

テレビ会議における対話者間の視線一致のための 目領域変換手法に関する検討

井上 卓弥[†] 高橋 友和[†] 平山 高嗣[†] 出口 大輔^{††} 井手 一郎[†]
村瀬 洋[†] 黒住 隆行^{†††} 柏野 邦夫^{†††}

[†] 名古屋大学 大学院情報科学研究科 〒464-8601 愛知県名古屋市千種区不老町

^{††} 名古屋大学 情報連携統括本部 〒464-8601 愛知県名古屋市千種区不老町

^{†††} 日本電信電話株式会社 NTT コミュニケーション科学基礎研究所
〒243-0198 神奈川県厚木市森の里若宮 3-1

E-mail: †inouet@murase.m.is.nagoya-u.ac.jp, ††takahashi@gifu.shotoku.ac.jp, †††ddeguchi@nagoya-u.jp,
†{hirayama,ide,murase}@is.nagoya-u.ac.jp, †††{kurozumi.takayuki,kashino.kunio}@lab.ntt.co.jp

あらまし 近年、ウェブカメラなどの普及により、映像を用いた遠隔コミュニケーションが手軽に行なえるようになってきた。一般にカメラはディスプレイの外部に設置されているが、ユーザはディスプレイに映った相手を見ながら対話するため、対話者間の視線が一致しない問題が発生する。本報告では、ディスプレイの上部に設置されたテレビ会議用カメラから得られた視線不一致画像を視線が一致するように画像処理により変換する手法を提案する。提案手法では、ディスプレイ上の対話相手の目を注視している画像とカメラを注視している画像を事前に撮影し、それらを用いて視線不一致画像の目領域のみを変換することで視線が一致した画像を生成する。

キーワード テレビ会議, 視線一致, 目領域変換

A Study on Image Transformation of Eye Areas for Synthesizing Eye-Contacts in Video Conferencing

Takuya INOUE[†], Tomokazu TAKAHASHI[†], Takatsugu HIRAYAMA[†],

Daisuke DEGUCHI^{††}, Ichiro IDE[†], Hiroshi MURASE[†], Takayuki KUROZUMI^{†††}, and Kunio
KASHINO^{†††}

[†] Graduate School of Informaion Science, Nagoya University
Furo-cho, Chikusa-ku, Nagoya-shi, Aichi, 464-8601 Japan

^{††} Information and Communications Headquarters, Nagoya University
Furo-cho, Chikusa-ku, Nagoya-shi, Aichi, 464-8601 Japan

^{†††} NTT Communication Science Laboratories, Nippon Telegraph and Telephone Corporation
3-1 Morinosato-Wakamiya, Atsugi-shi, Kanagawa, 243-0198 Japan

E-mail: †inouet@murase.m.is.nagoya-u.ac.jp, ††takahashi@gifu.shotoku.ac.jp, †††ddeguchi@nagoya-u.jp,
†{hirayama,ide,murase}@is.nagoya-u.ac.jp, †††{kurozumi.takayuki,kashino.kunio}@lab.ntt.co.jp

Abstract In recent years, the spread of Web cameras have facilitated video conferencing. Although a Web camera is usually located outside a display, the user looks at his/her opponent in the display, so there is a problem that they cannot make eye-contacts with each other. In this report, we propose a method for image transformation of eye areas for synthesizing eye-contacts. The proposed method captures eye images when a user is gazing at the camera and also those when actually talking to someone before the video conferencing, and transforms images of the eye areas to synthesize eye-contacts by using these images while video conferencing.

