

平成21年度

筑波大学第三学群情報学類

卒業研究論文

題目 ジェスチャを用いた
実世界ビデオメモシステム

主専攻 情報科学主専攻

著者 中井川峻

指導教員 高橋伸，志築文太郎，三末和男，田中二郎

要 旨

メモは、記憶を外化し未来の自分や他人に情報を伝えるために、現在もいたるところで利用されている。これらのメモの内、未来の自分に向けたメモについては多くの研究が行われ、様々な手法が提案されているのに対し、伝言メモに代表される他人に向けたメモについての研究は少ない。伝言メモの主要な記録方法としては未だに紙とペンが用いられているが、その情報量には限りがあり、記録に時間がかかってしまう、正確な情報伝達が出来ない、といった問題点がある。

本研究では、伝言メモを支援する手法としてオフィスなどで用いられる机を拡張し、ジェスチャにより音声と映像を用いた伝言メモの記録・閲覧を行うシステムを提案し、実装を行った。これにより、ペンや操作デバイスなどの道具に頼らず、音声や映像を用いた情報の素早い記録を可能とする。

目次

第1章	序論	1
1.1	メモの利用と現状	1
1.2	紙による伝言メモの利点と問題点	2
1.2.1	紙による伝言メモの利点	2
1.2.2	紙のメモによる伝言の問題点	2
1.3	本研究の目的	3
1.4	本研究のアプローチ	3
1.5	本論文の構成	3
第2章	関連研究	4
2.1	メモやメッセージを用いた研究	4
2.2	拡張机型実世界指向インタフェース	4
第3章	ビデオメモシステム	6
3.1	システムのイメージ	6
3.2	想定される利用シナリオ	7
3.3	伝言メモを記録する操作	7
3.4	伝言メモの表示	9
3.5	伝言メモの閲覧と編集を行う操作	9
3.5.1	再生・編集メニューの操作	9
第4章	実装	12
4.1	開発環境	12
4.2	システム構成	12
4.3	手と指先の認識	12
4.3.1	手領域の抽出	14
4.3.2	指先の認識	19
4.4	伝言メモの記録開始・終了ジェスチャの認識	20
4.5	伝言メッセージの記録	23
4.6	アイコンの選択	23
第5章	今後の課題・発展	24
5.1	試用とそこから得られた知見・課題	24

5.2	発展	25
5.2.1	Natural Storage との統合	25
5.2.2	個人の認証とネットワークの利用	25
第6章	結論	26
	謝辞	27
	参考文献	28

目次

1.1	メモの利用	2
3.1	システムのイメージ	6
3.2	伝言メモの記録操作の流れ	8
3.3	机に投影された伝言メモアイコン	9
3.4	メモの再生・編集メニュー	10
3.5	伝言メモの再生操作の流れ	11
4.1	ハードウェア構成のイメージ	13
4.2	構築したシステムの一部	13
4.3	システムの利用の様子とシステム内座標	14
4.4	画像処理の流れ	15
4.5	背景差分の結果	16
4.6	肌色サンプルの H, S 値	17
4.7	手領域に含まれる肌色領域と赤色領域	18
4.8	肌色抽出の結果	19
4.9	手の形状変化における重心の移動	21
4.10	矩形と手領域の接点と重心との距離	21
4.11	指認識の結果	22
4.12	カーソル位置の決定	23
5.1	試用している様子	24

第1章 序論

1.1 メモの利用と現状

オフィスや研究室，家庭など，いたるところで日常的にメモが利用されている．メモの内容は，思いついたアイデア，TODO，予定，気になった記事，伝言など多岐にわたるが，それらの用途は大きく二つに分けられると考えられる．

一つは，未来の自分のために記憶を外化することである．人は一度見聞きしただけでは，情報を完全に記憶することが難しい．忘れてしまう可能性が高い情報をメモすることにより，記憶を補助することが可能となる．

もう一つの用途は，自分の記憶を他人に伝えることである．自分の記憶として存在する情報を他人に伝えるためには，それを他人からアクセス可能な状態に加工する必要がある．伝える対象と会話出来る状態であれば言葉にして伝えることが出来るが，そうでない場合は情報を何らかの形に外化することが求められる．メモはその手段の一つとして用いられており，紙や音声などをメモとして保存することにより，非同期的な情報伝達が可能となる．

未来の自分に情報を伝えるためのメモの要件としては，持ち歩いて閲覧出来ること（ポータビリティ），思い立った瞬間に素早く情報を記録出来ること（即時記録性），閲覧時に瞬時に見渡せること（一覧性）などが挙げられる [1]．これらの要件を満たすために様々な研究，製品開発が行われており，図 1.1(a) に示すように，その性質と親和性の高い携帯電話などのモバイルデバイスに実装され，実用されている．

他人に情報を伝えるためのメモに関する研究，製品開発は GPS を利用して地図上に情報を残す，というようなものが主であり，新しいメモの形として実用が始まっている [2]．一方で，他人に情報を伝えるためのメモの内，伝言メモのように机の上などの小さな範囲を対象としたメモを支援する研究や製品は，ほとんど存在しない．そのため，図 1.1(b) のように，現在でも紙とペンによるメモが多く用いられている¹．

¹ 「企業における電話とユニファイド・コミュニケーション利用に関する実態調査」

URL : http://www.avaya.co.jp/corporate/pressroom/2009/09_03_24.shtml

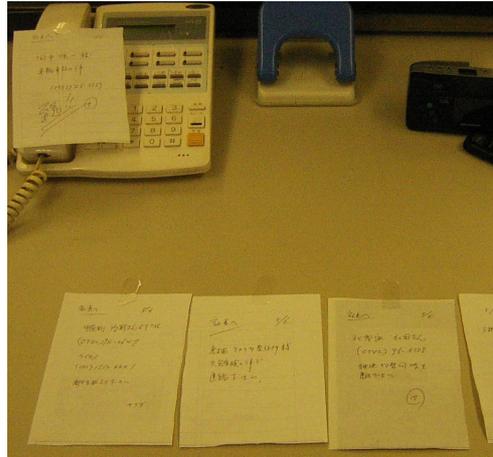
調査時期：2009年2月下旬

調査方法：インターネットによる調査（ヤフーバリューインサイト株式会社に調査委託）

調査対象：20歳以上で従業員数50名以上の事業所に勤務する，業務上で電話を使用する男女年代別計300名



(a)モバイルデバイスのメモソフト



(b)オフィスにおける伝言メモ

図 1.1: メモの利用

1.2 紙による伝言メモの利点と問題点

1.2.1 紙による伝言メモの利点

紙のメモは、物との空間的隣接性を利用した情報伝達を行うことが出来る。例えば、書類のチェックを頼む状況では、チェック箇所についてメモし、それを書類自体、もしくは付近に貼り付けることにより、書類の指定を省略することが出来る。また、すでに伝言した内容について追加の連絡があった場合、前回貼りつけたメモの付近に貼りつけることにより、続きの内容であることを示すことが出来る。

これらをメールや電話で行う場合、空間的隣接性を利用できないため、文字や音声のみによる位置の指定が必要となり、直接的な位置の指示を行う場合に比べ、情報伝達が困難になる。また、伝言情報を作業中に視認出来ず、伝言の存在自体を忘れてしまう可能性もある。

1.2.2 紙のメモによる伝言の問題点

オフィスにおいて紙の伝言メモを残す際には、当然ながら紙とペンが必要となる。また、メモが他の書類に隠れてしまうことや、人や空調の起こす風によって紛失してしまうことも考えられるため、テープや糊付き付箋などを用いて、視認しやすい位置に固定する必要がある。しかし、情報を残す対象が他人の机であるため、テープや糊付き付箋などの文具がその場に無い可能性も高い。したがって、自分の机からメモのための文具を探し、メモを記録し、それを持って伝言メモの対象となる机まで赴かなくてはならず、思い立った瞬間に素早く情報を記録することが難しい。また、手書きの文字や図といった比較的簡易な視覚情報しか用いることが出来ないため、考えたままの情報をそのまま書き留めただけでは文章が長すぎてし

