴリズムが高精細かつ高フレームレートのカメラ映像において利用可能であるか検証するために、処理時間の理論値を算出する。3.6.1項で述べたとおり、今回のアルゴリズムは計算量が $O(N)$ となるように構成されており、画素数が“2K”、“4K”の映像では、 640×480 pxを使用した今回の実験からそれぞれ3倍、6倍の処理時間がかかるものと考えられる。今回の実験ではおおむね2ms以下で処理が完了していたため、それぞれの処理時間は6ms、12msとなる。これらの値は60fpsの映像のフレーム間の16ms以下に抑えられていることから、実験で使用した端末は高精細・高フレームレートのカメラ映像に対しても利用できるのではないかと考えられる。

また、携帯端末として2.2節で触れたウェアラブル機器を用いた場合の実用性について論じる。実用性は、ウェアラブル機器のサイズから制約される計算資源を用いて提案方式が動作するかどうかで評価できる。Augmented TVにおいて特に有効と思われるウェアラブル機器は、HMD (Head Mount Display) の分類 [11] において「**「**脳メガネ」と呼ばれる透過型で両眼全体を覆うタイプのHMDだと想定される。文献 [11] でも例示されている製品としての脳メガネは、計算処理部分および電池がメガネとは別のパーツとして用意されている形態が主流である。4.1節で示したとおり、提案方式はタブレットで十分な余裕を持って動作していることから、現状の脳メガネの形態であればタブレットに対してサイズの制約が少ないので、表2に示した要求条件の下で提案方式は十分に動作すると考えられる。将来的に脳メガネは計算処理部などと一体型になると思われるが、その場合はあわせてサイズ対計算能力の向上が必要である。

5.2 枠の認識

4.2節で述べたとおり、枠認識はほとんどの場合において問題なく動作することが確かめられたが、まれに枠を誤検

出してしまふ問題があった。この問題を詳しく調べると、今回3頂点検出時に残りの頂点位置を誤って推定することが一因であることが分かった。この問題に対する改善を行うために、枠を誤検出した場合にそれまでとはビュー変換行列が大きく変化することに着目し、以下の工夫により対処した。文献 [10] では、スクリーン座標系において「**「**矩形の形状を表す四辺の長さの比」と「**「**矩形の大きさを表す周囲長」の連続性を評価することでそれまでの「**「**矩形」と同一か判定を試みている。今回扱うケースでは、カメラからTV画面を部分的に扱ったりフレームイン/アウトするときにおける連続性も問われているため、スクリーン座標系での判定を設けるのが難しい。そこで、TVと携帯端末間の距離に注目し、前フレーム推定時の距離と今回推定時の距離との比率の範囲を限定すること（たとえば0.5~2.0）により枠認識の成否を判定することとした。この判定により、枠の誤検出は低減され、大きな改善が見られた。

また、4.2節で述べたジャイロセンサのみによる推定が続いた場合に大きな誤差が蓄積し、再び枠を認識できなくなる問題について簡単に考察する。この問題は、特にタブレットを強く振ったときに起きたことから、加速度センサを用いて対処ができることを示唆している。たとえば、重力加速度以外の加速度を各軸ごとに積算し、その値によってジャイロセンサのみによる推定の信頼度を回転軸ごとに算出する。この信頼度に基づき枠の探索範囲を決定するといった対策が考えられる。

ジャイロセンサのみによる推定時において誤差が蓄積する問題や併進運動にともなう推定ズレについては、コンテンツ制作時の工夫により対処する方法がある。たとえば、画面から大きく外れた場所を演出上の注視点とするような場合には携帯端末をゆっくり動かしてもらるように演出することで、誤差の蓄積を軽減することができる。また、演出上高精度に推定すべき場面においては、視聴者に携帯端末のカメラをTV画面に向けてもらうように促し、枠認識により推定のキャリブレーションを行う。また、その時点で枠認識できない場合はTV画面が映っている領域のタップを促すことによりキャリブレーションにつなげることもできる。

今回開発した方式では、TV画面の検出の手がかりとしてTV画面内に白黒の枠を表示した。枠のデザインは、表現の安定性を第一に考えて枠の認識精度を優先した白と黒を用いたが、演出に応じて色やテクスチャを変えるとといった柔軟性が望まれる。今後の研究の方向性として、3.6.2項で述べた「**「**枠検出フィルタ」などを工夫することにより枠のデザインの多様性を確保する方法が考えられるが、安定性が犠牲にならないように考慮する必要がある。

6. 体験デモ展示

提案する方式が一般の視聴者にとって実際のコンテンツ



図 9 NHK サイエンススタジアムの様子
Fig. 9 State of “NHK Science Stadium”.

