2005－HI－112　(9)
2005／1／28

注目領域に基づく適応的な注釈情報の提示

竹村　知晃*1　　吉高　淳夫*1　　平嶋　宗*1

拡張現実感（AR）において注釈情報を提示する際に、ユーザが注目したオブジェクト内の特定領域に関する情報を提示することが考えられる。この提示を、特定領域にタグを付け、注目した領域のタグを読みとることにより実現しようとした場合、オブジェクトによってはタグを付けることにより外観を損ね、鑑賞の妨げとなることもあるため、実用的でない。本研究では、眼球運動を解析することにより、オブジェクトに手を加えることなく、ユーザが注目したオブジェクトのうち注目した領域を判別して、その特定領域に適応させた注釈情報を提示する手法を提案する。

Presenting Adaptive Annotation based on Gazed Region

Tomoaki Takemura *1　　Atsuo Yoshitaka*1　　Tsukasa Hirashima*1

Abstract

In case of presenting annotations in an Augmented Reality (AR) system, it is possible to present information that refers to a specific region in the object where a user gazes. If this presentation were implemented by attaching tags to specific regions and identifying the tag, it would not be practical for some objects like paintings. It is because the appearances of objects are spoiled and it becomes the hindrance of viewing. This paper proposes a method that identifies regions in the object where the user gazes by analyzing eye movement without altering the objects, and presents the annotation which is adapted to the regions.

1 はじめに

拡張現実感(Augmented Reality、以下 AR と略す)とは、テキストや CG 等の注釈情報を実世界のオブジェクトに重ねて、あるいは近い位置に提示することによって、実世界のオブジェクトに対して情報を付加する技術であり、これによりユーザはオブジェクトの外観が持つ以上の情報を得ることができる。AR システムでは、ユーザが興味を持つオブジェクトの情報を提示することを目的とするため、ユーザが注目しているオブジェクトを認識する必要がある。既存のシステムにおいて、オブジェクトの認識方法は、オブジェクトを認識するためのタグ付けを行う方法と、行わない方法に大別でき、それぞれの方法により研究が盛んに行われている。

オブジェクトにタグを付け、読みとることによりオブジェクトを認識し、関連付けられた情報を提示する方法では[1]-[4]、2 次元バーコードや RFID タグ、赤外 LED をオブジェクトに取り付ける方法や[1]-[3]、物理的なタグを貼り付けるのではなく、赤外線によりタグを投影す

*1: 広島大学大学院工学研究科
*1: Graduate School of Engineering, Hiroshima University

る方法が提案されている[4]。これらの方法では、注釈情報を提示するために、ユーザはタグリーダやカメラなどのデバイスによりタグを読みとるという明示的な操作が必要となる。

一方、タグ付けを行わずオブジェクトの認識を行い、関連付けられた情報の提示を行う方法では[5][6]、事前に獲得しておいたオブジェクトの画像を基にして、ユーザに装着したカメラに映っているオブジェクトの認識を行っている。これらの方法では、オブジェクトに関する情報を提示する際に、ユーザがそのオブジェクトに興味を持っているか否かの判断を行っていないため、ユーザにとって不必要な情報を提示する場合もある。

既存のシステムで行っているのはオブジェクトの認識であるため、オブジェクト全体に関する注釈情報を提示することは可能であるが、オブジェクトのうちユーザが注目した特定領域に関する注釈情報を提示することはできない。よって、オブジェクトのうち、ユーザが注目した特定領域に関する注釈情報を得るためには、注目したオブジェクトや、特定領域を判断することが必要となる。

ユーザがオブジェクトに対して注目していることを判断する手がかりとして、眼球運動に着目する。眼球運動は人間の注意や関心を表しており、静止している視覚情報に注目している場合は、約 300 ミリ秒間の固視状態と約 30 ミリ秒間に起こる跳躍運動を頻繁に繰り返すことが知られている[7]。この特徴を利用することにより、ユーザがオブジェクトに注目している状態を検出することが可能である。

本研究では、眼球運動からオブジェクトに注目している状態を検出して、そのオブジェクトの注釈情報を適応的に提示することを目的とする。注釈情報を適応的に提示するとは、オブジェクト内の注目している個所(以下、注目領域と記す)を判別し、その注目領域に関する情報を提示することをいう。本手法では、注目対象となるオブジェクトに関して、注釈情報を付加した領域(以下、注釈付加領域と記す)の位置情報と、その注釈情報をデータベースに定義しておく。ユーザの頭部に装着したカメラにより注目したオブジェクトを撮影し、カメラの画像に映っているオブジェクトの領域を抽出する。次に、そのときの瞳孔の位置と、注目した絵画に関するデータベース中の注釈付加領域の位置情報を基にして、注目領域を判別する。これにより、明示的にオブジェクトを選択することなく、注目したオブジェクトのみに、注目領域に適応させて注釈情報を提示することができる。

