

What's Virtual? —バーチャルリアリティーにおける視覚提示技術の虚実—

前 田 太 郎

1990年代初頭より提唱されたバーチャルリアリティー (VR) という言葉が日本語において「仮想現実」と訳されて言の端にのぼるようになってはや10年が過ぎた。この間の研究者らの不断の啓蒙にもかかわらず、その技術と言葉のもつインパクトが「仮想と現実の区別がつかなくなる不安」という語り口でマスコミが煽り文句のように煩雑に用いるようになってこの言葉のもつ意味がどんどん歪曲されていった感がある。

「Virtualとは仮想ではない」というのが、この分野の研究者たちの合い言葉にならざるを得ないのもこんな事情が背景にあるからである。それは本来の原語の意味を正確に伝えるなら「実物と等価な働きをする」といったほうが正しい。もちろん、実物そのものではないのであるから、完全に同じというわけにはいかないが、要はその差異について人間に気づかせなければよい、という工学的な決着点こそがVRにおける感覚提示技術の骨子である。そして、そのためには人間の知覚の隙を突く必要がある。本稿ではVR技術にみられる各種の視覚提示のための工夫がどのようにして人の感覚の隙を突いてvirtualたりえる臨場感を実現しているのかについて紹介してみたい。

1. 第0の隙～3D フォトリアリスティック CG では終われない

この最初の章だけは感覚の隙ではなく、技術の隙、とでもいうものである。VR技術はこの隙を埋める必要から生まれた。そもそも、視覚提示の観点からVR技術とはなんだろうといったときによくされる誤解が「立体に見えるリアルなCG技術」という理解である。しかし、立体に見

えるだけならそれは3Dグラフィックの提示技術であり、リアルに見えるというのは昨今ではフォトリアリスティック技術とよばれるCG加工の技法にすぎない。これらとVR技術を隔てる最大の違いは、実は観察者側に与える制約の違いにある。VR技術とは「観察者の能動的な行動を保証する」という一点において従来の提示技術と異なっているといてもよいだろう。観察者の運動に対応した提示というインタラクティブ性がVRをVRたらしめている要因であり、それゆえに視覚提示ひとつをとっても常に観察者の身体運動を計測しつつリアルタイムに提示情報を制御する必要に迫られる。これがVR技術の骨子となる。

2. 第1の隙～2視点画像で済ませる

本来、人間にとっての視覚とは自分の周囲の光線情報を知るためのものであるから、理想的には人間を取り巻く全周囲からくる光線群を再構成することができればそれは視覚的には実際と等価(=virtual)な視覚空間を提示できることになる。この考え方は波面再構成の考え方であり、この代表格がホログラフィーである。この技術は記録面上での波面の記録・再生そのものであり、現在もより高度な臨場感をもった記録再生のために多くの研究がなされているが、いかんせんその情報量の膨大さと取り扱うべき光学技術の分解能の高さのために、カラー化・動画化に関しては2次元映像の記録再生に比べてあまりに遅れが大きいのも事実である。むしろ2次元映像を多視点から記録再生することによってホログラフィーと等価な、つまりバーチャルホログラフィーを作ろうとする試みもある。しかしいづれにしても情報量の膨大さ自体は変わらない。

そこで多くのVR提示装置が利用している隙が「人間の目玉は2つだけ」という点である。つまり、視覚は両目の

東京大学大学院工学系研究科計数工学専攻 (〒113-8656 東京都文京区本郷 7-3-1)
E-mail: maeda@star.t.u.tokyo.ac.jp

