

自由視点テレビのための観視者の頭部検出

東 尊弘，藤井 俊彰，谷本 正幸

我々が提案している自由視点テレビでは，視聴時に視点を指定することができる．そのインターフェイスとして，視点と視聴者の頭部位置とを連動させたシステムを提案する．テレビ上部に設置したカメラ画像をフレーム間差分処理した後，高さ方向に画像を走査して最も高さのある物体の上部を顔位置とすることで頭部の座標を得る．さらに，時間方向に対して座標の不連続性を取り除くことで視点移動時の画像のぶれを防ぎ，自然な視点の変更が行えるようになった．

A Head Tracking System for Free Viewpoint Television

Takahiro Higashi, Toshiaki Fujii, Masayuki Tanimoto

The Free Viewpoint Television (FTV) is a new television system in which the user can change his viewpoint freely in the 3D-world. Here, we propose a new user interface for the FTV. The location of the user's head is detected and a view images at this location is successfully displayed in a 2D display in realtime.

1 はじめに

我々は，自由視点テレビ (Free Viewpoint Television) [1][2][3] と名付けた新たなテレビジョンシステムを提案している．これは，その名が示すとおり，自由な視点からテレビ番組を観ることができるというものである．

例えば，歌番組においてステージの中央にいる歌手を，正面からだけでなく，上から見たり後ろから見たり，あるいはバックダンサーに注目したりといったことが可能である．このシステムでは，現在テレビゲームなどをはじめ広く実用化されているモデルベースによる手法ではなく，光線空間法というイメージベースの技術を用いることにより 3 次元情報を格納している．実写ベースで任意視点画像を生成でき，スポーツ中継など様々なシーンでの活用が期待されている．

現在，テレビを視聴する際には主にリモコンによって選局などの操作を行っているが，自由視点テレビにおいては視点を選択する必要性が生まれるため，リモコンに機能を追加，もしくは新しいシステムによってその操作を実現しなくてはな

らない．

これに対して，頭部位置に対応した視点画像を提示するインターフェイスシステムが考えられる．窓から見える景色をイメージして欲しい．人が頭を移動させると，それに応じて窓に映る風景の見え方が変わる．近くにあるものほど見え方の変化が大きく，これによって物体と自分との位置関係を知ることができる．このような見え方の変化は運動視差と呼ばれる．

このような原理を搭載した「多視点立体ディスプレイ」が開発されている [4][5]．レンチキュラススクリーンを用いた 8 眼式立体ディスプレイや，パララックスバリアを利用した立体ディスプレイなどがこれにあたる．2 眼式の立体ディスプレイではどこから見ても同じ情報しか得られないが，多眼式では人が頭の位置を動かすごとに異なる情報が流される．

本研究では，これらの特殊な装置を用いることなく，センサーによって頭部の位置を検出してテレビ画面を連動させることによって，一般的なディスプレイで運動視差を実現する．また，多眼式

ディスプレイではその眼数が増えると情報の伝送量も肥大化してしまうため、予めセンサーによって人物位置を取得しておき、その人に合った情報を流すという方法を用いることで情報を削減することが望まれており、本システムの応用により解決が期待できる。

2 自由視点テレビ

2.1 概要

自由視点テレビとは、映像の製作者により視点やカメラワークがあらかじめ決められている従来のテレビとは異なり、ユーザが見たい視点やカメラワークを自由に指定して観察することができるテレビである。我々は、複数台のカメラによって3次元シーンを撮影し、イメージベース法によって映像を生成する方法で自由視点テレビを実現している。

2.2 モデルベース法とイメージベース法

自由視点画像を生成するためのアプローチには、大きく分けてモデルベース法とイメージベース法の2つの方法がある。これらについて説明する。

まずモデルベース法は、複数視点から物体を捕えた画像情報をもとに形状情報・表面反射特性・光源情報などシーンの3次元情報を復元し、物体のポリゴンモデルを生成する手法である。一度モデルを生成すればCG手法を用いて自由に視点移動を行なう事ができる。イメージベース法は、カメラで撮影した実写画像を用い、それらを変形したりパッチワークのようにつなぎ合わせたりすることにより、任意視点映像を生成する手法である。生成画像が非常にリアルである点、レンダリング処理を必要としないため高速な画像生成が可能であるという点などが特長である。