応用例の一つとして、美術館のような環境で展示されている絵画に、ユーザの注目領域に適応させて注釈情報を提示する。図 1 のように注釈付加領域を定義しておき、森に注目した場合、森に関する情報を提示する(図 2)。注釈情報は、ユーザが持っているノート型 PC の画面上に表示される。

本手法は、データベース中の注釈付加領域の位置情報に基づき、注目領域の判別を行っている。注目領域の判別は、オブジェクトや領域に 2 次元バーコードのような物理的なタグを付けることにより行うこともできる。しかし、絵画のように付けたタグが鑑賞の妨げになるオブジェクトに対しては、タグによる注釈情報の提示は適切な方法ではない。また、赤外線による非可視のタグを投影する方法も考えられる。しかし、タグの認識精度は投影面の色やテクスチャに影響を受けるため、絵画のようなオブジェクトには適切な方法ではない。これに対し、本手法はオブジェクトに手を加えることなく注目領域を判別し、その領域に適応させて注釈情報を提示することができる。

本稿では, まず 2 章でシステム構成について述べ, 3 章で注目の検出と注視点分布の決定方法について述べる. 次に 4 章では視界画像からオブジェクト領域を抽出する方法について述

べ，5 章では注視点分布に基づく注釈情報の提示方法について述べる．そして 6 章では提案方法の実験と評価を行い，最後に 7 章でまとめと今後の課題を述べる．

図 1. 注釈付加領域

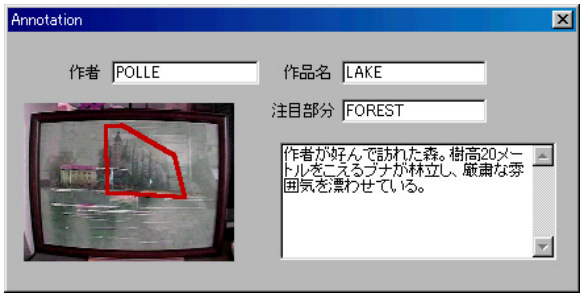


図 2. 注釈情報の提示

2 システム構成

システム構成を図 3 に示す。眼球を撮影する CCD カメラは眼球の下部に設置する。また、視界を撮影する CCD カメラは額に設置する。2 つのカメラの位置関係は、ユーザが正面を見ているときの眼球とそのときの視界の中心がそれぞれのカメラ画像において中心に映るようにする。入力画像の画像座標系は図 4 のようになる。カメラを装着したときの様子を図 5 に示す。

使用した PC は、CPU が PentiumⅢ、500MHz のノート型 PC であり、処理する映像のフレームサイズは 2 台のカメラとも 160×120[pixel]、処理速度は 10[fps]である。なお、眼球撮影画像は 256 階調グレースケール、視界画像は 24bit color である。