まう場合もあり，要点をまとめて記録するための時間や労力が必要になる．その過程で情報が抜け落ち，正確な伝達が出来ないという問題も発生する．

1.3 本研究の目的

本研究では，オフィスなどにおける机を対象とした伝言メモを，コンピュータを用いて支援することにより，スムーズな非同期的情報伝達を可能とし，オフィスコミュニケーションを円滑化することを目的とする．そのために，1.2.1 で述べた紙の伝言メモ手法における利点を継承した上で，1.2.2 で述べた問題点を解決する手法を提案し，それを実現するシステムの開発を行う．

1.4 本研究のアプローチ

オフィスにおいて他人の机に，思い立った瞬間に伝言メモを記録する，ということを考えた場合，道具を使用することなく全ての操作を行えることが好ましい．これに対し我々は，オフィスの机を拡張し，ハンドジェスチャを用いて実世界上にメモを残す手法を提案する．これにより，道具に直接触れることなく操作することが可能なだけでなく，紙のメモの利点である実空間上の物体との空間的隣接性を利用したメモの記録が可能となる．また，ハンドジェスチャを用いることにより，実世界上に投影されたメモをより直感的に操作することが可能であると考えられる．さらに，本研究ではメモとして残す情報の種類として，音声と映像を用いることを提案する．音声や映像は筆記に比べ即時記録性に優れ，考えたことをよりそのままに近い形で記録することが可能である．また，映像による視覚情報を付加することにより，身振り手振りなどを用いた素早く正確な情報記録の実現を図る．ただし，電話対応における伝言などを想定すると，視覚的な情報が必要ない場合も存在するため「音声のみ」，「音声+映像」を選択可能にする．

1.5 本論文の構成

本章では，既存の伝言メモについての問題点を挙げ，それを解決する手法の提案と研究の目的を述べた．続いて第2章では関連研究について述べる．第3章ではハンドジェスチャを用いた実世界ビデオメモシステムのコンセプト，第4章では実際に開発したシステムの実装について詳細に述べる．第5章でそのシステムの評価と議論を行い，第6章で今後の課題と発展について述べる．最後に，第7章で結論を述べる．

第2章 関連研究

2.1 メモやメッセージを用いた研究

Yonezawaらは、頭部に装着した3軸地磁気・加速度センサを用い、頭部方向に応じた音声付箋の貼り付けとその閲覧を行う手法を提案した[3]。この手法では、閲覧に時間がかかってしまうという音声の持つ問題を解決するため、音声メモに空間を活用した一覧性を与えるというアプローチをとっている。また、これを発展させた[1]では、よりポータビリティを持たせるため、iPhoneに搭載されたセンサを用い、モバイルデバイス上で動作するシステムを開発した。これらは、メモの即時記録性を高めるために音声を利用している、という点で本研究と関連するが、他人に伝えることを目的としていない点が異なっている。

TarumiらによるSpaceTag[4]は、GPSにより自分の現在位置を取得し、ポータブルPCから実世界空間へSpaceTagと呼ばれる情報付箋を張り付けたり、貼り付けられたSpaceTagをポータブルPCから閲覧したりすることが可能である。しかし、これらの情報は自分に対するものである。さらに、GPSの精度の問題から屋内の机上の位置を利用することは難しい。

sekai camera[2]はiPhoneアプリとして利用が始まっているサービスである。GPSや加速度センサを利用して実世界空間にエアタグと呼ばれる付箋のようなものを貼り付け、カメラ映像に重畳表示する。これを多数のsekai camera上で共有することにより、テキスト、音声、写真を用いた非同期なコミュニケーションを可能とする。しかし、GPSでは誤差が大きく、数cm単位の空間的隣接性を利用出来ないため、オフィスでの利用には適さないと言える。

StickySpots[5]は、タッチパネル式のディスプレイを用い、その操作により伝言メモを貼り付けることが出来る。家庭内での伝言メモを対象とし、伝言掲示板のように複数人で共有することにより、家庭内コミュニケーションを支援する。腕時計などに埋め込んだRFIDタグによる個人認証を行い、プライベートなメッセージを送ることも可能としている。

松原ら[6]は、家庭内の家事分担に焦点を当てた、家庭用総合支援システムの中に伝言メモシステムを統合している。ホワイトボードをイメージしてデザインされており、タスクのカテゴリに合わせたメモの形状や、天気やタスクが追加されてからの経過時間を基にしたタスクの推薦機能などを備えている。

2.2 拡張机型実世界指向インタフェース

この分野の先駆的な研究として、WellnerらのDigitalDesk[7]が挙げられる。DigitalDeskでは、机をカメラとプロジェクタによって拡張し、物理的なオブジェクトと仮想的なオブジェ

クトを関連付けて扱う手法や指先の認識，遠隔コミュニケーションなどを提案し，実装している．

Rekimoto らによる Augmented Surfaces[8] は，机だけでなく，壁やラップトップのディスプレイとの間で，仮想的なオブジェクトとして投影された情報を，スムーズに移動することが出来る．パンチルトカメラによる 2 次元コードの認識なども実現している．

Sato, Koike らによる EnhancedDesk[9] は，赤外線カメラを利用し，人の体温付近のみを撮影することにより，手領域の切り出しを高速に行う手法を提案した．また，指先をロバストにトラッキングすることにより，投影されたオブジェクトや 2 次元コードなどとの，より高度なインタラクションを可能としている．

これら既存の拡張机型実世界指向インタフェースは，自分の情報を扱うことに焦点を当てていた．これに対し，本研究では他人のための情報を扱うために，机拡張型インタフェースを用いる．

Chih-Sung らによる WikiTUI[10] は，書籍に対するアノテーションの重畳表示や，その情報の編集などを行うことができる．また，それらアノテーション情報を，多人数で共有することを想定している．実書籍を扱うことに重点を置いており，書籍を介した情報共有である点において，本研究とは異なる．

第3章 ビデオメモシステム

3.1 システムのイメージ

システムイメージを図3.1に示す。本システムは、オフィスなどで日常的に使用される机を拡張したインターフェースにより、道具やコンピュータを直接使用することなく、音声と映像を用いた伝言メッセージをメモとして記録・再生する。伝言メモの記録は、伝言が必要になった時に、その対象者の机に対し、記録ジェスチャをすることにより行う。記録されたメモは紙のメモを模した矩形のオブジェクトとして机の上に投影され、記録されてからの経過時間に応じて大きさを増すことにより、伝言メモの存在を伝言メモの対象者に気付かせる。そして、このように貼りつけられた伝言メモに対し、再生ジェスチャをすることにより、再生を行う。

