

ドーム型ディスプレイによる全方位ビデオのインタラクティブ表示

山口健[†], 波部齊^{††}, 松山隆司[†]

[†] 京都大学大学院情報学研究科知能情報学専攻

^{††} 京都大学学術情報メディアセンター

概要 ドーム型ディスプレイに全方位ビデオを映すことで、あたかも人がその映像を撮影した場にいるかのような感覚、没入感を得ることができる。この没入型映像表示システムでは、従来のように人が道具を使って指示を与え、システムがその指示に従うという関係では、没入感を阻害してしまい目的と整合しない。また、映像表示システムの側で自発的に表示する映像をコントロールすることで、より効果的・魅力的な映像を提示できると考えられる。そこで本稿では、システム側が人の自然な行動から意図を認識し、その認識した人の意図と映像の持つ特性、過去の履歴などから表示する映像を決定するインタラクティブ表示システムを提案する。

Interactive Visualization of Omnidirectional Video with Dome Display

Takeshi Yamaguchi[†], Hitoshi Habe^{††} and Takashi Matsuyama[†]

[†]Department of Intelligence Science and Technology,
Graduate School of Informatics, Kyoto University

^{††}Academic Center for Computing and Media Studies, Kyoto University

Abstract We can obtain immersive experiences by displaying omnidirectional video on a dome display. In the case of ordinary display systems, a viewer operates the display system with some mechanical interfaces such as a keyboard. However, in the case of an immersive display system, the operation of such tools is inconsistent with the immersive experience. In this paper, we propose an interactive visualization system of omnidirectional video on a dome display. Our system can perceive viewer's intentions via gesture recognition and generate effective and attractive video images based on the viewer's intentions and the characteristics of the original video.

1 序論

「百聞は一見にしかず」ともいわれるように、映像は人に対して多くの情報を提供する。テレビに代表される映像メディアは日常生活に欠かせないものになっており、さらに効果的・魅力的な映像メディアを生み出すための取り組みがなされている。そのような中で、映像メディアを人に対して提示するシステム（以下、映像表示システムと呼ぶ）と、それを見る人との関係を考えてみると、人からシステムに対して何らかの道具を使って指示を与え、その指示に対して映像表示システムが従順に従うという関係が常に成立してきたと言える。

しかし、映像メディアの発展により、この関係で

は必ずしも十分ではない状況が現れてきた。その一つとして、ドーム型ディスプレイと全方位ビデオを用いた没入型映像表示システムが挙げられる。この映像表示システムは人の視野を全て覆いつくす映像を提示し、人に対して映像の撮影された場にいるような感覚（没入感）を与える。すなわち、人の感覚に対して従来の映像表示システムでは考えられないような強い影響を与えるといえる。

このような映像表示システムにおいて、従来と同様に道具を用いた指示の伝達を行うと、没入感を阻害することになりシステムの最大の目的と整合しない。さらに、没入型映像表示システムでは人が得る視覚情報の全てをコントロールできるため、より効果的な映像を提示することができるが、従来は映

像表示システム側が自発的に表示する映像をコントロールする機能は考えられていなかった。

そこで、本研究においては、

- 映像表示システム側が人の自然な動作から意図を理解する。
- 理解した人の意図と、映像コンテンツの特性、過去の履歴などから、人にとって必要な映像は何か、あるいは人が本当にみたい映像は何か、を映像表示システムが判断して表示する。

ことを特徴とする没入型映像表示システムを提案する。本研究では、人と映像表示システムとの間のインタラクティブな情報のやりとりによってこれらの特徴を実現する。

本稿においては、2章で提案するインタラクションについて述べ、その有効性を論じる。さらに、この考え方に基いて開発したプロトタイプシステムについて3章で述べる。最後に4章においてプロトタイプシステムを用いた実験から得られた知見および今後の課題について述べる。

2 ドーム型ディスプレイによる全方位ビデオのインタラクティブ表示

本章ではまず、従来の映像表示システムにおける人とシステムの関係について述べる。これに対して、本研究で着目する全方位ビデオおよびドーム型ディスプレイについて述べ、さらに本研究で提案するドーム型ディスプレイによる全方位ビデオのインタラクティブ表示について考察する。

2.1 従来の映像表示システムにおける人とシステムの関係

従来の映像表示システムにおいては、人は自分の意図に基づいた指示を道具を使って与え、システム側はその指示に従う、という関係が成り立っていた。例えば、ビデオなどはリモコンを用いて再生や早送り、巻戻しなどの操作を全て人が行なう。図1はこの関係を一般的に示したものである。