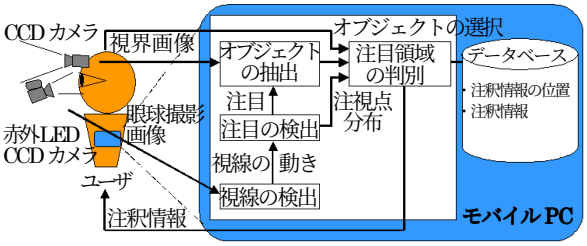


図 3. システム構成

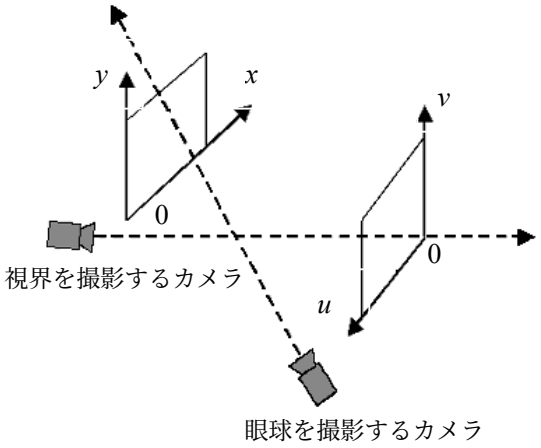


図 4. 画像座標系

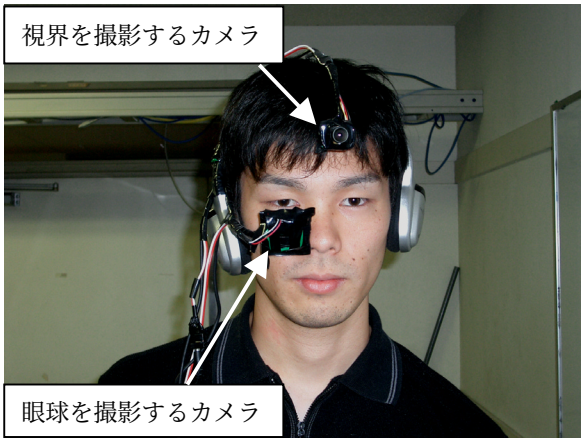


図 5. カメラ装着の様子

3 注目の検出

ユーザが視覚情報に注目していることを検出するために、眼球を撮影するカメラの画像(以下、眼球撮影画像と記す)から注目時の視線の動きを検出する。システムが注目時の視線の

動きを検出してから、その動きを連続して検出できなくなるまでを一注目動作とする。

まず、ユーザの視線を検出し、視線に対応した注視点座標を算出する。ここで、注視点座標とは、ユーザが視線を合わせていた点（以下、注視点と記す）を、視界を撮影するカメラの画像(以下、視界画像と記す)における座標で示したものとする。そして、一注目動作の注視点列を注視点分布とする。

3.1 視線の検出及び注視点の算出

(1) 視線の検出

視線の位置の検出は、瞳孔の位置を検出することにより行う。より正確な視線の位置を求めるため、眼球に赤外線を照射することで瞳孔と虹彩のコントラストを強調し、眼球撮影画像の二値化処理によって、瞳孔領域を抽出する。そして、瞳孔領域の中心座標($x(t), y(t)$)を時刻 t (フレーム)における瞳孔の位置とする。ここで、眼球撮影画像の水平方向を x 軸、垂直方向を y 軸とし、原点を画像の左下とする。また、フレームレートは 10fps である。

(2) 注視点座標の算出

注視点座標($G_u(t),\ G_v(t)$)を式(1)によって求める。視界画像における水平方向を u 軸、垂直方向を v 軸とし、原点を画像の左下とする。ここで、視界画像と眼球撮影画像の一部領域($40 \leqq x \leqq 120$、$40 \leqq y \leqq 80$)を 4×4 の 16 ブロックに分割する。$(D(i,j)_x, D(i,j)_y)(i,j=1,...,4)$は眼球撮影画像のブロック$(i,j)$の中心座標とし、$(Cv(i,j)_u, Cv(i,j)_v)$ は視界画像のブロック(i,j)の中心座標とする。また、$\alpha(i,j)$、$\beta(i,j)$は各ブロックにおける、瞳孔の中心の移動距離に対する注視点の移動距離の比を表す値である。

$$\begin{cases} G_u(t) = Cv(i,j)_u - (x(t) - D(i,j)_x) \times \alpha(i,j) \\ G_v(t) = Cv(i,j)_v + (y(t) - D(i,j)_y) \times \beta(i,j) \end{cases} \quad (1)$$

(3) 注目状態の検出

瞳孔の位置が 3 フレームの間(約 300 ミリ秒)動かないか、もしくは微小な動きであり、4 フレーム目で跳躍する状態を検出し、固視・跳躍とする。固視・跳躍が連続して 3 回以上検出されたとき、注目状態と判定する。ただし、3 フレーム間で、瞳孔の位置の移動が視角にして 2.1 ° 未満であれば固視とみなし、2.1 ° 以上移動すれば跳躍したとみなす。また、固視・跳躍が連続するとは、跳躍が生じてから 3 秒間で次の跳躍が生じることとする。

3.2 一注目動作から得られる注視点分布

一注目動作における各注視点座標は、ユーザが注目状態であったときの瞳孔の位置から算出しているため、注視点座標はユーザが注目しているオブジェクト上、またはその付近にあたる。そのため、一注目動作の注視点を含む最小凸多角形領域には、注目していたオブジェクト全体、またはその一部が含まれている。