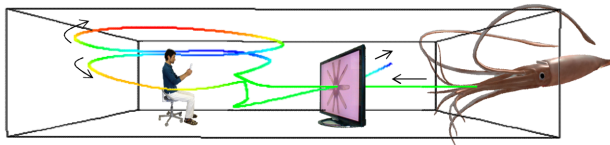


図 10 デモに用いたコンテンツのイメージ
Fig. 10 Schematic diagram of the demo content.

を用いた場合にも十分な精度であるかを検証する必要がある。そこで、一般の視聴者の主観的な評価を収集することも兼ねて、科学系イベントである「NHK サイエンススタジアム」[12]にて2日間にわたり体験デモ展示を行った(図9)。コンテンツのメインパートでは、図10のように、テレビからダイオウイカが飛び出し、体験者の周りを1周16秒間かけて2周回り、TV画面の前でいったん静止し、5秒間ほどかけてTV画面へ突入する姿勢になり、再びTV画面の中に入って行く。この間、体験者にはダイオウイカが泳ぐ姿を観察しやすいようにキャストを固定した回転椅子に座ってもらった。来場者は親子連れが多く、子供が体験している様を後ろから親が見ているというケースが大半であり、両日でおよそ150組の方が体験した。また、デモの環境は表2と同様である。

体験デモの結果、ダイオウイカが2周して戻ってきたときにTV画面にカメラを向けていれば、ほとんどの場合においてTV画面を再認識することができた。これは今回のコンテンツでは、ジャイロセンサによる推定誤差が、提案するアルゴリズムでTV画面を再認識できなくなるほど蓄積せず、TV画面にダイオウイカが突入する前にTV画面を再補足できたことを示している。また、デモの反応は、子供はダイオウイカが飛び出してくる様に驚いて目を見開いたり、終わってから「すごい！」と感動を表現する方が大半であった。このことから、ほとんどの体験者にとっては、ダイオウイカの遊泳やTV画面の出入りに関して、その技術的な精度に満足していたと考えられる。体験デモを通じて、提案する方式が驚きや実在感を与えることに成功したことを示している。

7. メディアとしての可能性

本章では、視聴者にとっての Augmented TV の魅力に



図 11 “3DTV”での検索結果画像のイメージ
Fig. 11 Model image of search results by a word “3DTV”.

ついて過去の作品などをもとに論じるとともに、将来の利用シーンを展望する。TV画面内のキャラクターが画面から飛び出す、あるいは入っていくという演出は、古くは50年前に水木しげるの漫画『テレビくん』[13]の中で登場する。その後も映像作品やCMなどにおける作中の演出としてたびたび効果的に使われている。中でも実写ホラー映画『リング』[14]において恐怖の対象である「貞子」がTV画面から出てくることで主人公達を恐怖に陥れるシーンは、その鮮烈さから有名である。

また、TV画面からキャラクターなどが飛び出している画像は、3DTVの商品宣伝などにおいてよく使われている。実際に画像検索サイトにおいて“3DTV”で検索すると、図11に示すような画像が多数得られた。これらの画像は既存の3DTVの魅力をイラストで表したものであり、便宜的にTVのベゼルを跨いだイラストになっているが、実際の3DTVでは画面の外側に及ぶ演出は技術的に表現できない。ベゼル外にも表現が可能な Augmented TV は、このような望まれながらも実現できなかった演出を可能にする。

現状ではTV画面に把持した携帯端末のカメラを向けなければならないが、将来的には、ウェアラブル機器として5.1節で述べた電脳メガネを用いればその必要がなくなる。電脳メガネでTV画面を見れば、キャラクターが飛び出す演出はより自然に見えることだろう。しかも電脳メガネごとに違う演出ができるので、人によって違うキャラクターを飛び出させたり、視聴者の動きに応じて反応を返すように設計してもよい。TV画面と電脳メガネの組み合わせには新しいメディアとして限りない可能性がある。

