

複数人の視線計測に基づく「場の注意」の推定

竹村憲太郎[†] 松本 吉央[†] 小笠原 司[†]

[†] 奈良先端科学技術大学院大学 情報科学研究科 〒630-0192 奈良県生駒市高山町 8916-5

E-mail: [†]{kenta-ta,yoshio,ogasawar}@is.naist.jp

あらまし 本研究は対面式会議等の状況において複数人の注視情報を利用し、注意の集中する位置(場の注意)の推定手法を提案する。本手法では複数人の視線情報(3次元ベクトル)から幾何学的に3次元空間中の交点を求め、場の注意を推定する。その結果、注意を物体や人単位ではなく、連続的な3次元座標として推定することが可能となる。本稿ではしりとりや対面式会議等の状況において、場の注意を推定し、その有効性を確認した。

キーワード 注視情報, 注意, 会議

Estimation of Focus of Attention Based on Measurement of Gaze Information of Multiple People

Kentaro TAKEMURA[†], Yoshio MATSUMOTO[†], and Tsukasa OGASAWARA[†]

[†] Graduate School of Information Science, Nara Institute of Science and Technology

Takayama-cho 8916-5, Ikoma-city, Nara, 630-0192 Japan

E-mail: [†]{kenta-ta,yoshio,ogasawar}@is.naist.jp

Abstract In this paper, we propose a method to estimate the focus of attention of multiple people. It utilizes the gaze information as 3D vectors and estimates the position of focus of attention defined as the intersection of the multiple vectors. As the results, the position of the focus of attention can be represented as a continuous 3D position unlike other researches where only discrete objects can be the focus of attention. As experiments, the focus of attention at word-chain game and meeting situation is estimated and the feasibility of the method is confirmed.

Key words gaze information, focus of attention, meeting

1. はじめに

近年、ネットワークの発展によりビデオ会議システム等の研究が盛んに行われている。遠隔地とのコミュニケーションが容易に行えるように様々な研究が行われているが、その中で注視情報を共有しコミュニケーションを円滑に行えるようにサポートするシステムが注目されている。注視情報を伝える会議システムとして代表的なものに The GAZE Groupware System [1] がある。これは仮想空間上に人を配置し、注視対象に合わせて仮想空間上の人の向きを動かすことで注視情報を伝えるシステムである。また、コンピュータの前に座っているユーザ同士でアイコンタクトができるようなシステム [2] [3] も提案されている。複数人が同一空間に存在する対面式会議等の場合では、顔向きより会議中の注意の推定が行われて

いる [4]。会議中の注意の推定は竹前らが提案するようにビデオ編集、カメラワークなどへの応用が期待される [5]。しかしながら、これらの研究は注視の対象を人等に限定している。対話中に注目を集める物は様々であり、それらをすべて予想することや対象物の位置を取得することは困難である。そこで注意を物体や人単位で推定するのではなく、視線情報から3次元座標として連続的に推定する必要がある。注視位置の推定は、測定した注視情報(視線や頭部方向)を既知である平面や環境上に投影する手法 [6] や、注視を特定したい機器にそれぞれカメラを装着するなどの手法 [7] がある。一般にアイカメラによって計測される視線情報は角度情報であることから、注視点を求めるためにはこのような工夫が必要となる。しかしながら、これらの手法では環境マップを事前に作成する必要があることや、環境マップ上にない物体に対する

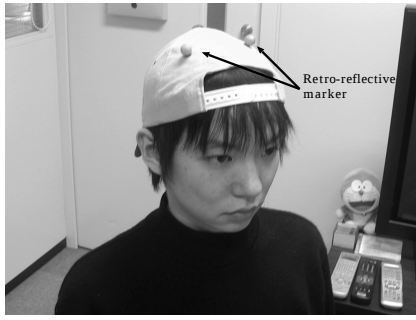


Fig. 1 Retro-reflective marker

注視を判定することが不可能である．また，輻輳角や頭部運動などを利用し，注視位置を推定する手法 [8] もあるが，一定時間注視している必要があり，移動する物体への適応が難しいなどの問題点がある．そこで本研究は複数人の注視情報を用いて，場で注意の集中する位置「場の注意」の推定を行う．複数人の注視情報から場の注意を求めることで，3次元座標として注意を推定することが可能となる．

