特別研究報告書

e-learningシステムにおける
学習者の顔情報表示のための画像合成

指導教員　　美濃 導彦 教授

京都大学工学部情報学科

中村 和晃

平成17年2月10日

e-learning システムにおける
学習者の顔情報表示のための画像合成

中村 和晃

内容梗概

現在の e-learning システムを単純な Web ベース学習と比較したときの最大の特徴は，学習ログを残すことができるという点である．学習ログには，各学習者の教材へのアクセス時間や設問の解答状況が保存されており，教師は各学習者の学習進捗状況を把握することができるようになっている．しかし，この学習ログで残されるものは学習進捗状況のみであり，学習の様子，すなわち表情や視線の情報は残されない．現実の講義においては，教師は学習者の表情や視線を見て，どのように講義を進行するかを定めており，これらは教師にとって重要な情報であるといえる．表情や視線が教師にとって重要である点は，e-learning システムにおいても同様であると考えられる．そこで本研究では，e-learning システムで学習している学習者の表情・視線の情報を学習ログとして管理し，教師の要請に応じてそれらを提示するシステムを提案する．

学習者の学習時の表情は，大きくは変化しないと考えられるため，そのような表情を機械が認識することは困難である．また，学習者の視線，すなわち教材に対する注視位置を追跡するためにはアイマークカメラ等の特殊な機器が必要であるが，このような機器を装着することは，e-learning システムで学習中の学習者にとって負担が大きい．そこで本研究では，学習者の表情や注視位置が教師自身に判別できるような画像を提示することを提案する．特に注視位置に関しては，注視している部分の教材内容も判別できるように，教材の画像も提示する．そして，これらの画像を用いて，学習者の表情と視線の情報を学習ログとして利用することのできるシステム，いわば “顔の見える e-learning システム” の構築を目指す．

表情が判別できる画像としては，学習者の顔を撮影した実写の画像をそのまま利用する．一方，注視位置が判別できる画像としては，学習者の顔とモニタ画面との位置関係が分かるような画像，すなわち，両者が共に写されている画像を用いればよい．しかし，このような画像を撮影する際，視点をモニタ画面の両脇や斜め前方に置くと，注視部分の教材内容がひずんで見えるため，視点

はモニタ画面に正対する位置に置く必要がある．また，学習者の顔が横方向や斜め方向から観測される位置に視点を置くと，画像上で学習者の視線方向が一方向に偏り，教師がその変化を感じにくくなるため，視点は学習者の顔にも正対させるべきである．以上の条件を満たすものとして，モニタの裏側からモニタ画面を透かして学習者の顔を観測したような画像を用いる．

このような画像は当然実写では得られないので，本研究では，モニタ枠の外側に設置されたステレオカメラ画像から仮想視点顔画像を生成し，その顔画像と教材の画像を重畳することで，モニタの裏側からモニタ画面を透かして学習者の顔を観測したような画像を合成する．具体的には，まず，学習者の顔の中心とモニタ画面の中心を結んだ直線上に仮想視点を置く．さらに，教材の提示されているモニタ画面を底面，学習者の顔中心を頂点とする錐体に従ってモニタ画面を一旦相似縮小し，このモニタ画面を通して見た顔画像を生成することで，学習者の視線方向とモニタ上の注視位置の関係を保存したまま，顔画像に対する教材画像の大きさを補正する．

この視点からの学習者の正面顔画像は，次のように合成する．あらかじめ学習者の顔を多面体で近似した 3 次元モデルを用意しておき，その上でステレオカメラで撮影した 2 枚の学習者の顔画像から，両目と口の中心位置を抽出し，その 3 次元位置を求める．これらの 3 次元位置に合わせて，用意したモデルの姿勢を定め，これを入力画像に逆投影し，モデルの各面に入力画像の対応する部分をテクスチャとして貼り付ける．これを上記の視点位置から観測し，画像を生成する．こうして生成された正面顔画像上で，相似縮小したモニタ画面が投影される領域に，教材の画像を，視線方向と注視位置の関係性が保存されるよう，左右反転させて重畳する．最後に，教材画像が左右反対にならないようにするため，もう一度画像全体を左右反転させる．

提案した合成画像から，学習者の注視位置がどの程度判別できるのかを調べるために，e-learning システムに WebCT を用いて被験者の顔画像を実際に撮影し，教材のページごとにその顔画像を並べた画面と，提案手法を用いて合成した画像を表示した画面とを作成した．そして，合成画像に写されている被撮影者が教材のどのオブジェクト（図・文章・数式）を見ていると思うかについて，5 人の被験者に回答を求めた．その結果，一つの図や一段落の文章を単一のオブジェクトと見た場合には，75%から 80%の正解率を得た．今後は，教材・顔間の距離推定に関するファクタを検討する必要があろう．

Image Synthsis for Presenting Facial Information of Learners in e-Learning System

Kazuaki NAKAMURA

Abstract

One of the major advantages of recent e-learning systems compared with the conventional web based training is that learning log files of learners are stored in the system. From the learning log files, teachers are able to know when each learner accessed to course materials and what answer each learner replied to quizzes. However, they are not able to know how each learner learned materials, in other words, which object in the material s/he focused on, with which facial expressions. At a real lecture, the teacher give a lecture on some subject by considering facial expressions and focus attentions of learners because that information is useful for estimating the degree of understanding and interest of the learners. This is also the case in learning by e-learning system. In this paper, we propose the system that stores the data of facial expression and focus attention of each learner and presents these data to the teachers.

It is not easy to recognize facial expressions appearing on the face of learners because those facial expressions do not have distinct difference with each other. For detecting focus attention of learners, particular devices such as eye-mark cameras are often used. Putting those devices during e-learning is inconvenient and frustrating for the learners. In this paper, we propose to present the facial images of the learners with which the teachers are able to understand their facial expressions and focus attentions. For presenting the focus attention by facial images, image of materials is also presented so that the teachers understand which object in the materials the learners focus on. We use these images as the learning log in which facial expressions and focus attentions of each learner are recorded.

For the images to present facial expressions of learners to the teachers, we use real facial images captured by a camera. For the images to present focus attentions, we use images in which the face of learners appears with the monitor so that their positional relation can be easily recognized. If we create these images when the viewpoint does not exist in front of the monitor, the content

of materials looks distorted. And, if the viewpoint does not exist in front of the face of learners, the lines of sight of the learner cannot be easily recognized from the images. Hence, we need to set the viewpoint behind the monitor so that the viewpoint must exist in front of the face of learners and the monitor, and observe the face of learners through the monitor.

Since these images cannot be obtained by the camera installed around the monitor, we create synthetic facial image with a virtual viewpoint from images captured by stereo-camera installed around the monitor, and overlay the image of the course material on the facial images. In this process, we first set the virtual viewpoint on the line passing through the center of the learner's face and the monitor. In order to make the facial image sufficiently large compared with the image of material by preserving the learner's focus on the material, the monitor scaled down along the pyramid with the apex at the center of the face and the bottom at the monitor by keeping the distance between the virtual viewpoint and the scaled monitor to the focal length of the camera.

The frontal face of the learner is recovered by adjusting the 3Dmodel that approximates the learner's face to the 3D positions of facial features including the centers of the eyes and the mouth extracted from images of the stereo-camera, and mapping each part of the stereo-camera images on the corresponding part of the model for the texture. By observing the model from the virtual viewpoint above, the frontal facial image is obtained. The image of the course material is reversed and overlaid on the frontal facial image so that the relation between the learner's focus and the course materials is preserved. Finally, the overlaid image is reversed again to restore the proper direction of the image of the material.

Using *WebCT* as an example of e-learning system, we created the facial images in order to verify how properly the teachers can recognize focus attentions of the learners by observing those images. On the monitor, there is a window that displays some real facial images of learners. And, a window that displays an overlaid image of the learner appears by clicking the real facial image. These windows are presented to the test subjects who were asked which object in the window the learner looked. As the result, correct answer was replied with 75% to 80% precision.

e-learning システムにおける
学習者の顔情報表示のための画像合成

目次

第1章　緒論

教育分野へのIT技術の導入に伴い，e-learningシステムが急速に普及している．このようなe-learningシステムは，当初，Webコンテンツとしての教材を，各学習者が個々に自身のパソコンを用いて閲覧し，学習を進めるというものであった．その利点は，学習が時間的・空間的に制約されない点にある．すなわち，学習者や教師が学習のために一堂に会する必要がなく，また，学習時刻・時間が同じである必要もないため，学習者は自身のスケジュールに合わせ，好きな時間に好きな場所で好きなだけ学習することができる．一方，教師の側には，指導方針の立案・改善などのために，各学習者の学習の様子や進捗状況を把握したいという要求があるが，これは，教師と学習者が対面している現実の講義に比べ容易ではなかった．しかし現在では，学習進捗状況把握の問題は改善され，教材への学習者のアクセスデータ，学習者の学習進捗状況といった情報をWebサーバで一括管理し，学習ログとして記録することができるシステムが開発されている．

このように学習ログを残せるという点は，現在のe-learningシステムの大きな特徴であるが，現在のe-learningシステムにおいても，学習者の学習の様子までを，学習ログとして残すには至っていない．現実の講義に目を向けてみると，講義中の学習者の様子は，教師にとって講義の進行を決定する重要な要因である．学習者の表情や視線はその最たるものであり，教師は，学習者の表情，視線などから全体的な，あるいは個々の学習者の，理解の度合い，興味の度合いなどを判断している．その結果教師は，理解の度合いが不十分と思われるところを再度重点的に説明したり，興味の度合いの低そうなところでは話題を転換したりなどして，効果的に講義を進めていく．このように，教師が学習者の表情や視線を通じて学習者の様子に関する情報を獲得できることは，教師と学習者が対面している現実の講義の大きな利点といえる．

e-learningシステムでの学習においても，現実の講義のこのような利点をシステムに導入することで，学習をより効果的なものとすることができると考えられる．教師が学習者の顔を観察して，その表情や視線を把握できれば，それぞれの学習者の学習の様子を的確に判断できる．その結果，現実の講義と同様に，理解度が不十分と思われるところや興味の度合いの低そうなところの文章や図構成などを，後で再検討することができると考えられる．また，学習者の

