

# モデルを位置合せに用いた注釈提示と 注視点による物体選択

島田 恭宏・河野 良一\*・島田 英之・濱本 高志  
大倉 充・宮垣 嘉也

岡山理科大学工学部情報工学科

\* 松下システムテクノ株式会社

(2002年11月1日 受理)

## 1 はじめに

現実世界へ仮想世界で生成されたオブジェクトを合成することで、現実世界に情報を付加する拡張現実感 (Augmented Reality: 以下 AR と略す) が新たな情報の提示・操作手法として注目されている。特に実環境 (現実情景) への注釈情報の重ね合せは、未知のオブジェクトへの解説の付与など、応用面で期待の高い AR 技術のひとつといえよう。

現実空間と仮想空間を融合させるためには、空間的ずれを解消せねばならない。これは、現実世界と仮想世界の空間座標を一致させ、観察者の位置・視点方向を正確に固定し、観察視点の移動にも追従できる必要がある。

これまでに現実世界と仮想世界の空間座標を一致させる手法として、人工的なマーカ群を実環境中に配置し、映像中から検出されるマーカの位置情報に基づいてカメラの位置と姿勢を用いる手法や実験環境のパノラマ画像を複数用意し、利用者が有するビデオカメラからの画像とマッチングを行う手法 [1] などが提案されている。また、注釈提示対象オブジェクトにユニークなマーカを添付し、それを抽出・認識することによって注釈提示を行うという位置合せを必要としない手法 [2][3] も提案されている。しかし、これらの手法は画像をベースとしており、計算コストが非常に高い。注釈提示システムはリアルタイム処理が前提であり、注釈提示時の時間遅れは空間的ずれと同様に致命的である。

また、単なる位置合せによる注釈提示は、ビデオカメラで撮影された情景中に複数の注釈提示オブジェクトが存在した場合、そのすべてが表示されることになる。HMD 等のデバイスの解像度は、ただだか 640 × 480 Pixel であり、このような低解像度のディスプレイに無数の注釈情報を提示することは利用者に認知的負担を強いることになる。筆者らは、注釈情報は単一の物体に対する表示が妥当と考える。

そこで本稿では、注視点による注釈提示対象オブジェクトの選択可能なモデルベースの注釈提示システムについて報告する。

## 2 モデルベースの注釈提示と注視点によるオブジェクト選択

### 2.1 モデルベースの注釈提示手法の概念

本稿では、注釈情報が添えられたモデルベースによる位置合せ手法を提案する。本来 AR における位置合せの問題は、観測者やカメラ視点の 3次元位置方位を求める問題に帰着できる。注釈提示システムではそれに加え、実環境の知識として注釈提示位置 (注釈提示対象オブジェクトの位置) と注釈情報が必要となる。本提案では、磁気式の 3次元トラッカを用い、観察者の視点位置を取得し、実環境の知識として、実験環境をそのまま仮想空間に CG (Computer Graphics) で再現したモデルを使用する。モデルは、実験環境の場、およびそこに存在する各オブジェクトを CG の図形プリミティブとして構成される。通常、この

ような3次元CGにより構成された任意のシーンを人間が観察するためには、ディスプレイ等の2次元平面上に描き出さなければならない。そのために透視投影変換やビューイング変換が行われる。詳細は専門書に譲るが(たとえば[4]など)、適当な視体積、視点位置、視線方向等を与えることによって、構成されたプリミティブ群(モデル)を任意の位置・方向から観察することが可能となる。

本実験においてCGにより作成されるモデルは、現実のものと同じサイズ、距離から構成される。ここで、そのモデルと対応する実験環境を観察する観察者がいるとする。透視投影変換において、人間が肉眼で知覚するシーンと同等の見えを実現する視体積を定義し、かつ、その観察者の視点位置、視線方向等が3次元トラックによりリアルタイムに取得可能ならば、これらの情報をCGのシーン描画に使用することで観察者が肉眼で観察する実環境のシーンとCGにより描画されたシーンを一致させることが可能である。当然、観察者の視点移動にも追従することもできる。たとえば、ここで生成されたCGシーンを光学シースルー型HMDに出力し観察者が観察するとすれば、透過して知覚される現実環境のシーンとCGにより生成されたシーンは完全に一致して知覚される(図1参照)。

実際の注釈提示では、モデルは観察者に提示せず、注釈情報のみを提示する。よって、モデルと空間座標が完全に一致した別の仮想世界を設け、ここに注釈情報のみを構成し、観察者に提供することで、光学的に知覚している実環境の情景に注釈情報のみを提示し融合させることが可能となる。また本提案では、CGにより描画された各オブジェクトは、そのまま注釈情報を収めたデータベースへのタグである。よって、何らかの方法で単一オブジェクトを選択すれば、注釈情報を選択することが可能となる。