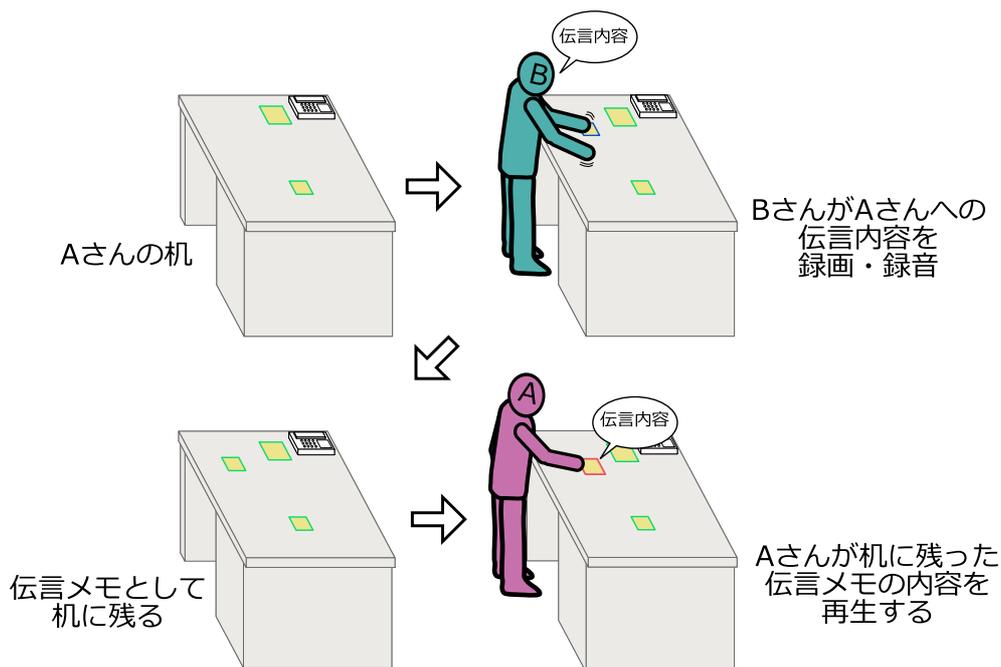


図 3.1: システムのイメージ

3.2 想定される利用シナリオ

研究室

研究室において、自分の参加しているチームプロジェクトに関連する論文を見ている時、とっさに良いアイデアが思い浮かび、それをチームメンバに紹介したいがそのメンバが席を外してしまっている、という状況を想定する。従来は、電話をする、アイデアを紙に記し、それを添付した論文を机の上に置いておく、直接会えるまで待つ、といった方法でしかアイデアを伝えることが出来なかった。電話を用いる場合は、視覚的な情報を使用せずに全ての情報を伝える必要があり、紙に記す場合、文章として情報をまとめる時間が必要である。さらに、直接会えるまで待つ場合に至っては、待つ時間は予測不能であり、かなりの時間を無駄にする可能性がある。

本手法を用いる場合、映像と音声を残すことが可能なため、論文を手に、チームメンバの机に赴き、思いついたアイデアについて、論文内の関係する部分を指で指しつつ、音声により素早く説明することが可能となる。メモを残されたメンバは、論文を閲覧しつつ、音声やメモにおいて指示されている箇所、仕草などを見ることにより、直接話すような感覚でアイデアの説明を受けることが可能となる。また、時間を浪費することなく、思い立った瞬間に伝言メモを残すことが出来る。

オフィス

オフィスにおいて、電話が鳴り、それに應對したが、電話の用件は現在席をはずしている社員へのものだった、という状況を想定する。従来は、伝言を紙にメモするため、受話器を片手に持ったまま紙とペンを探し、それに手間取っている間、相手を待たせ続けることになった。さらに、その伝言の内容が複雑な場合、書き取りに時間がかかり、また、内容を確認するために書き終わった後に全ての内容を復唱する、といった必要があった。

本手法を用いる場合、道具を手取ることなく音声を残すことが可能なため、相手の話に合わせて、ジェスチャにより録音を開始し、随時内容を復唱するだけで素早く伝言メモを記録出来る。さらに、手書きのメモの場合省略せざるを得ない情報や、抜けてしまう情報がより詳細に伝わる可能性が高くなる。

3.3 伝言メモを記録する操作

本研究における伝言メモを記録する操作は、映像・音声を用いた伝言メッセージを記録し、それを机の上に伝言メモとして貼りつける操作である。その手順はメモの開始、メモの終了、メモの種類を選択から構成される。

紙のメモを開始する場合を想定すると、メモ用紙を片手で押えるという動作を行う場合が多い。本研究では、この動作に着目し、伝言メモの開始ジェスチャとして導入する。机の上

に手を広げて叩くように置くジェスチャを行うことにより伝言メモを開始し、映像と音声
が保存され始める。

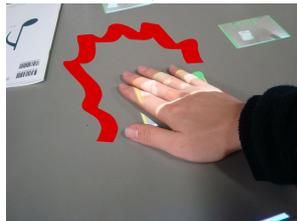
映像は机の真上から下向きに撮影される。そのため、表情は伝えにくい
が、机の上にあるものを指示した情報やその時の机の上の様子などをビデオメモとしてより詳細に記録することが可能である。また、音声は机の上に設置したマイクから取得しており、身振り手振りと組み合わせて音声メッセージを残すことが可能である。

紙のメモを記録した後、伝言メモとして利用する場合、伝言対象の机にそのメモ用紙を貼り付ける。本研究では、この動作を伝言メモの記録終了ジェスチャとして採用した。再び机を叩くように手を置くことにより、録音・録画を停止し、手を置いた位置へ伝言メモを貼り付ける。

伝言メモの記録操作の流れを図 3.2 に示す。



叩くようにして
手を机に置くことにより
伝言メモの記録を開始



再び叩くようにして
手を机に置くことにより
伝言メモを貼り付ける

図 3.2: 伝言メモの記録操作の流れ

3.4 伝言メモの表示

記録され、貼りつけられた全ての伝言メモは、図 3.3 に示すように紙のメモを模した矩形のアイコンとして、その最初のフレームを縮小したものが、天井のプロジェクタから机に投影される。

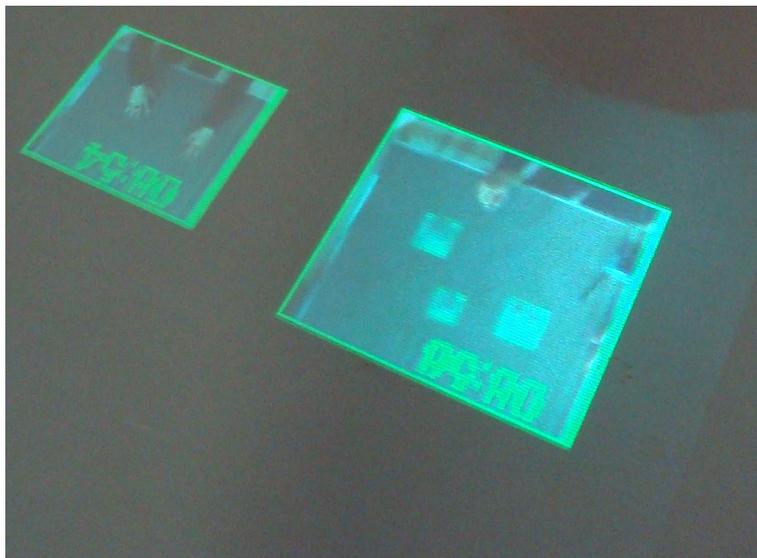


図 3.3: 机に投影された伝言メモアイコン

また、伝言メモは閲覧してほしい情報があるために貼り付けられるものであり、貼り付けた机の持ち主に気付いてもらう必要がある。そこで、本システムでは、伝言メモが貼り付けられた日時を記録し、サムネイル画像を時間の経過とともに拡大していくことにより、伝言対象者に対して伝言メモへの気づきを与える。伝言メモの大きさは、6 時間で最大となるように設定し、縦横とも元の大きさの 2 倍になる。拡大していくのは一度も閲覧されていない伝言メモのみであり、再生後は初期の大きさに戻る。

3.5 伝言メモの閲覧と編集を行う操作

3.5.1 再生・編集メニューの操作

貼り付けたメモを 1.5 秒間指差すことにより、メモの下に再生・編集メニューを表示する。それぞれのメニュー項目の選択についても、伝言メモアイコンの選択と同様に、1.5 秒間の指差しによって行う。メニュー項目は図 3.4 のようなアイコンから成り、それぞれのアイコンを選択した場合の動作は以下のとおりである。