以下，本報告は次のような構成となっている．第2節において場の注意の推定手法について提案する．第3節ではしりとり及び対面式会議中の注意の推定実験を行う．第4節では応用例として本手法によって得られた場の注意を映像の自動編集に適用する．最後に第5節でまとめと今後の課題を述べる．

2. 場の注意の推定

2.1 頭部情報の利用

頭部や視線方向の計測は盛んに行われ，画像処理による計測の精度も向上してきている．しかし，本研究では複数人の注視を計測するため，各人に視線計測システムを用意することは容易ではない．そこで複数人の動きがまとめて取り扱いやすい頭部方向のデータを使用し場の注意を推定する．なお，頭部方向を注視方向として取り扱うことの妥当性に関しては R.Stiefelhagen が頭部と視線の水平方向の向きに関しては 87%一致するというデータを示している [4]．

2.2 頭方向ベクトルの定義

本研究では Fig.1 に示すように頭部方向の測定を高帰性光学反射マーカを装着しモーションキャプチャシステム (VICON) を用いて行った．求められた座標から頭部方向を求める場合，正面方向は決定できるがピッチ方向の角度の調整が必要となる．ピッチ方向の調整は Fig.2 に示すようにマーカを装着し，一定時間キャリブレーション用のマーカを注視した際の頭部位置・方向及びキャリブレーション用マーカの位置データを用いて行う．

キャリブレーション用のマーカをランダムに動かし，そのときの頭部方向とキャリブレーション用マーカの距離に注目する．Eq.1 はピッチ方向の回転ベクトルを求

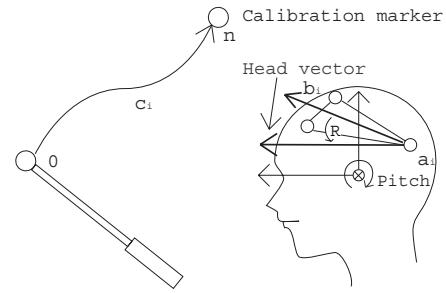


Fig. 2 Method of the optimization

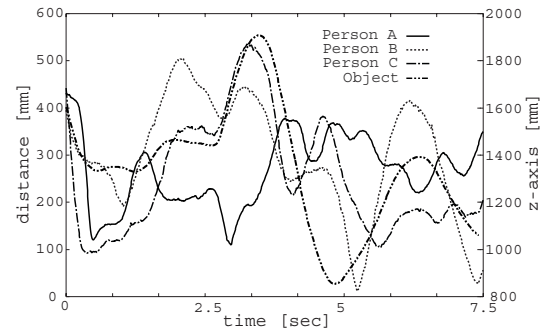


Fig. 3 Distance from head vector to object

めるための誤差関数であり D を最小にするような回転行列 R を最急降下法を用いて求める．

$$D = \sum_{i=0}^n \left\| \frac{(c_i - a_i) \cdot Rb_i}{\|Rb_i\|^2} Rb_i + a_i - c_i \right\| \quad (1)$$

誤差関数 D は一定時間 i の頭部ベクトルとキャリブレーション用マーカのユークリッド距離の和である． c はキャリブレーション用マーカの位置， Rb は頭部方向ベクトル，また a は頭部位置である．