4 視界画像からのオブジェクト領域の抽出

ユーザが注目状態であるときの視界画像に映っているオブジェクト領域を抽出する。本研究では、美術館のような環境で展示されている絵画をオブジェクトの一例として、注目時の視界画像に映っている絵画領域を抽出する方法を述べる。ただし、絵画の鑑賞はほぼ正面から行うものとする。

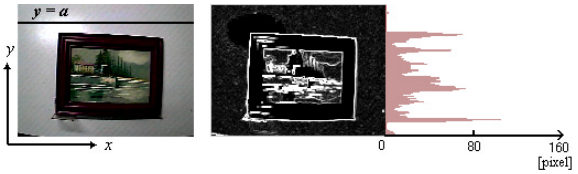
絵画が映っている視界画像において、壁と額縁の境界で水平方向、垂直方向に直線のエッジが現れる。また HSV 表色系の色相において、壁と額縁の境界で色相が変化する。このことを利用して、直線のエッジと色相の境界を基に、視界画像から絵画領域を抽出する。

まず、視界画像の濃淡画像から roberts フィルタによりエッジ画像を生成し、水平方向、垂直方向に現れる直線のエッジを検出する。さら

に、エッジ画像から水平方向、垂直方向のエッジの長さを算出する。例えば、水平方向に現れる直線のエッジを検出する場合、$y = a$（$0 \leqq a \leqq 119$）において、エッジとなる画素が連続している最大の数をエッジの長さとする。図 6(b)右のグラフは、図 6(a)の視界画像における水平方向のエッジの長さを表示したものである。この例では、求めたいエッジはほぼ水平に直線となるエッジであるので、$y = a$ においてある画素の上下($y = a \pm 1$)に位置する画素のいずれかがエッジ成分であれば、その画素もエッジ成分とした。垂直方向に対しても同様の処理を行い、垂直方向の直線のエッジを検出する。

次に、隣り合うエッジより長いエッジを検出し額縁のエッジ候補とする。額縁の境界として現れるエッジを検出するために、額縁のエッジ候補に隣接する画素の色相のヒストグラムから色差を算出する。例えば、水平方向の境界となるエッジを検出するために、額縁のエッジ候補を境界にして上下に 1 画素ずらした領域のヒストグラム $hist_{an}$, $hist_{bn}$ (n=1,...,7：色相の等級)を求め、色差 CS_e を式(2)により算出する。同様の処理を垂直方向にも行い、垂直方向において境界となるエッジを検出する。色相の等級数を 7 とすれば、視界画像に映っている壁、額縁、絵画の境界で色相が大まかに分かれることを実験により確認している。また、色相を 7 等級に分割する際に、360 等級の色相のヒストグラムにおいて、頻度が最大となる色相を中心に±25 ° の範囲を一等級とし、その一等級の範囲を基準として他の範囲を 6 等分した。CS_e の値が大きい上位の直線エッジを水平方向、垂直方向において 2 本ずつ検出し、これらの 4 本の直線エッジで囲まれた領域を、視界画像における絵画領域とする。

$$CS_e = \sum_{n=1}^{7} \left| hist_{an} - hist_{bn} \right| \qquad (2)$$



(a) 視界画像　　(b) エッジ画像と水平方向のエッジの長さ

図 6. 視界画像に映る絵画とエッジ画像に現れる直線エッジ

5 注視点分布に基づく注釈情報の提示

注視点分布に基づいて、オブジェクトにおける注目領域を判別し、注目領域に適応させた注釈情報の提示を行う。

5.1 注目領域の判別

注目時の視界画像から絵画領域を抽出した後に、一注目動作の注視点を含む最小凸多角形領域と、データベース中の注釈付加領域の位置情報を基にして、以下の手順により注目領域を判別する。

1. 視界画像から絵画領域を抽出した後、絵画領域の左下を基準として最小凸多角形領域の各頂点までの水平方向、垂直方向の距離を求め、絵画上での位置を算出する。
2. データベース中の絵画についても、絵画領域の左下を基準として注釈付加領域である凸多角形の各頂点までの水平方向、垂直方向の距離を算出する。
3. データベース中の絵画領域を基準とし、視界画像から抽出した絵画領域の大きさを正規化した後に、手順 1 で算出した絵画上での最小凸多角形領域の各頂点について、正規化後の位置を算出する。
4. 視界画像の絵画領域における最小凸多角形領域と、データベース中の各注釈付加領域との重なりが、最大となる領域の注釈情報を表示する。