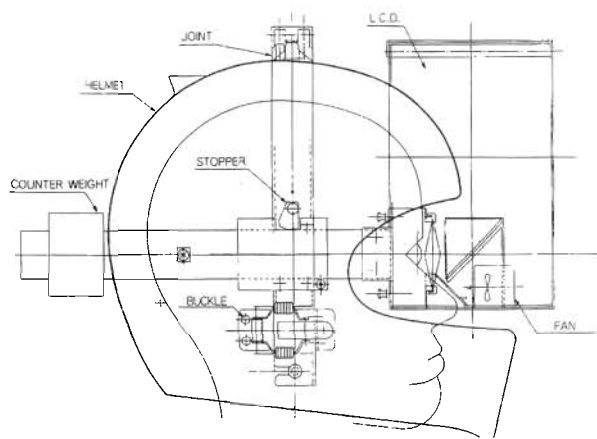


図1 HMD構成の一例。LCDパネルをミラーとレンズ系を通して覗き込む構造になっている¹⁾。

網膜像に頼っているのだから、各眼視点の2次元画像2枚で済ませてしまおうというものである。視点限定のパーティクルホログラフィーというわけだ。アイデア自体はホログラフィーより早く立体写真の昔からあるもので、両眼視による立体視が成立することで単なる「絵」であった写真が立体感という「リアリティー」を得るわけだが、当然写真では映像は固定視点からの静止映像でしかない。次に、映画の登場で立体映画が実現して動画となったが、これにしても観客の視点は撮影時のカメラ位置に縛られたままである。勝手に視点を動かそうものなら見える世界はその立体的な構造においてぐにゃぐにゃと歪んでしまう。

3. 第2の間～実時間生成で済ませる

この2視点画像提示に、リアルタイムに変更できる映像であるコンピューターグラフィックス技術を組み合わせることによってようやくバーチャルリアリティー技術は陽の目をみる。実際に提示されている映像は2次元画像2枚のみだが、頭部運動を計測することで視点位置を計算し、この動きを反映した画像をリアルタイムで合成・提示し続けるという仕組みである。

このために開発された画像提示デバイスがヘッドマウンティッドディスプレイ (HMD) であり、2枚の液晶画面と接眼光学系を頭部に搭載して常に左右眼に同一画面を安定提示する仕掛けである (図1)¹⁾。これを頭部の3次元位置・姿勢を計測する磁気センサーと組み合わせるのが一般的な構成であり、近年の小型化によってデザインも洗練されサイバーの代名詞化しているものの、当初は「蠅男がめくら踊りを踊っている」と称され、その奇矯な風貌を揶揄されたものであった (図2)²⁾。

HMD提示で用いられている画像生成技術は、移動視点にすでに存在するべき光線群を動いてしまってから計算・

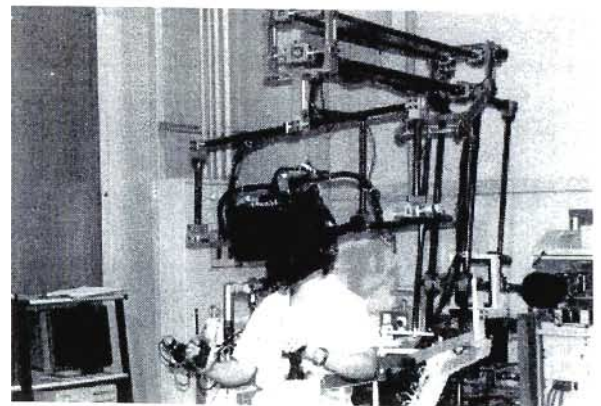


図2 HMDの装着状態。装着者にしか見えない世界に没入して作業をしている状態は傍目には奇異に映る²⁾。

生成するという光線情報の自転車操業ともいうべきものであり、そこには因果の逆転が生じている。それゆえに、この技術のもうひとつの本質は「因果の逆転を気づかせないほど短時間の遅れで済ませる」という「時間の隙」を突いていることにある。VRの世界での経験則としてこの時間遅れはおおむね100ms以下なら及第点、200msを超えるといわゆるVR酔いを引き起こすとされている。

一方でこのリアルタイム性さえ確保できるのであれば、実は生成される映像はコンピューターグラフィックに限る必要はない。実映像を撮影しているカメラを搭載したロボットが人間に追従してまったく同じ動きをしてくれるならば、観察者はあたかも自分がロボットの位置に存在しているかのような視覚を得ることができる。これがトレイグジスタンス技術である。この技術も起源は1960年代まで遡れるが、やはりその本質である「気づかせないほどの遅れで追従する」という要求を満たした完成型に至るには1989年のトレイグジスタンスシステムの登場をまつこととなる³⁾。

しかしながらこのリアルタイム性という要素の重要性は意外なほど浸透が遅く、数年前までは「秒間10フレームの更新レートですから100ms以内の遅れで済んでいます」と真顔で胸を張る技術者が珍しくなかった。問題は画像の更新レートではなく、装着者の動作から画像変化までの時間遅れなのだから計測時間遅れと提示時間遅れの和としてのスループットが問題なのであって、この認識のずれが多くての市販のVRシステムをして「立体的で綺麗だけど世界ががたがたしている」という評価につながり、その真価を理解されぬまま「アミューズメントでなら我慢できるかな」といった印象を残す結果となった。CGという技術がリアルタイム性と無縁のワークステーション上で発達してきた技術であっただけにその間隙を埋めるために要した

