

裸眼 3D ディスプレイのための複数人に対応した視聴者トラッキング A Multiple Face Tracking Technology for Glasses-Free 3D Displays

下山 賢一[†] 瀧本 崇博[‡]
Kenichi Shimoyama Takahiro Takimoto

1. はじめに

近年の 3D の普及は目覚ましく、「アバター」等に代表される 3D 映画がヒットし、家庭用の 3D テレビが多く販売されるようになってきている。現在、3D 映像を視聴する主流の方法はメガネを用いたものである。また、専用メガネを必要としない裸眼式の 3D テレビも発売され、より少ないユーザ負担で 3D 映像を楽しめるようになってきた[1]。しかし、裸眼式ディスプレイでは、図 1 に示すように立体視可能な範囲である「視域」が限定されてしまうという性質がある。そのため、視聴者はメガネ式とは異なりどこからでも 3D 映像が視聴できるわけではなく、自ら視域位置に移動する必要がある、その手間が煩わしいという問題がある。そこで、視聴者を検出し、その位置に視域を自動制御する「視聴者トラッキング技術」が必要となる。

上記問題に対する従来の解決策はいくつかある。たとえば濱岸ら[2]は、視域の中心にある視聴者が常に入るように視域の位置を制御した。この方法は視聴者が 1 人の時は問題ないが、複数人では誰か 1 人に視域が追従してしまうため、他の視聴者にとって映像が見づらくなる問題が生じる。また、[3]に代表されるような特殊な装置(スクリーンや特殊形状のディスプレイ等)を用いた視域を直接的に広げる手法は、複数人にも対応可能だが、ハードウェアの大幅な変更が必要である。

そこで本稿では、複数人の視聴者にソフトウェア的な処理のみで対応可能な「視聴者トラッキング技術」を提案する。以下では、まず視域の定式化と位置の制御方法について説明し、次に視聴者の立体視の良好度を表す満足度について述べる。さらに、満足度を用いた視域位置の評価基準について述べる。最後に、その評価基準に基づいて視域位置を最適化することを特徴とする“複数人に対応した視聴者トラッキング”手法を示す。

2. 視域

フラットパネル型の裸眼 3D ディスプレイでは、偏向光学素子としてレンチキュラシートやパララックスバリアを用いる。本稿ではレンチキュラシートを用いた場合について説明する。以下の説明において、ワールド座標系を図 2 に示すように、パネル面中央を原点とし、X 軸をパネル面水平右方向を正、Y 軸を鉛直上方向を正、Z 軸をパネル面法線方向表示面側を正として設定する。さらに N 視差の画像を表示するものとする。

2.1 視域の定式化

[†] 株式会社東芝 研究開発センター Corporate Research & Development Center, Toshiba Corporation

[‡] 株式会社東芝 デジタルプロダクツ&サービス社
プラットホーム&ソリューション開発センター Digital Products & Services Company, Toshiba Corporation

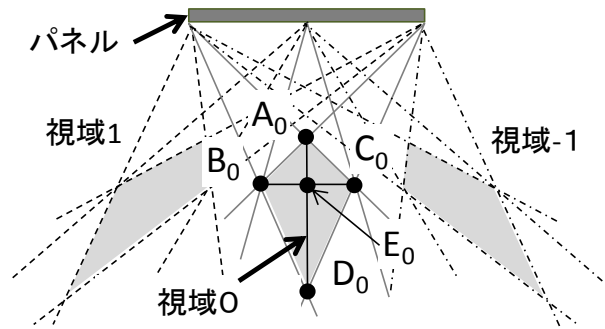


図 1 視域の様子

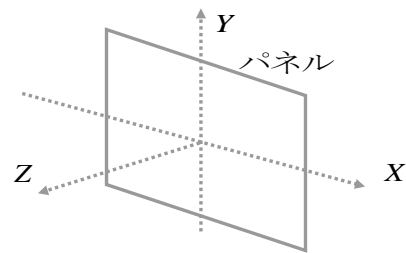


図 2 座標系

視域はパネル全面から全視差画像の光線が到達する範囲で図 1 のように複数存在する。その頂点位置はパネルのパラメータから幾何学的に計算することが可能である[4]。以下で、視域を形成する各頂点の座標を計算する方法について述べる。