Key words video conferencing, eye-contacts, image transformation of eye areas

1. まえがき

近年、ウェブカメラなどの普及により、テレビ電話やテレビ会議などが容易に行なえるようになってきた。一般に、テレビ会議を行なう際に、ユーザはディスプレイに映った相手を見ながら会話するが、カメラはディスプレイの外部に設置されているため、対話者間の視線が一致しないという問題が発生する。この問題に関して、テレビ会議システムにおける対話者間の視線一致の重要性について調査研究がなされている [1]~[3]。片山ら [1] は、テレビ会議において視線が一致する環境と一致しない環境が会話の質に及ぼす影響について調査している。この調査では、会話の円滑性と深い関わりがあることに着目し、映像対話における会話の円滑性を評価する尺度として身振りの頻度やパターンを用いた。その結果、視線が一致しない環境では、身振りの頻度が実対話の頻度に比べ半分になり、会話の円滑性が低下した。一方、視線が一致する環境では、実対話と身振りの頻度やパターンが類似したことを報告している。このように、テレビ会議における対話者間の視線一致は、対話者とのコミュニケーションにおいて重要な役割を果たすことが分かっている。しかし、現在のテレビ会議システムではこの問題を解決するための機能が不足しており、対話者との視線を一致させる機能の実現が望まれている。

テレビ会議における対話者間の視線一致を実現する手法として、ハードウェアによる手法とソフトウェアによる手法が提案されている。ハードウェアによる手法として、Kollaritsらはハーフミラースクリーンを利用して視線を一致させる手法を提案している [4]。この手法は、装置が大きいことや、設置に時間がかかることが問題である。ソフトウェアによる手法として、複数カメラを用いる手法が提案されている [5]~[7]。例えば、Yangら [5] は 2 台のカメラから得られた画像を View Morphing [8] により合成することで視線が一致した画像を生成する手法を提案している。この手法は、複数のカメラを設置する必要があることや、2 台のカメラで撮影された顔画像間の特徴点を密に対応付けるために、様々なアピアランスの変化にロバストで高精度な特徴点抽出が必要であることが問題である。また、可視光カメラとデプスカメラの 2 台のカメラを利用している手法も提案されている [9]。一方、単一カメラと顔モデルを併用する手法も提案されている [10], [11]。この手法では、顔モデルに表情の変化が含まれていないため、その変化に対応した顔画像を生成することが難しい。

そこで我々は、顔モデルを用いず、単一のカメラのみを用いて、対話者間の視線を一致させることを目的とし、その達成のために、ユーザがディスプレイ上の対話相手の目を見ている顔画像をテレビ会議用カメラで見ている顔画像に変換する手法を検討している。本手法では、ユーザが対話相手に送信する映像を変換するため、双方が本手法を使用することにより、対話者間の視線一致が実現できる。具体的には、事前にカメラを注視した画像とディスプレイ上の対話相手の目を注視した画像を撮影し、それらの画像を用いて、ユーザがディスプレイ上の対話相手の目を見ていると判定した場合にのみ顔画像の目領域を



(a) テレビ会議用カメラの配置 0 度 (b) テレビ会議用カメラの配置 15 度

図 1 テレビ会議用カメラの配置における顔画像の例

表 1 テレビ会議用カメラの配置と視線の一致を感じる割合

テレビ会議用カメラの配置 (度)	0	5	10	15
視線の一致を感じる割合 (%)	90	73	19	4

換して視線が一致した顔画像を生成し、それ以外の場合には変換を行わずにそのまま出力する。ここでは、ユーザが対話相手の目を見ていない場合は画像を変換しなくても会話の円滑性には影響を与えないという仮定をおいている。

以下、2 節でテレビ会議用カメラの配置と視線一致の関係について述べ、3 節でテレビ会議における対話者間の視線一致のための目領域変換手法について述べる。次に、4 節で、提案手法の有効性について調査した実験と結果について述べる。最後に、5 節でまとめと今後の課題について述べる。