Play アイコン

Play アイコンを選択すると、伝言メモの再生を行う。伝言メモを原寸大に拡大し、音

声とともにその映像を机全体に投影する．映像の再生は天井に取り付けたプロジェクタにより行い，音声の再生は机に設置したスピーカにより行う．また，再生中に伝言メモの記録を開始することは出来ないが，記録開始ジェスチャと同様の机を叩くように手を置くジェスチャを行うことにより，再生を中止することが可能である．

Move アイコン

Move アイコンを選択すると，伝言メモの移動を行うことが出来る．選択が確定されると，伝言メモアイコンは指の位置に追従し始める．その後，指の位置を2秒間固定することにより，移動状態を解除して伝言メモの位置を固定することが出来る．また，指が認識されないと，伝言メモアイコンは元の位置に戻る．これにより，紙のメモの貼り付け位置を変更することと同様の効果を実現する．

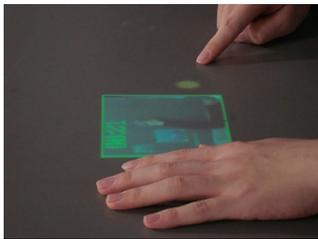
Delete アイコン

Delete アイコンを選択すると，伝言メモの削除を行うことが出来る．伝言メモアイコンは机の上に投影されなくなり，そのメッセージ内容も映像，音声共に削除される．これにより，紙のメモを剥がして捨てることと同様の効果を実現する．

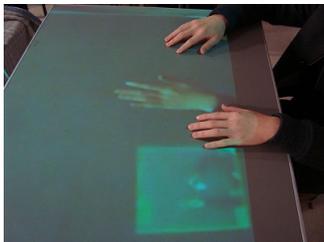


図 3.4: メモの再生・編集メニュー

伝言メモの再生操作の流れを図 3.5 に示す．



貼り付けられた
伝言メモアイコンを
指先によるカーソル操作で
選択



メニューからPlayを
選択することにより
メモを再生する

図 3.5: 伝言メモの再生操作の流れ

第4章 実装

4.1 開発環境

システムはC++によって実装し、画像処理ライブラリとしてOpenCV[11]を用いた。また、GPU向けのC言語開発環境であるCUDA[12]を一部使用している。CUDAを利用することにより、GPUの持つ多くのコアそれぞれにスレッドを割り当て、数十から数百のスレッドを用いたマルチスレッドによる高速処理が可能となる。これにより、CPUのみを利用した処理では困難な、実時間での処理を実現している。

4.2 システム構成

本システムのハードウェア構成のイメージ図を図4.1に、実際に構築したシステムの一部を図4.2に示す。机の上を撮影するためのカメラと画像を机に投影するためのプロジェクタを天井に設置し、音声入力のためのマイク、音声出力のためのスピーカを机に設置する。それぞれのデバイスは1台の計算機に接続し、その計算機によって画像の入出力、音声の入出力、画像処理、音声処理など全ての処理を行う。システムに対する操作は全てハンドジェスチャにより行うため、その他の操作デバイスは備えていない。

また、一般的なオフィスにおいて使用者が机に面する時、図4.3(a)のようにその方向は机の手前側に限られる。そこで、システムにはユーザが操作する方向を机の手前側に設定し、システム内の座標としては図4.3(b)のように、システムの左奥を原点とし、左から右向きにx軸、奥から手前方向にy軸を取る。伝言メモの投影方向などもここで設定するシステムの向きに応じて行う。

4.3 手と指先の認識

天井に設置したカメラから取得した画像を図4.4のように処理し、手と指先の認識を行った。まず、カメラからの入力画像に背景差分を適用し、前景を求める。前景画像にHSVによる肌色抽出を適用し、肌色領域を抽出する。そして、肌色領域に対し、ノイズ処理を行うことにより、手領域を求める。手領域のサイズから指差しを行っていると思われる手の存在を判定し、存在していれば指認識を行う。指認識では、その手領域内の点と重心との距離から指先を認識する。本節では、それぞれの処理の詳細について述べる。

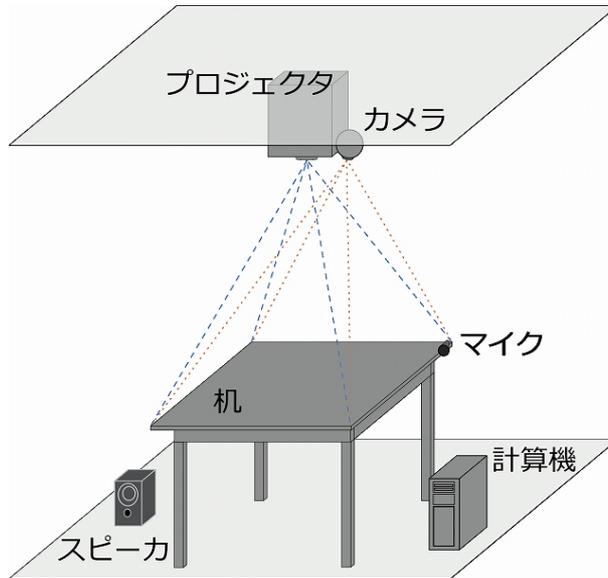
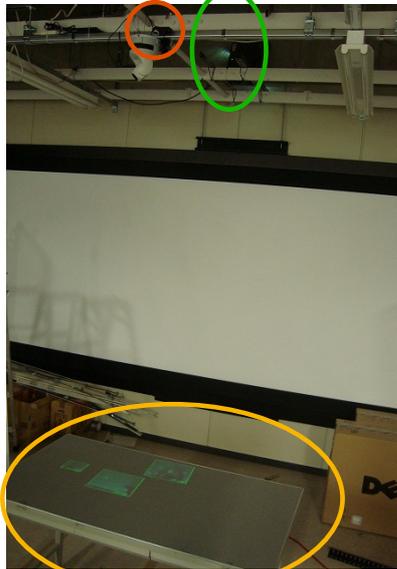


図 4.1: ハードウェア構成のイメージ

Webカメラ プロジェクタ



机

図 4.2: 構築したシステムの一部

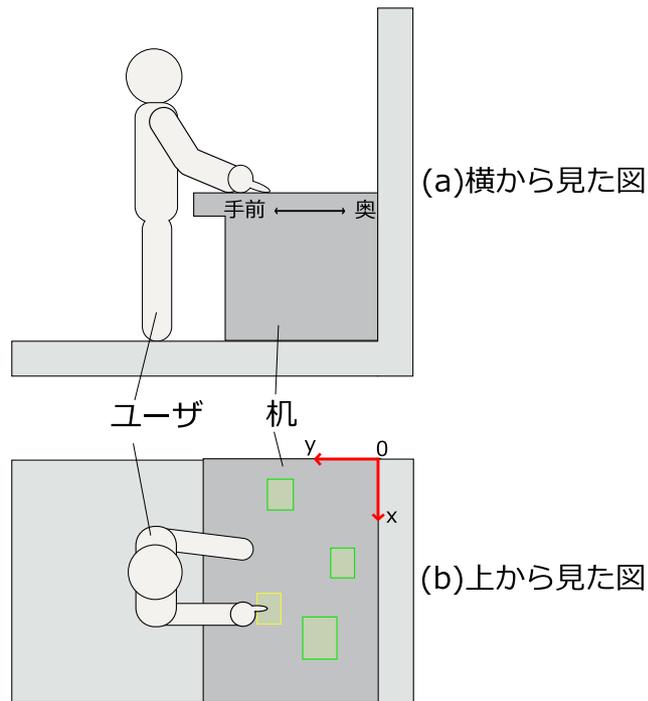


図 4.3: システムの利用の様子とシステム内座標

4.3.1 手領域の抽出

背景差分

初めに背景画像を取得し、現在のカメラ画像との差分を取ることで、人などの動く物体のみを抽出する背景差分法を用いた。これにより、下に述べる色の閾値を用いた肌色抽出処理で残ってしまうノイズの中で、最初に撮影した背景部分にあるものを取り除くことが可能になる。