頭部方向ベクトルの調整を行った後，ランダムに動く対象を注視する実験を被験者 3 人に対して行った．頭部方向ベクトルと注視対象との距離は Fig.3 に示ようになった．各人に対する平均誤差は 270mm 程度であった．注視対象の Z 軸座標と誤差の関係に注目すると人が座ったときの頭部の高さである 1400[mm] 付近に注視対象がある場合は比較的に誤差が小さいが，頭部の位置より注視対象が高い位置にある場合，誤差は大きくなる傾向がある．

2.3 場の注意の位置推定

本研究で提案する基本的なアイディアは Fig.4 に示すように注視対象が共有されている場合，頭部方向のベクトルの交差を考えることで，空間中の注意点を推定するというものである．複数人の注視情報を用いることで，従来必要とされていた平面や環境等の情報を利用せずに注視点を求めることができる．しかしながら，実際には 3次元空間において頭部方向ベクトルが交差することはほとんどない．そこで，各人の頭部方向ベクトルから注

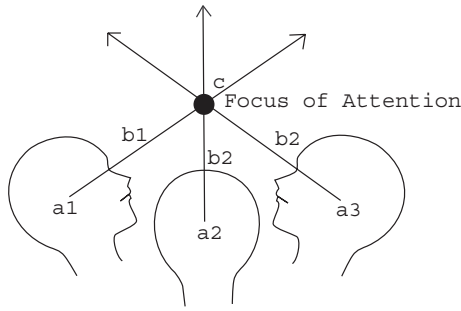


Fig. 4 Estimation of the focus of attention

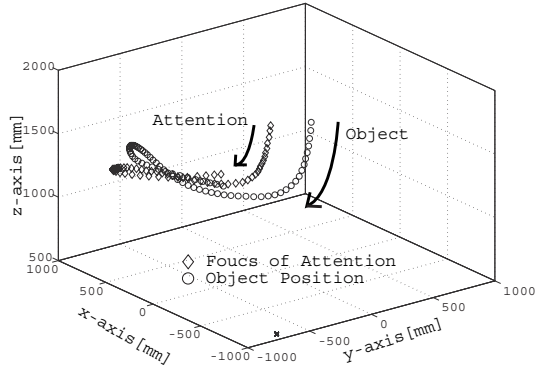


Fig. 5 Trajectory of the focus of attention

視点 c までのユークリッド距離を Eq.2 のように定義し、注意点 $\hat{c}$ は Eq.3 に示すように、最小自乗法を用いて求める。

$$\varepsilon_i = \left\| \frac{(c - a_i) \cdot b_i}{\|b_i\|^2} b_i + a_i - c \right\| \quad (2)$$

$$\hat{c} = \underset{c}{\operatorname{argmin}} \sum_i^n (\varepsilon_i)^2 \quad (3)$$

a は頭部位置、 b は 2.2 節で述べたようにピッチ方向の調整を行った頭部方向ベクトルとする。 ε_i は i 番目の人の注意点と頭部方向ベクトルのユークリッド距離であり、最小自乗法により求めた点 $\hat{c}$ を注意点として定義する。

3 人の被験者にランダムに動く対象を注視する実験を行い、精度評価を行った。そのときの注意点と注視対象の平均誤差は約 300mm となった。また、注意点と注視対象の軌跡を Fig.5 に示す。頭部方向のみの使用でも、注意点は注視対象の運動を十分に表現していることが確認できた。本手法を用いることで従来研究では難しかった空間上を移動する物体に対する注意を計測することが可能である。

2.4 ロバスト推定の利用

2.3 節にて提案した手法で問題となるのが、注意の共有が成立しない第 3 者への対応である。実際のコミュニケーションにおいては、全員が同一の物を常に注視するとは限らないため、この問題に対して考慮することは非常に重要である。2.3 節において述べた手法で推定を行うと異なる対象に注視を行っている人の情報にも大きく