観察者が実験環境を任意の視点位置から任意の方向で観察したシーンと同等のシーンをCGにより描画するゆえ、この時点で位置合わせは完了している。もっともこの手法では、観察者の視点位置・視線方向を検出するセンサの精度に位置合わせの精度は依存することになる。しかしながら、画像処理的手法を用いるような煩雑な処理は皆無であり、高速に処理が可能となる。

## 2.2 視点を用いた物体の選択

詳細は次章に譲るが、本実験では視点計測装置を用いている。これは、装置が有する視野カメラで観察者が見ているシーンと同等のシーンを映像化し、その眼球運動を解析して視点位置を視野カメラの映像上の座標として出力する。そこで、前節で示したモデルをディスプレイ上に描画する場合の透視投影変換において、視体積を視野カメラのそれと同等に定義し、視野カメラの視点位置・視線方向等によりビューイング変換を行うことで、モデルは視野カメラ映像と同等のシーンが描画できることになる。前述のように、視点計測装置は観察者の視点位置を視野カメラ映像のX-Y座標として出力することができる。視野カメラ映像とモデルが一致していれば、この座標を使用して、モデル内における単一オブジェクトを選択することが可能である。

## 2.3 注釈の表示

ARに使用されるHMDは通常、両眼式のもので使用される。本実験は、シースルー型HMDの使用を前提とした注釈提示システムの構築を指向したものである。しかし現在の両眼式HMDでは、これを装着し、HMD越しに視点を計測することが不可能である。よって本実験では、単眼式HMDを使用することにした。前述のように、従来ARでは、実環境と注釈情報の合成が行われるが、本実験では実環境のシーンと注釈情報を合成せず、単に選択されたオブジェクトの注釈情報のみを単眼式HMDに表示する。

# 3 注釈提示システムの構築

## 3.1 視点検出装置

本実験では、視点計測にナックアイマークレコーダEMR-8(nac社製)を使用した(以後EMR-8と称する)。EMR-8は、コントローラ、ヘッドユニット、アイマーク検出ユニットから構成され、ヘッドユ

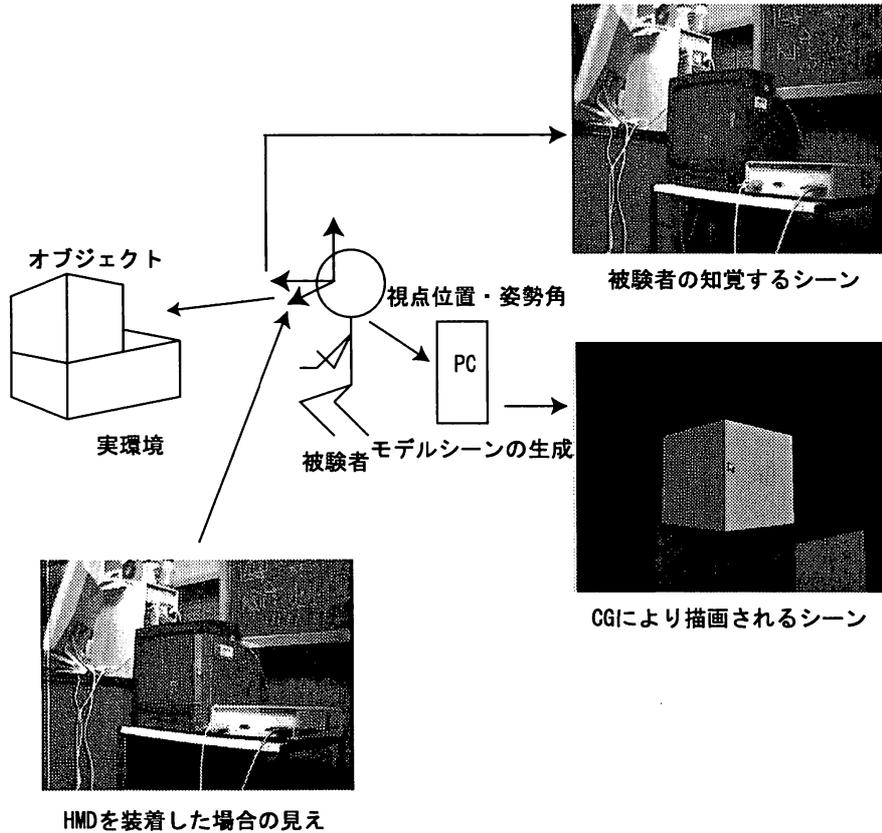


図 1: モデルベースの注釈提示手法の概念

ニットに存在する視野カメラの映像に対し、観察者がどこを見ているかをアイマーク検出ユニットで眼球の動きを検出し、視点座標（アイマーク座標）を視野カメラ映像の X-Y 座標として出力する。検出間隔は 1/60 秒である。

