3次元講義空間の動的状況把握による映像生成法

亀田 能成 †　美濃 導彦 †
†京都大学 総合情報メディアセンター

Abstract: We work on constructing an information media environment where users can not only get 3D and 2D visual information in a scene but also browse supplement information related to objects in the scene. In the environment, the system acquires situation information in the scene so that it can generate video image that are suitable to view and understand the scene.

1　はじめに

ある1つの空間に人が集まり，さまざまなコミュニケーション活動をする状況は，現代社会において一般的である．例えば，集会や会議，スポーツ，音楽会などの娯楽イベント，学校での講義室，会社のオフィスなどがそうした空間の例である．

このような空間ではさまざまな活動が行われるが，計算機の高性能化，通信の高速化のおかげで，現実にこの空間に存在している人だけでなく，遠隔地にいる人々が何らかの形でこの空間を観察したり，そこで行われている活動に参加したりすることが技術的に可能になりつつある．このような環境を我々は情報メディア環境と呼んでいる．情報メディア環境では，利用者は対象空間の動的状況を把握できるだけでなく，その空間内の物体を動的状況に合わせて観察したり，関連情報を得たりすることができる．すなわち，対象空間をその状況に合わせて映像化できるようになる．

遠隔地にいる人が情報メディア環境にアクセスする場合，映像情報が重要な意味をもつと考えられる．映像情報にも2種類あり，現実世界をそのまま撮影して得られる映像の他に，現実世界の3次元形状を取得し仮想的な3次元世界を構築した後，仮想カメラを介して得られる映像が考えられる．

情報メディア環境で動的状況に応じてこうした映像を提示するためには，対象空間の動的状況を把握することと，動的状況に対して適切な映像生成法を確立することが必要である．特に映像生成を動的状況に合わせて行うため，処理は実時間で行わなければならない．つまり，本研究では

1. 対象空間の映像ないし3次元形状を計算機内に取り込む
2. 対象空間内にどのような物体が存在し，どのような状態であるかを認識し，動的状況を把握する
3. 動的状況に合わせて利用者に映像を提示する（必要であればネットワーク上の関連情報ともリンクして利用者に提示する）

ことを実時間処理で行うことが研究の目標となる．

1番目については，画像からの3次元空間再構築を目指してコンピュータビジョンの分野で研究が進められてきた．これらの技術を活用し，3次元世界の形状を動的に取り込むための試みがなされつつある[1, 2]．しかし，これらのシステムでは，処理を実時間で実現するための手法が提示されていない．我々はこれを分散協調視覚の枠組みを用いて解決する．すなわち，コンピュータネットワーク上に分散配置されたカメラ付き観測ステーションと計算機を利用することで，実時間での3次元形状取得を実現する．

2番目及び3番目によって，対象空間がその動的状況に合わせて映像化されるようになる．対象空間の状況に合わせた映像生成のための最も単純な方式は対象空間のどの部分に対して映像生成するかを利用者が自分で決定する方式である．この利用者依存方式は利用者が自分の興味に合わせた映像生成ができるのが利点であるが，対象空間内で何が生じているのかという動的状況を何らかの方法で常に把握して，目標とする部分空間にカメラを適切に向ける操作を行う必要があり，利用者に非常な負担を強いる．これに対して，計算機が対象空間の動的状況を把握していれば，それに合わせて自動的に映像生成が行える．具体的には，動的状況に合わせてカメラワークを変更したり，複数用意したカメラの中で撮影に利用するカメラを切り替えたり，物体に関連情報を付帯して可視化することなどが挙げられる．

本稿では，情報メディア環境の実現に対して，解決

Video Image Generation of 3D Lecture Space by Interpreting Dynamic Situation
Yoshinari Kameda†, Michihiko Minoh†
†Kyoto University
http://www.imel1.kuis.kyoto-u.ac.jp/

しなければならない基本的な問題を整理し，それらの問題が分散協調視覚の枠組みでうまく対処できることを示す．

なお，我々は現在，このような問題に設定に対して，講義空間を対象として，実験による実証を行っている．

2 動的状況把握に基づく映像生成

2.1 動的状況

対象空間の映像生成を行うためには，システムが対象空間内の活動がどのような状態にあるかを認識し，利用者の希望と合わせた形でどの部分空間をいつ見せるかを決定しなくてはならない．この処理を映像生成の制御と呼ぶ．