2.3 光線空間法

自由視点テレビにおいて我々は、イメージベース法の一つである光線空間法[6][7]の利用を提案している。ここでは、その利用によって可能になる3次元映像処理を挙げる。

(1) 任意視点画像の生成

任意視点画像は光線空間の部分空間で与え

られる。従って、視点位置に応じて光線空間を2次元平面（あるいは曲面）で切断し、その断面を表示すればよい。図1に光線空間からのステレオ画像、多視点画像の生成の例を示す。指定された視点に対応する光線空間内の直線に沿ってデータを取り出して表示することにより実現される。

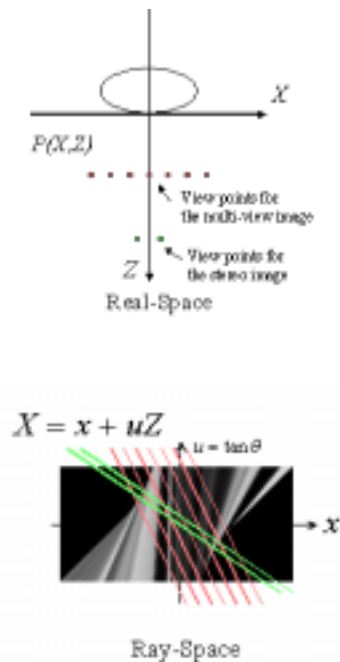


図1 光線空間からの多視点画像生成

(2) 可変焦点画像の生成

上記(1)においてピンホールの代わりに有限の開口を持ったレンズを仮定する。レンズの作用は、光線空間データに対してレンズの焦点距離に応じた方向に低域通過フィルタを施すことにより実現される。その後、(1)と同様に光線空間を2次元平面で切断し、その断面を表示すればよい。

(3) カメラワークの実現

カメラワークとしては、パン・チルト・ズーム等が挙げられる。これらのカメラ操作は視点を固定したままで行う操作である。従って、まず(1)の2次元断面像を切り出し、次に

パン・チルトに応じて表示する範囲（視野）を，ズームに応じて標準化密度を決め，表示する画像を生成する．焦点距離が異なる場合，画像の解像度を一定に保ったまま視野角を変える処理をすればよい．

(4) 視差補間

ここで言う視差補間とは，実際には撮影されていない視点位置からの画像を作り出す処理である．これは光線空間を θ 方向に高解像度化する処理に他ならない．従来の2次元画像処理で用いられている高解像度化の手法を利用することもできるが，光線空間データが直線群から構成されるという特徴を生かすことによって，精度の高い補間が可能となる[8]．

3 人物検出を用いたインターフェイス

3.1 頭部位置の取得方法

本稿では，テレビモニターの上部に人物の位置を検出するためのセンサーを取り付け，その情報を元にモニターに映し出す画像を決定するというシステムを提案する．利用者が違和感なくシステムを使用できるためには，

「リアルタイムでの動作」

「意図通りの動作」

を行える必要がある．前者に関しては，現在のテレビ放送のフレームレートが29.97fpsであることから，これよりも高速に人物位置を検出し，提示画像が生成できることが望まれる．また後者については，観視者にストレスを与えず，安定に動作することが期待される．

頭部位置の取得を行う方法は，大きく分けて次の三つがある[9]．

1. 赤外線照射による方法
2. 磁場計測による方法
3. 画像処理による方法

赤外線による手法では，赤外線を人物に照射し角膜からの反射光をカメラで撮ることで位置を検出する．この方法では，照射光によりドライアイなどの眼精疲労を引き起こすといわれており，長時間利用が考えられるテレビにおいてはあまり

好ましくない．磁場による方法では，磁場の変化をコイルで捉えることにより測定するが，一般に精度が高くないとされている．

これらに対し，画像処理による手法は計算機やカメラなど，コストが高くなることが問題点であるが，これらの価格は低くなってゆくことが見込まれており有望な手法であるといえる．

画像処理によって頭部の位置を検出する手法としては，色情報に基づいた処理やパターンマッチングなどの方法がある．しかし，これら静止画処理に基づく手法では処理時間が長くなるため，リアルタイム動作が難しい．また周辺環境の影響を受けやすいため，安定な検出が難しい．今回は，このことを考慮しフレーム間差分による検出を利用しシステムを構築した．