いままでの映像表示システムは図1のような関係のもとで設計されてきた。従来の映像メディアの枠の中ではこの関係で十分であったが、この関係では不十分な、あるいはこの関係を拡張することでより

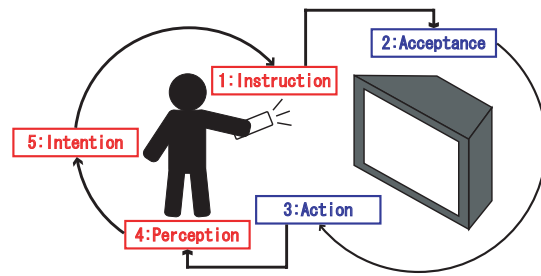


図 1: 従来の人と映像表示システムの関係

効果的な表示を行なうことができる映像メディアが現れてきた。次節以降ではその代表例として、本研究で提案するドーム型ディスプレイによる全方位ビデオのインタラクティブ表示について述べる。

2.2 全方位ビデオおよびドーム型ディスプレイ

全方位ビデオとは、あるひとつの視点位置から全ての方向の視覚情報を記録した映像である。全ての方向のうち任意の方向の視覚情報を選択して取り出すことができるため、複数の人にそれぞれ必要とする方向の映像を提示したり、ロボットのナビゲーションに利用するなど、さまざまな応用が可能である [1]。

次に、ドーム型ディスプレイとは IPD (Immersive Projection Display) の一種で、球状のスクリーンとそれに映像を照射する為のプロジェクタからなる。見ている人の視野を全てスクリーンが覆うため、あたかも人が映像の撮影された場にいるような感覚、没入感を得ることができる [2]。本研究で用いたドーム型ディスプレイを図2に示す。半球状のスクリーンは人の正面方向に設置され、人が正面を向いたときに没入感が得られるように設計されている。



図 2: ドーム型ディスプレイ

この2つを組み合わせることで、それぞれの特徴を最大限に生かすことができる。すなわち、ドーム

型ディスプレイに必要な広範囲の視覚情報を全方位ビデオから得ることで、全方位ビデオを撮影した場にいるような没入感を人に対して与えることができる。このようなシステムとしては円筒状スクリーンをもつ全周型景観提示システムに全方位ビデオを表示するシステムが既に提案されている [1]。

さらに、このシステムでは映像表示システムが人が得る視覚情報のすべてをコントロールできるため、映像をそのまま表示するだけでなく表示システム側で適切に加工することで、人の知覚の特性を利用した新たな効果を生み出すことが期待される。

2.3 ドーム型ディスプレイにおけるインタラクション

2.2 節で述べたドーム型ディスプレイを用いた全方位ビデオの表示システムにおいて、人とシステムとの間のインタラクションを考える。2.1 節で述べたように、従来の映像表示システムでは人はリモコンやキーボード、ジョイスティックのような道具を操作して意図をシステムに伝えていた。しかし、そのような操作はドーム型ディスプレイが提供する没入感から引き離される動作であり、システムの目的にそぐわない。

そこで、本研究においては、ドーム型ディスプレイが人の動作から意図を認識するインタラクションシステムを提案する。動作による意図の伝達は、先述の道具を利用した意図の伝達に比べて通常の状態に近いと考えられ、没入感に対する親和性が高いと考えられる。

さらに、本システムにおいては、動作を通じて認識した意図に加えて、映像のもつ特性、人が視覚を通じて得る感覚などの情報を利用してドーム型ディスプレイがより効果的な映像を表示する。

これらの特徴により、従来の映像表示システムにはない、新たな映像メディアの表示システムが構築できることが期待される。

ここで、人とドーム型ディスプレイとのインタラクションとは図 3 のような流れに基づく。つまり、

1. 人が何らかの意図を行動で表現する。
2. ドーム型ディスプレイがそれを認識する。
3. 人の行動の認識結果と、映像のもつ特性（撮影時の状況など）、人が視覚を通じて得る感覚、過去の操作履歴などの事前知識から、その状況に適した映像をドーム型ディスプレイが判断する。