システムは起動時に取得する 50 枚の画像から、各画素に対し標準偏差を求める。求めた標準偏差に 5 をかけたものをそれぞれの画素の閾値とし、現在のカメラ画像との差分をとることにより背景を取り除く。これにより得られた背景差分結果を図 4.5 に示す。腕以外の物体は全て背景として取り除かれるように背景取得を行っており、ほぼその通りに、腕を切り出していることが分かる。腕の領域であるにもかかわらず、取り除かれてしまっている部分があるが、これは背景の色と同様であると判断されてしまうためである。この誤認識は予備実験を行った際に、操作を行う上で大きな認識誤差を引き起こす原因とはならなかったため、問題ないとした。

肌領域の抽出

手領域の抽出を行うためには、熱画像を用いる手法 [13][14]、赤外線発光面を用いる手法 [15] 一般的な Web カメラ画像からの肌色抽出を用いる手法 [16][17] など様々なも

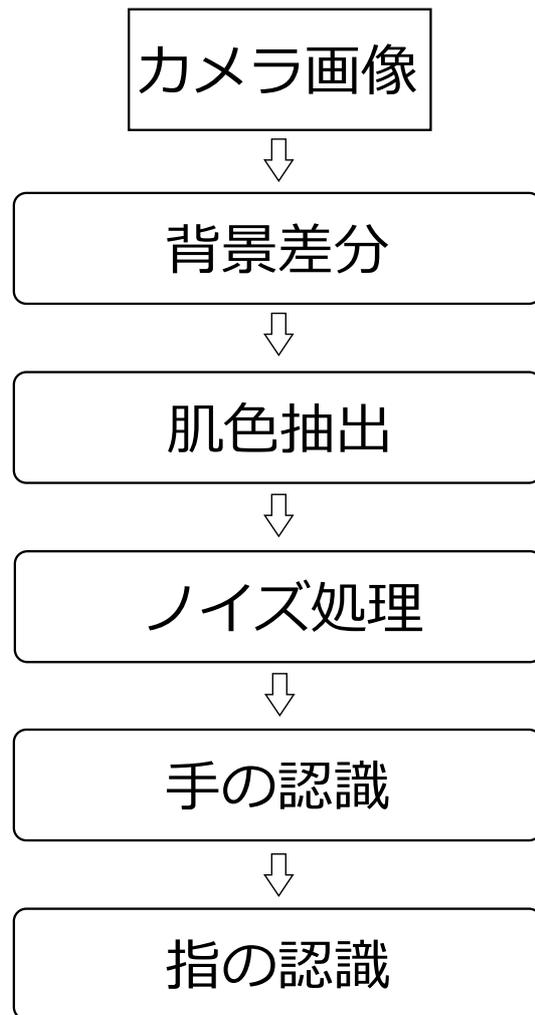
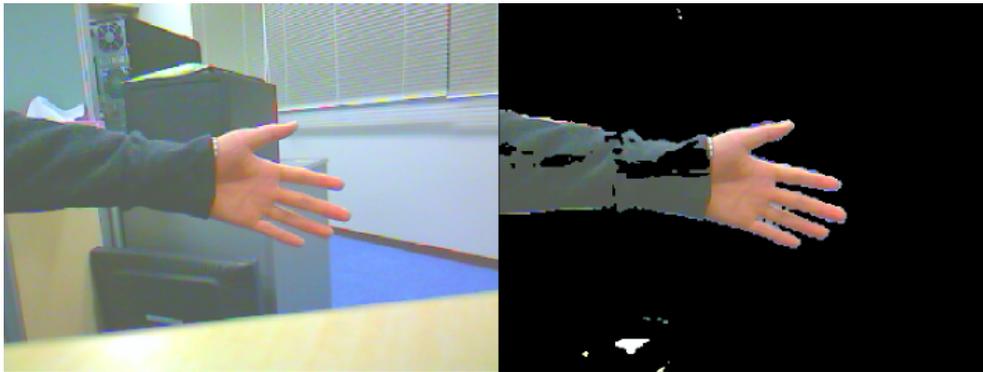


図 4.4: 画像処理の流れ



入力画像

背景差分法適用後

図 4.5: 背景差分の結果

のが提案されている。

熱画像を用いる手法は人体が熱を持つことを利用する。人の平熱に近い温度を持つ物体のみを特殊なカメラで撮影し、人物領域を抽出することが可能である。しかし、カメラの価格が比較的高い。また、机拡張型である本システムでは、机に置いた温かい飲み物や、ノートパソコンなどを誤検出してしまう可能性がある。よって本システムには適さない。

赤外線発光面を用いる手法は、面に対し投光された赤外線を、前景となる物体が遮ることにより、それらの物体のみを抽出することが出来る。しかし、この手法は、赤外線を投光するためのデバイスが必要であるため、オフィスの机などを拡張することに適さないと思われる。また、人以外の物体も抽出してしまうため、手領域のみを抽出する処理には向かないと言える。

一般的な Web カメラ画像からの肌色抽出を用いる手法は、人の肌の色の持つパラメータに注目し、肌色とそうでない色に画像を二値化することが出来る。二値化の際には、色のパラメータを用いた閾値を設定する。環境によって閾値を正しく設定することにより、正確な検出を行うことが可能となる。また、通常の Web カメラを用いることが可能なため、比較的安価に実装可能である。これらの利点から、本研究では一般的な Web カメラ画像からの肌色抽出を用いる。

本研究で用いる色空間は HSV である。色相 (H) と明度 (V) が独立しているため、画像の明るさに対してロバストに肌色検出を行うことが可能となる。図 4.6 は多数の肌色サンプルの H, S 値を色相環にプロットしたものであるが、これに見られるように H が 0 ~ 30 のとき肌色に近い色であると言える [18]。つまり、カメラから取得した RGB 画像を HSV 色空間に変換し、この値を基に設定した閾値により二値化することで、肌色が抽出出来る。しかし、一般的に他の色空間の表す色は完全に表現できるものではなく、通常の HSV への変換式では、本来再度が低いはずの黒に近い色が高い彩度 (S) を示して

しまうという問題が発生し，彩度による肌色領域抽出の精度を低下させる恐れがある．これに対し，本研究では，松橋ら [19] によって提案された修正彩度 (S_M) を用いることによりこの問題を回避している．

RGB から修正彩度を用いた HSV への変換式は以下のようなになる．

$$V = \text{MAX}(R, G, B)$$

$$S_M = \sqrt{\left(\frac{(R - G) + (R - B)}{2}\right)^2 + \left(\frac{\sqrt{3}(G - B)}{2}\right)^2}$$

$$H = \cos^{-1}\left\{\frac{(G - B) + (G - R)}{2\sqrt{(G - B)^2 + (G - R)(B - R)}}\right\}$$