映像生成の制御が行えるように，情報メディア環境では動的状況の記述を蓄積していく．つまり，映像生成の制御は動的状況の記述に基づいて行われる．動的状況の記述とは，時系列順に物体の識別とその属性記述（物体の状態を表現する特徴量）を組にして表現したものである．

2.2 映像生成

本節では，情報メディア環境における映像生成をするための各要素技術について説明する．図 1に映像生成の全体像を示す．

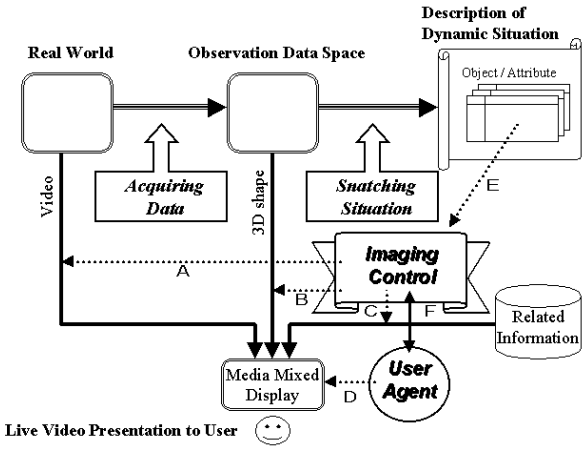

図 1: 動的状況の把握に基づく映像化

動的状況を把握するための前段階として，計算機内部に対象空間の観測データを取り込む（Acquiring Data 部）．動的状況の把握は，このステップによって取り込まれた観測データ空間（Observation Data Space）の観測データに基づいて行われる．また，取り込んだ観測データをそのまま利用者に提示することも可能である．

次に，観測データに基づいて，動的状況の把握を行う（Snatching Situation 部）．これは，動的状況の記述を時系列に沿って蓄えていくことである．対象空間が固定されていることから，出現しうる物体が限定されるという制約を導入し，物体の識別や属性記述の抽出をモデルマッチングで行う．また，静物体については，時間軸に沿ってその出現や属性が変化しないので，予め記述を蓄積しておき，実時間での処理からは除外する．

こうして動的状況の把握ができると，得られた動的状況の記述に基づいて映像生成の制御を行う（Imaging Control 部）．図 1の矢印 Eに沿って動的状況の記述を参照し，各経路 A, B, Cのように各データ型に対してメディアの取得・映像化方法を指示する．その内容は以下のようなものである．

- A：対象空間を直接光学的に映像化する際の制御．複数の観測ステーションからどの観測ステーションを選択するのか，観測ステーション上の首振りカメラにおける首振り角度やズーム値をどのように設定するべきかといった情報が制御情報となる．
- B：観測データを利用して映像化する際の制御．我々は現在，観測データとしてボクセル表現に基づく 3 次元形状を利用している．この場合，ボクセルに基づいた仮想空間が構成できるので，その中での形状の精度やボクセルの粒度，仮想カメラの視点位置などが制御情報となる．
- C：関連情報を付帯させて映像化する際の制御．物体の識別結果やその状態に合わせて，システム内に予め蓄えておいたどの関連情報を提示するかといった指示が制御情報となる．

このように，映像生成の制御を実映像・3 次元形状（ボクセル）関連情報の各データ型で行ったあと，最後にこれらを融合させて映像として利用者に提示することで，映像生成が実現されることになる．これらの融合と最終的な映像の提示（図 1の矢印 D）は，利用者の希望を具現化する利用者エージェント（User Agent 部）が行う．利用者エージェントの中にある利用者のモデルは，Fに示す経路を伝って映像生成の制御に影響を及ぼす．

図 1に示すように，映像生成において利用者が視ることになるデータ型は，実映像か 3 次元形状ないし関連情報であり，主に前 2 者である．我々はこの両方について取り組んでいるが，続く 3節では，3 次元形状