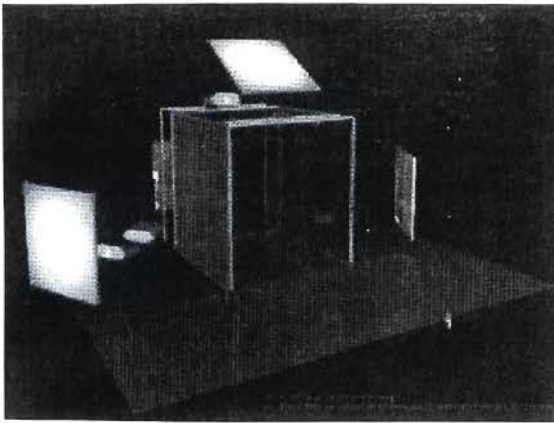


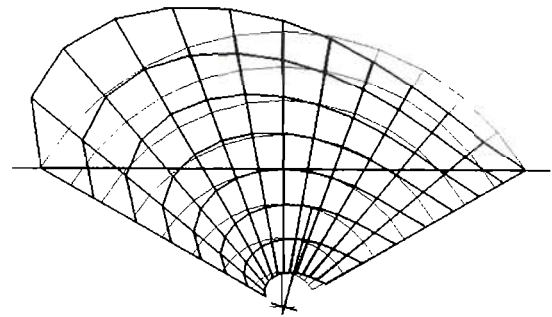
図3 包囲型ディスプレイの代表例：CAVE⁴⁾。

時期ではあったのだが、その間に最も強くリアルタイムの要請を受けてこのギャップを埋めて発展していったのがまさしくアミューズメント分野でのゲーム用CG技術であったことは学術分野からみれば皮肉な話である。

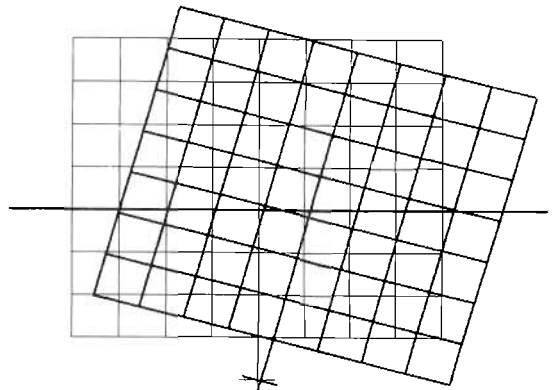
4. 第3の隙～歪みで済ませる

この時間遅れへの対処が技術的なブレイクスルーを迎えないまま遅れていたことが、VR用視覚提示系の開発の方向性に意外な影響を与えることになる。HMD系でのリアリティーが頭打ちとなっている間に、旧来技術であった立体映画のマルチスクリーン化をVR的に発展させた包囲型ディスプレイ (immersive display) が台頭してきたのである (図3, <http://www.evl.uic.edu/EVL/VR/> より転載)⁴⁾。これは理想的には6面の立方体の中に観察者を入れ、全方向の外周から各面を透過型スクリーンとして6組の立体映像を投影しようというものである。HMDに比べて観察者への拘束が少ないのが利点とされるが、やはり頭部追従によって映像を更新するため、観察者は1名のみであり、頭部計測センサーや3D映画同様の偏光または液晶シャッター眼鏡なども必要となるため完全な非装着というわけではない。そもそも、HMDが開発された背景にはこの包囲型システムがあまりにも空間と画像のコストを無駄に消費することに対する反省から必要十分な視野だけに画像を生成するために考案されたという経緯がある。しかしながら、CG技術がデスクトップユースのコンピューター技術の急速な進歩に支えられて恐竜的進化を遂げた結果、リアルタイム性を向上させる方向性ではなく、たとえ無駄でも全方向映像をあらかじめ提示しておくだけの画像生成能力を提供することを可能にしたわけである。