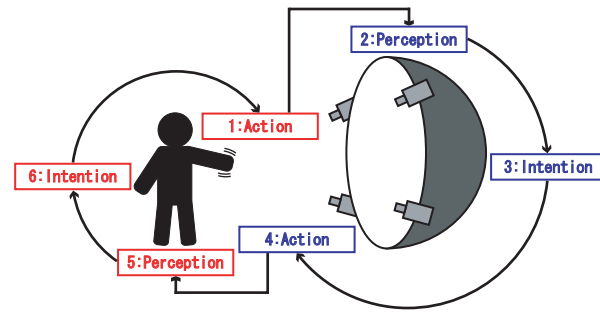


図 3: 人とドーム型ディスプレイとのインタラクション

4. その判断に従って映像を表示する。
5. 映像を見ることによって人が新たな感覚を得る。
6. そこから人が何らかの意図をもち、必要があれば 1. に戻ってそれをディスプレイに伝達する。

という順序であり、これを繰り返すことで効果的な映像表示を実現する。

以下ではこのインタラクションモデルについて詳細に述べる。

2.3.1 人からの意図の伝達

まず、人が意図をドーム型ディスプレイに対して伝達する方法について考察する。これは図 3 では 1 の部分に相当する。

2.1 節で述べたように、現在までの映像表示システムでは、例えばリモコンなどの道具を利用して人が指示を伝えていた。しかし、上で述べたように、我々が通常生活している状況から考えると、道具を使うのは特別な状況であるということが出来る。没入型映像表示システムの目的は、人があたかも映像の中に自然にいるような環境を作り出すことであり、その目的から考えると、リモコンなどの道具を使った指示の伝達は不適當であるということが出来る。従って、通常的生活環境で人同士の意思伝達に使用する、身体動作などによって意図を伝達することが望ましい。

さらに、道具を使用した場合には、人は“指示すること”を明示的に意識していることになる。これは、人が意識しない限りその意図が伝わらないことを意味しており、人から映像表示システムへ伝わる情報量を制限することにつながる。これに対し、身体動作などによった場合には、人が無意識のうちに

とった行動などからその意思を推し量ることが可能となり、映像表示システムが得る情報量が増すことになる。

以上の理由で、本研究で提案するインタラクションシステムでは、人が日常生活で意図の伝達に使用する手段を使用して、人からドーム型ディスプレイに対して意図を伝達させる。具体的には、身体動作・音声などの手段が考えられるが、それぞれ以下のような特徴をもつ。

身体動作 定量的な表現は難しいが、人が見たものを指し示すなど、直感的な指示には適している。

音声 定量的かつ具体的な表現が可能であるが、身体動作に比べて直感的な指示には適していない。

実際のシステムを設計する際には、以上のような特徴を踏まえた上で、伝達したい意図からその手段を選択する必要がある。

2.3.2 ドーム型ディスプレイによる人の意図の認識と映像の生成

次いで、ドーム型ディスプレイが人の身体動作や音声を通じてその意図を認識する。これは、図3の2の部分に相当する。

身体動作や音声を通じて伝えられた人の意図は、ドーム型ディスプレイに取り付けられた各種センサによって認識される。認識を行うためには、カメラなどによって人の姿勢を計測したり、マイクなどによって音声を計測する必要がある。その計測結果から人がなにを意図しているを認識する問題は一般に大変困難であるが、タスクを限定して実用化したシステムが多く提案されている。例えば、手指を認識して机上に投影した情報を直接操作するシステムが提案されている [3]。

本研究でとりあげるドーム型ディスプレイによる全方位ビデオの表示システムにおいても、人がシステムに伝える意図はある程度限定することができ、動作などの計測結果からそれを認識することも比較的容易であると考えられる。

次いで、その認識結果からディスプレイシステムが表示する映像を選択・生成する。これが図3の3の部分に相当する。その際には、人の意図にただ従順に従うのみではなく、映像が撮影された状況（カメラ自体の運動）や過去の操作履歴などを考慮に入れて、必要であれば映像を補正することで効果的な映像を提示する。例えば、カメラが揺れている場

合にそのまま映像を表示すれば、映像が人の視野すべてを覆っているために、人に不快感を与えることが考えられる。このような場合には、表示システム側で自動的に揺れを補正した後で表示すれば、人に不快感を与えることなく映像を表示することができる。

3 プロトタイプシステムの開発

この章では第2章で提案したドーム型ディスプレイと人とのインタラクションモデルに基づき、そのプロトタイプとしてドーム型ディスプレイに実装したシステムについて述べる。

3.1 システム構成

プロトタイプシステムのシステム構成を図4に示す。ここでは、人からドーム型ディスプレイに対して意図を伝える動作として指差し動作を選び、指差しの方向に応じてディスプレイに表示する視線方向を変化させるようにした。指差し動作は人が見たいと思う場所や方向を指し示すための基本的な動作であり、人が映像を見たときに感じる「あれが見たい」や「もっと右がみたい」といった意図をドーム型ディスプレイに対して伝達するのに適している。