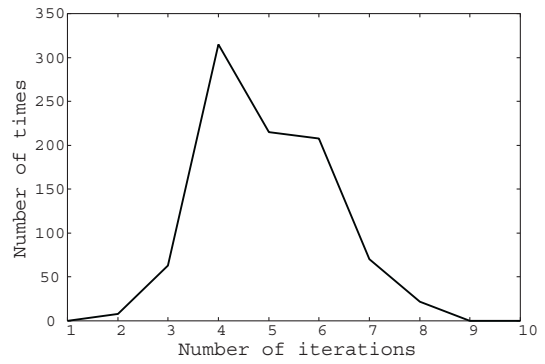


Fig. 6 Convergence criterion

影響を受けるため、求めた注意点と実際の注視対象物の位置の誤差が大きくなってしまふ。そこで本研究では異なる対象に注視を行っている人の注視情報の影響を少なくするため M 推定を用いる。 M 推定を用いると Eq.4 に示す様に、重み付き最小自乗問題として注意点を求めることができる [9]。

$$\hat{c} = \underset{c}{\operatorname{argmin}} \sum_i^n \left\{ \sqrt{\omega(\varepsilon_i^{(k-1)})} \cdot \varepsilon_i \right\}^2 \quad (4)$$

ω は重み関数であるが、本研究では Welsch の重み関数を使用する。 ε_i は Eq.2 にて示したように i 番目の人の頭部方向ベクトルと注視対象のユークリッド距離である。また、 k は繰り返し回数であり、 $k = 0$ の時、 $\varepsilon_i^{(0)} = 0$ を初期値として、重み $\omega(\varepsilon_i^{(k-1)})$ を更新し、注意点 $\hat{c}$ を求める。収束判定を行い繰り返し回数の決定を行うが、繰り返し回数 k は Fig.6 に示すように、それほど大きくないため画像処理技術を用いて頭部の位置を推定した場合にはフレームレートでの処理が十分に可能である。

M 推定を用いて、異なる対象を注視している人が存在する場合を想定し、シミュレーションを行った。ランダムに動く対象を注視する 3 人のデータに任意に設定したベクトルを加えて注意点を算出した。Fig.7 に示すように注視対象との距離が大きい任意のベクトルを加える。 M 推定を用いていない場合は任意のベクトルの影響を受け、誤差が大きくなっているが、 M 推定を用いている場合は、任意のベクトルによる影響があまりなく、安定した結果を出していることが分かる。

3. 対話中の場の注意

本稿では対話中の注意に注目し「しりとり」及び「対面式会議」中の注視行動の計測を行い、場の注意の推定を行った。

3.1 しりとり

Fig.8 に示すような状況で「しりとり」を行った際の場の注意を推定した。推定した注意点の軌跡は Fig.9 に示す通りである。実験結果は予想されるように、注意点は順番・方向を示すように円を描いた。また、テーブル中央付近を原点とした座標系において注意点の方向は Fig.10 に