ただし、注目領域と判断された注釈付加領域の 50%以上が、一注目動作の注視点から得ら

れる最小凸多角形領域に含まれていない場合は、ユーザはその注目動作において、いずれの注釈付加領域も見ていないと判断し、注釈情報は何も提示しない。

5.2 注釈情報の提示方法

注目領域を判別すると、図 7 のように注目領域に適応させた注釈情報を、ノート型 PC の画面に提示する。図 7 左には判別した注目領域を囲い線で表示し、その領域に関する注釈情報を図 7 右に表示する。注釈情報がノート型 PC の画面上に提示されている間は、システムは眼球運動の解析を行わない。注釈情報を一定時間提示した後に、自動的に眼球運動の解析を開始する。このように、注目領域に適応させて注釈情報を提示することができる。

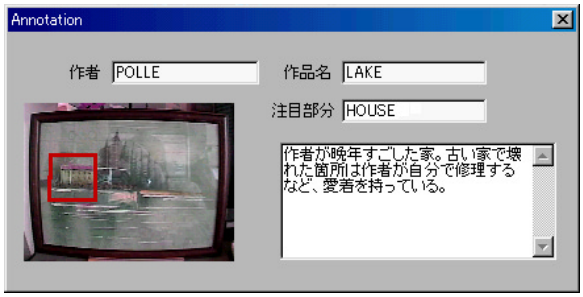


図 7. 適応的な注釈情報の提示

6 実験と評価

システムにより判別された注目領域が、実際に注目した領域であるかを評価するために実験を行った。A3 サイズの 20 種類の絵画を額縁に入れ、単一色の壁に掛けた状態で実験を行った。20 種類の絵画の内訳は、人物画や動物画が 10 枚、風景画が 10 枚である。各絵画を注目する際、システムにおいて被験者(4 人)に注目する絵画を選択してもらう。また、4 で述べた手法により、視界画像に映っている絵画領域の面積のうち、90%にあたる領域を抽出可能であることを、実験により確認している。

プロトタイプ使用時の様子を図 8 に、絵画抽出の様子を図 9 に示す。図 9 の上部 5 枚の画像において、下段右の画像の赤線の交点は注視点を表す。

図 8. システム使用時の様子

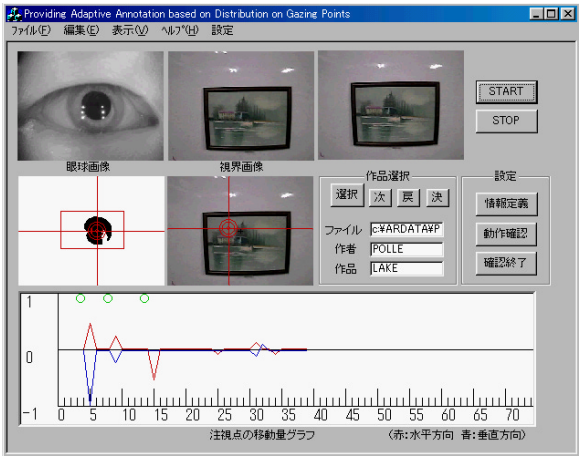


図 9. 視線と絵画の抽出

6.1 注目領域の判別精度

4 で述べた方法により視界画像から絵画領域を抽出し後に、5.1 で述べた方法により判別した注目領域が、実際に注目した領域であるかを評価した。20 種類の各絵画に対して、構成要素である人や建物などの領域を注釈付加領域として 3, 4 箇所定義しておく。その際に、領域の大きさから算出した視角が 6 ° 以上、領域間の最短距離から算出した視角が 1 ° 以上になるようにした。このように定義すれば、ユーザと絵画の距離が 1.0m の場合に、領域の判

別精度が 99%であることを実験により確認している。

各被験者に合計 60 箇所の注釈付加領域を 1 回ずつ注目してもらい、システムにより判別された注目領域と、実際に注目した領域が同じならば判別成功とした。被験者 4 人の結果の平均を表 1 に示す。

ユーザと絵画の距離が近ければ注目領域の判別精度は良くなっている。この理由として、距離が近ければ、絵画領域の抽出精度や、注視点の抽出上の精度が良いことが挙げられる。

判別に失敗した領域は、絵画の周辺に定義している注釈付加領域が多かった。この原因として、視界画像の周辺における注視点の抽出上の精度が、中心付近における精度と比べ低下することが考えられる。この注視点の抽出上の精度の低下は、眼球撮影画像のフレームサイズや解像度が十分でないために起こる。そのため、絵画の周辺に定義する注釈付加領域を大きくし、領域間の間隔を広げる必要がある。