具体的な計算方法について述べる前に、図 3 を用いて表示画素とレンズに関するパラメータについて定義する。図 3 上は、パネル上での表示画素とレンズの位置関係を示したもので、 $X=0$ 位置のレンズ番号を 0 とし左右に M 個のレンズが分布する様子を示す。 m 番のレンズ中心の X 座標を $L(m)$ と表す。さらに各レンズの後方には表示画素がそれぞれ N 個ずつ存在し、その左右端の画素中心の X 座標を $G_L(m)$ 、 $G_R(m)$ と表す。図 3 下は、1 つのレンズと対応する N 個の表示画素を拡大したものである。この N 個の表示画素群を要素画素と呼ぶ。 N は視差数でもある。

さて、レンズの幅を N_p 、要素画素の幅を P_p 、レンズ-画素間の空気換算距離(ギャップ)を g と表すと、 $L(m)$ 、 $G_L(m)$ 、 $G_R(m)$ は、

$$L(m) = N_p * m$$

$$G_R(m) = P_p * (m + \beta) = N_p * \alpha * (m + \beta)$$

$$G_L(m) = P_p * (m - \beta) = N_p * \alpha * (m - \beta)$$

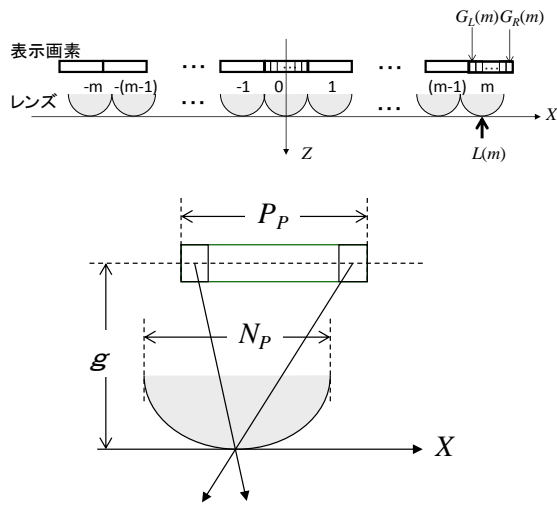


図3 表示画素とレンズと表現できる。ここで、 α と β は

$$\alpha = \frac{P_p}{N_p} = \frac{N * (L_0 + g)}{N_p * L_0}$$

$$\beta = \frac{1}{2} - \frac{P_p}{2} * j$$

で、 pp は画素の幅を示す。 L_0 は視域の一番幅の広い部分の Z 座標 (図1のB,Cの位置) であり、 j は逆視領域の幅を決める値である。クロストークの全くない理想的なパネルの場合 $j=0$ となり、クロストークの多いパネルの場合 j は大きくなる。

ここで、 m 番目のレンズと $m+k$ 番目の要素画素の左右端の画素が形成する光線の式を $f_{L,m,m+k}^R$ (L,R は左右端それぞれの式を表す) とすれば、

$$f_{L,m,m+k}^R(X) = g * \left\{ \left(\frac{1}{N_p} * \frac{1}{(1-\alpha)*m - \alpha*(k \pm \beta)} \right) * X - \frac{m}{(1-\alpha)*m - \alpha*(k \pm \beta)} \right\}$$

と表せる。したがって、 M 番目のレンズと対応する要素画素の両端の表示画素および、 $-M$ 番目のレンズと対応する要素画素の両端の表示画素の光線は、

$$f_{-M,-M-k}^R(X) = g * \left\{ \left(\frac{1}{N_p} * \frac{-1}{(1-\alpha)*M + \alpha*(-k + \beta)} \right) * X - \frac{M}{(1-\alpha)*M + \alpha*(-k + \beta)} \right\}$$

$$f_{-M,-M-k}^L(X) = g * \left\{ \left(\frac{1}{N_p} * \frac{-1}{(1-\alpha)*M + \alpha*(-k - \beta)} \right) * X - \frac{M}{(1-\alpha)*M + \alpha*(-k - \beta)} \right\}$$

$$f_{M,M-k}^R(X) = g * \left\{ \left(\frac{1}{N_p} * \frac{-1}{(1-\alpha)*M - \alpha*(-k + \beta)} \right) * X - \frac{M}{(1-\alpha)*M - \alpha*(-k + \beta)} \right\}$$

$$f_{M,M-k}^L(X) = g * \left\{ \left(\frac{1}{N_p} * \frac{-1}{(1-\alpha)*M - \alpha*(-k - \beta)} \right) * X - \frac{M}{(1-\alpha)*M - \alpha*(-k - \beta)} \right\}$$