解答時の自信の有無の様子を把握し，解答の評価の際の参考とすることもできると考えられる．あるいは，学習者がアクセスしている間たしかに教材を閲覧しているか，アクセスした学習者とは別の人間が解答を行っていないか，などをチェックすることにも役立つであろう．

e-learning システムによる学習にこのような利点を持たせるためには，学習者の学習時の顔情報を管理し，教師の要求に応じて各人の表情や視線の情報を表示するシステムが必要となる．しかし，視線の情報，すなわち教材に対する注視位置とその部分の内容に関する情報を機械的に獲得するには，特殊な機器が必要である．例えば，人間の注視位置を機械的に追跡する手法として，満上らは，アイマークカメラを用いて両眼の視線方向を計測し，それと両眼の輻輳角を用いて注視点を求める手法 [1] を用いている．しかし，e-learning システムを利用している学習者にとって，アイマークカメラ等の特殊な機器の装着は負担が大きい．故に注視位置を機械的に追跡することは困難である．また，学習者が一人で学習しているとき，その表情は大きくは変化しないと考えられるため，顔画像などから表情を機械的に認識することも困難であろう．そこで本研究では，学習者の学習時の顔画像をカメラで撮影して，それを直接提示することで，教師自身が学習者の表情を判別できるようにする．また，学習者の顔とモニタの位置関係，および教材の内容が把握できるような画像として，モニタ画面を透かして学習者の顔を観測したような画像を提示することで，教師自身が学習者の注視位置とその部分の教材内容を判別できるようにする．そして，これらの画像を学習ログとして利用することのできるシステム，言わば “顔の見える e-learning システム” の構築を目指す．

以下本論文では，**2** 章で，学習者の表情と注視位置を教師自身が判別できるような画像として，上に述べたような画像を合成する手法について述べる．**3** 章では，**2** 章で提案した手法による画像合成に必要となる正面視点からの顔画像を，実写の画像から合成するための正面顔復元手法について述べる．**4** 章で提案手法を用いて行った実験について述べ，その結果を検討する．最後に **5** 章で結論を述べる．

第２章　顔情報表示のための画像合成

2.1　合成の概要

本研究では，学習者の表情と視線の情報を，教師自身が判別できるような形で提示することを検討する．表情は，学習者の顔を観察することで判別できると考えられる．視線，すなわち学習者の教材に対する注視位置は，学習者の顔とモニタの両方を観察し，それらの位置関係を把握することで判別できると考えられる．従って，それらを観察できるような画像を提示すれば，教師には学習者の表情や注視位置が判別できると考えられる．ただし注視位置に関しては，同時にその位置における教材の内容も判別できなければならない．

学習者の顔を観察できるような画像としては，撮影された顔画像をそのまま用いればよい．一方，学習者の顔とモニタの両方を観察できるような画像としては，それらの位置関係が保存された状態で，両者がともに写されているような画像を提示すればよいが，位置関係が保存されたままで顔が写るように撮影すると，モニタは画面のある側ではなく裏側が写ってしまい，教材の内容が判別できない．そこで，モニタに関しては，教材の内容が表示されている画面部分のみを撮影対象とし，その画面を透かして学習者の顔を撮影したような画像を提示する．このような画像ならば，顔とモニタの位置関係が保存された状態で両者が写されており，かつ教材の内容も表示されているため，教師は，学習者の注視位置とその部分の教材内容をともに判別できる．しかしこのような画像は，当然実写では得られない．そこで本研究では，モニタの写されていない顔画像に対して，実際にはモニタ画面が存在するはずの領域に教材画像を重畳することで，モニタ画面を透かして学習者を観測したような画像を合成する手法を提案する．そして，その合成画像と顔画像を用いて，学習者の表情，教材に対する注視位置，およびその部分の教材内容が判別できるような学習ログ，具体的には，教材のページごとに，それを見ている学習者の顔画像が並べて表示され，さらにその顔画像をクリックすることで，モニタ画面を透かして学習者を観測したような合成画像が表示される学習ログを構成する．

モニタ画面を透かして学習者を観測したような画像を合成する際に，モニタを横方向や斜め方向から観測することになる位置に視点を置くと，教材がひずんで見え，内容が判別しにくいため，視点はモニタに正対する位置をとることが望ましい．また，学習者の顔を横方向や斜め方向から観測することになる位

置に視点を置くと，画像上で学習者の視線方向がモニタの写っている側に偏るため，その変化を感じにくくなる．従って，視点は学習者の顔にも正対させるべきである．さらに，画像上の顔の大きさが，重畳する教材画像の大きさに比べ大きすぎたり小さすぎたりすると，学習者の注視位置が判別しにくくなるため，視点はそのような写り方のしない位置に置かなければならない．

2.2 視点位置の設定

図 1 の視点 A のように，学習者の顔の中心とモニタ画面の中心を結んだ直線上に視点を置いた状況を想定する．学習中の学習者はモニタの方向を向いていると考えられるので，この状況では，視点は学習者の顔に正対していると考えられる．ここでさらに，モニタ画面が顔中心とモニタ画面中心を結んだ直線に直交しているならば，視点はモニタ画面にも正対していることになる．この場合，単純に考えれば，視点からモニタ画面中心までの距離を，学習者を観測するカメラの焦点距離 f に等しくすることで，モニタ画面を透かして顔を観測したときと同じように見える画像を撮影できると考えられる．このとき，モニタ画面の投影像の大きさは，モニタ画面の実際の大きさに等しくなる．

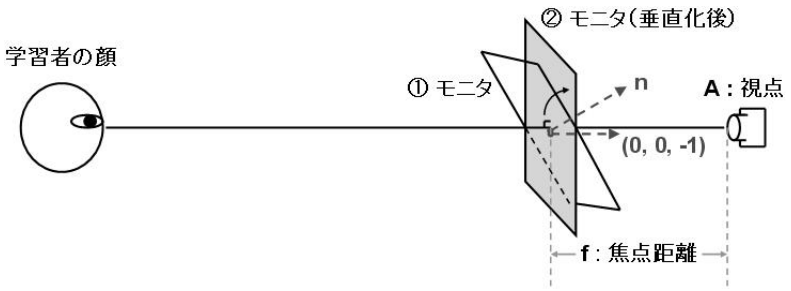


図 1: 注視位置判別のための画像の視点

しかし，一般には図 1 の① に示すように，モニタ画面は顔中心とモニタ画面中心を結ぶ直線に垂直ではない．そこでモニタ画面に回転を加え，図 1 の② に示すように垂直化する．図 1 の位置 A を原点，A からモニタ画面の中心へ向かう方向を z 軸と定める．z 軸とモニタ画面の走査線方向のベクトルにより張られる平面上にあり，かつ z 軸に垂直となる方向を x 軸と定める．y 軸は，z 軸と x 軸から，直交座標系の右手系をなすように定める．このような座標系において，モニタ画面の法線ベクトル $\boldsymbol{n} = (n_x\ n_y\ n_z)^T$ $(||\boldsymbol{n}|| = 1)$ と zx 平面がなす角

を ϕ，$\boldsymbol{n}$ の zx 平面への投影 $(n_x\ 0\ n_z)^T$ と z 軸のなす角を θ とすると，図 2 のようになる．

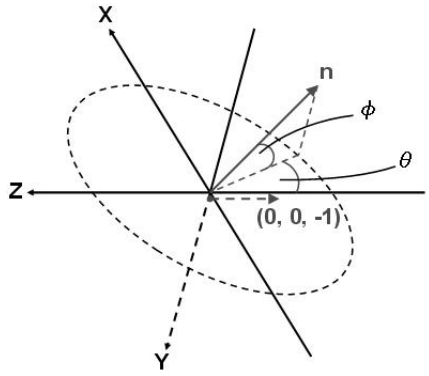


図 2: モニタの回転

このとき，次のように極座標表示される．

$$\begin{cases} n_x = \cos\phi \sin\theta \\ n_y = -\sin\phi \\ n_z = -\cos\phi \cos\theta \end{cases}$$

この座標系では，顔中心からモニタ画面中心へ向かう方向のベクトルは $(0\ 0\ -1)^T$ であるから，モニタ画面上の各点を，y 軸周りに θ 回転し，その後 x 軸周りに ϕ 回転させることで，モニタ画面を顔中心とモニタ画面中心を結ぶ直線に対して垂直化できる．垂直化により，モニタ画面は視点 A に対して正対する位置に来る．このとき，視点 A から学習者を観測する人間には，横方向や斜め方向からモニタを見た場合と比較して，画面に表示された教材の内容が最も判別しやすくなると考えられる．

視点 A から学習者を観測する人間は，学習者に関する情報として，目・口・鼻などの顔特徴の位置関係から顔の向きを，目中心（白目まで含めた目領域全体の中心）と瞳中心（黒目領域の中心）の位置の差異から顔に対する眼球の向きを，それぞれ推定できると考えられる．そしてこの両者から，学習者の視線方向を推定できると考えられる．また，画像上の顔の大きさと，顔の実際の大きさに関する前提知識から，学習者の顔とモニタ画面の間の距離を推定できると考えられる．視線方向の情報，および，学習者の顔とモニタ画面の間の距離から，人間には学習者の注視位置を判別できると考えられる．

従って，図 1 の A から垂直化したモニタを透かして学習者を観測した画像は，教材内容と学習者の注視位置の両方が判別可能なものとなる．しかし，この方法では，視点をモニタの後方で，かつモニタに近い位置に置くことになるため，学習者の顔画像がモニタ画面の像に対して相対的に小さくなりすぎ，注視位置が判別しにくくなる．この問題に対し，モニタ画面の像を固定した上で，単純に顔画像全体を拡大することで対応する方法が考えられる．しかしこの方法は，顔とモニタの大きさ，および，モニタ・視点間の距離を一定にしたまま，顔を視点に近づけることと等価であるため，図 3 のように注視位置の情報が保存されなくなる．