そのアルゴリズムを図2に示す．差分画像に対し高さ方向への走査を行い，その結果に基づいて頭部位置を特定し，さらに結果が安定なものとなるように工夫した．

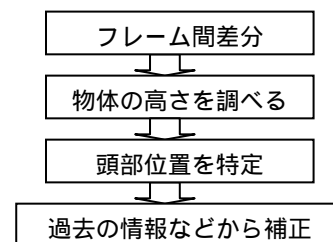


図3 頭部位置特定のフローチャート

3.2 フレーム間差分

フレーム間差分法は動画像において移動物体を抽出する際，非常に有効であり広く利用されている[10]．時間的に環境が変化するには特に有用であり，今回のケースにも適しているといえる．動画像 $f(x,y,t)$ の，2枚の差分画像 $g(x,y,t)$ は式(3.1)で定義される．

$$g(x,y,t) = |f(x,y,t) - f(x,y,t-1)| \quad (3.1)$$

このようにして計算された差分画像は，服の模様などに依存したものとなる．また，ランダムなノイズが入るため，収縮・膨張処理によりノイズカットおよび線の強調をする．この結

果を図2に示す。



図3 膨張・収縮処理後の差分画像

3.3 頭部位置の特定

まず、センサーより取り込んだ画像に対しフレーム間差分処理をする。次に画像を高さ方向に走査し、動物体の候補が検出された座標の最大値および最小値を求める。

図4がこの様子である。これにより、検出された物体の高さ $Y_{\max} - Y_{\min}$ を得る。高さの値が最も大きく、かつあらかじめ設定された閾値を越えている場所の X の値を人物の位置 X_h とみなす。

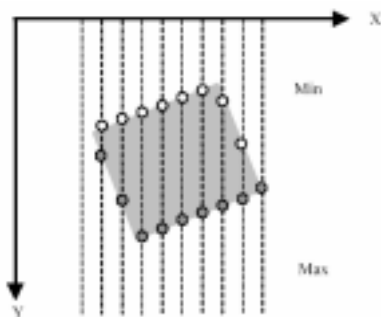


図4 物体の走査による高さ検出

また、 X_h における最小の Y の値を頭部の座標

Y_h とすることにより、頭部座標 (X_h, Y_h) を得る。

3.4 検出値の補正

フレーム間差分法は計算コストが低く、アルゴリズムもシンプルであるため広く使われているが、目的とする物体（この場合は人間）以外の動物体をも誤検出してしまうという欠点がある。また、同様にランダムノイズも物体として検知するため $g(x,y,t)$ に対して適当な閾値を用いることで

これを取り除くが、この際に正しい信号をも取り除いてしまうことがある。つまり、この方法では厳密な頭部座標を得ることが難しい。

よって、この仕組みを用いてシステムを設計すると視点移動が不連続・不自然に行われ、観視者に不快感を与える原因となる。このため、人物移動の軌道を平滑化する方法により位置情報の補正を行う。

まず、先ほどは最も高さのある X の値を人物の位置 X_h としていたのに対し、今度は高さのある順に上位数カ所の X を選び、その平均値を X_h とする。また、時刻 t および $t-1$ での頭部座標の値を比較し、連続性の比較を行い、ある閾値以上の変化がある場合は誤検出とみなし無視することとする。

さらに、時間方向にも時刻 t から $t-4$ までの平均値をとることにより、軌道を安定化する。

これらの処理により、視点移動はスムーズに行われ、観視者の負担を軽減できると考える。

4 システム構成

4.1 動作仕様

目的とするシステムは、観視者の頭部が移動すると自由視点テレビにおける視点が移動するというシステムである。構成としてはモニター（テレビ）、CCDカメラ、およびデータを処理するPCとなり、一般的な機器で構成されている。

CCDカメラから画像を逐次取り込み、人物位置を検出する。人物が1人いる場合、その位置情報を元に提示画像を生成し、モニターに表示する。人物の位置が移動した場合には提示画像をその都度更新する。位置検出はカメラに対して左右方向および上下方向に対して行われ、前後方向には行っていない。しかし、前に移動した場合はカメラに移る人物の頭部は上方向へ移動し、提示画像も上から見たものとなるため、前後への移動に対しても運動視差を与えることができる。