### 3.2 視点計測実験

EMR-8 より計測されるアイマーク座標の精度を調べるために、予備実験を行った。実験環境を図 2(A) に示す。実験は視野レンズの画角 44 度を選択した EMR-8 を装着させた被験者に 1m 離れた黒板に書かれた 36cm × 26cm の長方形オブジェクトの頂点を (1) 左上 → (2) 右上 → (3) 右下 → (4) 左下の順に見てもらい、各頂点におけるアイマーク座標を計測した。実験では、被験者の頭部は強制的に固定しておらず、頭部を動かさないようにして実験に臨むよう依頼しただけである。取得するアイマーク座標は、被験者が注視したと考えた任意の時刻におけるその時点でのものである。実験は、同一被験者による計測を 2 回試行した。結果を図 2(B)(C) に示す。

図は、視野カメラ映像に獲得したアイマーク座標をプロットし、各座標を視点の移動と同じ順番に線分で結んだものである。図からわかるように、計測されたアイマーク座標はオブジェクトの頂点と正確に一致していない。その差は各水平垂直座標において最大 22 pixel を有している。そこで、注視時（3 秒間）の視点の経時変化を測定してみた。その結果を図 3 に示す。

各フレームごとに取得座標を比較すると、数 pixel 程度の範囲で振動している。これは、振顫 [5] によるものと考えられる。振顫とは、人間の眼球の生理的機構であり、注視している場合にも生ずる眼球の振動である。EMR-8 は角膜反射像（プルキニエ像）の位置と、瞳孔中心位置の相対的な距離からアイマーク