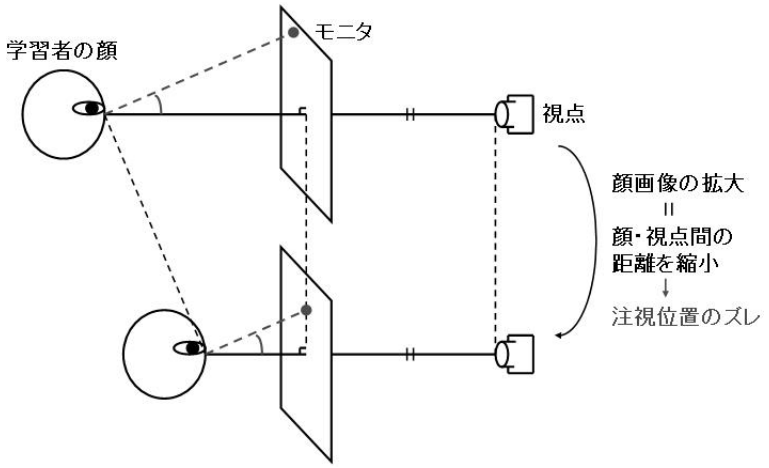


図 3: 画像拡大による注視位置のずれ

そこで，図 4 のように，モニタを本来の位置から図中の B の位置に相似縮小し，その上で視点を図中の C の位置に置く．このとき，縮小後における視点・顔間や視点・モニタ間の距離推定を縮小前と同条件で行えるようにするために，モニタ画面の画像上の縮小割合を実際の縮小割合に一致させなければならない．故に，視点 C は B から距離 f のところとなるよう定め，視点・モニタ間の距離を縮小前のそれに等しくする．これにより，モニタ画面の実際の大きさに対する投影像の大きさの割合は，相似縮小の割合 α に等しくなる．視点は顔に近づいているため，画像上の顔の大きさは図 1 の場合より大きくなる．以上のことから，顔画像のモニタ画面に対する相対的なサイズは大きくなる．また，この場合，顔中心の位置を錐体の頂点にして縮小しているため，顔画像を単純に拡大するときのような注視位置のずれは起こらない．従って，注視位置の情報を

保存したまま，学習者の顔画像のモニタ画面に対する相対的な大きさを一定以上に保つことが可能となる．

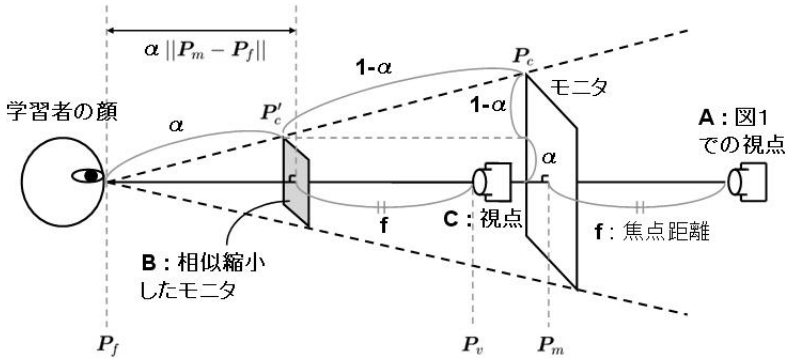


図 4: 注視位置判別のための画像の視点（改良）

2.3 顔画像と教材画像の重畳

図 4 の視点位置 C から学習者の顔を観測した画像を実写で得ることは，焦点距離 f が大きい場合にはカメラをモニタの裏側におく必要があるため不可能であり，f が小さい場合でも，カメラをモニタ画面中心の前方に設置すると学習者の学習の妨げとなるため不可能である．そこでこのような画像を正面顔復元手法により生成する．正面顔復元手法に関しては，**3** 章で詳述する．復元された正面顔画像上で，図 4 のようにして相似縮小されたモニタ画面の投影像が占める領域に，教材の画像を重畳する．重畳する際には，注視位置の情報を保存するために，モニタを裏側から観測している状況となることを考慮し，教材の画像を左右反転させる．このままでは教材の文字・数式・図表等が左右反転しており，表示内容が判別しにくいので，最後に重畳済み画像全体を左右反転させる．

画像の重畳は次のように行う．まず，モニタ画面の 4 隅の 3 次元ワールド座標をあらかじめ測定しておく．モニタ画面の中心の 3 次元ワールド座標は，4 隅の 3 次元ワールド座標の重心として与える．次に，ステレオカメラで撮影された入力画像から **3** 章で提案する手法を用いて正面顔復元を行う．その際，モニタ画面中心の 3 次元ワールド座標 $\boldsymbol{P}_m$ と顔中心の 3 次元ワールド座標 $\boldsymbol{P}_f$ を結ぶ直線上の点から，仮想カメラの視点の位置 $\boldsymbol{P}_v$ を

$$\boldsymbol{P}_v \;=\; \boldsymbol{P}_f \;+\; \left(\alpha + \frac{\mathrm{f}}{||\boldsymbol{P}_m - \boldsymbol{P}_f||}\right)(\boldsymbol{P}_m - \boldsymbol{P}_f) \tag{1}$$

により与える．また，モニタ画面を図 4 のように相似縮小したときの，モニタの左上隅の 3 次元ワールド座標 $\boldsymbol{P}_c'$ を

$$\boldsymbol{P}_c' \ = \ \boldsymbol{P}_f \ + \ \alpha(\boldsymbol{P}_c - \boldsymbol{P}_f) \tag{2}$$

により与える．$\boldsymbol{P}_c$ は縮小前のモニタ左上隅の 3 次元ワールド座標である．右上，左下，右下の隅についても，同様にして相似縮小後の 3 次元ワールド座標を与える．このとき，**2.2** 節で示した方法で，モニタ画面を，顔中心とモニタ画面中心とを結ぶ直線に対し垂直化しておく．縮小率 α には，顔画像がモニタ画面の像に対して十分な大きさで撮影されるように，適当な値を選ぶ．相似縮小後のモニタ画面を仮想視点画像に投影したときにできる長方形が，モニタ画面の投影像の占める領域となる．教材の画像を，この領域の大きさに合うよう縮小し，さらに注視位置が保存されるよう左右反転させて顔画像に重畳する．最後に，教材の表示内容の向きを元に戻すため，重畳後の画像全体を左右反転させる．

第 3 章　正面顔復元手法

3.1　復元に用いる枠組み

2 章で，教材の内容と，その教材に対する学習者の注視位置の 2 点が判別可能であるような画像を合成する手法を提案した．この手法では，学習者の顔画像を撮影するための視点を，学習者の顔中心とモニタ画面中心を結んだ直線上に置く必要があるが，モニタを見る妨げになるため，実際にはカメラをモニタ画面の中心付近に設置することはできない．そこで本章では，このような視点から撮影した仮想視点画像を，モニタの周囲に設置したステレオカメラによる実写の画像から合成するための手法について述べる．

複数のカメラの画像から仮想視点画像を生成する手法には，Transfer Based 手法 [2] と Model Based 手法 [3] がある．Transfer Based 手法は，入力画像同士の対応関係を弱校正によって画像のみから求め，これに基づいて仮想視点画像を生成する手法である．この手法は，誤差を生じやすいカメラの強校正が必要ではないため，劣化の少ない画像を生成できるが，仮想視点での対応点の位置を入力視点間の線形内挿により求めるために，仮想視点の位置が入力視点間に限られるという問題がある．これに対して，Model Based 手法は，強校正ステレオに基づいて対称物体の 3 次元形状を復元した上で，これを形状モデルとし，そのモデルに入力画像をテクスチャとして貼り付け，それを仮想視点から撮影

することで，仮想視点画像を生成する手法である．この方法には，仮想視点を自由に設定できるという利点があるが，他のカメラや物体までの距離が大きいとカメラの強校正の誤差が大きくなりやすいという欠点がある．本研究では，撮影対象である学習者からカメラまでの距離は小さいが，仮想視点を置くべき位置が入力視点間に存在するとは限らないので，Model Based 手法で正面顔復元を行う．

Model Based 手法を正面顔復元に適用する方法として最も単純なものは，顔の 3 次元形状を単一の平面で近似すること，すなわち顔を平面モデルで記述する手法である．しかし，本研究では顔からカメラまでの距離が小さいため，顔はどのように平面近似しても，単一平面から外れる部分が多い．故に，平面モデルによる手法では，仮想視点画像中のひずみが大きくなる．そこで，顔を多面体で近似した 3 次元モデルを用いて，正面顔復元を行う．

3.2 ステレオ視による三次元計測

3.2.1 投影方程式

本論文では，2 次元座標 $\boldsymbol{m} = (u\ v)^T$ や 3 次元座標 $\boldsymbol{M} = (X\ Y\ Z)^T$ で表される点の斉次座標を，$\tilde{\boldsymbol{m}} = (u\ v\ 1)^T$ および $\tilde{\boldsymbol{M}} = (X\ Y\ Z\ 1)^T$ で表すものとする．また，行数が同じである 2 つの行列 $\boldsymbol{F}$, $\boldsymbol{G}$ に関して，$\boldsymbol{F}$ の右に $\boldsymbol{G}$ を連接した行列を，$(\boldsymbol{F} \mid \boldsymbol{G})$ と表記する．

カメラの結像系において，ワールド座標 $\boldsymbol{M} = (X\ Y\ Z)^T$ で与えられる点が，画像座標 $\boldsymbol{m} = (u\ v)^T$ に結像するとき，次の投影方程式が成り立つ．

$$w\tilde{\boldsymbol{m}} \ = \ \boldsymbol{A}(\boldsymbol{R} \mid \ -\boldsymbol{R}\boldsymbol{T})\tilde{\boldsymbol{M}} \tag{3}$$