のデータ型に基づく対象空間の映像生成方法について説明し，4節では，実映像のデータ型に基づく対象空間の映像生成方法について述べる．

3　3次元形状を基本とする映像生成

3次元形状のデータ型に基づく映像生成では，観測データを3次元形状の形式で得るので，その観測データを仮想空間上で表現することは容易である．この結果，この方法の利点として，視点が仮想空間内の仮想カメラなので任意に設定できることや，物体の表面の曲率分布など立体的な対象の特徴を観測データから得やすいことが挙げられる．他方で，問題となるのは実時間で3次元形状を取り込むことが一般的に困難であることである．我々はこれに対して，分散環境下の計算機群のもとで，分散環境に適した3次元形状再構成手法を提案し，この問題の解決を目指している．

3.1　処理の概要

カメラをセンサーとする場合，観測する対象空間が3次元世界であるので，観測に基づいて3次元形状を取得しても，その形状は空間において開空間をなしてしまう．ところが，現実の物体は閉空間で構成されている．そこで，カメラによる観測によって，得られる3次元形状が全体で閉空間となるように，カメラを搭載した観測ステーションを複数設置する（図2）．

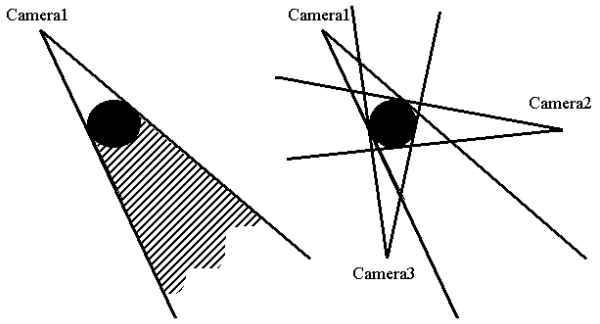

図 2: 開空間と閉空間

観測ステーションが複数台存在すること，映像データはデータ量が多いこと，実時間で処理しなければならないこと，などを考慮すれば，

- 複数台の観測ステーションから得られるデータをそれぞれ処理するエージェントを考え，
- 複数のエージェントが協調して3次元世界を構成する枠組を考える

ことが必要となり，分散協調視覚の考え方が適する．全体の処理の流れを図3に示す．3次元形状を基本とする映像生成では，各処理は以下のようになる．

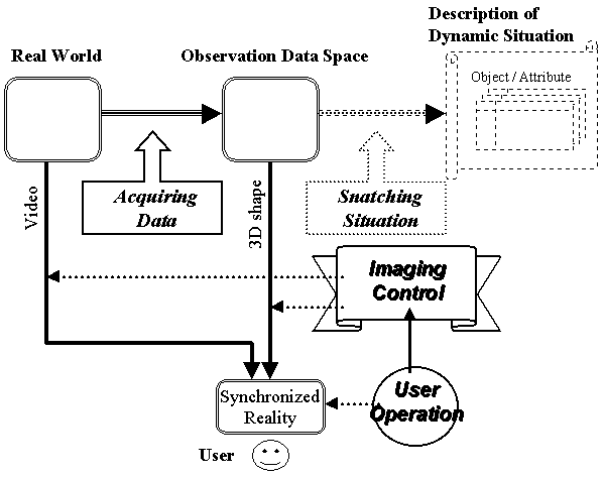

図 3: 3次元形状に基づく映像生成

観測データ空間の生成 ：現実空間の実映像を複数の観測ステーションにより撮影し，画像処理により3次元形状を計算し，結果を観測データとして取り込む．

動的状況の把握 ：計算機内部の3次元形状モデルと観測データをマッチングし動的状況の把握を行う．

映像生成の制御と利用者エージェント ：観測データと実映像とを融合して仮想的な3次元空間を合成する．この仮想空間を現実同期空間と呼び，その中では実時間と同じ時間の流れが存在する．利用者エージェントを介して利用者は希望する仮想視点を入力し（User Operation），利用者エージェントは仮想視点情報を映像生成制御に渡すとともに，仮想視点が観測ステーションのカメラ位置と一致したときには実映像を現実同期空間内に重畳させて利用者に提示する．

それぞれの処理で技術的に解決しなければならない課題は次のようなものである．

A. 複数の観測ステーションの分散協調による観測データの生成

複数の観測ステーションが設置されている状況では，それぞれの観測ステーションから得られる映像そのものを一個所に集めて処理する方法は，データ量，ネットワークのトラフィックを考えると現在の計算機の処