2. テレビ会議用カメラの配置と視線一致の関係

佐藤ら [12] はテレビ会議におけるカメラの配置が対話者間の視線一致に与える影響を調査した。その結果、テレビ会議用カメラの位置が正面から上方向に 2 度よりも大きくなると視線が一致していないと感じ、12 度よりも大きくなると実際の対話と比べてかなり不自然と感じることが分かった。佐藤らの調査は 1968 年に行われたものであり、現在のデバイスでも同じような結果が得られるかを確認するため、我々もテレビ会議用カメラの配置が視線一致に与える影響を調査した。本調査では、テレビ会議用カメラは被撮影者の正面から上方向に 0 度、5 度、10 度、15 度の 4 つの位置に配置した。図 1 に本実験で使用した画像の例を示す。これらを 11 人の被験者に対して提示し、「画像の人物と視線が一致しているか否か」という問いへの回答を得た。結果を表 1 に示す。テレビ会議用カメラの位置が高くなるにつれ、視線の一致を感じる割合は低下していき、10 度では多くの人が視線の不一致を感じ、15 度ではほとんどの人が視線の不一致を感じた。テレビ会議用カメラの配置が 15 度の場合は、22 型のディスプレイの上部にカメラを設置し、ディスプレイからユーザが 60cm 離れた状況に相当する。これはテレビ会議の一般的な利用状況と言える。つまり、視線の不一致を感じる状況は多く存在すると考えられる。

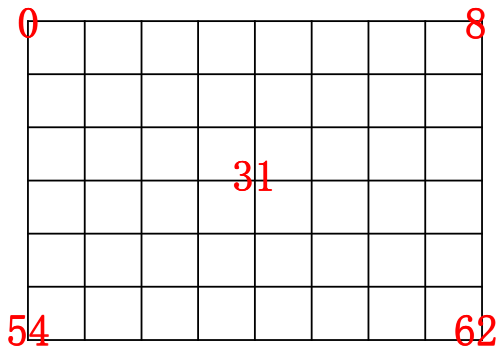


図 2 撮影に使用した格子点

12	27	35	31	35	65	31	12	8
12	27	62	58	62	50	27	23	8
27	23	65	73	77	62	50	15	4
27	46	77	92	92	88	58	23	4
35	58	85	100	92	92	58	31	12
31	54	81	92	96	92	73	31	12
27	46	65	73	85	73	35	8	8

図 3 実験結果 (視線の一致を感じる割合 [%])

3. テレビ会議における対話者間の視線一致のための目領域変換手法

日常的な会話において、ユーザが対話相手と視線を一致させようとしていない状況でも、視線方向が対話相手の目領域周辺に向いている場合、対話相手は視線が一致していると感じる場合がある。テレビ会議システムでもこのような場合を考慮して、視線が一致するように目領域を変換する必要がある。そこで、対話相手がユーザと視線の一致を感じる視線方向の範囲を調査する。その結果を踏まえて、目領域を変換するか否かを決定する目領域変換判定を用いたテレビ会議システムを提案する。

3.1 対話相手が視線の一致を感じる範囲の調査

本実験では、どれぐらいの範囲で視線がずれた場合に、対話相手が視線の一致を感じるか否かについて調査した。

3.1.1 実験条件

実験を行なうにあたり、データセットを作成した。被撮影者には顔を固定したまま視線だけを動かし、60cm 離れた位置にある 2cm 間隔で配置した 9×7 個の格子点を 0 番から 62 番まで順に見てもらいながら画像を取得した。図 2 に撮影に使用した格子点を示す。カメラは 31 番の位置に設置したため、31 番が正面に相当する。これらの画像を 15 人の被験者に対して提示し、「テレビ会議をしている想定で、画像の人物と視線が一致しているか否か」という問いへの回答を得た。

3.1.2 実験結果と考察

図 3 に実験結果を示す。図中の数字は視線の一致を感じた被験者の割合である。実際に相手の目を見ているのは黒枠で囲まれた部分であるが、相手が視線の一致を感じるのはそれよりも広い範囲であることが分かる。つまり、目領域の変換をある範