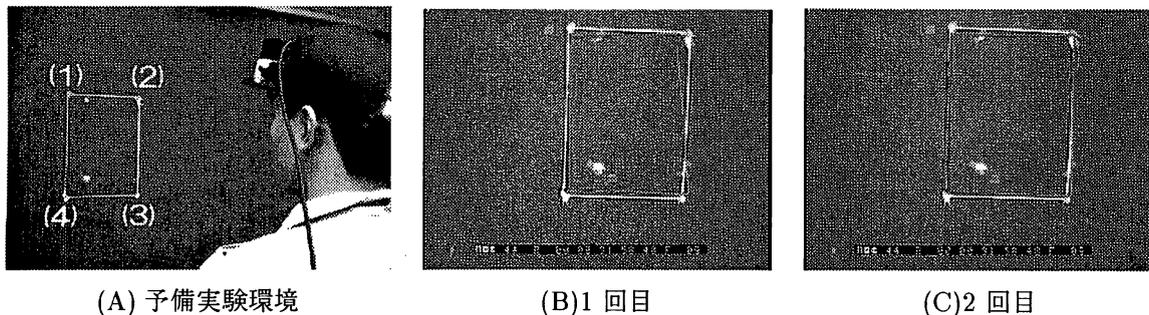


図 2: 予備実験環境と計測結果

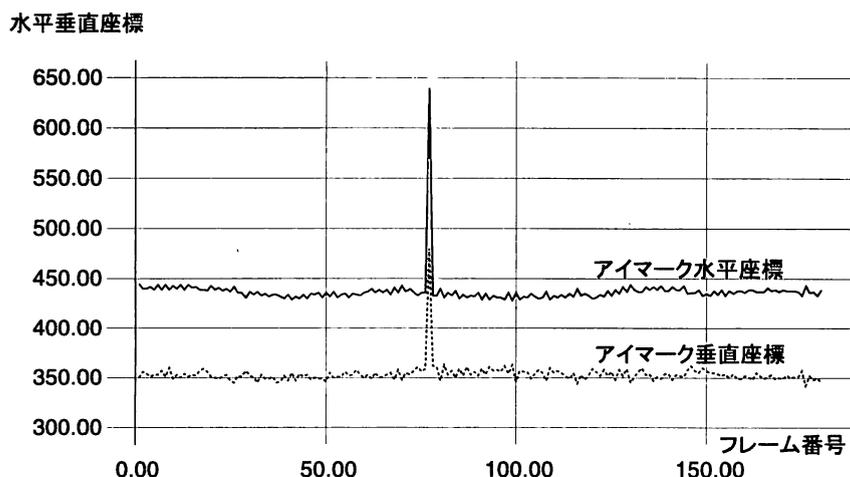


図 3: アイマーク座標の経時変化

を検出している。そのため、振顫による眼球運動を検出してしまい、アイマーク座標が振動しているものと考えられる。また、計測値は微小ではあるが「うねり」を伴っている。EMR-8 のサンプリング間隔は  $1/60$  秒と比較的短い。それに対して人間の動作は遅く、それによるアイマークの変化は時間的には比較的長いスパンで出現すると考えられる。計測時、被験者の頭部は固定していない。そのため、このような緩やかな変化として頭部の動きが影響したものと考えられる。結果、視野映像に対する注視点の大きなずれの原因は、フレーム間でのアイマーク座標のずれがただか数 pixel であることから、頭部の動きが影響していると考えられる。

なお図中、77 フレーム目に不連続な座標が計測されている。これは、被験者の眼鏡のレンズの反射や外光の影響により EMR-8 がうまく眼球の動きを計測できずに発生したエラーであることがわかっている。

### 3.3 注視点の決定

前節の検討から、視野映像と注視点のずれは、計測時の頭部の動きによるものと判断される。この予備実験のように被験者の動きを制限した場合には、視野内に基準点を設け、ずれを補正することは容易であろう。しかし、想定するシステムでは被験者は実験環境内を自由に動き回ることになり、そのような環境内に基準点を無数に用意することは現実的ではない。

そこで本実験では、複数の時刻におけるアイマーク座標より注視点を決定することとした。まず、注視という動作を以下のような手法で規定する。EMR-8 より得られるアイマーク座標を 1 秒ごとに区切り、各

時間内における座標の平均値を求める。この平均値を3セット求め(3秒分のデータ)各々の差を検出し、その差がしきい値以下となった場合に、注視したと断定することとした。また、そのときの注視点座標は、先に求めた3つアイマーク座標の平均値より重心を求めることで得ることとした。

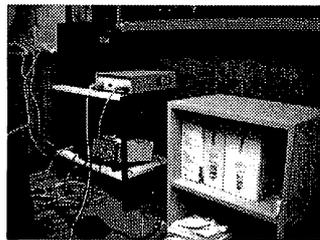
### 3.4 実験環境と作成したモデル

本実験では、注釈を表示するオブジェクトとして、棚、カラーボックス(2種類)、テレビ、机、パソコンを対象とした。これらを配した実験環境の模式図を図4に示す。実験環境は約 $3m^2$ である。矢印で示した視点1~3において観測できるシーンも同図に示している。また、図中の計測フィールドとは、後述する3次元トラックの計測域を示している。本来であれば、実験環境すべてを包含する計測フィールドが理想的であるが、3次元トラックの性能からこのような範囲で実験を行うことにした。

モデルの作成は、各々の物体の大きさを実験環境中での配置位置を計測し、OpenGLを使用して作成した。図5に仮想空間上に作成したモデルを示す。また、モデルの描画時の大きさは、 $640 \times 480$ である。