となる。視域の各頂点 ABCD はこれら光線の交点であるから、視域 k の各頂点と対角線の交点の座標は、

$$\left. \begin{aligned} X_{Ak} &= N_p * \frac{\alpha(-k)M}{(1-\alpha)M - \alpha\beta}, & Z_{Ak} &= \frac{-Mg}{(1-\alpha)M - \alpha\beta} \\ X_{Bk} &= N_p * \frac{\alpha(-k + \beta)}{(1-\alpha)}, & Z_{Bk} &= \frac{-g}{(1-\alpha)} \\ X_{Ck} &= N_p * \frac{\alpha(-k - \beta)}{(1-\alpha)}, & Z_{Ck} &= \frac{-g}{(1-\alpha)} \\ X_{Dk} &= N_p * \frac{\alpha(-k)M}{(1-\alpha)M + \alpha\beta}, & Z_{Dk} &= \frac{-Mg}{(1-\alpha)M + \alpha\beta} \\ X_{Ek} &= N_p * \frac{\alpha(-k + \Delta)}{(1-\alpha)}, & Z_{Ek} &= \frac{-g}{(1-\alpha)} \end{aligned} \right\} \text{数式 1}$$

で計算することが可能である。なお、視域の位置を AD と BC の交点 E の位置で表すものとする。

2.2 視域制御

ここでは視域位置をソフト的に制御する方法について述べる。

2.2.1 視域の前後位置の制御

視域の前後位置 (Z 方向) の制御は [4] の方法を用いる。図4は要素画素のピッチを変化させることによって視域位置が前後に移動する様子を示している。レンズ幅 N_p に対して要素画素のピッチを大きくしていくと、パネル端で要素画素の位置がレンズに対し外側になる。その結果、光線の方向が内向きになり、視域は前方向に移動する。反対に、 P を小さくすると視域は後ろに移動する。具体的には、距離 L に視域を設定するには、要素画素のピッチ P_p を、

$$P_p = \frac{N * (L + g)}{L}$$

とする。数式 1 において $\alpha = P_p/N_p$ であるから、実際には α を制御すればよい。

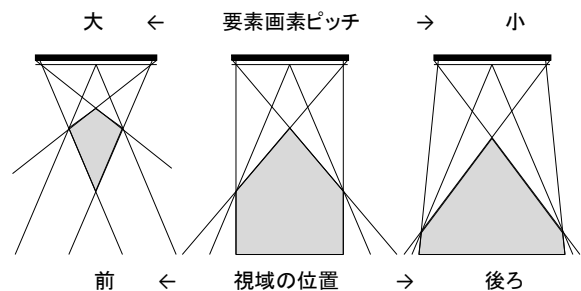


図4 視域の前後位置の制御

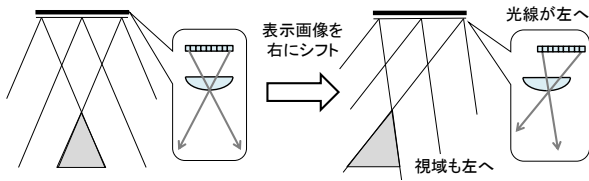


図 5 視域の横位置の制御

2.2.2 視域の横位置の制御

視域の横位置(X 方向)の制御は画像をシフトさせることによって行う。図 5 にその様子を示す。左図は右図に対して表示画素を右にシフトさせた状態を示す。表示が右にシフトした結果、光線が左へ向き、視域が左へと移動する。つまり、画像シフトの反対方向に視域は移動する。画像のシフト量は Δ と表す。

画像が Δ シフトしたときの m 番目のレンズと $m+k$ 番目の要素画素が形成する光線の式 f を、

$$f_{L,n,n+k}^R(X, \Delta) = g * \left\{ \left(\frac{1}{N_p} * \frac{1}{(1-\alpha) * m - \alpha * (k \pm \beta + \Delta)} \right) * X - \frac{m}{(1-\alpha) * m - \alpha * (k \pm \beta + \Delta)} \right\}$$