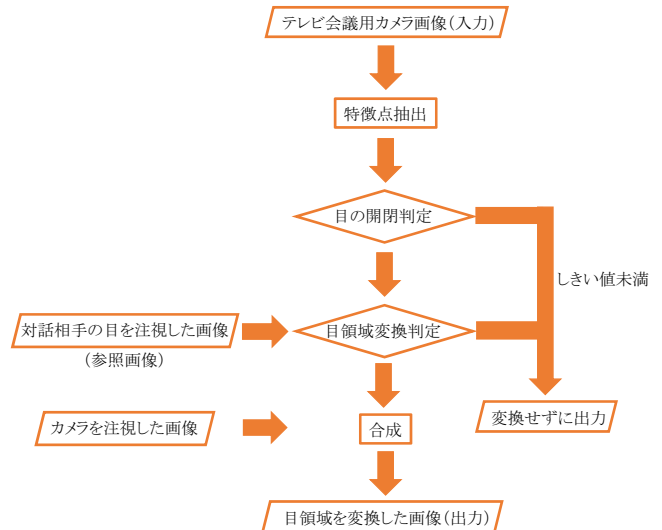


図 4 提案手法の処理の流れ

囲において行なう必要があると考えられる。

3.2 目領域変換判定を用いたテレビ会議システム

ここでは、対話者間の視線一致のための目領域変換手法について述べる。図 4 に処理の流れを示す。テレビ会議用カメラがディスプレイの上部に設置されている場合を想定し、事前にディスプレイ上の対話相手の目を注視している図 6 のような画像 (以降では参照画像と呼ぶ) とカメラを注視している図 7 のような画像を撮影する。提案手法は大きく分けて、(1) 特徴点抽出、(2) 目の開閉判定、(3) 目領域変換判定、(4) 目領域画像の合成の 4 つの処理からなる。以下、各処理について詳しく述べる。

3.2.1 特徴点抽出

本手法では、図 5 に示すような左右それぞれの目領域の輪郭部分から 8 点の特徴点を抽出する。ただし、特徴点の位置が連続するフレーム間で少しずつずれていると、動画像にした際に目領域がゆらぎ、不自然に感じることがある。この問題を避けるため、前フレームとの各特徴点の距離の 2 乗和 d を式 (1) に示すように定義し、この値がしきい値よりも小さければ、前フレームと特徴点の位置がほとんど変化していないとみなし、前フレームの特徴点の位置を使用する。ここで、 $\mathbf{x}^{(t)} = (x^{(t)}, y^{(t)})$ であり、 $x^{(t)}$ と $y^{(t)}$ はそれぞれ t 番目のフレームの特徴点の x 座標、 y 座標を示す。

$$d = \sum_{i=1}^8 \|\mathbf{x}_i^{(t-1)} - \mathbf{x}_i^{(t)}\|^2 \quad (1)$$

3.2.2 目の開閉判定

まばたきの際、目が完全に開ききっていない状態では、ユーザは特定の場所を注視していないと考えられる。また、目の開きが不十分な状態で合成すると不自然になる場合がある。そのため、参照画像を基準として、目の開閉度を計算し、しきい値よりも大きい場合は次の処理に進む。しきい値よりも小さい場合は変換を行わず、入力画像をそのまま出力する。

3.2.3 目領域変換判定

目領域変換判定は入力画像の目領域を変換するか否かを決定

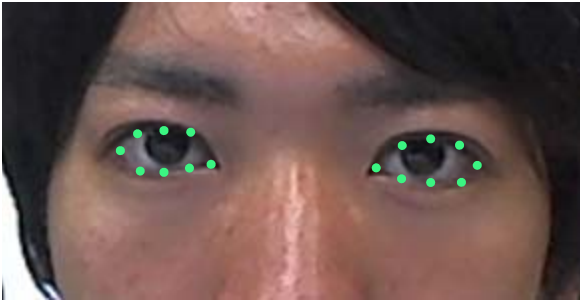


図 5 抽出した特徴点

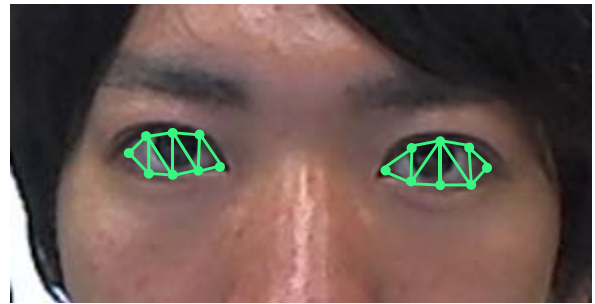


図 8 三角形分割の例



図 6 ディスプレイ上の対話相手の目を注視している画像



図 7 カメラを注視している画像

する処理であり、参照画像と入力画像の目領域のテクスチャの類似度を用いて行なう。類似度算出の際に三角形パッチ分割を用いることで、参照画像と入力画像の目の形が多少異なっても、その変形に対応することができ、判定を正確に行なうことができる。