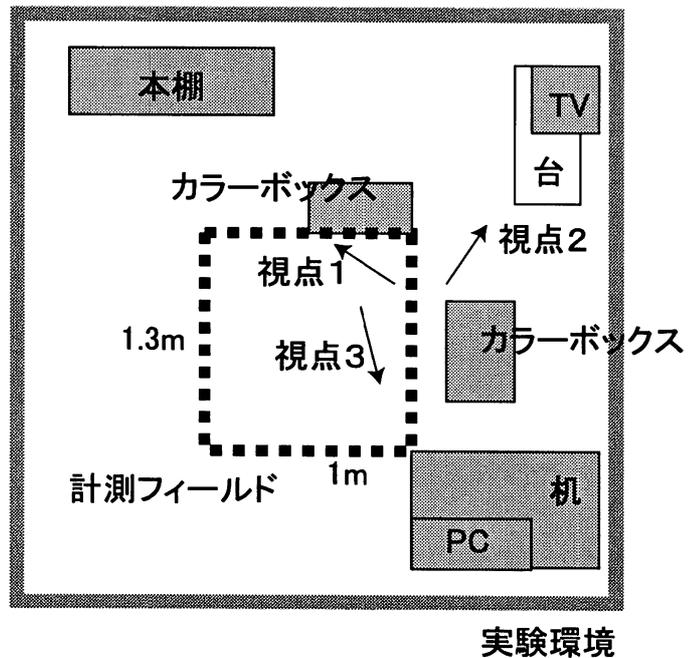
視点1



視点2



視点3



実験環境

図4: 想定した実験環境

### 3.5 視野映像とモデルのキャリブレーション

EMR-8の計測座標は、視野映像のピクセル座標 $640 \times 480$ の座標系で計測されるため、モデルの描画ウィンドウサイズも $640 \times 480$ の大きさで作成した。さらに、仮想空間上の物体には、カメラの画角(EMR-8の視野カメラの画角44度に合わせる)及び、スクリーンのアスペクト比(縦横比)を使用して透視投影変換を行う。よって、図6のように視野映像のピクセル座標系とモデルを表示するウィンドウのデバイス座標が一致することになる。

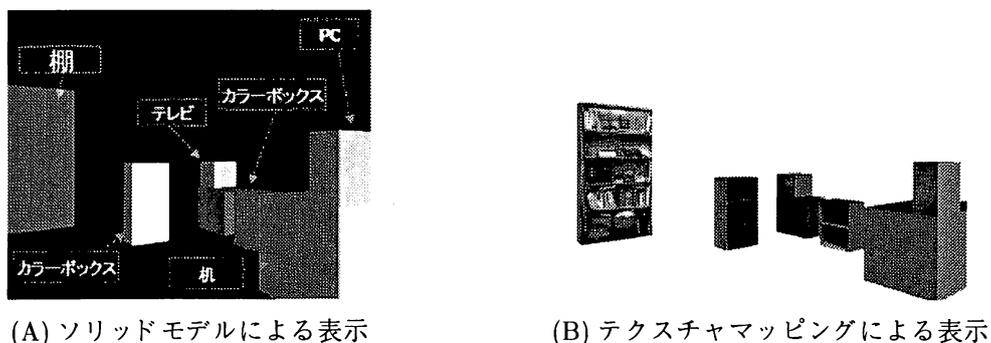


図 5: 作成したモデル

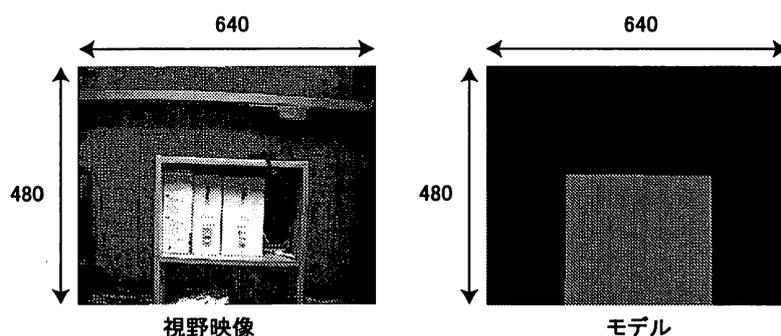


図 6: 座標系の一致

視野映像とモデルのキャリブレーションのため、3次元トラッカ (POLHEMUS 社製 3 SPACE ISOTRAK II) のレシーバを EMR-8 のヘッドユニットに装着する。この 3次元トラッカは、3次元位置座標値 (X, Y, Z), 及び姿勢角 (azimuth, elevation, roll) を計測することができる。3次元トラッカから計測される 3次元位置座標値 (X, Y, Z) より、観察者の実環境における視点位置、そして、姿勢角 (azimuth, elevation, roll) により視野の方向を取得できる。ここで roll 角の変化は特殊な場合と考えられるゆえ (頭部を左右にかしげた状態)、簡単化のために roll 角は考慮しないものとした。

### 3.6 視点によるオブジェクト選択

視点によるオブジェクトの選択は、マウス操作 (マウスカーソルの移動とマウスボタンのクリック) を視点の移動と注視という眼球運動によりエミュレーションすることで実現する。まず、マウスカーソルの移動であるが、EMR-8 から計測されるアイマーク座標を作成したモデルのウィンドウ画面が表示されている PC のマウスポインタに与える。3.5 節より、ピクセル座標系とデバイス座標系が一致しているゆえ、マウス・ポインタの位置と EMR-8 から計測される視点の位置は一致する。よって、視野映像 (視野カメラの映像) のアイマーク座標が PC 上のマウスポインタの座標となる。これによりモデル内の単一オブジェクト上にマウスカーソルを移動させることが可能となる。マウスボタンのクリックは、3 秒間注視することでマウスボタンのクリックイベントを発生させ実現する。つまり、観察者は注視を提示させたいオブジェクトを 3 秒間注視するだけでこれらの操作を行うことができる。具体的には、注視によりクリックイベントが発生した時点でのマウスカーソルの座標を取得し、その後、OpenGL のセレクションメカニズムを使用してオブジェクトの特定を行う。各オブジェクトにはラベルが付されており、オブジェクト選択の結果として、このラベルが得られる。ラベルは注視情報へのタグであり、注視はこのラベルから選択される。