この方法はあらかじめ全方位の映像が準備されていることによって先述の「因果の逆転」を一部取り除くことに成



(a)包囲型ディスプレイの場合



(b)HMDの場合

図4 包囲型ディスプレイおよびHMDにおける世界の歪み⁵⁾。これは応答時間遅れによって15度の頭部回転の間に画像が更新されなかった場合に視空間がどのように歪んだように感じられるかについて示したもの。細線が提示されるはずの世界像、実線が観察者に見えている世界像である。(a)の包囲型ディスプレイの場合、世界像が方向について安定している代わりに奥行き方向に伸縮して歪んでいる。これと対称的に(b)のHMDの場合、世界像全体が頭部について動いてしまう代わりに像自体に立体的な歪みは発生していないことがわかる。

功している。すなわち視線の方向変化に伴う提示領域の変更についてはまさしく因果関係が正常に戻っているからである。まさに視線方向変化に対するバーチャルな視野の獲得に成功していることになる。しかし、本質的な時間遅れの問題を解決していない以上、その影響は残る。それが視点の位置移動に関する成分である。HMDと違い、頭部に固定されていないスクリーンとの観察者の視点位置の相対運動は常にリアルタイムに提示映像に反映されていなければならない。これはHMDとはまったく逆の事情である。この結果、時間遅れの影響は図4のように現れる⁵⁾。

要約すれば、映像の更新が視線の方向移動に遅れることによって包囲型ディスプレイでは世界が歪み、HMDでは世界がゆれるのである。別な観点からはこれらはそれぞれ回転と並進に関してバーチャルな視野を提供するデバイスであると整理することもできる。

よく誤解を招くいい方として「包囲型ディスプレイは広

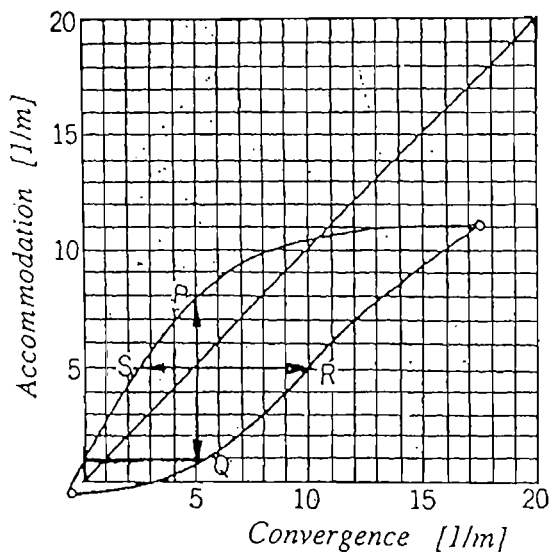


図5 水晶体調節と輻輳の食い違いによる融像限界⁶⁾。縦軸と横軸はそれぞれの手がかり提示距離の逆数。

視野で世界の広がりを感じさせる。HMDは没入感で対象物をリアルにみせる」という評価があるが、実際には提示領域・解像度の問題もさることながら、包囲型ディスプレイが環境提示に適しているといわれるのはこの世界像の安定感のためであり、HMDが対象物提示に適しているといわれるのはこの対象物提示の低歪率のためなのである。ことほどさようにリアルタイム性と臨場感の関係はまだまだ浸透が遅れている。

5. 第4の際～固定焦点で済ませる

さて、ここまではやや恣意的に立体視、と一言で片づけてきた両眼提示の手法であるが、これ自体がバーチャルな要素を含んでいることに目を向けておこう。本来の生理的な奥行き手がかりを考えた場合に、まず現行の立体ディスプレイが対応していないのが焦点調節の要素である。人間の目は水晶体調節機能をもつレンズ眼であり、網膜像の焦点ぼけを検出することで対象物の奥行きを検出することができる。しかしながら、これをディスプレイにおいて再現することはかなり困難である。単一奥行きに位置する2次元ディスプレイ上において映像に合わせて複数の焦点手がかりをリアルタイムに分布させる光学系を構成することがどれほど困難かは本誌の読者であれば容易に想像がつくであろう。そこで次善の策として接眼系において全体の焦点を一様に変動させる方法が考えられる。この場合、注視点の位置に焦点を合わせる制御戦略が一般的であるが、この場合は輻輳系-焦点調節系間の生得的な反射機能の存在を考慮するべきであり、輻輳角から焦点位置を決定するのが適切と考えられている。しかし、この制御系を実現する