ここで $\boldsymbol{A}$ はカメラの内部パラメータ行列，$\boldsymbol{R}$, $\boldsymbol{T}$ はワールド座標系に対するカメラ座標系の回転行列と平行移動ベクトルである．また，w は 0 でない定数であり，斉次座標においては $(u\ v\ 1)^T$ と $(wu\ wv\ w)^T$ は同一の点を表す．

3.2.2 3 次元位置の算出

ステレオカメラ $\mathcal{C}_1$, $\mathcal{C}_2$ により撮影された画像をそれぞれ $\mathcal{I}_1$, $\mathcal{I}_2$ とし，$\mathcal{C}_1$, $\mathcal{C}_2$ の内部パラメータ行列を $\boldsymbol{A}_1$, $\boldsymbol{A}_2$，ワールド座標系に対する $\mathcal{C}_1$, $\mathcal{C}_2$ のカメラ座標系の回転行列および平行移動ベクトルを $\boldsymbol{R}_1$, $\boldsymbol{R}_2$, $\boldsymbol{T}_1$, $\boldsymbol{T}_2$ とする．これらの行列及びベクトルは，ステレオカメラの強校正により測定できる．ワールド座

標が $\boldsymbol{M}$ の点が，$\mathcal{I}_1,\ \mathcal{I}_2$ 上で $\boldsymbol{m}_1,\ \boldsymbol{m}_2$ に結像するとき，式 (3) から，

$$w_1\tilde{\boldsymbol{m}}_1 \;=\; \boldsymbol{A}_1(\boldsymbol{R}_1 \mid -\boldsymbol{R}_1\boldsymbol{T}_1)\tilde{\boldsymbol{M}} \tag{4}$$

$$w_2\tilde{\boldsymbol{m}}_2 \;=\; \boldsymbol{A}_2(\boldsymbol{R}_2 \mid -\boldsymbol{R}_2\boldsymbol{T}_2)\tilde{\boldsymbol{M}} \tag{5}$$

が満たされる．ただし $w_1,\ w_2$ は 0 でない定数である．

式 (4)(5) から $\tilde{\boldsymbol{M}}$ を消去すると，

$$w_1\boldsymbol{R}_1^{-1}\boldsymbol{A}_1^{-1}\tilde{\boldsymbol{m}}_1 \;-\; w_2\boldsymbol{R}_2^{-1}\boldsymbol{A}_2^{-1}\tilde{\boldsymbol{m}}_2 \;=\; \boldsymbol{T}_2 - \boldsymbol{T}_1 \tag{6}$$

となる．$\mathcal{C}_2$ のカメラ座標系に対する $\mathcal{C}_1$ のカメラ座標系の回転行列 $\boldsymbol{R}$ および平行移動ベクトル $\boldsymbol{t}$ は，

$$\begin{aligned}\boldsymbol{R} &= \boldsymbol{R_1}\boldsymbol{R_2}^{-1}\\ \boldsymbol{t} &= \boldsymbol{R_1}(\boldsymbol{T}_2 - \boldsymbol{T}_1)\end{aligned}$$

で表される．これを用いると式 (6) は

$$\begin{aligned} & w_1\boldsymbol{A}_1^{-1}\tilde{\boldsymbol{m}}_1 \;-\; w_2\boldsymbol{R}\boldsymbol{A}_2^{-1}\tilde{\boldsymbol{m}}_2 \;=\; \boldsymbol{t}\\ \Longleftrightarrow\;\; & (\boldsymbol{A}_1^{-1}\tilde{\boldsymbol{m}}_1 \mid -\boldsymbol{R}\boldsymbol{A}_2^{-1}\tilde{\boldsymbol{m}}_2)\begin{pmatrix} w_1 \\ w_2 \end{pmatrix} \;=\; \boldsymbol{t} \end{aligned} \tag{7}$$

と書ける．

ステレオカメラの画像から，対応する二点の座標 $\boldsymbol{m}_1,\ \boldsymbol{m}_2$ を得たとき，$\boldsymbol{A}_1,\ \boldsymbol{A}_2,\ \boldsymbol{R},\ \boldsymbol{t}$ は測定済であるから，式 (7) は未知変数 2 つに対して拘束式 3 つの線型方程式となる．この方程式は，$\mathcal{C}_1$ の光学中心と画像点 $\boldsymbol{m}_1$ を結んだ直線と，$\mathcal{C}_2$ の光学中心と画像点 $\boldsymbol{m}_2$ を結んだ直線とが，ただ一点で交わる場合に唯一の解をもつが，対応点の計測誤差やカメラの校正誤差などの理由で，一般には解を持たない．そこで次のような行列を用いて近似解を求める．すなわち，

$$\boldsymbol{B} \;=\; (\boldsymbol{A}_1^{-1}\tilde{\boldsymbol{m}}_1 \mid -\boldsymbol{R}\boldsymbol{A}_2^{-1}\tilde{\boldsymbol{m}}_2)$$

とおくと，式 (7) は次のように変形できる．

$$\begin{pmatrix} w_1 \\ w_2 \end{pmatrix} \;=\; (\boldsymbol{B}^T\boldsymbol{B})^{-1}\boldsymbol{B}^T\boldsymbol{t} \tag{8}$$

ここで $(\boldsymbol{B}^T\boldsymbol{B})^{-1}\boldsymbol{B}^T$ を $\boldsymbol{B}$ の “擬似逆行列” と呼ぶ．式 (8) により得られる w_1 と式 (4) から，$\boldsymbol{M}$ は次式で求められる（左辺は $\tilde{\boldsymbol{M}}$ ではない）．

$$\boldsymbol{M} \;=\; w_1\boldsymbol{R}_1^{-1}\boldsymbol{A}_1^{-1}\tilde{\boldsymbol{m}}_1 + \boldsymbol{T}_1 \tag{9}$$

こうして求めた $\boldsymbol{M}$ は，$\mathcal{C}_1$ の光学中心と画像点 $\boldsymbol{m}_1$ を結んだ直線までの距離と，$\mathcal{C}_2$ の光学中心と画像点 $\boldsymbol{m}_2$ を結んだ直線までの距離の和を，最小にするような点である．2 直線が交わる場合には，真の解であるその交点が求まる．

3.3 顔の多面体近似による正面顔復元

3.3.1 処理の概要

まず，学習者ごとに，顔を多面体で近似した 3 次元モデルを用意する．このモデルはサーフェスモデルであり，顔特徴点の三次元座標の集合と，それらの顔特徴点を頂点とし，かつ顔表面を近似する三角形の集合からなる．次に，ステレオカメラにより撮影された 2 枚の学習者の顔画像から，用意したモデルに含まれる顔特徴点のうち 3 点の 2 次元位置を抽出し，その 3 次元座標を **3.2** の手法により計算する．これらの 3 次元位置に，その 3 点の顔特徴点の位置が適合するようにモデルの姿勢を定め，モデルの他の頂点を入力画像に逆投影する．これにより入力画像が，モデルの各面に対応した三角形の集合に分割される．その三角形ごとに，入力画像をモデルの対応する部分にテクスチャとして貼り付ける．このモデルを仮想視点から観測し，画像を合成する（図 5）．

この方法を用いるためには，ステレオカメラの画像から，顔上の特徴点を 3 点抽出しなければならない．これらの点は，画像処理によって抽出しやすいものであり，かつ，モデルの姿勢を決定するために，その位置ができる限り同一直線上から外れたものである必要がある．本研究では，このような 3 点の顔特徴点として，右目中心，左目中心，口中心を用いることにする．このとき，顔の 3 次元モデルには，右目中心，左目中心，口中心が頂点として含まれていなければならない．

以下では，**3.3.2** でモデルの姿勢を決定する方法を，**3.3.3** で仮想視点からのモデルの観測に必要となる仮想カメラのカメラパラメータの決定法を，**3.3.4** でテクスチャ貼り付けによる入力画像と仮想視点画像との対応点関係の決定法を，それぞれ詳述する．

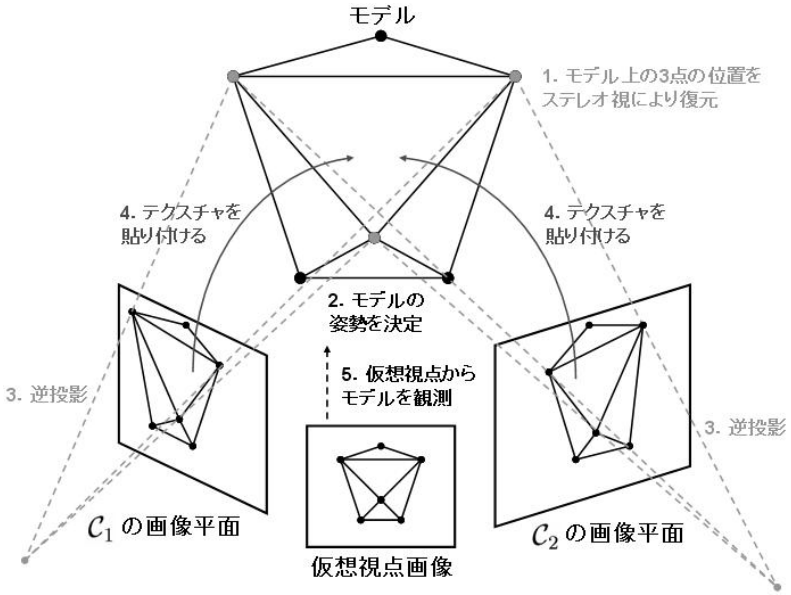


図 5: 正面顔復元の概要

3.3.2 モデルの姿勢の決定

モデルの右目中心，左目中心，口中心の 3 次元座標を $\boldsymbol{P}_1,\ \boldsymbol{P}_2,\ \boldsymbol{P}_3 \in \Re^3$ とする．これに対し，ベクトル $\boldsymbol{e}_1,\ \boldsymbol{e}_2,\ \boldsymbol{e}_3$ を

$$\begin{aligned} \boldsymbol{e}_1 &= \boldsymbol{P}_2 - \boldsymbol{P}_1 \\ \boldsymbol{e}_2 &= \boldsymbol{P}_3 - \boldsymbol{P}_1 \\ \boldsymbol{e}_3 &= \boldsymbol{e}_1 \times \boldsymbol{e}_2 \end{aligned}$$