と表す。画像が Δ シフトしたときの M 番目のレンズと対応する要素画素の両端の表示画素および、 $-M$ 番目のレンズと対応する要素画素の両端の表示画素の光線は、

$$f_{-M,-M-k}^R(X, \Delta) = g * \left\{ \left(\frac{1}{N_p} * \frac{-1}{(1-\alpha) * M + \alpha * (-k + \beta + \Delta)} \right) * X - \frac{M}{(1-\alpha) * M + \alpha * (-k + \beta + \Delta)} \right\}$$

$$f_{-M,-M-k}^L(X, \Delta) = g * \left\{ \left(\frac{1}{N_p} * \frac{-1}{(1-\alpha) * M + \alpha * (-k - \beta + \Delta)} \right) * X - \frac{M}{(1-\alpha) * M + \alpha * (-k - \beta + \Delta)} \right\}$$

$$f_{M,M-k}^R(X, \Delta) = g * \left\{ \left(\frac{1}{N_p} * \frac{-1}{(1-\alpha) * M - \alpha * (-k + \beta + \Delta)} \right) * X - \frac{M}{(1-\alpha) * M - \alpha * (-k + \beta + \Delta)} \right\}$$

$$f_{M,M-k}^L(X, \Delta) = g * \left\{ \left(\frac{1}{N_p} * \frac{-1}{(1-\alpha) * M - \alpha * (-k - \beta + \Delta)} \right) * X - \frac{M}{(1-\alpha) * M - \alpha * (-k - \beta + \Delta)} \right\}$$

となる。したがって、画像が Δ シフトしときの視域の各頂点の座標は、

$$\left. \begin{aligned} X_{Ak} &= N_p * \frac{\alpha(-k + \Delta)M}{(1-\alpha)M - \alpha\beta}, Z_{Ak} = \frac{-Mg}{(1-\alpha)M - \alpha\beta} \\ X_{Bk} &= N_p * \frac{\alpha(-k + \Delta + \beta)}{(1-\alpha)}, Z_{Bk} = \frac{-g}{(1-\alpha)} \\ X_{Ck} &= N_p * \frac{\alpha(-k + \Delta - \beta)}{(1-\alpha)}, Z_{Ck} = \frac{-g}{(1-\alpha)} \\ X_{Dk} &= N_p * \frac{\alpha(-k + \Delta)M}{(1-\alpha)M + \alpha\beta}, Z_{Dk} = \frac{-Mg}{(1-\alpha)M + \alpha\beta} \\ X_{Ek} &= N_p * \frac{\alpha(-k + \Delta)}{(1-\alpha)}, Z_{Ek} = \frac{-g}{(1-\alpha)} \end{aligned} \right\} \text{数式 2}$$

となる。

以上より、視域位置はパラメータ α や Δ によって制御できることを示した。 α は Δ は画像表示をソフト的に変更させるだけのパラメータである。視域を位置 J に設計するパラメータ α と Δ の組を S_J と表記する。

3. 複数人トラッキングアルゴリズム

第 2 章で述べた方法を用いれば、視域をある位置 J に設定すること自体は難しいことではない。また、図 1 に示すように視域は複数存在するため、複数人の視聴者に対してそれぞれ視域を割り当てていけばいいように思われる。しかしながら、数式 2 などからわかるように、全ての視域は同じ距離にしか設定できないし、常に同じ間隔にしか配置されない。そのため、視聴者個別に視域を割り当てることは困難であり、複数人の視聴者に対して最適な視域を設定するという問題は非常に難しいものとなっている。本章では、この問題に対する解決策である“複数人に対応した視聴者トラッキング”を提案する。

図 6 に、視聴者トラッキング技術の処理の流れを示す。まずカメラを用いて視聴者を検出し、その視聴者位置を求める。次に、その視聴者位置に最適な視域位置を決定し、決定した視域位置に合わせた画像を表示する。

ここで重要になるのは、“最適な視域位置”をどんな基準で決定するかということである。そこで我々は、視聴者がどの程度良好な立体視ができていのかを数値化し、視域位置を評価する方法を提案する。以下では、処理の流れに沿って提案アルゴリズムについて述べていく。