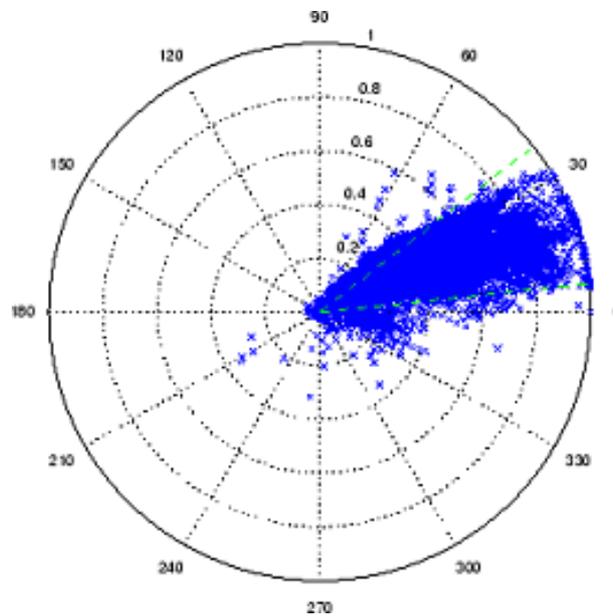


図 4.6: 肌色サンプルの H, S 値

さらに我々は人の手の色の閾値設定を試行する内，図 4.6 に示した H の範囲に含まれない，赤に近い色が手に含まれることに気付いた．これは，図 4.7 のように，指先や爪，手の平の手首付近に多くみられる部分である．この部分を別の領域として抽出し，前述した肌色の範囲の領域に加算することにより，より正確な手の形状を抽出することが可能となる．正確に手の形状を抽出することは，

この処理のため，H の最小閾値 (H_{\min})，H の最大閾値 (H_{\max})，S の最小閾値 (S_{\min})，S の最大閾値 (S_{\max})，V の最大閾値 (V_{\max}) を肌色用，赤色用として 2 つずつ用意し，閾値セット 1 の全ての範囲に入った画素を肌色，閾値セット 2 の全ての範囲に入った画

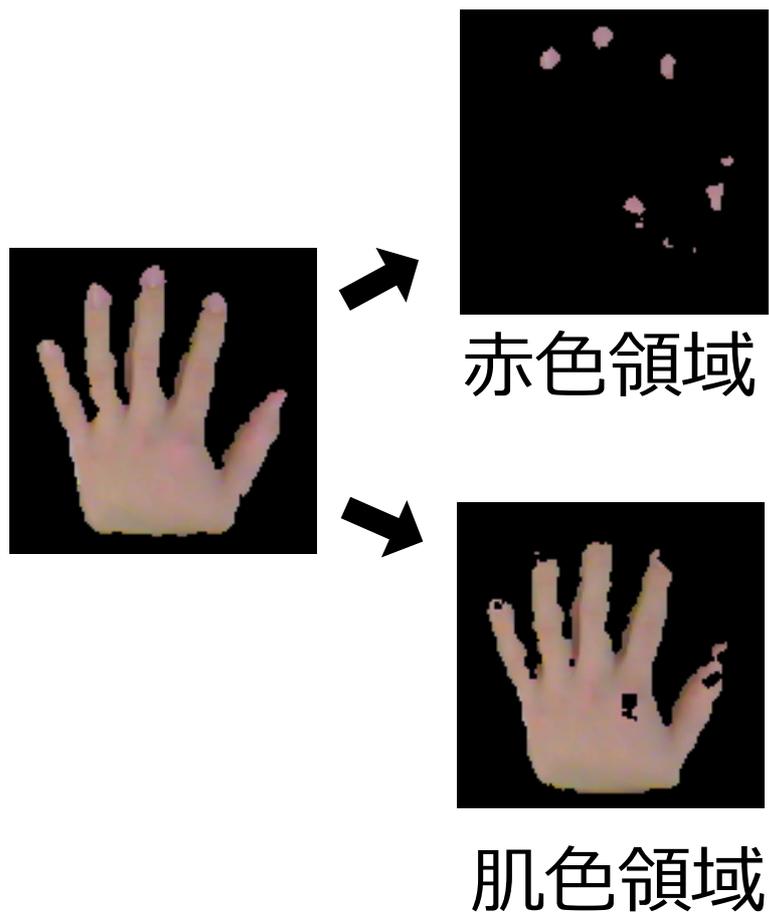


図 4.7: 手領域に含まれる肌色領域と赤色領域

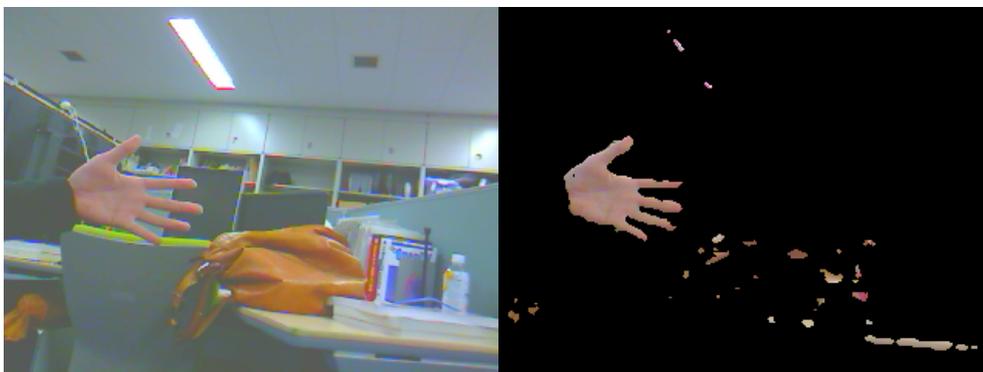
素を赤色とした．ある画素を修正彩度を用いた HSV に変換したのちの各チャンネルの値を， H' ， S' ， V' とすると，肌色である条件は以下のように定めた．

$$(H_{min1} < H' < H_{max1} \text{ and } S_{min1} < S' < S_{max1} \text{ and } V' < V_{max1})$$

or

$$(H_{min2} < H' < H_{max2} \text{ and } S_{min2} < S' < S_{max2} \text{ and } V' < V_{max2})$$

これにより得られた肌色抽出の結果を図 4.8 に示す．一般的に肌色抽出を用いた手法では，茶色の物体や蛍光灯の光などを誤検出しやすい．本研究で用いた手法では，茶色の物体や蛍光灯の光は小さく誤検出しているものの，ほぼ正確に肌色を抽出出来ていることが分かる．



入力画像

肌色抽出適用画像

図 4.8: 肌色抽出の結果

ノイズ処理

人の手の大きさには個人差があるとしても誤差数 cm 程度である．よって，固定されたカメラに映る，一定の高さにおける手のサイズはある程度絞ることが出来る．本システムでは二値化した画像における各白領域のサイズを求め，予備実験によって求めた最小閾値と最大閾値を基に，最大閾値以下かつ最小閾値以上である部分を抽出する．これにより，肌色抽出で誤検出してしまう茶色の物体などを取り除き，手領域を正確に抽出することが可能となる．また頭部にも肌色領域が含まれるが，手と比較しサイズが大きいため，閾値処理により取り除くことが可能となる．

4.3.2 指先の認識

肌色領域のみを抽出した二値画像からの指の認識については多くの研究が行われ，ロバスト性の高い手法が提案されている．岡ら [20] は拡張机型インタフェースのための複数指先認

識手法を提案しており、白い円状のテンプレート画像を用いたパターンマッチングやカルマンフィルタを用いた動きの予測など、複雑な処理によって複数の指先の高精度な認識を行っている。しかし本研究において指先はアイコンを選択するために使用するものであり、単数の指先を高速に処理することが求められるため、処理速度が求められるこれらの手法は適さないと思われる。