と定義する．$\boldsymbol{P}_1,\ \boldsymbol{P}_2,\ \boldsymbol{P}_3$ が一次独立であれば，すなわちこの 3 点が一直線上に存在しなければ，$\boldsymbol{e}_1,\ \boldsymbol{e}_2,\ \boldsymbol{e}_3$ も一次独立となり，これらは 3 次元実空間の基底をなす．両目と口の位置関係から，$\boldsymbol{P}_1,\ \boldsymbol{P}_2,\ \boldsymbol{P}_3$ は自明に一次独立であるので，モデルの他の点の 3 次元座標を $\boldsymbol{P}$ とすると，$\boldsymbol{P} - \boldsymbol{P}_1$ は

$$\begin{aligned} \boldsymbol{P} - \boldsymbol{P}_1 &= \gamma_1 \boldsymbol{e}_1 + \gamma_2 \boldsymbol{e}_2 + \gamma_3 \boldsymbol{e}_3 \\ &= (\boldsymbol{e}_1\ \ \boldsymbol{e}_2\ \ \boldsymbol{e}_3) \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \end{pmatrix} \end{aligned} \tag{10}$$

と表せる．ここで $\boldsymbol{P} - \boldsymbol{P}_1$ の $\boldsymbol{P}_1,\ \boldsymbol{P}_2,\ \boldsymbol{P}_3$ に対する位置，向き，大きさは，モデルの姿勢によらず一定であるので，これらを基準として構成された基底 $\boldsymbol{e}_1,\ \boldsymbol{e}_2,\ \boldsymbol{e}_3$ で $\boldsymbol{P} - \boldsymbol{P}_1$ を表す限り，$\gamma_1,\ \gamma_2,\ \gamma_3$ はモデルを回転・並進させても一定である．

従って，入力画像からステレオ視により計算された左目中心，右目中心，口中心の 3 次元ワールド座標を $\boldsymbol{P}'_1,\ \boldsymbol{P}'_2,\ \boldsymbol{P}'_3$ とし，同様に基底 $\boldsymbol{e}'_1,\ \boldsymbol{e}'_2,\ \boldsymbol{e}'_3$ を構成したとき，その姿勢でのモデルの各点の座標 $\boldsymbol{P}'$ は，同じ $\gamma_1,\ \gamma_2,\ \gamma_3$ を用いて

$$\begin{aligned} \boldsymbol{P}' - \boldsymbol{P}'_1 &= \gamma_1\boldsymbol{e}'_1 + \gamma_2\boldsymbol{e}'_2 + \gamma_3\boldsymbol{e}'_3 \\ &= (\boldsymbol{e}'_1\ \ \boldsymbol{e}'_2\ \ \boldsymbol{e}'_3)\begin{pmatrix}\gamma_1\\ \gamma_2\\ \gamma_3\end{pmatrix} \end{aligned} \tag{11}$$

で表される．式 (10)(11) より，撮影時の顔に姿勢を合わせたモデルの各点の座標 $\boldsymbol{P}'$ は，

$$\boldsymbol{P}' = (\boldsymbol{e}'_1\ \ \boldsymbol{e}'_2\ \ \boldsymbol{e}'_3)(\boldsymbol{e}_1\ \ \boldsymbol{e}_2\ \ \boldsymbol{e}_3)^{-1}(\boldsymbol{P} - \boldsymbol{P}_1) + \boldsymbol{P}'_1 \tag{12}$$

で求められる．この処理を行うことで，モデルの姿勢を決定した後の各頂点の 3 次元ワールド座標を求めることができる．右目中心，左目中心，口中心を除くすべての頂点に対しこの処理を行うことで，結果的にモデルの姿勢を定めたことになる．

3.3.3 仮想視点のカメラパラメータの決定

姿勢を定めたモデルを，仮想視点から観測した画像を合成するために，仮想視点に置く仮想カメラのカメラパラメータを決定する必要がある．仮想カメラの内部パラメータ行列を $\boldsymbol{A}_v$，カメラ座標系のワールド座標系に対する回転行列と平行移動ベクトルを $\boldsymbol{R}_v,\ \boldsymbol{T}_v$ とする．まず，$\boldsymbol{A}_v$ に関しては，ステレオシステムに用いたカメラと同じカメラで撮影すると考えれば，カメラ $\mathcal{C}_1$ の内部パラメータ行列をそのまま用いることができる．故に，

$$\boldsymbol{A}_v = \boldsymbol{A}_1 \tag{13}$$

とする．

次に，外部パラメータについて考える．仮想視点のワールド座標を $\boldsymbol{P}_v \in \Re^3$ とすると，

$$\boldsymbol{T}_v = \boldsymbol{P}_v \tag{14}$$

であるから，これにより平行移動ベクトル $\boldsymbol{T}_v$ が与えられる．また，**2** 章の議論により，仮想カメラは顔の中心を向くように設置される．ただし本研究では，顔中心を右目中心，左目中心，口中心からなる三角形の重心と定義する．顔中心

のワールド座標を $\boldsymbol{P}_f \in \Re^3$ とする．仮想カメラのカメラ座標の z 軸正の向きがワールド座標の $\boldsymbol{P}_f - \boldsymbol{P}_v$ の向きに一致するように，$\boldsymbol{R}_v$ を定めてやればよい．ここで $\boldsymbol{P}_f - \boldsymbol{P}_v$ 方向の単位ベクトルを

$$\frac{\boldsymbol{P}_f - \boldsymbol{P}_v}{||\boldsymbol{P}_f - \boldsymbol{P}_v||} = (g_x \ g_y \ g_z)^T$$

と置き，さらに $(g_x \ g_y \ g_z)^T$ を次のように極座標表示する．

$$\begin{cases} g_x = \cos\phi \sin\theta \\ g_y = -\sin\phi \\ g_z = \cos\phi \cos\theta \end{cases}$$

ただし，ϕ は $(g_x \ g_y \ g_z)^T$ とワールド座標系の ZX 平面のなす角，θ は $(g_x \ 0 \ g_z)^T$ とワールド座標系の Z 軸のなす角である（図 6）．このとき，回転 $\boldsymbol{R}_v$ は，ワールド座標系の Y 軸の周りに $-\theta$ 回転し，その後，同じくワールド座標系の X 軸の周りに $-\phi$ 回転する変換に相当する．

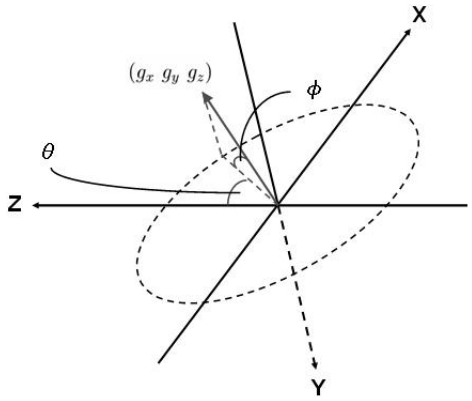


図 6: 仮想カメラの回転 1

ここで，Y 軸周りの $-\theta$ 回転を $\boldsymbol{R}_y(-\theta)$，$X$ 軸周りの $-\phi$ 回転を $\boldsymbol{R}_x(-\phi)$ とすると，これらはそれぞれ

$$\boldsymbol{R}_y(-\theta) = \begin{pmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{pmatrix} = (g_x^2 + g_z^2)^{-\frac{1}{2}} \begin{pmatrix} g_z & 0 & -g_x \\ 0 & 1 & 0 \\ g_x & 0 & g_z \end{pmatrix}$$

$$\boldsymbol{R}_x(-\phi) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\phi & \sin\phi \\ 0 & -\sin\phi & \cos\phi \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & (g_x^2 + g_z^2)^{-\frac{1}{2}} & -g_y \\ 0 & g_y & (g_x^2 + g_z^2)^{-\frac{1}{2}} \end{pmatrix}$$

と表される．従って，$\boldsymbol{R}_v = \boldsymbol{R}_x(-\phi)\boldsymbol{R}_y(-\theta)$ とすることで，仮想カメラは顔中心を向く．

ここまでの操作のみでは，画像平面を z 周りに回転することによる自由度がまだ残っているため，仮想カメラ座標系の x 軸，y 軸の向きが定まってない．そこで，モニタ画面の左上隅の y 座標と右上隅の y 座標が一致するように，仮想カメラ座標系の y 軸を定める．$\boldsymbol{R}_v = \boldsymbol{R}_x(-\phi)\boldsymbol{R}_y(-\theta)$ としたときの，モニタ画面の左上隅のカメラ座標を $\boldsymbol{l} = (l_x\ l_y\ l_z)^T$，右上隅のカメラ座標を $\boldsymbol{r} = (r_x\ r_y\ r_z)^T$ とする．$\boldsymbol{l}$, $\boldsymbol{r}$ を xy 平面に投影して考えると，$\boldsymbol{R}_v = \boldsymbol{R}_x(-\phi)\boldsymbol{R}_y(-\theta)$ で与えられるカメラ座標系を，さらに z 軸の周りに次式を満たす角 ψ だけ回転させれば，y 軸は上の条件を満たすようになる（図 7）．

$$
\begin{aligned}
\sin\psi &= \frac{l_y - r_y}{((l_x - r_x)^2 + (l_y - r_y)^2)^{\frac{1}{2}}} \\
\cos\psi &= \frac{-l_x + r_x}{((l_x - r_x)^2 + (l_y - r_y)^2)^{\frac{1}{2}}}
\end{aligned}
$$