図4に示したシステムの具体的な動作は以下のとおりである。

1. あらかじめ人の両肩・指先にマーカーを貼りつけておき、指差し動作によって人が意図をドーム型ディスプレイに伝える。
2. ドーム型ディスプレイの外周部に取り付けた複数のカメラによって人の動作を撮影する。
3. マーカーを貼った両肩、指先を画像から検出する。
4. ステレオ視を行ない、両肩と指先の3次元位置を算出する。その算出した位置により、指差ししている方向を求める。
5. 指差し方向に応じて全方位ビデオの中から表示する部分（視線方向）を選んで表示する。
6. 人がその映像の変化を見て必要であれば新たな指示を行う。

以降では上記の流れの中でステレオ視による指差し動作の認識手法とその結果による映像表示の制御について述べる。

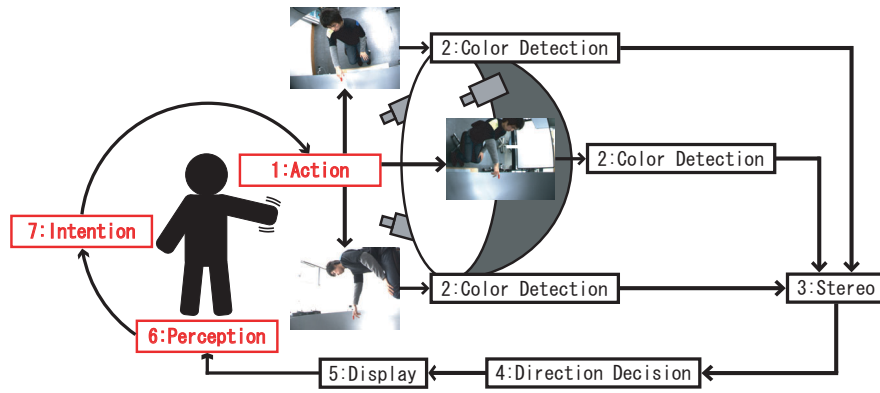


図 4: システム 構成図

3.2 指差し動作の認識

今回のプロトタイプシステムでは指差し動作を認識するために、図 4 に示すように、ドーム型ディスプレイの外周部にカメラを取り付け、そのカメラで撮影した映像から指差し方向を認識することにした。指差し動作の認識の流れを以下に述べる。

1. カメラとそのキャリブレーション プロトタイプシステムでは 640×480 の解像度を持つデジタルカメラ¹³ 3 台をドーム型ディスプレイの外周部に設置した。カメラの内部・外部キャリブレーションは Zhang により提案された方法 [4] を用いた。
2. 両肩・指先のマーカー検出と位置計測 人の両肩・指先には異なった色のマーカーをあらかじめ貼ってあるので、デジタルカメラで撮影した映像の RGB 情報から、マーカーの位置を検出した。1. においてカメラ間の位置関係もキャリブレーションされているため、複数のカメラで同一のマーカーを検出することで、その 3 次元位置が計測できる。
3. 指差し方向の決定 次に指差し方向を求める。ある基準点と指先とを結んだ直線の上に、人が指示している対象があるものとし、その基準点を指示基準点と呼ぶ。指示基準点は、個人差はあるが、指示している腕側の肩から頭にかけて存在していることが確かめられている [5]。プロ

¹³PointGrey 社製 Dragonfly

トタイプシステムにおいてもこのことを利用した。

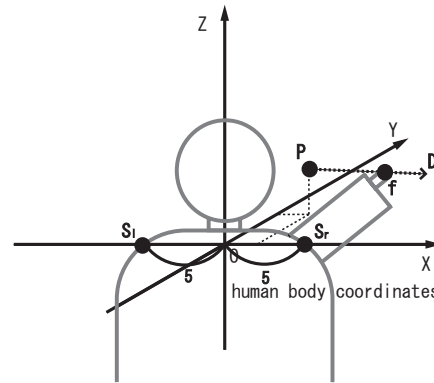


図 5: 人体に対する指差し基準点

図 5 のように、両肩の中点を原点とし、左肩から右肩へ結んだ直線を X 軸、原点から頭に向かう直線を Z 軸、そして原点から人の正面に向かって X 軸と Z 軸に垂直な直線を Y 軸とする人体基準座標を考える。3 次元計測の結果求めた左肩のマーカー位置を S_l 、右肩のマーカー位置を S_r 、指先のマーカー位置を f とする。両肩の位置から指示基準点 P を定めることができる。そして、指示基準点から f まで向かって引いた方向が指示方向となる。