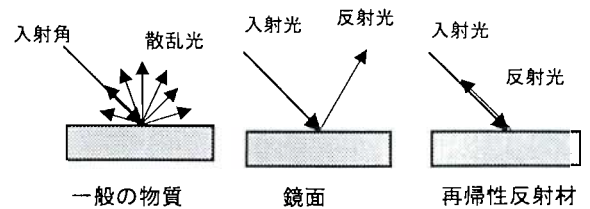


図6 再帰性反射材の反射特性。

には眼球運動の計測と焦点の自動制御が可能な接眼系を装備する必要があり、頭部に搭載するにはかなりの困難が伴う。このために、一般的なHMD設計では画面の焦点距離を1mに固定している。これは前述の反射系による融像限界(図5)から決定された値で⁶⁾、無限遠から20cmの近傍まで融像が可能となる焦点距離である。輻輳と調節のずれは違和感を生じさせるものの、融像の崩れは決定的な問題となるためにこれを避けることによって実用的な決着を狙った設計指針である。ちなみに、包囲型ディスプレイであるCAVEにおいてもおおむね1辺が2~3m程度に設定されており、中央付近に視点を置く限りはおおむねこの条件を満たすことができることになる。

一方で、固定焦点であっても結像の焦点深度を十分に広くとることができれば、この違和感を軽減することが可能になる。この理想的な光学系がマクスウェル光学系であるが、その原理は強い絞りの存在を前提とするために提示系の光量とのバランスが問題となってくる。この問題を解決したのが、再帰性反射材と同軸光学系の組み合わせによるクリスタルビジョン⁷⁾である。

再帰性反射材は入射光をその入射方向にそのまま反射する性質をもつ(図6)のために、これをスクリーンとして用いると図7のような同軸光学系をとおして反射光を観察する場合、通常スクリーンに比べてその反射効率は等価的にきわめて高くなる。これによって、強い絞りをとおしたために通常スクリーンでは観察不可能な光量となった映像を鮮明に観察することが可能になる。このため結果的に映像の焦点深度の広がりが拡大されるとともに、スクリーン部での映像の視認性を確保しつつ非スクリーン部への投射映像の映りこみが観察されることもなくなる。同システムはこの構成によって、従来オーグメンティッドリアリティーと称される実物体と仮想物体の混在した環境の視覚提示において致命的な問題であった遮蔽問題を部分的に解決することに成功しているといえる。また、同軸光学系であることによって、実物体によって生じた影がスクリーン上に落ちる投影上の遮蔽問題も回避することに成功しており、遮蔽問題の回避という点において現段階で最も進歩したオーグメンティッドリアリティー提示環境を提供している

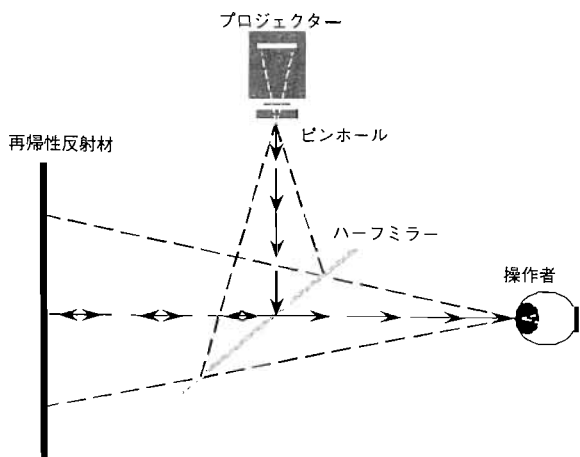


図7 クリスタルビジョンの構成。ピンホールによってバーチャルな瞳を構成し、これとおとしてハーフミラーで同軸光学系において投影された映像は再帰性反射材によって逆行し、眼球に収束する。



図8 クリスタルビジョンの実装例。実物体と投影映像の遮蔽関係が実現されていることがわかる。



図9 画像処理による視点再構成手法：virtualized reality⁸⁾。

(図8)。

6. 第5の間～水平視差で済ませる

さて、ここまではCGによる映像提示を前提としてきたために、立体視のためには両眼における幾何学的に正しい映像を提示するものとして議論してきた。しかし、実画像の撮像と再生をVR的な自由度で行いたいとする要求も当然のようにある。VR空間利用の遠隔会議などが、この代表格であり、この場合単なるCGでは相手の表情や視線、仕草などの非言語情報を読み取ることができなくなることが問題とされる。このため、なるべく実映像による映像情報の伝達が望ましいとされるが、これを実現しようとするならば常に必要とされる視点からの画像撮像と同時にそれを他の使用者に裸眼立体視で提示することが要求されることになる。これを実現する方法としては、全方向に多数のカメラを配置し、複数の固定視点の映像から任意視点の映像を合成するという方法があり、オフラインでの映像の再構成法としてはある程度実用の段階にある(図9, <http://www.cs.cmu.edu/afs/cs/project/VirtualizedR/www/VirtualizedR.html>より転載)⁸⁾。しかし、このためには撮像対象すべての3次元情報を画像処理によって計算した後、