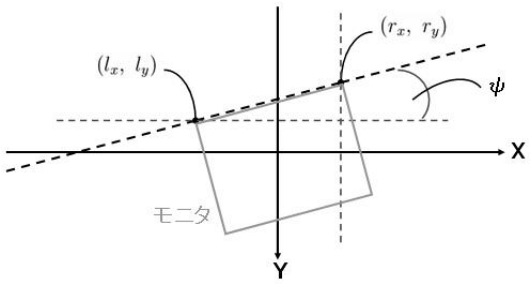


図 7: 仮想カメラの回転 2

z 軸周りの ψ 回転 $\boldsymbol{R}_z(\psi)$ は

$$
\begin{aligned}
\boldsymbol{R}_z(\psi) &= \begin{pmatrix} \cos\psi & -\sin\psi & 0 \\ \sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{pmatrix} \\
&= ((l_x - r_x)^2 + (l_y - r_y)^2)^{-\frac{1}{2}} \begin{pmatrix} r_x - l_x & r_y - l_y & 0 \\ l_y - r_y & r_x - l_x & 0 \\ 0 & 0 & ((l_x - r_x)^2 + (l_y - r_y)^2)^{\frac{1}{2}} \end{pmatrix}
\end{aligned}
$$

と表される．結局 $\boldsymbol{R}_v$ は,

$$
\boldsymbol{R}_v = \boldsymbol{R}_z(\psi)\boldsymbol{R}_x(-\phi)\boldsymbol{R}_y(-\theta) \tag{15}
$$

で与えられる．

3.3.4 テクスチャマッピングによる対応点関係の決定

入力画像をモデルに貼り付け，そのモデルを仮想視点から観測したときの画像を生成するために，入力画像と仮想視点画像との間の対応点関係を，顔モデルを構成する各三角形ごとに与える必要がある．

仮想カメラのカメラ座標系で，モデル上の任意の三角形の頂点の 3 次元座標が $\boldsymbol{M}_1,\ \boldsymbol{M}_2,\ \boldsymbol{M}_3 \in \Re^3$ であるとする．この三角形の法線を

$$(a\ b\ c)^T \ = \ \boldsymbol{N} \ = \ (\boldsymbol{M}_1 - \boldsymbol{M}_2) \times (\boldsymbol{M}_1 - \boldsymbol{M}_2)$$

とおく．また，$d = \boldsymbol{M}_1 \cdot \boldsymbol{N}$ とおく．このとき，三角形上の点 $\boldsymbol{M} = (X\ Y\ Z)^T$ は，

$$\begin{aligned} & \boldsymbol{M} \cdot \boldsymbol{N} = (\boldsymbol{M} - \boldsymbol{M}_1) \cdot \boldsymbol{N} + \boldsymbol{M}_1 \cdot \boldsymbol{N} = \boldsymbol{M}_1 \cdot \boldsymbol{N} \\ \Longleftrightarrow \quad & aX + bY + cZ = d \end{aligned} \tag{16}$$

を満たす．ここで，モデルの姿勢は **3.3.2** 節の方法で，仮想視点の位置と向きは **3.3.3** 節の方法で定まっているので，モデルの各三角形を仮想視点画像に逆投影することで，仮想視点画像上の任意の点 $\boldsymbol{p}$ がモデルのどの三角形上の点の像であるかを知ることができる．従って，対応する三角形の法線ベクトルの値に応じて式 (16) を拘束として用いることで，$\boldsymbol{p}$ に結像した点の元の 3 次元ワールド座標は，斉次座標を用いて次のように求められる．

$$\begin{pmatrix} \boldsymbol{R}_v & -\boldsymbol{R}_v\boldsymbol{T}_v \\ \boldsymbol{0} & 1 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ \frac{a}{d} & \frac{b}{d} & \frac{c}{d} \end{pmatrix} \boldsymbol{A}_v^{-1}\tilde{\boldsymbol{p}} \in \Re^4$$

この点は，入力画像 $\mathcal{I}_1$ では，投影方程式により斉次座標

$$\boldsymbol{A}_1(\boldsymbol{R}_1 \mid -\boldsymbol{R}_1\boldsymbol{T}_1) \begin{pmatrix} \boldsymbol{R}_v & -\boldsymbol{R}_v\boldsymbol{T}_v \\ \boldsymbol{0} & 1 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ \frac{a}{d} & \frac{b}{d} & \frac{c}{d} \end{pmatrix} \boldsymbol{A}_v^{-1}\tilde{\boldsymbol{p}}$$

$$= \boldsymbol{A}_1\boldsymbol{R}_1(\boldsymbol{R}_v^{-1} \mid \boldsymbol{T}_v - \boldsymbol{T}_1) \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ \frac{a}{d} & \frac{b}{d} & \frac{c}{d} \end{pmatrix} \boldsymbol{A}_v^{-1}\tilde{\boldsymbol{p}} \in \Re^3$$

の位置に結像する．従って，仮想視点画像上の点 $\boldsymbol{p}$ と，それに対応する入力画像 $\mathcal{I}_1$ 上の点 $\boldsymbol{p}_1$ との間の対応点関係は，パッチごとに行列

$$\boldsymbol{H}_1 = \boldsymbol{A}_1\boldsymbol{R}_1(\boldsymbol{R}_v^{-1} \mid \boldsymbol{T}_v - \boldsymbol{T}_1) \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ \frac{a}{d} & \frac{b}{d} & \frac{c}{d} \end{pmatrix} \boldsymbol{A}_v^{-1}$$

を用いて

$$\tilde{\boldsymbol{p}_1} \approx \boldsymbol{H}_1\tilde{\boldsymbol{p}} \tag{17}$$

で与えられる．ここで記号 $\approx$ は右辺の定数倍が左辺に等しいことを意味する．入力画像 $\mathcal{I}_2$ 上の点 $\boldsymbol{p}_2$ との対応点関係についても同様である．仮想視点画像上の位置 $\boldsymbol{p}$ のピクセルの色は，入力画像 $\mathcal{I}_1$, $\mathcal{I}_2$ 上の位置 $\boldsymbol{p}_1, \boldsymbol{p}_2$ のピクセル色の平均値を与える．

第4章　実験

4.1　実験の概要

3 章で提案した正面顔復元手法により合成される顔画像に，どの程度ひずみが生ずるのかを調べるために，実際にカメラで人間の顔を撮影し，仮想視点を顔に正対する位置において正面顔画像を復元した．復元した顔画像を，仮想視点と同じ視点から撮影した実写の顔画像と比較することで，本正面顔復元手法を評価した．また，**2** 章で提案した手法により合成される画像を見て，どの程度学習者の注視位置が判別できるのかを調べるために，復元した顔画像と WebCT の学習ログデータとを照合して重畳する教材画像を選択し，実際に画像を合成した．合成した画像を見て，被撮影者が教材のどの部分を見ていると思うかを，数人の被験者に回答してもらい，実際とのずれを確認することで，本合成手法を評価した．ただし本実験では，あらかじめ被撮影者に注視する部分を指定しておき，その位置を実際の注視位置とした．

4.2 準備

4.2.1 実験環境

本実験に使用した実験環境を図 8 に示す．ステレオカメラをモニタの枠部分の左上部と右上部に設置し，さらに顔の照明条件を少しでも向上させるため，モニタの上部には蛍光灯を設置した．この環境において，ステレオカメラで同一のチェスボードを数回撮影し，Zhang の手法 [4] を用いてステレオカメラの校正を行った．これにより，**3.2** 節の三次元位置計測における $\boldsymbol{A}_1$, $\boldsymbol{A}_2$, $\boldsymbol{R}_1$, $\boldsymbol{R}_2$, $\boldsymbol{T}_1$, $\boldsymbol{T}_2$ を与えた．

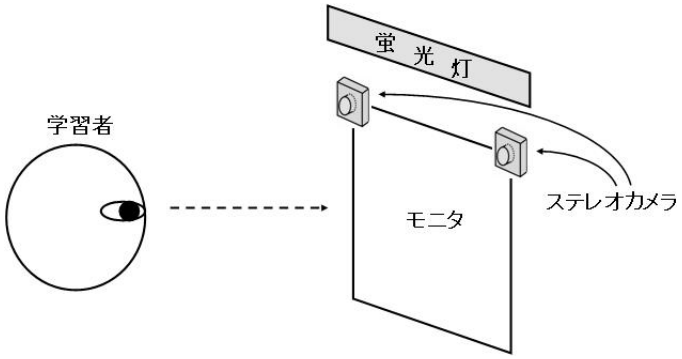


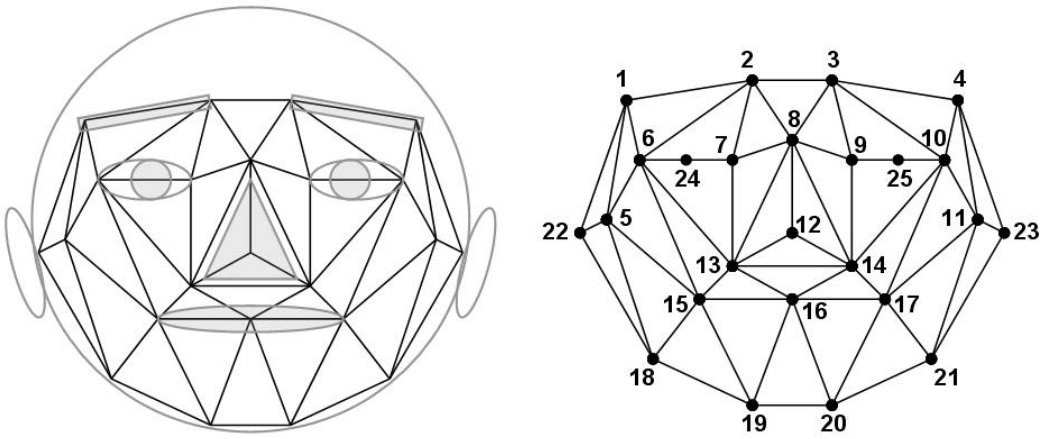
図 8: 実験環境

e-learning システムには高等教育機関向けプラットフォーム “WebCT” を用い，この環境で被験者がモニタに表示された教材を見ているときの顔を撮影した．撮影された顔画像にはタイムスタンプを装入した．このとき，撮影する側の PC の時刻を，Windows 2000 の Windows Time サービスを用いて，Simple Network Time Protocol(SNTP) により WebCT サーバの時刻に合わせることで，WebCT のサーバと顔画像撮影処理とを同期させた．