人物が2人以上いる場合には、はじめに動きを検出した人物に対応して動作することとなる。自由視点テレビは現在1人用のシステムとして研究が進められており、本システムでもそれに準ず

ることとする。



図5 本システムでテレビ鑑賞している様子

4.2 機材環境

本システムは、表1のような機材で構成されている。

表1 実験装置一覧	
モニター	Pioneer PDP-A503HD
CPU	Pentium4 2.8GHz
OS	Windows XP Professional
コンパイラ	C++ Builder 5
CCDカメラ	東芝 IK-SM43H
画像処理ハードウェア	Hicos IP7000BD

CCDカメラの画素数は41万画素であり、これをIP7000BDを用いて取り込む。取り込み画像サイズはグレースケールの512×480としハードウェアにより処理を行う。

4.3 ネットワーク上からの利用

頭部位置を測定するための計算機と、提示画像を生成するための計算機を同一なものにすると、どちらかの処理が重くなると他方にも影響を及ぼす。自由視点テレビでは、圧縮したデータの復号処理[11][12]や画像補間などの処理を行うため、提示画像生成のための計算量が大きくなることが予想される。よって、これらの計算機を別々に用意するのが望ましい。

このため、ネットワーク上のPCが、人物の位置情報を自由に参照し利用できるようにした。この様子を、図6に示した。位置検出用のマシンをサーバー、画像生成用のマシンをクライアントと

し、LAN経由でクライアントは自由にサーバーから情報を得ることができる。

実用化にあたっては、サーバーPCを各家庭に、クライアントPCをコンテンツプロバイダーのセンターなどに用意するのが良いと考えている。

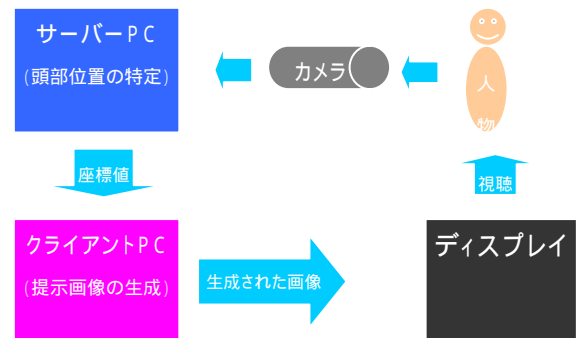


図6 ネットワーク上から位置情報を参照

5 実験

5.1 実験内容

環境変化の少ない室内において、カメラから1mほど離れた位置を身長1.7m程度の人物がカメラに対し左右および上下にゆっくり(通常の歩行速度程度で)移動し、位置情報がどのように測定されるかを調べた。

人物が着ている服は図7のように、柄やプリントなどのないシャツとジーンズであり、フレーム間差分法を利用するにはやや悪条件となる。

なお、今回は提示画像として自由視点テレビシステムで提案している光線空間画像ではなく、OpenGLによって作成したCGオブジェクトを用いた。

5.2 実験結果

取り込む画像サイズを512×480とし、環境変化の少ない室内において実験を行った。

平滑化処理を行う前および後での座標の変化の様子を図9に示す。グラフより、人物移動の軌跡が滑らかになっていることがわかる。本手法は高さによる閾値処理を行っているため、垂直方向に関して平滑化の前後で座標値のとり範囲が大きく異なるが、このことは観視者は垂直方向には水平方向ほど大きく動かないことを利用したものであり、実用上の問題はない。

処理速度は平均13fpsを超えるので、目標である30fpsには及ばないものの、体感としては不自

然な印象はない。また、カメラからの取り込み画像サイズを 320×240 などに縮小することで改善の余地がある。ただ、時間方向に平均を取っている都合上、提示画像の更新が少し遅れているような印象を受ける。グラフからも数フレームの遅延が読み取れる。