図 6 処理の流れ

3.1 視聴者位置

視聴者位置の取得方法について述べる。視聴者位置はディスプレイ付近に設置されたカメラを用いて行う。カメラ画像から顔を検出・追跡しその顔位置を視聴者位置とする。顔検出には三田らの[5]の特徴量を用いた、この手法は、局所的な明暗パターンの共起性を特徴量(Joint Haar-like 特徴量)として識別を行い、明度変化に頑強であるという特徴をもつ。図 7 に顔検出の結果(点線四角)と特徴量の様子を示す

カメラ画像から実空間への座標変換は用いたカメラによって異なる。ステレオカメラを用いた場合は 3 角測量の原理によって求めることができる。また、単眼カメラ



図 7 Joint Harr-like 特徴を用いた顔検出

の場合、カメラ画像上の顔サイズから Z を推定し、 Z から X, Y を計算することにより視聴者位置を取得することができる。

3.2 視域位置の最適化

最適な視域位置を決定するには、その視域位置を評価する必要がある。評価は各視聴者の立体視の良好度合いを数値化することによって行う。各視聴者の立体視の良好度合いを満足度と呼ぶ。

満足度は、①立体視可能なパネル面積や②光線の密度、などによって求める。全ての評価基準を[0-1]の範囲で表し、満足度をこれらの積で求めることにする。

まず、①は立体視可能なパネル面の割合によって評価する。位置 P から視聴したときのパネル上の正視・逆視の分布は、視域頂点の座標を利用して求めることができる。図 8 にその様子を示す。位置 $P(X, Y, Z)$ とパネルの両端 P_L, P_R を結ぶ 2 本の直線が、 BC の延長線から切り取る線分 $B'C'$ 上の視域内・視域外の分布と、パネル上の正視・逆視の分布は一致する。これを利用して P から正視可能なパネルの面積を求めることができる。具体的には、 $B'C'$ 上にある視域の長さ割合が正視度になる。視域 J の条件下における位置 P での正視度を $H(P, S_J)$ と表す。正視度 H は、 B', C' と数式 2 を用いて視域 k の B_k, C_k 座標を求めることによって計算することができる。

次に、②について述べる。視差が一定の画像を観察するとき、視聴者位置の光線密度は立体感に大きな影響を与える。光線密度とは、図 9 に示すような各視差画像の光線が空間上でどれくらいの密度で存在するのかわかるものである。

右図のように光線密度が十分である場合、すなわち光線間隔が眼間幅程度以下である場合、視聴者の左右眼には異なる視差画像が提供されるため、立体感が得られる。しかし、左図のように密度が疎である場合、すなわち光線間隔が眼間より大きな場合、視聴者の左右眼に異なる視差画像を提供することが難しくなり、立体感が低下する。

そこで、位置 P における光線密度による評価値 $D(P)$ は、眼間距離を d 、1 つの要素画素が 1 つのレンズを通る時の光線の広がり角度を 2θ とすると、

$$L_\mu = \frac{d * N}{2 \tan \theta} * \mu$$

$$D(P) = \begin{cases} 1 & \text{if } Z < L_\mu \\ \frac{L_\mu}{Z} & \text{otherwise} \end{cases}$$