任意視点からの2次元情報を再構築するという過程をとるために計算負荷が大きく、リアルタイム性を確保するためには大規模な専用の計算用ハードウェアが必要となる。そこで、この3次元の視点情報すべての再生を諦め、想定される視点位置を遠隔会議などの状況で最も多用される被写体と同一の水平面内に限定して考えるならば、計測情報もまた同一水平面内における水平視差に限定することができる。この観点から開発されたのが図10のTWISTERと称されるシステムである⁹⁾。この装置は使用者を取り囲むブースを形成し、使用者への視覚提示と使用者自身の実映像の撮影を全周囲から継続的に行うためのシステムである。図11のように撮像視点と表示点を物理的に移動させることにより、離散的な空間分解能を連続的な時間分解能によって補完することができるため、現行のデバイス技術による双方向の水平面内全方向VR環境を実現することに成功している。

7. 第6の間～テクスチャーで済ませる

一方で、CGの質感を向上させるために実画像を部分的に利用するCG技術として一般的に利用されている技術にテクスチャーマッピングがある。これは3DCGの基本とな

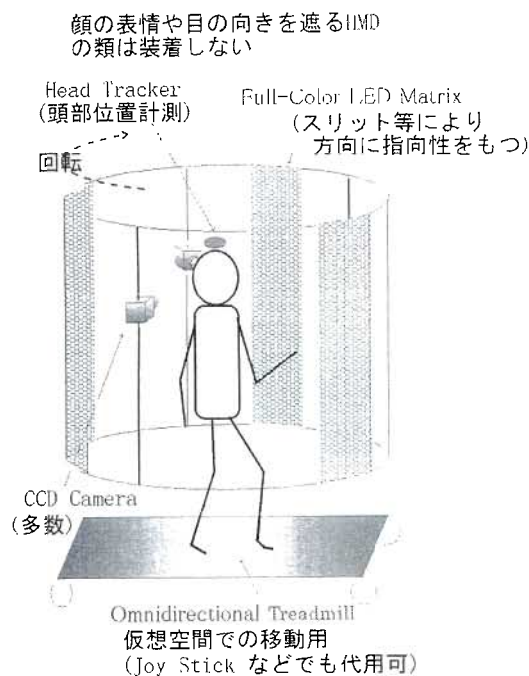


図10 TWISTERシステムの構成概念図。

るポリゴンの表面にテクスチャーパターンとして2次元映像を貼りつけるという手法で、通常のポリゴンCGとほぼ同程度の処理負荷でフォトリアリスティックな立体感のある映像が得られるのが特徴であるが、これが有効なのは単眼視提示の場合である。両眼視との併用を考えた場合、テクスチャーによるポリゴン面上での凹凸表現と両眼視差情報との食い違いが生じ、「テクスチャー写真を貼りつけただけの荒いポリゴン」という提示手法の実体のほうが認識されてしまうために本来のバーチャルな意味をなさない。これに対する従来の解決法は、ポリゴン数を増やし曲面近似をポリゴンの段階でより高精細にすることで対応しようとするものであった。しかし、この方法では本来テクスチャーマッピングによって減らそうとしていた処理負荷である描画ポリゴン数を増大させてしまうことになり、描画



図11 TWISTERシステム試作1号機作動情景。

自体のリアルタイム性が大きく損なわれてしまう。そこで、テクスチャー内においても視差を設定することにより、描画ポリゴン数を増やすことなく処理負荷の増加を最小限に抑えたままで、滑らかな3次元構造を提示することを狙ったのが図12のテクスチャー座標シフトである¹⁰⁾。

この手法は現行の実際のグラフィックエンジン上において従来の平面的なテクスチャーマッピングの手法に対して同程度の速度で実装することができ、ポリゴン描画を利用せずに同様の効果を狙う従来法であるイメージベースドレンダリングに比して10倍以上の速度を実現することに成功している。