図 7-1 左からモニターを見ている様子



図 7-2 右からモニターを見ている様子



図 7-3 低い位置からモニターを見ている様子

図 7 人物移動による提示画像の変化

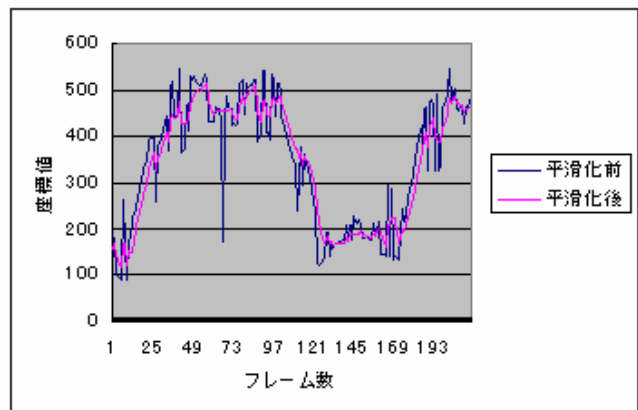


図 9-1 水平方向の軌跡

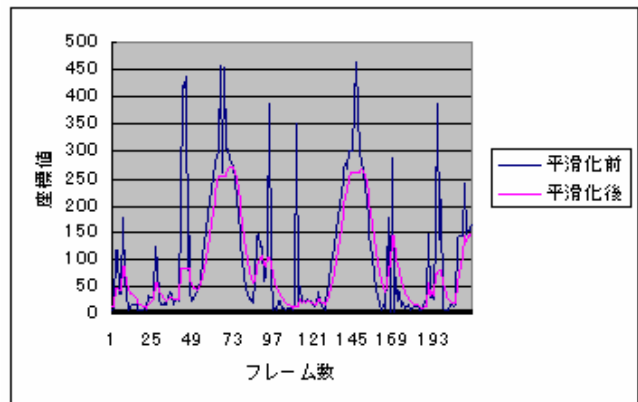


図 9-2 垂直方向の軌跡

図 9 頭部位置座標の補正前後での変化

6 まとめ

本稿ではフレーム間差分に基づく頭部位置検出に関して述べた。補正処理を行うことで、より自然な画像を提示することが可能となった。今後の課題として、人物の奥行き方向の位置を検出できるようにすることなどが挙げられる。

参考文献

- [1] Masayuki Tanimoto and Toshiaki Fujii, "FTV ISO/IEC JTC1/SC29/WG11, M8595, Jul. 2002.
- [2] Toshiaki Fujii, Masayuki Tanimoto, "Free-Viewpoint TV System Based on Ray-Space Representation", SPIE ITCOM, Vol. 4864-22, pp.175-189, Aug. 2002.
- [3] Purim Na Bangchang, Toshiaki Fujii, Masayuki Tanimoto, "Experimental System of Free Viewpoint Television", IS&T/SPIE Electronic Imaging 2003, Vol. 5006-66, Jan. 2003.

- [4] 志水 英二,岸本 俊一,“ここまできた立体映像技術”,工業調査会,pp.80-90,2000
- [5]高木 康弘,”変形 2 次元配置した多重テレセントリック光学系を用いた 3 次元ディスプレイ”,映像情報メディア学会誌 Vol.57,No2,pp-293-300,2003
- [6] T. Fujii, T. Kimoto, and M. Tanimoto, “Ray Space Coding for 3D Visual Communication”, PCS '96, 2, pp. 447-451 (1996.3).
- [7] 藤井俊彰,木本伊彦,谷本正幸,“光線群表現における 3 次元空間情報の圧縮”, 3 次元画像コンファレンス'96, pp. 1-6 (1996.7).
- [8]中西 敦士, 藤井 俊彰,木本 伊彦,谷本 正幸,”E P I 上の対応点軌跡を用いた適応フィルタによる光線空間データ補間”,映像情報メディア学会誌,Vol.56,No8,2002
- [9]飯尾 淳,谷田部 智之,比屋根 一雄,”顔の位置と頭部位置の 3 次元情報を利用したユーザインタフェースの試作”, 画像センシングシンポジウム SSII2002
- [10]東 尊弘,藤井俊彰,谷本正幸,”マルチカメラ人物追跡システムにおけるカメラ配置に関する検討”,映像メディア処理シンポジウム IMPS2002,pp 105-106
- [11]M.Kawaura,T.Ishigami,T.Fujii,T.Kimoto and M.Tanimoto, ”Efficient Vector Quantization of Epipolar Plane Images of Ray Space by Dividing into Oblique Blocks”, Picture Coding Symposium 2001,TP1-11,pp.203-206
- [11] Yutaka Hayashi, T.Fujii and M.Tanimoto,”Coding for Ray Space based on SPECK”, Picture Coding Symposium 2003,339-342