三角形パッチ分割は、前の処理で抽出した図 5 に示す 8 点の特徴点を使用する。まず、8 点の特徴点を用いて、図 8 に示すように入力画像と参照画像中の目領域を三角形に分割する。次に、分割した三角形それぞれにおいて、入力画像の三角形をアフィン変換により変形し、参照画像中の対応する三角形の形状に合わせる。これにより、目の形状の変形に対応することができる。

類似度は、三角形パッチ分割により生成した画像と参照画像



(a) まぶたの内側



(b) まぶたの外側

図 9 変換する際に使用した特徴点

のそれぞれの目領域のみをグレースケールに変換し、正規化相互相関 (Normalized Cross-Correlation) を式 (2) に従って計算する。ここでは、参照画像の画素値を $f(\mathbf{x}_i)$ 、入力画像をアフィン変換した後の画素値を $g(\mathbf{x}_{i'})$ で表す。また、参照画像で類似度計算に用いる画素数を I とする。この類似度がしきい値よりも大きい場合は、目領域を合成する処理に進む。

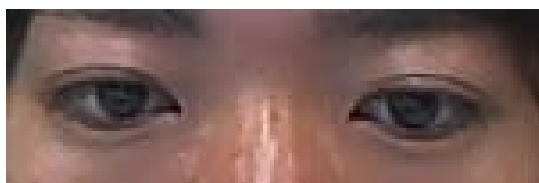
$$NCC = \frac{\sum_{i=1}^I f(\mathbf{x}_i)g(\mathbf{x}_{i'})}{\sqrt{\sum_{i=1}^I f(\mathbf{x}_i)^2 \sum_{i=1}^I g(\mathbf{x}_{i'})^2}} \quad (2)$$

3.2.4 合 成

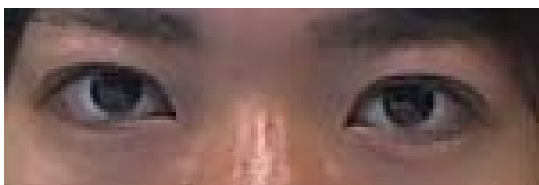
目の開閉度と類似度が共にしきい値よりも大きい場合、入力画像に対して事前に得たカメラを注視した画像の目領域を合成する。合成には 3.2.3 で述べた三角形パッチ分割を利用する。以下の実験では、変換する目領域として図 9(a) に示すようなまぶたの内側までを変換する方法と、得られた特徴点を移動させ、図 9(b) に示すようにまぶたの外側までを変換する方法を検討する。まぶたの外側より内側に特徴点を設定の方が抽出が容易であるが、図 10 に示すように、視線方向によってまぶたの見え方も変わるため、まぶたの領域を変換することが有効であると考えられる。

4. 実 験

提案手法の効果を評価するために行なった実験とその結果に対する考察について述べる。4.1 節では、画像間類似度を用いた目領域変換判定の精度について調査した結果について述べ、



(a) ディスプレイ上の対話相手の目を注視している目



(b) カメラを注視している目

図 10 視線方向に対するまぶたの見えの違い

0.97	0.97	0.97	0.97	0.97	0.97	0.94	0.95	0.93
0.96	0.96	0.98	0.98	0.99	0.97	0.97	0.96	0.94
0.96	0.97	0.97	0.98	0.99	0.97	0.96	0.97	0.94
0.96	0.97	0.97	0.98	1	0.98	0.96	0.96	0.93
0.95	0.96	0.98	0.97	0.99	0.98	0.97	0.97	0.96
0.96	0.96	0.97	0.98	0.97	0.98	0.96	0.95	0.94
0.95	0.96	0.94	0.97	0.96	0.97	0.97	0.95	0.94