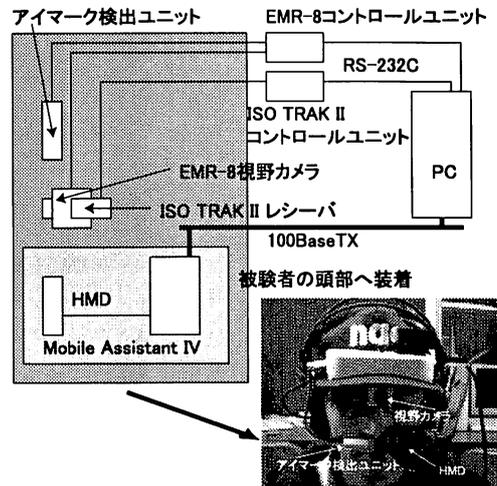


図 7: 構築したシステムの構成

## 4 実験と考察

### 4.1 構築したシステムの構成

図7にシステム構成を示す。観察者はHMD（左眼）とEMR-8を装着（右眼）し、実験環境中に位置するものとする。また3次元トラックのレシーバは、EMR-8の視野カメラ付近に設置する。3次元トラックとEMR-8のコントローラはRS-232CでPCと接続され、随時3次元トラックからは視野カメラの実験環境中での3次元座標と姿勢角を、EMR-8からはアイマーク座標を取得する。PC(EPSON DIRECT製, CPU: Pentium 733MHz, OS: Windows Me)には事前に実験環境のモデルが構築してあり、3次元トラックからの計測データをもとに、リアルタイムでモデルの描画を行う。それと同時にEMR-8からのアイマーク座標をモデルに反映させ、注視と判断された場合にオブジェクトの選択を行う。オブジェクトの選択が発生した場合には、ネットワークを介して接続されたHMD(ザイブナー社製, VGA)に選択されたオブジェクトのラベルを伝達し、HMDはそのラベルに対応する注釈情報を表示する。なおHMDは、ウェアラブルコンピュータ Mobile Assistant IV(ザイブナー社製, CPU: MMX Pentium 233MHz)を使用した。このHMDは、単体で他のPCと組み合わせて使用できないため、コンピュータシステム一式を注釈表示用に使用した。

### 4.2 実験結果と検討

実験空間内を観察者が移動し、任意の物体を注視することで注釈提示対象を選択し情報を提示する実験を行った。まず図8に実験中の視野映像とそれに対応したモデルの描画結果を示す。図中、円で示した部分が各々の映像における注視点を示している。図9に同様の環境において、注釈提示を提示した結果を示す。図は観察者が実験フィールド内で“テレビ”を観察した場合の視野映像、モデル画像、注釈画像の各々を示している。

空間的整合性については、モデルの描画、および、注視点の位置のずれが生じているものの、視線で物体を選択することができた。現実世界の物体とモデルの間で、誤差が生じた原因として、3次元トラックの計測誤差と、モデルの透視投影変換による誤差が考えられる。

3次元トラックより計測される値は変動し一定の値が計測できない。よって、計測された値を物体に与えるとモデルと現実世界の物体の間で誤差が生じてしまう。また、3次元トラックの誤差を大きくする要因として、HMDなどの機器から発生する電磁波などによる影響や計測フィールドにおける金属の影響が挙げられる。本実験で使用した3次元トラックは、トランスミッタにより生成される磁場をそのフィールド中に存在するレシーバが感知し位置座標と姿勢角を出力する。そのため、電子機器などから発せられる