理能力，ネットワークのデータ転送速度からみて現実的ではない．特に，画像処理における 3 次元形状の計算は計算量が多い．本研究では実時間処理を行わなくてはならないという制約があることから，個々の観測ステーションで得られる映像情報毎にある程度独立して情報を抽出するエージェントを実現し，個々のエージェントが抽出した情報をもとにデータ通信を行って処理を進める分散協調の枠組みが適すると考えられる．

エージェント間でデータ通信を行う際，データ粒度を制御できるように，対象空間の 3 次元表現形式としてボクセルを使用する．処理の高速化のために，再構成の処理をどのようにしてエージェントに分割するのか，エージェント間の同期のメカニズムをどのように実現するのか，が重要な課題となる．

これについて，我々は，観測データ空間の実時間構築を分散協調視覚の環境下で実現する方法を提案している [4]．3.2節及び 3.3節でその内容について述べる．

B. 動的状況の把握

ある対象空間について考えるとき，そこに現れる物体は種類が限られると考えられるので，計算機内部にあらかじめ対応するの 3 次元形状モデルを用意し，このモデルと観測データのマッチングをとる．このとき，観測データがボクセルであるので，ボクセルベースの高速なモデルマッチング手法が求められる．

C. 現実同期空間と利用者に合わせた映像生成

利用者が，仮想視点を現実の観測ステーションのカメラと一致する場所に指定したとき，ボクセル表現の観測データと実映像とが同時に利用者に提示される．このとき，実映像と，観測データの取得に用いた映像とが同じ時刻に撮影されたものを利用しないと，実映像と観測データが映像上で一致しないという問題が発生する．

3.2 講義室を対象とする映像生成

我々は，大学の講義室を対象空間とし，主たる対象物体として人間を想定して研究を行っている．講義室の 3 次元モデルを予め作成し，複数台の観測ステーションを，全体として講義室内が完全に観測できるような位置に設置して，図 4に示すような試作システムを構築している．

計算機システムは空間の 3 次元形状を動的に抽出し，ボクセル表現の観測データを取得する．その結果を現実同期空間内に可視化して利用者に提示する．

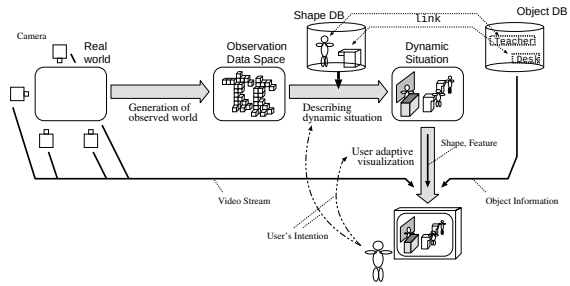

図 4: 講義室に対する試作システム

3.3 時空間の分割とビデオ画像のパイプライン処理による高速三次元再構成

対象空間内に分散配置されたカメラを用いた 3 次元形状復元において，最大の問題点は，データおよび処理量が膨大なため実時間処理を実現するのが困難な点である．我々は，この問題を分散処理により解決し，数台のビデオカメラと分散環境にある複数の計算機を用いて，対象空間の 3 次元形状を再構成する [4]．この方法では，計算機を複数用意することで，

1. ビデオ画像処理の過程を分割，各処理過程のパイプライン化によるスループットの向上
2. 観察中の対象空間を時空間と捉えて分割し，分散環境にある計算機群で分割した時空間の再構成をそれぞれ独立に行うことによるレイテンシの削減

を実現し，利用者に実時間で対象空間の 3 次元形状を可視化する．図 5に可視化の様子を示す．右図中の白点が一つ一つのボクセルを表している．

映像　　現実同期空間

図 5: 3 次元形状の再構成

分散処理により対象空間の 3 次元形状を観測空間上に効率よく再構成するためには，再構成のアルゴリズムが分散処理に適したものである必要がある．つまり，プログラムとデータの分割により各プロセスの処理時間を等しくすることができるアルゴリズムで，しかもプロセス間の通信の少ないものが必要である．すなわち，空間の再構成に使用されるデータは，時間と空間

に対して強い局所性を持っている必要がある．空間に対する強い局所性に基づいて，プロセス間の通信を抑えたまま各プロセスが処理する空間を分割する．これにより，各プロセスの処理量を少なくすることができる．また，時間に対する強い局所性に基づいて，各プロセスの処理量を時間に応じて変更する．