4.2.2 顔の 3 次元モデルの作成

正面顔復元処理に必要となる顔の 3 次元モデルは，被験者ごとに次の方法で作成した．まず，顔を近似する多面体を，図 9 のようなサーフェスモデルで定義した．頂点は 23 点あり，図 9(b) 中の番号と各点の位置との対応は表 1 のとおりである．24 および 25 は三角形の頂点としては用いていないが，提案手法による復元には右目中心と左目中心の位置情報が必要になるため，モデルに加えた．

次に，図 8 の環境で被験者の顔を撮影し，得られた画像上で，図 9 の多面体の各頂点の 2 次元位置を手動で抽出した．ただし，画像から目視により直接抽



(a) モデルと顔の関係　　(b) 顔の 3 次元モデル

図 9: 実験に用いた 3 次元モデル

出することが困難である頂点 5,11,18,19,20,21,22,23 については，撮影時に顔の対応する部分にマーカを貼付し，これを手がかりに画像中の 2 次元位置を抽出した．このようにして得られた顔特徴上の 2 次元位置から，その 3 次元位置を **3.2** 節の手法に従ってステレオ計測し，被験者の顔の 3 次元モデルの頂点位置とした．

4.2.3　モニタ位置の計測

顔画像に教材画像を重畳する際に必要となるモニタ画面の 4 隅のワールド座標は，次のようにして求めた．まず，マーカを 3 点貼付した平面状のプラスチック板（鏡の代わりとして用いた）を，**4.2.1** の実験環境のステレオカメラで撮影した．得られた画像には，貼付したマーカと，プラスチック板で反射したモニタ画面の像が写っているので，マーカの 2 次元位置と，モニタ画面の反射像の 4 隅の 2 次元位置を，手動で抽出した．次に，それらの 3 次元ワールド座標を，**3.2** 節の手法にしたがってステレオ計測した．マーカは 3 点存在するので，それらの座標値からプラスチック板の 3 次元位置と向きが求まる．このプラスチック板をはさんで，モニタ画面の反射像の 4 隅の位置と対称な位置をそれぞれ求め，それらをモニタ画面の 4 隅のワールド座標とした．ただしこのとき，4 隅のワールド座標に対し，それらを最適に近似する平面を最小二乗法により求め，各点をその平面に射影することで，4 点の座標が同一平面上にのるようにした．さらに，モニタ画面，すなわち 4 隅の座標からなる四角形が長方形となるよう，4 隅のワールド座標を手動で修正した．

番号	頂点の位置	番号	頂点の位置
1	右眉の右端	14	鼻・左辺の下端
2	右眉の左端	15	口の右端
3	左眉の右端	16	口の中心
4	左眉の左端	17	口の左端
5	右のほお骨	18	顔輪郭・右ほおの下部
6	右目の目尻	19	顔輪郭・左ほおの下部
7	右目の目頭	20	顔輪郭・あごの右端
8	眉間の下部中央	21	顔輪郭・あごの左端
9	左目の目頭	22	右耳の元
10	左目の目尻	23	左耳の元
11	左のほお骨	24	右目の中心
12	鼻の先端	25	左目の中心
13	鼻・右辺の下端		

表 1: 番号と頂点の位置との対応

4.3 正面顔復元実験

4.3.1 正面顔復元

図 8 の環境で，被験者の顔をステレオカメラで撮影した．それぞれの画像上で右目中心・左目中心・口中心の 2 次元位置を抽出する際には，オムロン社の顔画像処理ソフト “OKAO Vision” を用いた．抽出された目・口の 2 次元位置から，その 3 次元位置を **3.2** 節の式に従って算出し，**3.3** 節の方法により正面顔を復元した．

このとき，ステレオカメラでは，15(frame/sec) で画像を撮影したが，その中の数〜十数フレームは，OKAO Vision による右目中心・左目中心・口中心の 2 次元位置の抽出に失敗した．この原因としては，顔が横向きになりすぎて顔造作の位置関係が想定外のものとなった，照明環境により顔にかかるハイライトが強くなりすぎた，などが考えられる．このようなフレームに対しては，目・口中心の 3 次元位置を次のような線形補間により補間した．x を目・口中心の 2 次元位置の抽出に失敗したフレームの番号，a, b を x の前後で抽出に成功した

フレームの番号，$\boldsymbol{M}_i$ をフレーム i の左目中心（あるいは右目中心，口中心）の 3 次元ワールド座標とすると，$\boldsymbol{M}_x$ を，

$$\boldsymbol{M}_x \;=\; \frac{b-x}{b-a}\boldsymbol{M}_a \;+\; \frac{x-a}{b-a}\boldsymbol{M}_b \;(a<x<b)$$

として求めた．

また，目・口中心が抽出されたフレームであっても，顔のわずかな動きによって抽出位置が変動し，その結果，これに基づいて算出された 3 次元位置にノイズが含まれてしまう場合が多い．そこで，モニタを見ている被験者は，顔の位置・向きの変動が滑らかであると仮定し，左目中心・右目中心・口中心の各フレームでの 3 次元ワールド座標を，そのフレームと前後 1 フレーム，計 3 フレームの平均をとることで平滑化した．この平滑化処理は，一度の撮影で得られた画像系列あたり 10 回行った．

4.3.2 評価

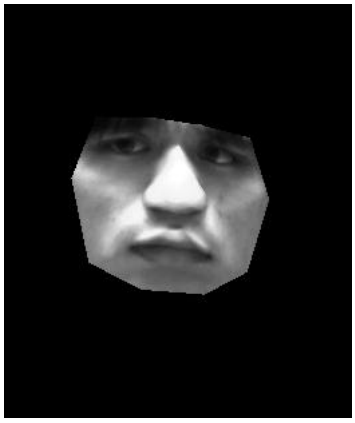
本実験では，提案した復元手法の評価を行うため，ステレオ計測に用いたものとは異なるカメラを別に設置した．そのカメラで被験者の顔を実写で撮影する一方，そのカメラの位置に仮想視点を置いて被験者を観測した場合の画像を正面顔と同様の方法で復元し，両者を比較した．結果を図 10 に示す．図の (a) は被験者から見てモニタの左上隅に設置したステレオカメラの画像，(b) は被験者から見てモニタの右上隅に設置したステレオカメラの画像，(c) はステレオカメラとは別の位置に設置した別カメラの画像，(d) は (a)(b) を用いて (c) のカメラと同じ視点での画像を復元した結果，(e) は (c) と (d) の差分画像，(f) は (a)(b) を用いて復元した正面顔画像である．

図 10(e) を見ると，本手法によって，全体的に実写画像との違いの少ない画像が復元されていることが分かる．なお，(d)(f) の復元結果では，鼻の下部が肌色一色で，鼻孔が移っていないが，これは入力として用いたステレオカメラの画像 (a)(b) がモニタの上部から撮影されたものであるため，そもそもこれらの画像に鼻孔がはっきりと写っていないからである．(e) において，鼻孔の周辺と上唇で差分が大きくなっているのも，同様の理由による．また，(d)(f) で，鼻のみが他の部位と比較して極端に明るいが，これは照明条件がまだ完全ではないためである．実際，入力画像 (a)(b) でも，鼻を除いて顔に十分な量の光が当たっていない．さらに，左右のほお全体にかけて，一様に微小な差分があるが，これは，本論文で提案する手法では，(c) での光の当たり方までを復元すること

(a) 左上ステレオカメラの画像

(b) 右上ステレオカメラの画像

(c) 別に用意したカメラの画像



(d) 別カメラと同じ視点での復元画像

(e) 中段の 2 枚の画像の差分

(f) 復元した正面顔画像

図 10: 正面顔復元実験 結果

はできないからである．

4.4 注視位置・教材内容が判別可能な画像の合成

4.4.1 教材画面の選択・重畳

WebCT に学習ログとして残っているアクセスデータから，ユーザごとの各ページへのアクセス時刻を割り出し，これと顔画像の撮影時刻とを照合することで，顔画像に重畳する教材画像を選択した．これに対して，**2.2**，**2.3** で提案した手法を用いて，各フレームでの復元顔画像に，対応する教材画像を，両画像の対応するピクセルの色の平均をとることで重畳した．このとき，顔画像の大きさが教材画像の大きさに対して小さくなりすぎないように，また，教材画像が十分な解像度で表示されるように，モニタ画面の相似縮小の割合を考慮した．この結果，式 (1)(2) における α の値を $\alpha = 0.38$ とした．

教材画像をスケーリングする際，教材に使われる文字フォントは，一般に線が細いため，縮小アルゴリズムを適用するだけでは文字が読めなくなる．このため本実験では，教材の画像に一旦 3×3 の移動平均をかけて文字の線をぼやけさせ，次いで画像の情報を最大限に残せるよう，バイキュービック法 [5] を用いてスケーリングした．

4.4.2 表示画面の作成

4.4.1 節で合成した画像を用いて，**2** 章で述べたような学習ログ，すなわち，教材のページごとに，それを見ている学習者の顔画像が並べて表示され，さらにその顔画像をクリックすることで，視線の情報が表示されるような画面を合成した．合成画像の作成時に入力画像として用いた右上ステレオカメラ画像を，WebCT のアクセスデータと照合して教材のページごとに並べ，そのページを見ている人の顔の一覧ページとした．一覧ページに表示されている顔画像をクリックすることで，**4.4.1** で作成した合成画像が表示されるようにした．

4.4.3 評価

4.4.2 で述べた画面の合成結果を図 11 に示す．さらに，その画面に表示されている顔画像をクリックしたときに表示される重畳画像の例を，図 12 に示す．

図 12 の合成画像で，注視位置がどの程度正しく判別できるかを調べるため，合成画像の時系列データを表示した映像を 5 人の被験者に提示して，被撮影者が教材のどのオブジェクトを見ていると思うかについて，回答を求めた．その結果と，被撮影者が実際に見ていた箇所とを比較し，正誤率および位置のずれ