表 1. 注目領域の判別精度

	ユーザと絵画の距離	
	1.0m	1.5m
判別精度	0.85	0.79

6.2 視覚情報注目時のユーザ状態の検出精度

本研究で提案した手法により、被験者の注目状態を正しく検出できているかを検証した。被験者(4 人)に絵画(20 枚)を 1 枚ずつ、1 分間鑑賞してもらう。実験開始から終了までの被験者の眼球撮影画像と視界画像をビデオで撮影しておく。ビデオを実験後に被験者に提示し、各絵画に対して注目した領域を申告してもらう。また、システムにより判断された注目領域に実際に注目したか否かの評価を行う。

実験結果を表 2 に示す。N_r は実際に被験者が部分的に注目していた状態の回数、N_s はシステムが部分的に注目していた状態と判断した回数、N_c はシステムが部分的に注目していると判断したもののうち、正しく判断された回数とし、$Recall=N_c/N_r$, $Precision=N_c/N_s$ とする。Nr は記録した眼球撮影画像と視界画像を被験者に提示して、各絵画において注目した部分を申告してもらい求めた。Nc はそれぞれの注目領域に対して、被験者に申告してもらった注目領域と一致しているか判断することにより求めた。

Recall を低下させた失敗原因の内訳は、注目状態の未検出が 44%となり、失敗原因のおよそ半分の割合を占めた。この値は、ユーザが注目する領域において、より長く注目することにより注目状態が検出され、*Recall* が向上すると考えられる。また、システムにより判別された注目領域は、79%の割合で実際の注目領域であり、本手法の有効性が示されたと考える。

表 2. ユーザ状態の検出の実験結果

$Recall(N_c/N_r)$	$Precision(N_c/N_s)$
0.71(89/125)	0.79(89/112)

7 まとめと今後の課題

本稿では、眼球運動からユーザの注目状態を検出し、そのときの注視点分布から注目領域を判別することにより、注目したときにのみ、注目領域に適応させて注釈情報を提示する方法を提案した。ユーザがオブジェクトに注目するという動作から、注目したオブジェクトに関する情報を提示することが可能となり、さらに、注目領域に適応させて注釈情報を提示することが可能となった。また、応用例として絵画を対象とした注釈情報の提示を行った。これにより、タグ付けすることが適切でないオブジェクトに対しても、注目領域に適応させて注釈情報を提示することができた。

今後の課題としては、視界画像から絵画領域を正規化する際に現れる、注視点の抽出上の精

度の影響を解決する必要がある。また、現在は視界画像に映っているオブジェクトはひとつだけとしているが、複数のオブジェクトが映っている場合に、注目領域に応じて注釈情報を提示できるようにすることが考えられる。

参考論文

[1] 小林元樹, 小池英樹, “電子情報の表示と操作を実現する机型実世界インターフェース「EnhancedDesk」”, インタラクティブシステムとソフトウェアⅤ:日本ソフトウェア科学会 WISS1997, pp.167-174, 1997.

[2] 椎尾一郎, 増井俊之, 福地健太郎, “FieldMouse による実世界インタラクション”, インタラクティブシステムとソフトウェアⅦ:日本ソフトウェア科学会 WISS1999, pp.125-134, 1999.

[3] 青木恒, “カメラで読みとる赤外線タグとその応用”, インタラクティブシステムとソフトウェアⅧ:日本ソフトウェア科学会 WISS2000, pp.131-136, 2000.

[4] 白井良成, 松下光範, 大黒毅, “秘映プロジェクタ：不可視情報による実環境の拡張”, インタラクティブシステムとソフトウェアXI:日本ソフトウェア科学会WISS2003, pp.115-122, 2003.

[5] T. Kurata, T. Okuma, M. Kourogi, T. Kato, and K. Sakaue, “VizWear: Toward Human-Centered Interaction through Wearable Vision and Vis-ualization”, The Second IEEE Pacific-Rim Conference on Multimedia, pp.40-47, 2001.

[6] T. Jebara, B. Schiele, N. Oliver, A. Pentland, “DyPERS: Dynamic Personal Enhanced Reality System”, M.I.T. Media Lab. Perceptual Computing Section Technical Report, No. 468, 1998.

[7] 池田光男, “眼はなにを見ているか”, 平凡社, 1998.