以上の流れのうち、1. を前もって行っておき、2, 3. を繰り返すことで指差し方向の計測を行う。なお、2. において、マーカーの検出に失敗した場合や、3 次元計測結果が人の指差し動作に対して不自然な場

合は、指差し動作をしていないものと判断する。

3.3 表示映像の生成

前節で述べた指差し動作の認識結果を用いて、表示する映像を生成する。ドーム型ディスプレイでは、ある視線方向 (θ, ϕ) に対してその方向に人が向いているような映像を表示することができる。ここで、 θ は Z 軸周りの角度、 ϕ は X 軸周りの角度である。

指差し動作の認識の結果得られた指差し方向は、 $D = f - P$ であるので、それと Y 軸とのなす角を ψ と定義する。この角度 ψ から生成する映像の視線方向を、

$$\theta(t + \Delta t) - \theta(t) = \alpha\psi(t), \quad (1)$$

に従って制御する。ここで、 $\theta(t)$ はある時刻 t での視線方向の水平方向成分の角度を示し、 α はスケールファクタである。

なお、指差し動作の認識に失敗した場合は視線方向を変化させない。

3.4 プロトタイプシステムの評価

ここまで述べてきたプロトタイプシステムを開発し、数人の被験者に対して実験を行った。その結果、指示動作に対してシステムの反応が遅く、違和感を感じるなどの感想があった。今回開発したシステムでは、人による指示動作から映像の変化までの時間が 500msec 程度かかっているため、人によっては違和感を感じる場合があることが明らかになった。

4 結論

本稿では、ドーム型ディスプレイを用いた全方位ビデオの表示システムにおける、人と映像表示システムのインタラクションモデルを提案した。そのモデルにおいては、映像表示システムが人間の身体動作や音声などを通じて人の意図を認識してそれに即した映像を生成・表示する。従来用いられていた、道具による意図の伝達では、ドーム型ディスプレイがもたらす没入感を損なってしまうという欠点があったが、身体動作などの自然な方法によって意図を伝達することで、より効果的な没入型映像表示システムが実現できることが期待される。

さらに、本稿で提案したインタラクションモデルでは、映像表示システムは人の意図に従うだけではなく、過去の履歴や映像コンテンツの特質から判断してより効果的な映像を生成・表示する。これは、

没入型ディスプレイが人の得るほとんどすべての視覚情報を与えることを利用したものである。

今回開発したプロトタイプシステムでは、人の意図を伝達する手段として、指差し動作に着目した。プロトタイプでの評価の結果、人が指示を与えてからシステムが反応するまでの時間が長いと違和感を感じる場合があり、その程度が人によって異なることが明らかになった。これは、応答時間によって人が受ける印象が大きく異なる、すなわち、人と映像表示システムとのインタラクションを考えると時間に時間の概念が不可欠になることを暗示している。

今後の課題としては、応答時間と人が受ける感覚との関係を明らかにし、それを反映したインタラクションモデルを構築することがあげられる。また、今回開発したシステムは指差し動作に限定したものであったので、もっと様々な動作によって人の意図を伝達できるようにすることや、より客観的な性能評価を行うことも課題としてあげられる。

謝辞

本研究は科学研究費補助金特定領域研究 13224051 の補助を受けて行った。

参考文献

- [1] 横矢直和, 山澤一誠, 竹村治雄, 全方位ビデオカメラを用いた視覚情報メディア, 情報処理学会論文誌, Vol.42, No.SIG13, pp.59-70, 2001.
- [2] 柴野伸之, 畑中智行, 中西弘泰, 澤田一哉, 野村淳二, 都市環境ヒューマンメディアにおける球面型没入ディスプレイ, 松下電工技報, No.74, pp.56-61, 2001.
- [3] 小池英樹, 小林貴訓, 佐藤洋一, 机型実世界指向システムにおける紙と電子情報の統合および手指による実時間インタラクションの実現, 情報処理学会論文誌, Vol.42, No.3, pp.577-585, 2001.
- [4] Zhengyou Zhang, A flexible new technique for camera calibration, IEEE Trans. PAMI, Vol.22, No.11, pp.1330-1334, 2000.
- [5] 田中宏一, 和田俊和, 松山隆司, 3次元人体形状計測に基づく指差し動作の解析, 情報処理学会研究報告, 2002-CVIM-133, p.125.