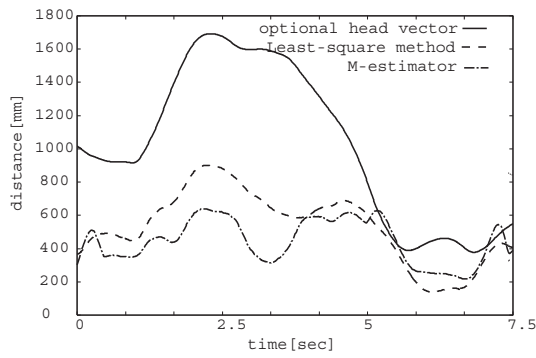


Fig. 7 Simulation of M-estimator



Fig. 8 The situation of word-chain game

示すように変化する．斜線や点線で囲まれた領域は発話区間であり，発話の順番は A から D となっている．発話区間は音声信号処理ツール wavesurfer [10] を用いてラベリングを行った．注意の方向が階段状に変化しているのは，発話の順番が決まっていることや，発話が行われる際に話者に対して注意点が停留することなどが理由として考えられる．各人の発話区間に注目すると，停留している区間の注意の方向は話者の方向とほぼ一致していることが分かる．例えば A の発話区間に注目するとどの区間も $40[\text{deg}]$ 付近に停留している．その他の人の場合も同様にある一定の角度で停留していることが分かる．また，区間 P のように発話より先に注意方向に変化が生じる場合や区間 Q のように階段状に変化せずに順番と逆方向への変化があった．区間 P のようなケースは「しりとり」のように発話の順番が決まっている場合，次に注意すべき対象を予測していることから生じると考えられる．区間 Q では他者の発話が観測されていることから，それによって一時的に注意が移ったことを表している．また，区間 R のように他者の発話が行われても注意点にほとんど変化のない場合もあった．

「しりとり」のように注意対象が人であり，位置の移動がない場合には注意点の停留に注目することで，Fig.11 に示すように注意対象の遷移を表すことが可能である．Fig.11 は停留状態を四角，遷移状態を円で表している．

3.2 対面式会議

Fig.12 に示すように実際のミーティングにおいても，

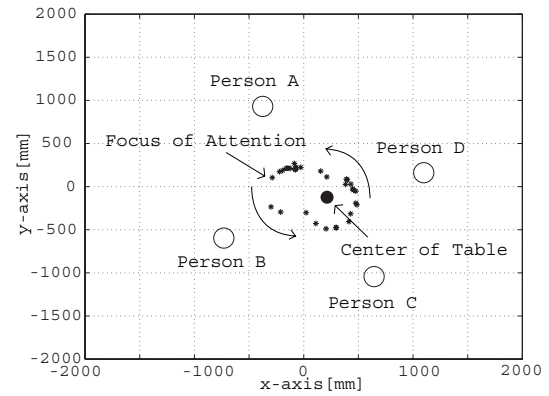


Fig. 9 Trajectory of the focus of attention in the word-chain game

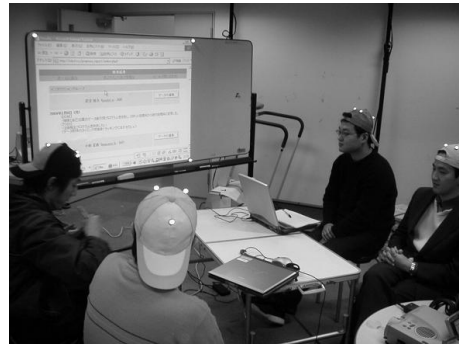


Fig. 12 The situation of meeting

場の注意の推定を行った．Fig.13 は注意の推定結果であり，左からホワイトボードへの注意，人への注意，その他の対象への注意となっている．右図では注意対象を特定することはできなかったが，全員の頭部方向から見ても明らかに共通の対象へ注意が行われている．このように予想範囲外の対象に対する注意も本手法では空間中の座標を求めることから考慮することができる．

3.3 考察

「しりとり」及び「対面式会議」において場の注意の推定実験を行ったところ，物体や人単位の注意の推定と異なり，予想範囲外の対象への注意も考慮することができた．また，複数人の発話が同時に起こっている場合には音声では注意の集中している箇所を特定することは困難であるが，複数人の注視情報を用いることからそれを実現した．

問題点としては，実際の対話状況では 2.3 節で行った評価実験と同等の精度で注意点を推定することができなかった．この理由として考えられることは，本システムは注意対象を点として取り扱っているために誤差が生じたと考えられる．特にホワイトボードへの注視の場合は，大きな誤差になってしまっている．また，ランダムに動く対象を注視する場合と異なり，注意すべき対象の位置が予想できるため，頭部の運動があまり生じなかったのではないかと考えられる．この問題に対しては視線計測