図 11 実験結果 (目領域変換判定)

4.2 節では、特徴点をまぶたの内側に設定した方法と外側に設定した方法が合成顔画像の視線一致に及ぼす影響について調査した結果を述べる。

4.1 画像間類似度を用いた目領域変換判定の精度

本手法では、目領域変換判定に画像間類似度を用いている。本実験では、画像間類似度を用いた目領域変換判定の精度について調査した。

4.1.1 実験条件

実験を行うにあたり、データセットを作成した。被撮影者には顔固定したまま視線だけを動かし、60cm 離れた位置にある図 2 に示すような 2cm 間隔で配置した 9×7 個の格子点を 0 番から 62 番まで順に見てもらった。カメラは正面から上方向に 10 度の位置に配置した。31 番を注視している場合がディスプレイ上の対話相手の目を注視している場合に相当する。31 番を注視している目領域画像と 0 から 62 番を注視している目領域画像との類似度を算出した。ここでは、3.2 節で述べた類似度計算を用いた。

4.1.2 実験結果と考察

実験結果を図 11 に示す。図中の数字は 31 番を注視している画像と 0 から 62 番を注視している画像との類似度である。本手法では、この値を用いて、対話相手が視線の一致を感じているか否かを判定した。本結果より、類似度 0.98 以上で対話相手の目を注視した場合の周囲で目領域を変換できていることが分かる。

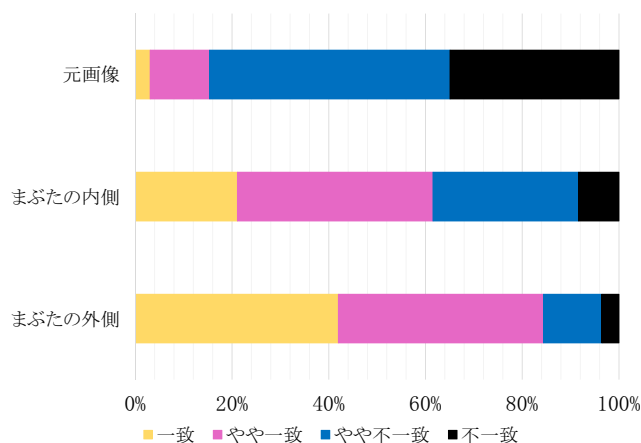


図 13 実験結果 (特徴点の位置)

4.2 特徴点をまぶたの内側に設定した方法と外側に設定した方法が視線一致に与える影響

本実験では、特徴点をまぶたの内側に設定するか外側に設定するかが、合成した顔画像の視線一致に与える影響を調査した。

4.2.1 実験条件

実験を行なうにあたり、データセットを作成した。解像度が 1920×1200 である 24 インチのディスプレイの上部にウェブカメラを設置し、ディスプレイからユーザが 50cm 離れたところで 640×480 画素の解像度で撮影をした。まず、対話相手の目を注視した画像とカメラを注視した画像を事前に撮影した。その後、対話相手の目を注視する場合、対話相手の目から視線を外す場合を含む動画を撮影した。本実験では、変換する目領域として図 9(a) に示すようにまぶたの内側までを変換する方法と、図 9(b) に示すようにまぶたの外側までを変換する方法を用いた。撮影した動画の各フレーム画像の目領域を変換し、その中の 105 枚の画像を用いてデータセットを構築した。図 12 にデータセットの画像例を示す。これらを 15 人の被験者に対して提示し、「テレビ会議で会話している想定で、画像の人物と視線が一致しているか」という問いに対して、「一致」、「やや一致」、「やや不一致」、「不一致」の 4 つから選択してもらい回答を得た。