このような方法として，視体積交差法を利用し，これを分散処理に適応させる．視体積交差法では，同一時刻に撮影された複数枚の画像を利用して，物体の形状を決定する．ボクセルで空間を表現すると，視体積交差法では各ボクセルに対応する空間が物体内かどうかの判定が，別の時刻や別の位置にあるボクセルの値によらず決定される．この局所性を利用して，観察中の時空間を分割する．

視体積交差法で動物体の位置や形状を求めるためには，各ビデオ画像上で動領域を求めなくてはならない．よって，カメラでの撮影から空間の再構成までに次の 3 ステージの処理が必要となる．

1. ビデオ画像の取り込み
2. ビデオ画像上の動領域の抽出
3. 視体積交差法による空間の再構成

本研究で提案する手法では，この 3 ステージの処理をビデオサーバ ·エクストラクタ ·3D コンポーザの 3 種のプロセスが行う．

1 台のカメラで撮影される画像を一つのビデオサーバで計算機に取り込み，一つのエクストラクタで処理すると，ビデオサーバとエクストラクタはカメラの数だけ必要である．3D コンポーザは観察を行う時空間の分割数だけ必要である．これらのプロセスはそれぞれ異なる計算機上で実行され，必要なデータはプロセス間通信によって得る．これら多数のビデオサーバ ·エクストラクタ ·3D コンポーザでパイプラインを構成することで，スループットとレイテンシを向上させる．

本方法では，複数のカメラで映像を取得するタイミングや，分割空間の統合のタイミングなど，プロセス間の同期制御が重要である．そこで，プロセス間の同期をとるプロセスを用意し，これをスケジューラと呼ぶ．スケジューラは，このほかにシステム内外の各プロセス間で行われる通信や各プロセスで行われる処理の進行も管理する．これにより，システム全体の同期制御と処理のタイミングに合わせたデータ転送を実現する．

また，我々は対象空間が固定であるとの前提をおいているので，対象空間内に存在する静物体については，モデルをあらかじめ計算機に持たせておくことができる．静物体が占有する空間には動物体が存在することができないので，その部分空間では 3 次元形状を求める必要がない．このことを利用し，3 次元再構成にかかる計算量を削減する．

本提案方法により，我々は試作システム上で 4 台のカメラを使用して，レイテンシ 0.56 秒で毎秒 7.2 フレームの観測データを生成することに成功した．この観測データは 1 フレーム当り約 70 万ボクセルであり，1 ボクセルは 5cm 立方である．

4 実映像を基本とする映像生成

情報メディア環境の構築において，最終的に提示するのが映像であることからすると，基本となるメディアを映像とするのは妥当である．しかしながら，一般に対象空間に自由にカメラを設置することは不可能であり，何らかの制約が付かざるを得ない．

これに対して，どの時刻においてどの物体の映像生成を行うべきかを動的状況の把握結果に基づいて自動的に決定することで，制約のあるカメラ配置条件下でもその時々に応じた重要な対象の映像生成を行う方法 [3] を研究している．

時々に応じた重要な対象が何かを決定する基準として，対象空間のコンテキストという概念を導入する．つまり，映像生成の制御を行うプロセスは，動的状況の記述から現在の状態がどのコンテキストに相当するかを判断し，コンテキストから想定される最も重要な対象の映像生成が行われるように制御を実行する (図 6)．

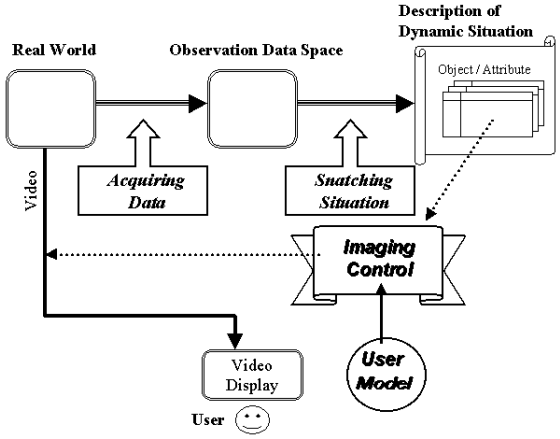

図 6: 実映像に基づく映像生成

4.1 カメラ制御と映像切り替えによる講義空間の映像生成

講義室を対象空間として捉え，講義室以外にいる利用者に対して映像を提供し遠隔地からの受講を可能にすることを考える．

講義を映像化する際に，撮影対象ごとに撮影エージェントを配置し，それらに講義のコンテキストに応じて動的にカメラを割り付けると共に，対象ごとに撮影されたビデオから１つのビデオを選択する．こうすることで，講義のコンテキストに応じて最も重要な対象を，最も良い条件のカメラで撮影することが出来るようになる．