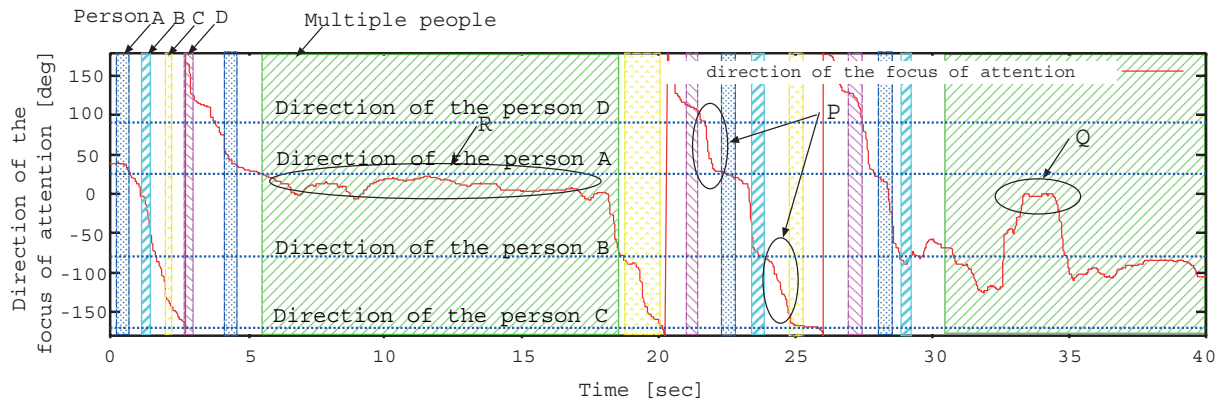


Fig. 10 Direction of the focus of attention

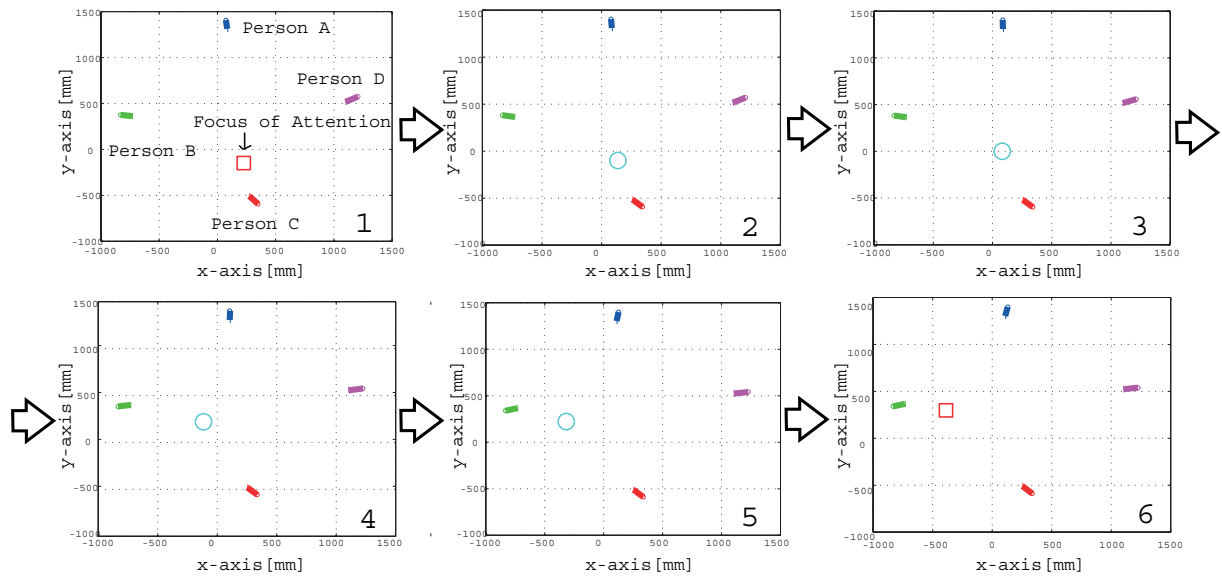


Fig. 11 Transition of the focus of attention

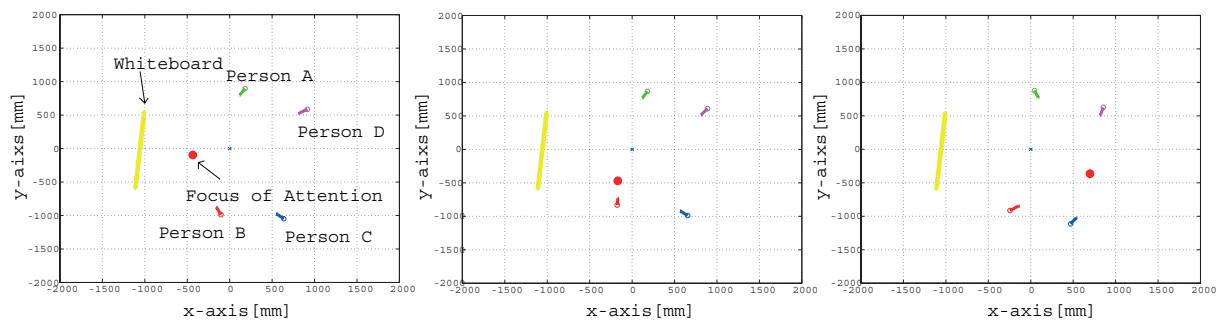


Fig. 13 Focus of attention in meeting situation

を用いると解決することができるため、今後取り組んでいきたいと考えている。

4. 場の注意に基づく映像自動編集

本稿では、場の注意の応用例として、映像の自動編集へ適用を行った。場の注意を捉えることで、状況把握しやすい映像への編集が可能である。従来研究ではシーンを自

動撮影する場合など、注目対象をあらかじめ設定する必要 [11] や音声情報などを利用し撮影が行われている [12]。