8. おわりに

Augmented TV において、携帯端末のカメラとジャイロセンサを用いてTV画面の位置・姿勢を高速に推定する方式を開発した。推定はジャイロセンサをベースとするが、カメラでTV画面の一部分をとらえた場合でも画像処理による認識を活かし、高精度に推定することができる。また、実装評価により処理速度を評価し処理負荷の軽い方式であることを示すとともに、体験デモにより主観的な評価を通じて演出面においても実用的な方式であることを示した。

今後は TV 画面の一部をとらえた場合における定量的な推定精度の評価を行う。

Augmented TV は、TV 画面の「枠」を文字通り越えて視聴者に驚きや実在感をともなうコンテンツを提供する新しいメディアである。将来的に TV 画面は、単に人が直接視聴するためのものにとどまらず、ウェアラブル機器なども含めた携帯端末を介することで、豊かな表現のための「アイコン」となるように研究・開発を進めていきたい。

参考文献

- [1] 川喜田裕之, 中川俊夫, 佐藤 誠: Augmented TV: 携帯端末内蔵カメラを用いて TV の映像を画面外へ拡張するシステム, 日本バーチャルリアリティ学会論文誌, Vol.19, No.3, pp.319-328 (2014).
- [2] 川喜田裕之, 中川俊夫, 佐藤 誠: カメラとジャイロセンサーを用いた TV 画面の位置と姿勢の推定方法, 2014 年映情学年次大, 13-1 (2014).
- [3] 川喜田裕之, 中川俊夫, 佐藤 誠: カメラで一部分を捕えた TV 画面の位置と姿勢の推定方法, 2014 年映情学冬季大, 4-12 (2014).
- [4] 加藤博一: 拡張現実感システム構築ツール ARToolKit の開発, 信学技報, Vol.101, No.652, pp.79-86 (2002).
- [5] 山本 奏, 安藤慎吾, 筒口 拳, 片山 淳, 谷口行信: モバイル動画透かし技術で実現する映像同期型 AR: Visual SyncAR, 第 41 回画像電子学会年次大会, R6-2 (2013).
- [6] Klein, G. and Murray, D.: Parallel Tracking and Mapping for Small AR Workspaces, *6th IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR)*, pp.1-10 (2007).
- [7] OpenCV ver2.4.10, available from (<http://opencv.jp>).
- [8] 片山 淳, 中村高雄, 山室雅司, 曾根原登: 電子透かし読取りのための i アプリ高速コーナ検出アルゴリズム, 電子情報通信学会論文誌, D-II, Vol.J88, No.6, pp.1035-1046 (2005).
- [9] 北原 亮, 片山 淳, 中村高雄: 辺と隅の幾何学的特徴を用いた高速矩形領域抽出手法, FIT2005, J-027 (2005).
- [10] 北原 亮, 中村高雄, 片山 淳, 安野貴之: 携帯端末上における幾何補正のためのリアルタイム矩形追跡手法, 信学技報, Vol.106, No.351, pp.1-6 (2006).
- [11] 塚本昌彦: ウェアラブル機器の現在と未来; ウェアラブル最前線, 日経 BP 社, pp.5-33 (2014).
- [12] NHK サイエンススタジアム, 入手先 (<http://www.nhk-p.co.jp/event/detail.php?id=423>).
- [13] 水木しげる: テレビくん (漫画), 講談社 (1965).
- [14] 中田秀夫 (監督): リング (映画), 東宝 (1998).



中川 俊夫

1989 年東京大学大学院工学系研究科電気工学専攻修士課程修了。同年, NHK 入局, 放送技術研究所に配属。編成局を経て 2012 年より放送技術研究所勤務, 現在ハイブリッド放送システム研究部部长。



佐藤 誠

1973 年 3 月東京工業大学工学部電子物理工学科卒業。1978 年 3 月, 同大学大学院博士課程修了。同年 4 月より同大情報工学科助手。1986 年 3 月東京工業大学精密工学研究所助教授を経て, 現在, 同大学精密工学研究所教授。博士 (工学)。コンピュータビジョン, パターン認識, ヒューマンインタフェース, VR の研究に従事。ストリング型ディスプレイ SPIDAR を提案, 開発。元日本 VR 学会会長。



川喜田 裕之

2004 年東京工業大学大学院理工学研究科修了。同年, NHK 入局。NHK 広島放送局にて放送波の受信技術業務に従事し, 2007 年放送技術研究所に配属。現在, 東京工業大学社会人博士課程在籍。放送通信連携における端末連

携の研究に従事。