本研究では、高速かつロバストな認識のため、以下のことに注目した。

- メモなどを書く際に行う「手を添える」動作
- 1本の指を伸ばした際に起こる手領域の面積変化
- 1本の指を伸ばした際に起こる手領域の重心変化

人は何かを書く際、紙に手を添える、紙を押さえるといった動作を行うことが多い。これを利用し、「2つの手領域を検出した」ということを指の認識を行うトリガとする。また、同一の高さにある場合、1本の指を伸ばした手領域の面積に比べ、全ての指を伸ばした手領域の面積が大きくなる。2つの手領域が検出された際、面積が小さい手領域のみに対し指の認識を行うことにより、全ての手領域に対する指認識を行う場合に対し、高速かつ安定した認識を可能とした。

さらに、全ての手が伸ばされた時に比べ1本の指が伸ばされると、手領域における重心が手首側に移動し、指先から遠ざかることを利用する。図4.9は実際のシステムにより手領域を抜き出し、それぞれの領域を包含する矩形と手領域の重心を描画したものである。手領域の中心付近に描画された白い点が重心であり、全ての指を伸ばした場合と1本指を伸ばした場合の矩形と重心に注目すると、矩形の高さが変化していないのに対し、重心は指先から遠ざかっていることが分かる。これを基に、本システムでは矩形と手領域の全ての接点と重心との距離を求め、その距離が最大となる接点を求めた。図4.10において赤い線で示すように、伸ばした指先の先端と重心との距離が最も大きくなり、指先座標を認識することが可能となる。

これまで述べた指先の認識手法による認識結果を図4.11に示す。指を上下左右に向けて認識を行っており、その向きに対してロバストに指先を認識出来ていることが分かる。また、試行を重ねた結果、衣服の袖が捲られて肌色領域が大きな場合でも十分な認識精度を得られることが分かった。

4.4 伝言メモの記録開始・終了ジェスチャの認識

本システムは、普段利用するオフィスデスクに設置することを想定しており、カメラにはその機の所有者の作業する姿が写り続ける。そのため、動きのみによる単純なジェスチャを何らかの操作に割り当てた場合、作業において手や顔などを動かした時に誤作動する可能性がある。伝言メモとして音声と映像を保存するが、これらのデータは圧縮しても大きなデータ領域を必要とするため、メモの記録ジェスチャが誤認識されることを避ける必要があった。

それに対し、我々は音声によるトリガを実装した。システムは常に机に設置したマイクから音声を取得しており、閾値を越える音量を検出した時に、その時間が短ければ机が叩かれ

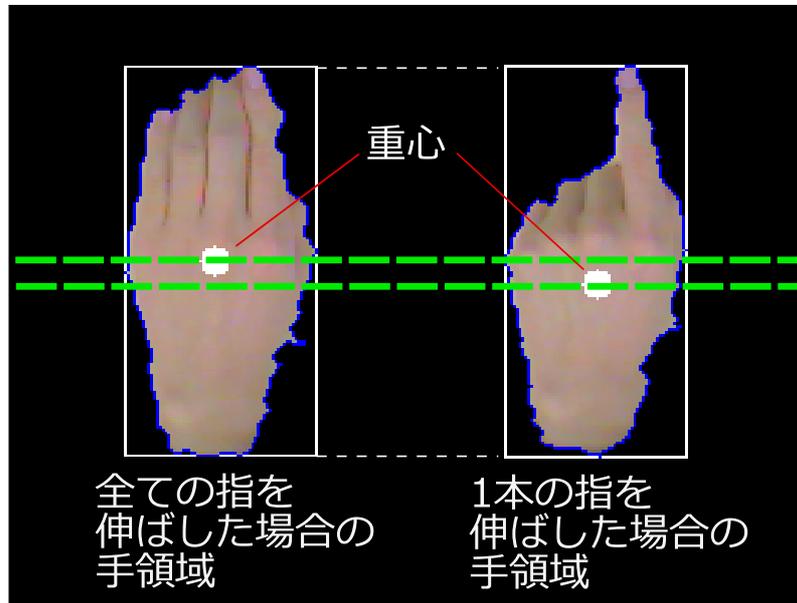


図 4.9: 手の形状変化における重心の移動

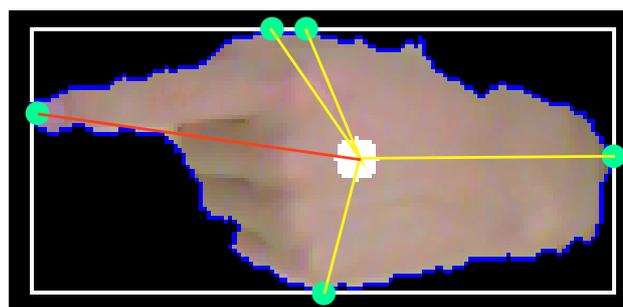
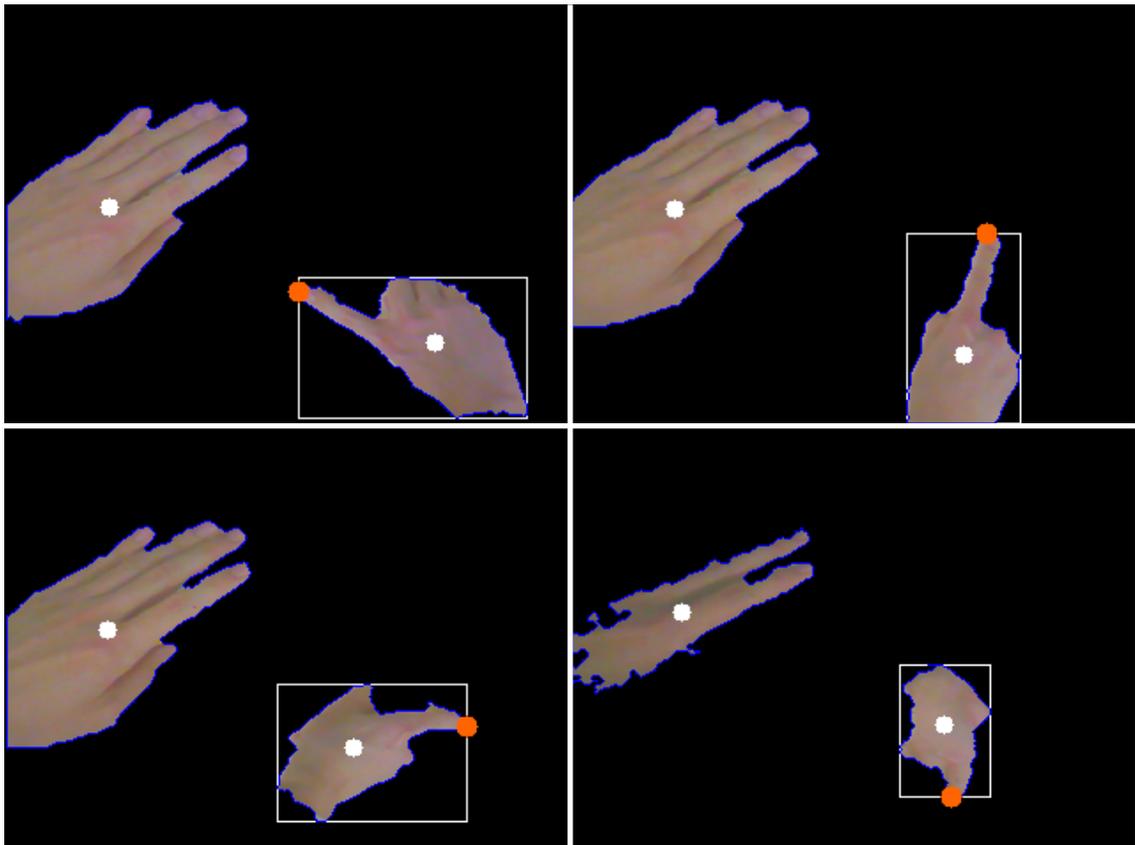