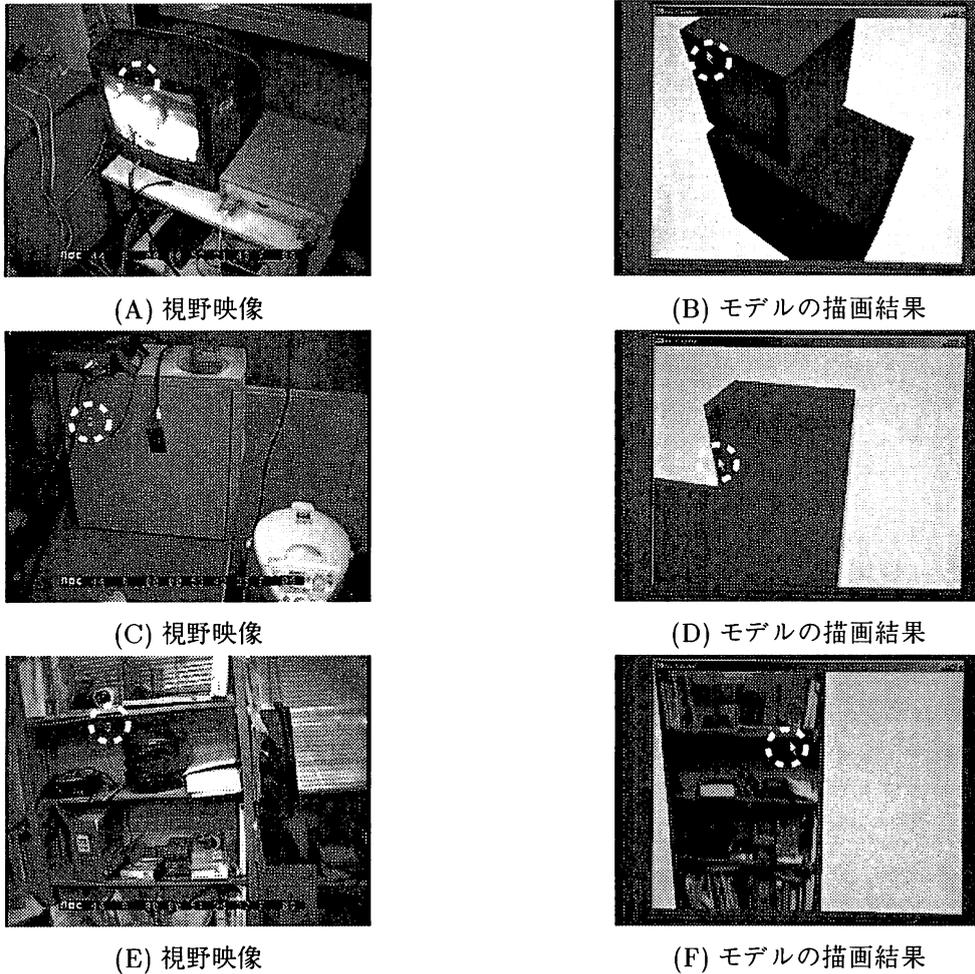


図 8: 実空間とモデルの位置合せ結果

電磁波やフィールド近辺に存在する金属物に影響されやすく、これらが原因となって誤差が発生したのと考えられる。これらの原因により生じた誤差は、仮想空間上に作成したモデルの視点を変化させ補正を行った。しかし、3次元位置センサの値が変動してしまうため完全な補正を行うことができなかった。モデルの透視投影変換による誤差に関しては、モデルの視点を変えることにより補正を行っている。しかし、3次元位置センサの計測値が変動してしまうため、変動した値に対して随時補正を行わなければならない、完全な補正を行うことができなかった。

このように、誤差が生じたにもかかわらず物体を選択できた理由として、本研究では、対象とした物体が、本棚、PC、机といった比較的大きな物体を対象としたことがあげられる。このように大きな物体を対象としたため、モデルと物体の間で誤差により完全に位置ずれを起こすことはなかった。そのため、一致している部分に関しては物体を注視することにより、選択を行うことができた。また、この誤差は、観察者が物体に近づくとつれて、視野カメラ中の物体の大きさが大きくなり、それに伴いモデルも画面に対して大きくなるため、物体に近づけば近づくほど誤差が小さくなり物体を選択することが容易である結果が得られた。3次元トラックの値を理想値でシミュレーションを行った実験では、確実に物体を選ぶことができた。そのため、3次元位置センサの計測誤差を改善することができれば物体を確実に選択することができると思われる。