となる。 L_μ は光線間隔が眼間 d の μ 倍になる距離を示す。 μ はパネルによって異なるが、経験的に 2~3 の値を設定するのが良い。

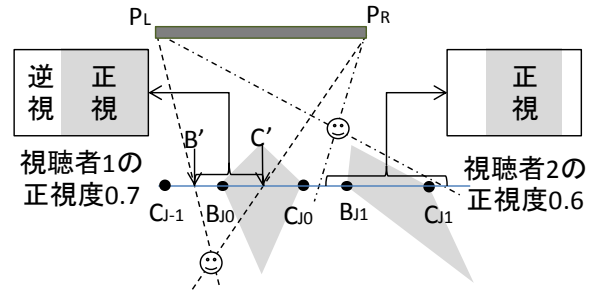


図 8 正視度

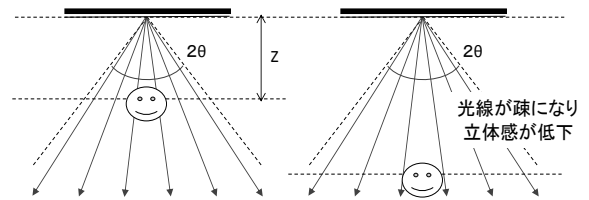


図 9 光線密度

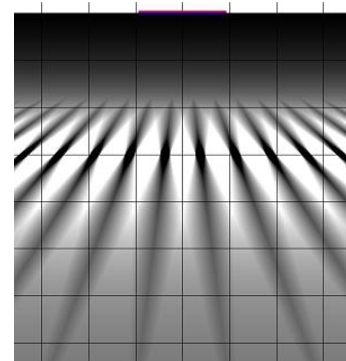


図 10 満足度の分布の様子

以上より、視域 J に関する位置 P における満足度を、

$$w(P, S_J) = H(P, S_J) * D(P) \quad \text{数式 3}$$

で計算する。図 10 は P を変化させたときの w の分布の様子を示した満足度マップの一例である。マップ中の明るい部分ほど満足度が高い。

3.3 最適な視域位置

数式 3 を用いると視域 J の評価が可能となる。視域 J の評価は、全視聴者それぞれ位置 P_i における満足度の総和で求める。 Q を全視聴者の位置の組 $Q = \{P_1, P_2, \dots, P_I\}$ とし、視域 J の評価値を $W(Q, S_J)$ と表せば、

$$W(Q, S_J) = \sum_{P_i \in Q} w(P_i, S_J) \quad \text{数式 4}$$

となる。したがって、視聴者配置 Q のときの最適な視域パラメータ $S_{J_{best}}$ は、数式 4 を最大とする S_J である。

$$S_{J_{best}} = \arg \max_{S_J} (W(Q, S_J)) \quad \text{数式 5}$$

数式 5 は S_j の全探索、すなわち α と Δ を適当に量子化して解を求める。量子化はパネルが表示可能なステップ等に合わせた方がよい。

3.4 アルゴリズム

我々の提案する複数人に対応した視聴者トラッキングアルゴリズムは以下になる。

- (1) カメラを用いて視聴者の顔を検出し、実空間上での座標を求める。
- (2) 全視聴者配置 Q と、設定可能な視域パラメータ S_j を用いて数式 5 を解き S_{jbest} を求める
- (3) パラメータ S_{jbest} にしたがって画像を表示する。

図 1 1 に 3 人の視聴者に対して視域制御を行った様子を示す。制御前は 1 人しか視域内にいなかったが、制御によって 3 人とも視域内に収まっていることがわかる。

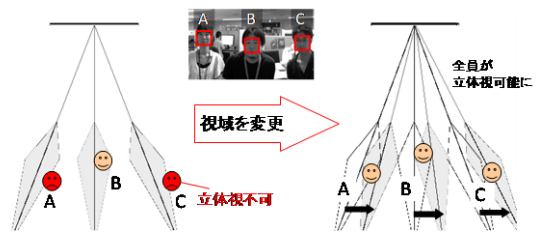


図 1 1 3 人の視聴者の視域を制御する様子

4. 評価実験

提案手法と従来手法[2]の比較実験を行う。提案手法は 3.4 で述べたものである。比較はシミュレーション実験によって行う。

4.1 実験条件

シミュレーションによって性能を検証する。以下の 2 つの実験を行った。

- <実験 1 >
- 50 インチ、9 視差、4K2K(3840x2160)、 $g=2\text{mm}$ 、 $j=3.0$ のパネル
 - 水平画角 60° のカメラ画像内、 $Z=[1.5\text{M}\sim 4\text{M}]$ の範囲に 5 人の視聴者をランダムに配置
 - 他の視聴者との距離は 30 cm 以上離れている
- <実験 2 >
- 30 インチ、6 視差、HD(1920x1080)、 $g=2\text{mm}$ 、 $j=2.0$ のパネル
 - 水平画角 60° のカメラ画像内、 $Z=[1.25\text{M}\sim 3\text{M}]$ の範囲に 3 人の視聴者をランダムに配置
 - 他の視聴者との距離は 30 cm 以上離れている

という条件の下、それぞれの手法で視域位置の最適化を実行する。図 1 2 は、実験 1 においてランダムに視聴者を配置した様子を示す。評価は 10000 回のランダム試行において、全員の満足度の平均と 2 人目以降それぞれの満足度の平均で行う。