しかし、本システムは注意点を捉えるようにカメラをコントロールするため、注意対象の設定の必要がないなどの利点がある。対面式会議等の状況において使用可能なカメラシステムは複数のカメラを用いる方法、アクティブカメラを使用する方法など様々なシステムがあるが、

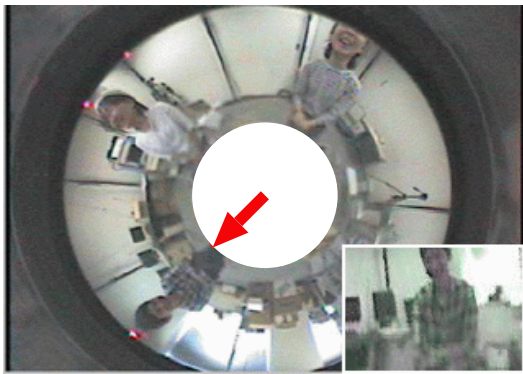


Fig. 14 Omnidirectional and panoramic Image

本研究では容易に複数カメラの切り替えやコントロールと同様の効果を得るために全方位カメラを用いて記録を行った。全方位カメラはテーブルの中央付近に設置した。注意点の位置情報は3次元の情報を持っているが、本研究ではX-Yの2次元の情報を映像に反映させる。本研究では注意の方向を画面中央に表示するように映像に反映した。注意点を映像に反映させた結果をFig.14に示す。全方位画像上の矢印は推定した注意点の方向であり、その方向を基準として展開を行った。評価として話者に注目したところ、展開画像の中央付近に話者が提示されるように画像が自動的にスクロールされた。