講師 ・板書 ・生徒の３種類の対象を即座に利用者に提示できるように，各対象に撮影エージェントを設け，各対象を常に撮影する．撮影エージェントは観測ステーション上の首振りカメラを動的に制御し，それらの映像が対象をフレーム内に確実に捉えているようにする．

複数の撮影エージェントに対して首振りカメラ（観測ステーション）は有限個であるので，各撮影エージェントへのカメラの割り付けが問題となる．そこで，講義のコンテキストに応じて各撮影エージェントにカメラの割り付けを行う．

講義のコンテキストは，講師 ・生徒の状態から判断する．コンテキストに基づいて，講師 ・板書 ・生徒の講義に対する重要度を順位づけする．コンテキストの切り替えは講師 ・生徒の状態を知ることで抽出することができるので，講師 ・生徒の状態を観測し動的状況の記述を行う観測エージェントを設ける．

この結果，３種類の対象の映像が取得される．一般に利用者が同時に３つの映像を視ることは困難であるので，自動的に１つの映像を選択し利用者に提示することを考える．このとき，講義のコンテキストのみに応じて選択を行うと，同一コンテキストでは必ず同じ対象の映像提示となってしまい，利用者の興味を損なう恐れがある．そこで，利用者の興味モデルを反映させることで，利用者の興味の持てる映像提供を行う．

試作システムでは講義室に８台のカメラを設置し，４台を観測エージェントが使用し，残り４台で対象の撮影を行っている．利用者に対する映像提示の様子を図 7に示す．

5 おわりに

本稿では，１つの固定した空間を対象として，観測空間を生成し動的状況を把握することで，情報メディア環境における映像生成がどのように行えるかについて示し，それが分散協調視覚に基づいて実現できることを述べた．

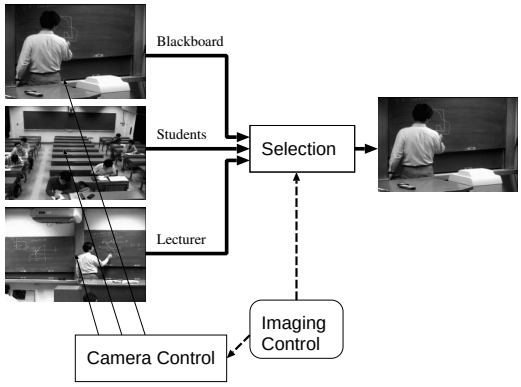

図 7: 講義の映像提示

我々の研究は現在，３次元形状を基本とする映像生成と実映像を基本とする映像生成について平行して研究している．今後は，この２つの研究を融合させ，さらにネットワーク上の関連情報と合わせた現実同期空間の生成や，動的状況に基づく仮想カメラのナビゲーション方法について取り組んでいく予定である．

謝辞 本研究の一部は科学研究費基盤研究 B-09558034, 学術振興会未来開拓プロジェクト 「分散協調視覚」 (JSPS-RFTF 96P00501) の支援を受けている．ここに感謝の意を記す．

参考文献

[1] Takeo Kanade and Peter Rander, “Virtualized Reality: Constructing Virtual Worlds from Real Scenes,” IEEE MultiMedia, vol.4, no.1, pp.34-47, 1997.

[2] Arun Katkere, Saied Moezzi, Don Y. Kuramura, Patrick Kelly, and Ramesh Jain, “Toward Video-based Immersive Environments,” Multimedia Systems, vol.5, No.2, pp.69-85, 1997.

[3] 宮崎英明, 亀田能成, 美濃導彦, “講義室における複数カメラからのビデオ送出法”, 情処 MIRU’98 講演論文集, July,1998 (掲載予定).

[4] 亀田能成, 太尾田 健男, 角所 考, 美濃導彦, “高速三次元再構成のための時空間の分割とビデオ画像のパイプライン処理”, 情処研資 DPS 89 回研究会, June,1998 (掲載予定).