4.2 結果と考察

<実験 1 > <実験 2 > の結果を、それぞれ表 1 と表 2 にまとめる。

表 1 から、従来手法では全員の平均が 0.55 であったのに対し、提案手法では 0.83 と 52% も見えが改善されることが分かった。特に 3 人目では提案手法が 0.92 と非常に高い値を維持しているのに対して従来手法は 0.53 程度でしかないことが分かる。さらに、表 2 から、従来手法では全員の平均が 0.57 であったのに対し、提案手法では 0.88 と 54% も見えが改善されることが分かった。特に 2 人目では提案手法が 0.97 と非常に高い値を維持しているのに対して従来手法は 0.49 程度でしかないことが分かる。以上より、提案手法は多様な設定のパネルに対して従来手法よりも優れた視域位置制御方法であることがわかる。

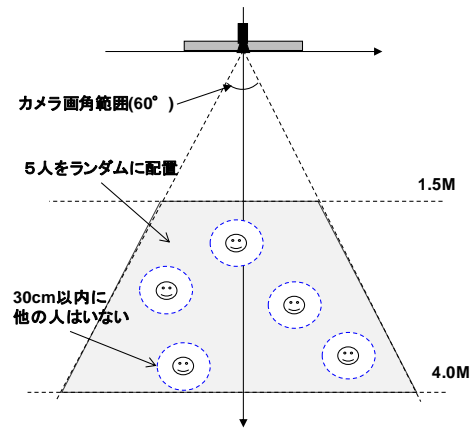


図 1 2 シミュレーション条件
視聴者をランダムに配置する様子

提案手法の実験 1、2 の結果を比較すると 2 人目までは同程度の性能だが、実験 2 のほうが 3 人目の結果が悪い。これはパネル性能が異なることに起因する。実験 2 のパネルのほうが実験 1 のパネルに比較して視域のサイズが小さいことが原因と思われる。

また、実験 1 では 4 人目以降、実験 2 では 3 人目の結果が他と比べると悪い。これは、人数が増えていくと視聴者の散らばりが大きくなり、どのようなパラメータの組み合わせで視域を設定しても、全ての視調者を必ずしも視域内に収められないことを示唆している。その原因は第 3 章の冒頭でも述べたように、視域は一見して複数あるように思われるが、全てが周期的に連動して移動することにある。

表 1 実験 1 の結果

	5人の平均	1人目	2人目	3人目	4人目	5人目
提案手法	0.83	1.00	0.98	0.92	0.76	0.50
従来手法	0.55	1.00	0.74	0.53	0.33	0.13

表 2 実験 2 の結果

	3人の平均	1人目	2人目	3人目
提案手法	0.88	1.00	0.97	0.66
従来手法	0.57	1.00	0.49	0.21

5. まとめ

我々の提案する“複数人に対応した視聴者トラッキング”手法は、視聴者の立体視の良好度を満足度として数値化することにより、最適な視域を客観的な評価指標で選択することを特徴とする。これにより、従来は難しく

った裸眼式 3 D ディスプレイの複数人視聴者の同時視聴に関して、提案手法を用いるとハードウェアの変更なく、複数の視聴者が同時に快適な 3 D 映像を視聴することが可能になることを示した。また、シミュレーション実験により提案手法は従来手法よりも 52~4% も性能が高いことを示した。

参考文献

- [1]伊藤 眞一,他,“先進の 3D テレビ グラスレス 3D レグザ(REGZA)”, 東芝レビュー, 66 巻 5 号,(2011 年 5 月号)
http://www.toshiba.co.jp/tech/review/2011/05/index_j.htm
- [2]濱岸 五郎,坂田 政弘,山下 敦弘,増谷 健,井上 益孝,志水 英二, “ヘッドトラッキング搭載 15 型高精細メガネなし 3D ディスプレイの開発”, 電気学会論文誌. C, 電子・情報・システム部門誌, 921-927, 2001-05-01 電気学会
- [3]吉田俊介, 矢野澄男, 安藤広志, “周囲から観察可能な光線野再生型テーブル様裸眼立体ディスプレイ”, 3 次元画像コンファレンス 2010, pp.51-54, 2010.7.
- [4]Rieko Fukushima, Kazuki Taira, Tatsuo Saishu, Yuzo Hirayama, “Novel viewing zone control method for computer-generated integral 3-D imaging”, Proc. SPIE 5291, 81 (2004)
- [5]三田 雄志, 金子 敏充, 堀 修. “顔検出に適した共起に基づく Joint Haar-like 特徴”. 電子情報通信学会論文誌 D-II Vol.J89-D-II No.8 pp.1791-1801, 2006