5. おわりに

本論文は複数人の注視情報を利用した注意点の推定手法について述べた。本手法では推定した注意点と注視対象の平均誤差は300[mm]程度となっている。また、M推定を用いて注意が成立しない者に対する考慮も行い、シミュレーションによってその効果を確認した。従来研究と異なり人や物体に対する注意ではなく、3次元座標として注意点を求めることを可能とした。これによって従来研究では難しかった空間上を移動する物体に対する注視や予想範囲外の対象に対する注意も考慮することが可能となった。実際に「しりとり」及び「対面式会議」中に4人の頭部運動から注意点を求めた。注意対象が動的でない場合には、注意点の停留に注目し注意の遷移を求め、表示することが可能である。推定した場の注意の応用例として映像に適用し自動編集を行った。発話と注意点の方向に注目したところ、注意点の方向は発話に非常に依存していて、話者は自動的に画面中央部分に表示されるようにスクロールされたことから、本手法の有効性を確認した。また、注視情報を用いたことから、新しい注意点に発話などよりもはやく切り替わるなどの効果があることが確認できた。

今後は頭部情報でなく視線情報も考慮し、注意点の精度向上を行う。注視対象が頭部位置よりも高い位置にある場合、誤差が大きくなったが視線計測を用いて、その

ような状況においても安定した結果が得られるように改良をしていく予定である。また、カメラシステムについてはミーティングに最適なシステムを検討していく必要がある。本研究では簡単化のため全方位カメラを用いたが、映像を記録・提示する場合はより高解像度であることが望まれるため、カメラシステムの再検討を行う。また、映像の提示方法に関しては自動スクロール以外の方法も比較検討し、臨場感や理解度に対する影響についても取り組んでいきたい。現在は、モーションキャプチャのデータを利用し、オフラインで計算、映像の再生を行っているが、今後は画像処理技術を応用し、非侵襲なシステムでのオンライン計測の実現に取り組む予定である。

文 献

- [1] R.Vertegaal. The gaze groupware system: Mediating joint attention in multiparty communication and collaboration. *ACM SIGCHI Conference on Human Factors in Computing Systems*, pages 294–301, May 1999.
- [2] R. Yang and Z. Zhang. Eye gaze correction with stereovision for video-teleconferencing. *Microsoft Research Technical Report*, 20:1–19, 2001.
- [3] R.Vertegaal, I.Weevers, C.Sohn, and C.Cheung. Gaze-2: conveying eye contact in group video conferencing using eye-controlled camera direction. *ACM SIGCHI Conference on Human Factors in Computing Systems*, pages 521–528, May 2003.
- [4] R. Stiefelwagen. Tracking focus of attention in meetings. *Proc. 4th International Conf. on Multimodal Interfaces*, pages 273–280, Oct 2002.
- [5] N.Mukawa Y.Takemae, O.Kazuhiro. Video cut editing rule based on participants' gaze in multiparty conversation. *Proc. of ACM Int. Conf. on Multimedia*, pages 303–306, 2003.
- [6] K.Takemura, J.Ido, Y.Matsumoto, and T.Ogasawara. Drive monitoring system based on non-contact measurement system of driver's focus of visual attention. *IEEE Intelligent Vehicles Symposium*, pages 581–586, June 2003.
- [7] J.S.Shell, T.Selker, and R.Vertegaal. Interacting with groups of computers. *Communications of the ACM*, 46(3):40–46, March 2003.
- [8] I.Mitsugami, N.Ukita, and M.Kidode. Estimation of 3d gazed position using view lines. *12th International Conference on Image Analysis and Processing*, pages 466–471, September 2003.
- [9] Z.Zhang. Parameter estimation techniques: A tutorial with application to conic fitting. *International Journal of Image and Vision Computing*, 15(1):59–76, January 1997.
- [10] S.Kare and B.Jonas. Wavesurfer - an open source speech tool. *Proc. of International Conference on Spoken Language Processing*, 4:464–467, October 2000.
- [11] M.Ozeki, Y.Nakamura, and Y.Ohta. Automated camerawork for capturing desktop presentations – camerawork design and evaluation in virtual and real scenes. *Proc. of 1st European Conf. on Visual Media Production*, pages 211–220, March 2004.
- [12] M.Onishi, T.Kagebayashi, and K.Fukunaga. Production of video images by computer controlled camera. *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, 2:131–137, December 2001.