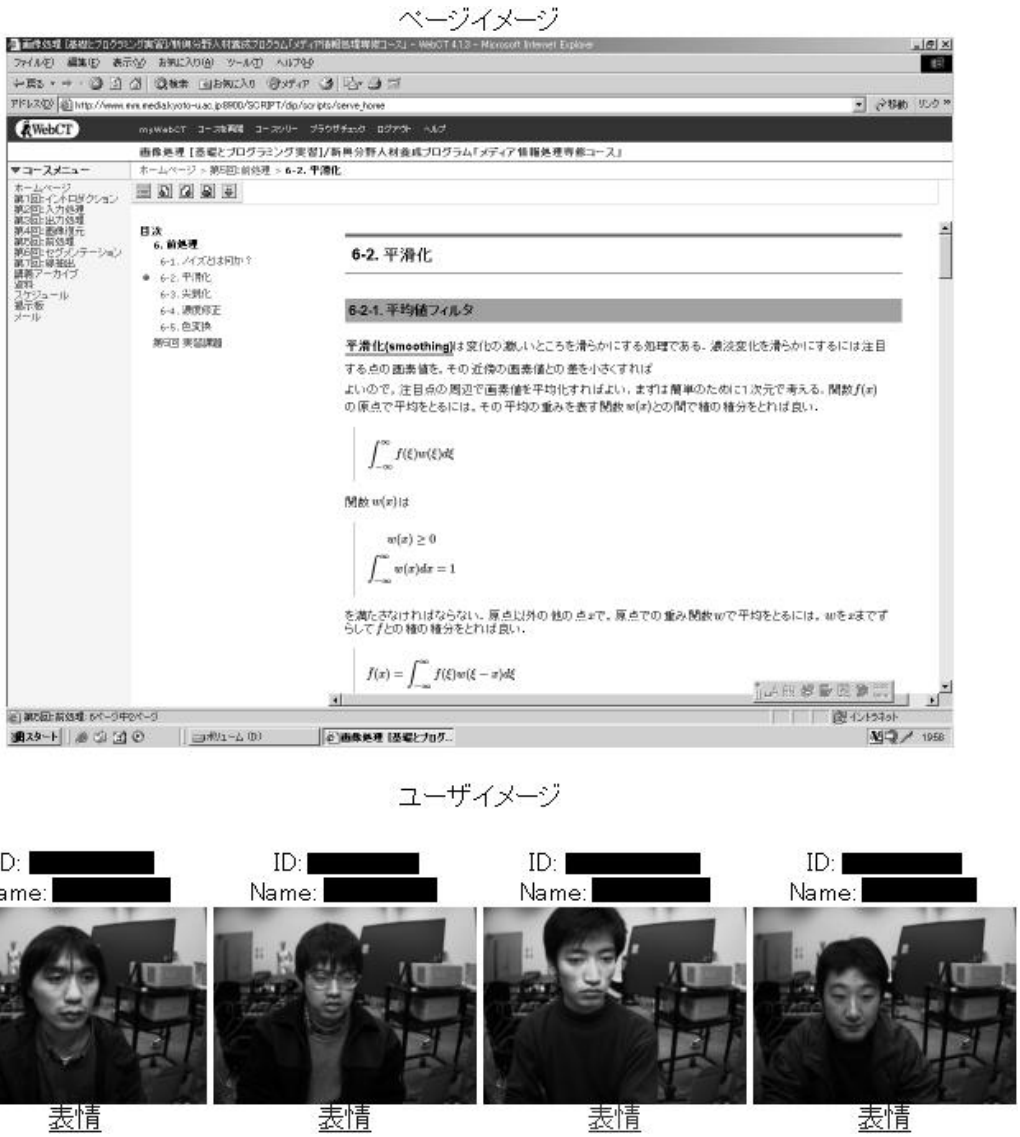


図 11: 学習ログ（一覧）の例

を検証した．用いた教材画像を図 13 に，得られた結果を表 2 に示す．

表 2 において，教材とは，提示した合成画像のベースとなった教材画像，総事例数とは被験者に提示した映像の総数，正解数とは被験者の推定したオブジェクトが実際の注視オブジェクトに一致していた事例数，不正解数とは被験者の推定したオブジェクトが実際の注視オブジェクトとは違っていた事例数，正解率とは正解数／総事例数である．また，ずれの平均とは，不正解であった事例を対象に，被験者の推定したオブジェクトと実際の注視オブジェクトとの間の距離の平均を求めたものである．ただしここでは，2 つのオブジェクト A, B 間

図 12: 画像合成 結果

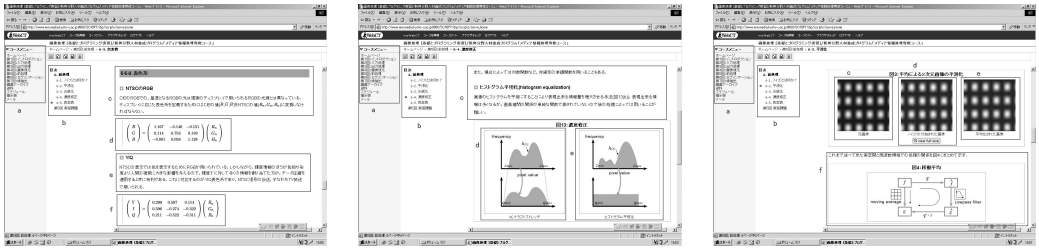
教材	総事例数	正解数	不正解数	正解率 (%)	ずれの平均
A	20	15	5	75.0	1.2
B	20	16	4	80.0	1.0
C	20	16	4	80.0	1.0

表 2: 学習者の注視位置判別の正誤率

の距離 $d(A, B)$ を次のように定義している．

- A と B が隣り合っているとき，$d(A, B) = 1$
- それ以外のとき，$d(A, B) = d(A, C) + 1$．ただし C は B と隣り合うオブジェクトで，A からの距離が最小であるもの

正解率はオブジェクト分割の仕方にもよると考えられるが，教師の立場で考えれば，意味のあるまとまりを一つのオブジェクトと見なすべきである．従って今回の検証では，図 13 において枠で囲まれた部分を，それぞれ単一のオブジェクトと見なした．その結果，表 2 のように，それぞれの教材で，75%から



A（題材:色変換）　　B（題材:濃度修正）　　C（題材:平滑化）

図 13: 注視位置判別に用いた教材

80%の精度で正しく被撮影者の注視位置が判別できることを確認した．また，不正解であった事例でも，正解とのずれは 1.0〜1.2 であり，正解のオブジェクトに隣り合わないオブジェクトと誤ることは稀であることが分かる．不正解事例の多くでは，被撮影者は，教材画面の外縁部に配置されたオブジェクトを注視していた．この場合に正解率が低下する原因としては，外縁部にあるオブジェクトでは，被撮影者の顔と教材の写っているモニタ画面との間の距離の推定に誤差が含まれたとき，その誤差に対応する画像上のユークリッド距離が大きくなるため，注視位置推定のずれ幅が大きくなりやすい，ということが考えられる．このような距離推定誤差は，モニタ画面の相似縮小時にその縮小割合を画像上でも保存することで，少なくなるように工夫しているが，さらに，教材画像に対する顔画像の大きさや，画像重畳時の色混合の割合などを視覚心理に基づいて再考することで，より小さくできると考えられる．

第 5 章　結論

本研究では，教師が生徒の表情や視線を把握できるという講義室での講義の利点を，e-learning システムでの学習に持たせるために，学習者の表情と注視位置が判別できるような画像を e-learning システムの学習ログとして利用できるようにすることを提案した．表情の分かりやすい画像としては，合成画像の不自然さを考慮し，実写の画像を用いて対応した．一方，注視位置を判別する際には，その注視部分の内容が判別できる必要があるため，注視位置と教材内容の両方が判別できる画像を学習ログにしなければならない．これを満足する画像として，学習者とモニタに正対する位置を視点とし，モニタを透かして学習者の顔を観測したような画像を合成することを提案した．このような視点から

の画像は，カメラの設置可能位置の制約上，実写から直接得ることはできないので，顔を多面体で近似して正面顔を復元し，それに教材の画像を重畳することで合成した．これらの画像を用いて，学習者の表情と注視位置が分かりやすい学習ログを作成した．

今後の課題として，注視位置判別の正解率をさらに向上させるために，学習者の顔とモニタ画面との間の距離推定に関係するファクタを，視覚心理に基づいて検討することが挙げられる．また，表情・注視位置に加え，身振りを学習ログとして利用できるようにすることや，合成画像において，顔のみという不自然さを解消するために，髪や胴体部分などを追加することなどが挙げられる．

謝辞

本研究を行うにあたり多くの御教示，熱心な御指導を賜りました美濃 導彦教授に深く感謝致します．日頃より数多くの助言を頂きました美濃研究室の角所 考助教授に深く感謝致します．また，本研究第 4 章の実験，および本報告書の執筆に際して多大な御指導・ご協力を賜りました伊藤 淳子氏をはじめとするコミュニケーショングループ，並びに美濃研究室の皆様に深く御礼を申し上げます．

参考文献

[1] 満上 育久，浮田 宗伯，河野 恭之，木戸出 正継．“両眼視線情報を利用した対象の位置推定とその応用”．人工知能学会 第 16 回全国大会，3C4-04，(2002)．

[2] Steven M.Seitz, Charles R.Dyer. “View Morphing”. Proceedings of SIGGRAPH ’96 (1996), pp.21-30.

[3] T.Kanade, P.J.Narayanan, P.W.Rander. “Virtualised reality: concepts and early results”. Proceedings of IEEE Workshop on Representation of Visual Scenes (1995), pp.69-76.

[4] Zhengyou Zhang. “A Flexible New Technique for Camera Calibration”. Microsoft Research Technical Report MSR-TR-98-71 (Dec 1998).

[5] 川崎 高志．“画像拡大手法に関する考察”．http://mikilab.doshisha.ac.jp/dia/monthly/monthly01/20011222/personal_kawasaki.pdf