今回紹介したVR技術は視覚的な提示技術に焦点を当てたものであったが、VR世界のリアルタイム性を保ちたいと思う場合、実際には空間記述の問題に突き当たる。もしもすべての物体を剛体近似して記述する数値化された物理空間を仮定するなら、剛体間の相互作用計算が剛体数の増加に伴ってNPコンプリートで増大するという演算量上の問題が知られている¹¹⁾。本格的にリアルなバーチャル世界に遊ぶためには、計算方法自体をバーチャルに実現す

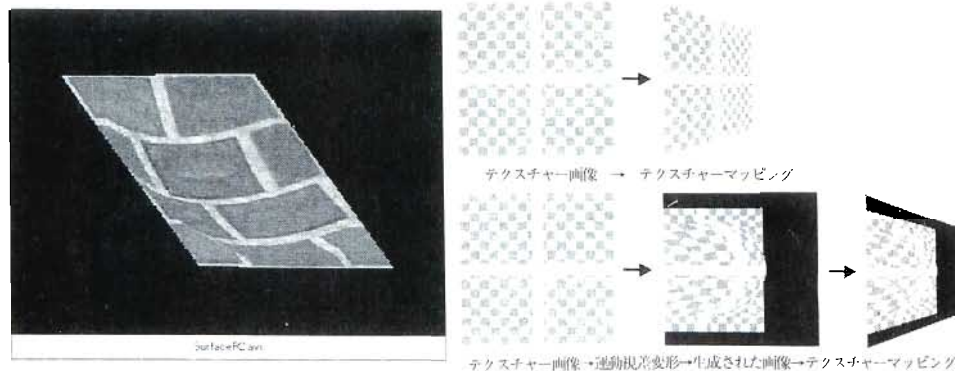


図12 テクスチャー座標シフト：ポリゴン表面のテクスチャー内に視差を設定することで、運動視差、両眼視差に関してテクスチャー表面の立体手がかりを正しく提示する。

る方法に挑まなければならないだろう。

文 献

- 1) 前田太郎, 荒井裕彦, 舘 暉: “頭部運動追従型両眼視覚提示装置の設計と評価”, 日本ロボット学会誌, 10 (1992) 99-109.
- 2) S. Tachi and K. Yasuda: “Evaluation experiments of a teleexistence manipulation system,” *Presence*, 3 (1994) 35-44.
- 3) S. Tachi, H. Arai and T. Maeda: “Tele-existence visual display for remote manipulation with a realtime sensation of presence,” *Proceedings of the 20th International Symposium on Industrial Robots*, Tokyo (1989) pp. 427-434.
- 4) C. Cruz-Neira, D. J. Sandin, T. A. DeFanti, R. V. Kenyon and J. C. Hart: “The CAVE: Audio visual experience automatic virtual environment,” *Commun. ACM*, 35 (1992) 65-72.
- 5) 柳田康幸, 前田太郎, 舘 暉: “固定スクリーン型視覚提示装置を用いたトレイグスタンス視覚系の構築手法”, 日本バーチャルリアリティ学会論文誌, 4(1999) 539-547.
- 6) 渡辺 毅: 視覚の科学 (写真工業出版, 1975).
- 7) M. Inami, N. Kawakami, D. Sekiguchi, Y. Yanagida, T. Maeda, S. Tachi and K. Mabuchi: “Head-mounted projector for projection of virtual environments on ubiquitous object-oriented retroreflective screens in real environment,” *Conference Abstracts and Applications of Siggraph '99*, Los Angeles (1999) p. 245.
- 8) H. Saito, S. Baba, M. Kimura, S. Vedula and T. Kanade: “Appearance-based virtual view generation of temporally-varying events from multi-camera images in the 3D room,” *Proceedings of Second International Conference on 3D Digital Imaging and Modeling*, Ottawa (1999).
- 9) Y. Kunita, M. Inami, T. Maeda and S. Tachi: “Real-time rendering system of moving objects,” *Proceedings of 1999 IEEE Workshop on Multi-View Modeling and Analysis of Visual Scenes*, Colorado (1999) pp. 81-88.
- 10) 高幣俊之, 稲見昌彦, 柳田康幸, 前田太郎, 舘 暉: “両眼立体視のためのリアルタイム・テクスチャマッピング手法”, 日本バーチャルリアリティ学会第5回大会予稿集 (2000)
- 11) D. Baraff: “Coping with friction for non-penetrating rigid body simulation,” *SIGGRAPH '91*, Las Vegas (1991) *Comput. Graphics*, 25 (1991).

(2001年3月6日受理)