4.2.2 実験結果と考察

図 13 に実験結果を示す。元画像では視線が一致していると回答した割合は 3%で、やや一致と合わせても 20%に満たなかった。それに対し、変換後の画像では、まぶたの内側までの領域を変換した場合、視線が一致していると回答した人は 20%で、やや一致と合わせると 60%となった。また、まぶたの外側までの領域を変換した場合、視線が一致していると回答した人は 40%で、やや一致と合わせると 80%を超えた。まぶたの内側までの領域を変換するよりもまぶたの外側までの領域を変換した方が視線の一致を感じる割合が高かったのは、後者の方が、カメラを注視している目とその周辺より自然な見えを再現することができたためと考えられる。



(a) 元画像



(b) まぶたの内側を合成した例



(c) まぶたの外側を合成した例

図 12 データセットの画像例

5. ま と め

顔モデルを用いず、単一カメラのみを用いたテレビ会議における対話者間の視線一致のための目領域変換手法を提案した。提案手法では、事前に撮影した画像を用いてユーザがディスプレイ上の対話相手の目を見ていると判定した場合、目領域を変換し、視線を一致させた画像を生成する。被験者実験により、視線が一致していない顔画像から視線が一致した顔画像を生成できたことを確認した。今後の課題としては、目領域変換判定の高精度化などが挙げられる。

謝辞 日頃より熱心に御討論頂く名古屋大学村瀬研究室諸氏に深く感謝する。本研究の一部は、科学研究費補助金による。

文 献

- [1] 片山 貴信, 武川 直樹, 徳永 弘子, 湯浅 将英, “多人数映像電話における話し手の身振りとアクティビティの関係—視線一致と不一致環境により会話の質はどのように変わるか?—,” 信学技報, HCS2008-78, Mar. 2009.
- [2] A. Kendon, “Some functions of gaze-direction in social interaction,” *Acta Psychologica*, vol. 26, pp. 22–63, Feb. 1967.
- [3] 藤田 かおり, “視線一致映像装置を用いた発話交替の行動分析,” 信学技報, MVE2007-91, Mar. 2008.
- [4] R. Kollarits, C. Woodworth, and J. Ribera, “An eye-contact cameras/display system for videophone applications using a conventional direct-view LCD,” *Digest of 1995 SID Int. Symposium*, pp. 765–768, May 1995.
- [5] R. Yang and Z. Zhang, “Eye gaze correction with stereovision for video-teleconferencing,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 26, no. 6, pp. 956–960, July 2004.
- [6] M. Ott, J. Lewis, and I. Cox, “Teleconferencing eye contact using a virtual camera,” *INTERACT’93 and CHI’93 Conf. Companion on Human Factors in Computing Systems*, pp. 109–110, Apr. 1993.
- [7] J. Liu, I. Beldie, and M. Wopking, “A computational approach to establish eye-contact in videocommunication,” *Proc. Int. Workshop on Stereoscopic and Three Dimensional Imaging*, pp. 229–234, Sep. 1995.
- [8] S.M. Seitz and C.R. Dyer, “Toward image-based scene representation using view morphing,” *Proc. 13th Int. Conf. on Pattern Recognition*, vol. 1, pp. 84–89, Aug. 1996.
- [9] C. Kuster, T. Popa, J.C. Bazin, C. Gotsman, M. Gross, “Gaze Correction for Home Video Conferencing,” *Proc.*

ACM SIGGRAPH Asia, *ACM Trans. Graphics*, vol. 31, no. 6, pp. 174:1–174:6, Nov. 2012.

- [10] M. Jones, “Multidimensional morphable models: A framework for representing and matching object classes,” *Int. J. of Computer Vision*, vol. 29, no. 2, pp. 107–131, Aug. 1998.
- [11] J. Gemmell, C.L. Zitnick, T. Kang, K. Toyama, and S. Seitz, “Gaze-awareness for videoconferencing: A software approach,” *IEEE Multimedia*, vol. 7, no. 4, pp. 26–35, Oct. 2000.
- [12] 佐藤 利喜夫, 三浦 彰, 永田 邦一, “映像電話における撮像管の位置に関する検討,” 昭和 42 年電気四学会連合大会講演論文集, p. 2316, Apr. 1968.