時間的整合性に関しては、本実験での位置合せの処理が比較的簡単なことからビデオレートとまでは行

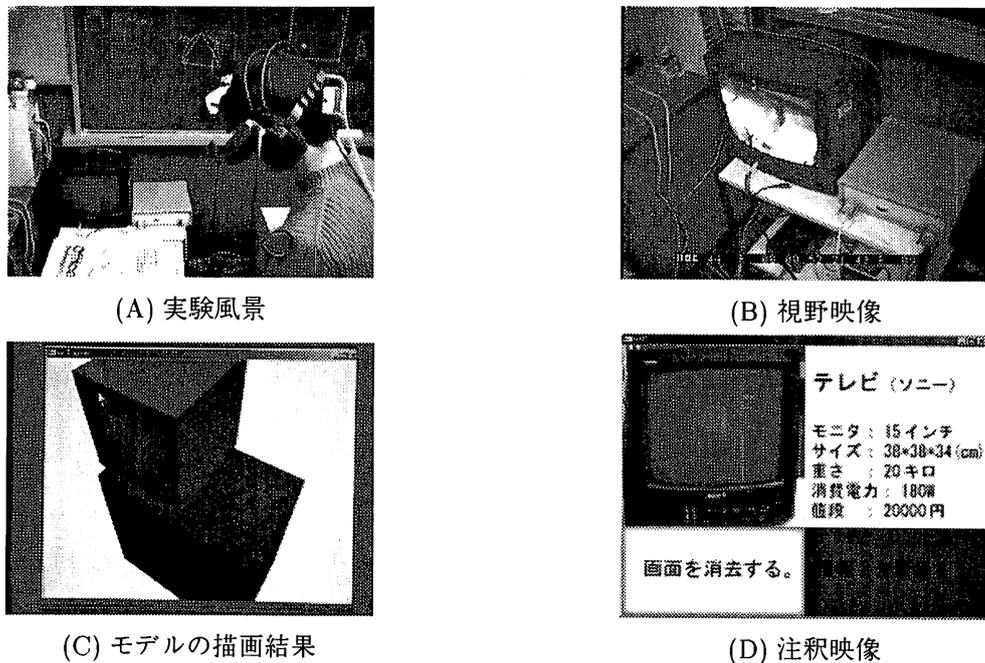


図 9: 注釈提示結果

かないまでも視点移動に充分追従して再描画でき、かつ注視点を求めることが確認できた。

## 5 むすび

本稿は、モデルを位置合せに用いた実環境に対する注釈提示システムと注視点による物体選択について述べたものであり、構築したシステムにより物体選択・注釈提示実験を行い、その結果を示した。空間的ずれの解消にはまだ不完全さが残るものの、実環境中の単一オブジェクトを注視点で選択可能であり、かつ、すべての処理がリアルタイムに動作可能なことが確認できた。

本手法は、モデルを使用して物体を特定しているため、物体の位置が変わらない限りどの方向から物体を見ても物体を特定する処理には影響を及ぼさず高速に処理できるという特徴を有している。反面、モデルに基づくため移動体やオブジェクトの移動には対応できない、モデルの作成に対する負荷が高いなどの欠点も存在する。今後は、空間的位置ずれを改善するためにロバストな視点位置の決定方法を考慮せねばならない。

### 参考文献

- [1] 興梠正克, 蔵田武志, 坂上勝彦, 村岡洋一: パノラマ画像群を位置合せに用いたライブ映像上への注釈提示とその実時間システム, 電子情報通信学会論文誌, Vol.J84-D2 No.10 p.2293-2301(2001).
- [2] Jun Rekimoto: Augmented Interaction: Toward a New Human-Computer Interaction Style Based on Situation Awareness, Interactive Systems and Software II (WISS'94 Proceedings), pp.9-17, 1994.
- [3] Jun Rekimoto and Katashi Nagao: The World through the Computer: Computer Augmented Interaction with Real World Environments, User Interface Software and Technology (UIST '95), 1995.
- [4] 技術編 CG 標準テキストブック編集委員会: 技術編 CG 標準テキストブック, 財団法人画像情報教育振興協会 (1999).
- [5] 松田隆夫: 視知覚, 培風館 (1995).

## A Model-Based Annotation Overlay on Real World and Object Selection using Gaze Point of the Eye

Yasuhiro SHIMADA, Ryouichi KOHNO\*, Hideyuki SHIMADA,  
Takashi HAMAMOTO, Mitsuru OHKURA and Yoshiya MIYAGAKI

*Department of Information & Computer Engineering,  
Faculty of Engineering,  
Okayama University of Science,  
Ridai-cho 1-1, Okayama 700-0005, Japan*

\* *Matsushita Systems and Technology Co., Ltd.*

(Received November 1, 2002)

In this paper, we report on the development of an interface using the gaze point of the eye, which can be used in the real world or in a HMD. Previous research on a system that gives specific information about an object has been conducted in various fields including augmented reality (AR). In AR research, the line of sight of the HMD user is determined by the direction of a camera mounted on the user's head, or using a measuring device that detects the direction of the head. Using these methods, the system can determine what the user is observing, and offer details about the object. However, these methods suffer problems. For example, when multiple objects appear in the field of view, these methods cannot be used. On the other hand, if we can determine the line of sight of the user from the movement of the eye, we can decide what the user is observing. In this paper, we investigate this approach.