図 4.10: 矩形と手領域の接点と重心との距離

左を指した時

上を指した時



右を指した時

下を指した時

図 4.11: 指認識の結果

たと判断し、ジェスチャの認識を開始する。その瞬間、手を認識できなかった場合、音声トリガ検出状態に戻る。手を認識できた場合、伝言メモ開始ジェスチャとして認識し、システムは映像・音声の記録状態に入る。同様に、映像・音声の記録中に音声トリガを検出した瞬間に手が認識されると、伝言メモ終了ジェスチャとして認識され、記録を終了する。

4.5 伝言メッセージの記録

伝言メモの記録ジェスチャが行われ、システムが記録状態に入ると、映像・音声の記録を開始する。映像を撮影するカメラは手の認識を行うためのものと同様である。映像は OpenCV の関数を用いて、DivX による圧縮を行い avi ファイルとして保存する。

音声は、音声トリガと同様に、机に設置したマイクから取得する。音声取得用のスレッドを設け、バッファを2つ用意するダブルバッファリングを用いることにより、音声の取得処理と wav ファイルへの書き込み処理を同時に行い、連続的な音声取得を行っている。

4.6 アイコンの選択

アイコンの選択を行うため、マウスのような役割をするカーソルを実装した。投影された色による手認識のミスを減少させるため、カーソルの投影位置は、認識した指先座標に直接重ならないようにした。その位置は、図 4.12 に赤い点として示すように、重心座標と指先座標を通る直線を求め、その直線上とすることにより、指の延長線上にある感覚を与える。このカーソルをアイコンに重ねることにより、伝言メモの再生などを行うことが可能となる。

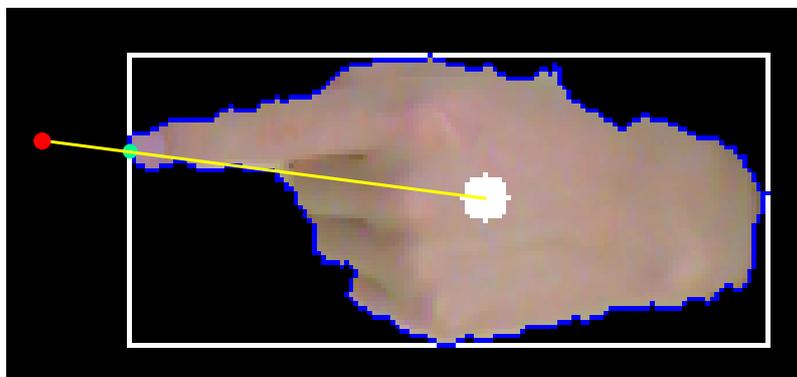


図 4.12: カーソル位置の決定

第5章 今後の課題・発展

5.1 試用とそこから得られた知見・課題

今回開発した実世界ビデオメモシステムを試用した．その様子を図 5.1 に示す．



図 5.1: 試用している様子

試用を通じて，即時記録性の高さを確認出来た．手で叩くようにして机に手を置く，という動作はメモを取ろうと思い立った後，即座に行うことが可能なため，瞬時に伝言メモの記録を開始することが出来る．即時記録性の高さは，より多くの情報をより素早く記録できることを意味するため，本システムの利用により，オフィスコミュニケーションを円滑化出来ると考えられる．

また，指先によるカーソル操作も安定して行うことが出来た．これにより，ペンとメモ用紙を探し，メモを書き始めるためにかかる時間よりも素早く，伝言メモの再生を行うことが可能であった．これも，オフィスコミュニケーションの円滑化に繋がるものと考えられる．

一方で，問題点も発見した．今回の実装では，認識のロバストさや設置の手軽さなどを考慮した結果，手の認識手法として通常の Web カメラによる肌色抽出を採用した．しかし，プロジェクタにより投影された映像が手に映ると認識が困難になり，その影響は予想以上に大きなものだった．投影された色に対しロバストな認識を行うためには，本研究により適した手法について再検討する必要がある．

また，紙のメモを書くように，文字を書く，絵を描くといった機能が必要だと試用中に感じ

た．映像や音声だけではなく，手描き機能を付加することにより，さらに豊かな表現を行い，柔軟な情報伝達を行うことができると考えられる．手描き機能の実装は今後の課題である．

5.2 発展

5.2.1 Natural Storage との統合

今回提案したメモシステムは，他人に対する伝言メモを支援するものであり，ポータビリティを持ち合わせていないため自分のための外部記憶には適さないと言える．逆に言えば，ポータビリティを付加することにより，自分のためにも，他人のためにも汎用的に使用することの出来るシステムになることが考えられる．ユビキタス環境下で情報のポータビリティを扱う手法として，Iwabuchi らによる Natural Storage[21] が挙げられる．Natural Storage は人の体をストレージとして用いる感覚を使用者に与え，コンピュータを意識することなく，情報の保存と閲覧を可能とする．本システムに NaturalStorage を統合することにより，有益だと判断したメモを自分の体に保存し，持ち運ぶことでいつでも情報を閲覧することが可能となる．また，体から取り出した情報を別な場所にメモとして貼り付けることで，ビデオメモを用いたより高度なコミュニケーションが可能になると考えられる．

5.2.2 個人の認証とネットワークの利用

伝言メモのコミュニケーションは非同期的であるため，伝言の対象者がメモを見るかどうかは分からない．そのため，伝言メモにより情報を伝えるには，メモを行うだけでなく，正しく伝わったかどうかを確認するというステップが必要になってしまい，それは利用者の負担となると思われる．これに対し，本システムに個人認証機能とネットワークによるメッセージ送信機能を実装することにより，誰がメモを行ったのかを明確に記録し，そのメモを閲覧したことを伝えるメッセージを伝えることが可能になると考えられる．

第6章 結論

本研究では，オフィス机を媒介とした伝言メモをコンピュータを用いて支援する手法の提案し，その手法を実現するシステムの開発を行った．本システムは，机をカメラとプロジェクタにより机を拡張することにより，他人の道具やコンピュータに直接触れることなく，ハンドジェスチャのみを用いた操作により，映像や音声を用いた伝言メモによるメッセージを残すことを可能にした．さらに，ハンドジェスチャのみを用いた操作により，残された机の持ち主がその再生・停止を行うこと可能にした．これらの機能により，伝言メモを用いた非同期的なオフィスコミュニケーションを行うことが出来る．今後は客観的な評価を行い，さらにそのフィードバックから改善を行っていきたい．

謝辞

本研究を進めるにあたり，指導教員の田中二郎教授をはじめ，高橋伸講師，三末和男准教授，志築文太郎講師には適切なご指導をいただきました．特に，Ubiquitous チームの担当としてご指導くださった高橋伸講師には，日頃から細やかな助言をいただきました．心より感謝を申し上げます．また，相談に乗っていただいた IPLAB の皆様や，研究生生活を支えていただいた家族にも深く感謝致します．

参考文献

- [1] 米澤朋子, 山添大丈, 寺澤洋子. ユーザ周囲の仮想ポータブル相対空間に貼って剥がせる音声の付箋. 17th Workshop on Interactive Systems and Software (WISS 2009) 論文集, pp. 149–150, 2009.
- [2] Tonchidot Corporation. Sekai camera support center. <http://sekaicamera.com/>.
- [3] Tomoko Yonezawa, Hirotake Yamazoe, and Hiroko Terasawa. Portable recording/browsing system of voice memos allocated to user-relative directions. In *Pervasive2009 Adjunct Proceedings (Demonstration)*, pp. 241–244, May 11-14, 2009.
- [4] Hiroyuki Tarumi, Ken Morishita, Megumi Nakao, and Yahiko Kambayashi. Spacetag: An overlaid virtual system and its applications. Proc. International Conference on Multimedia Computing and Systems (ICMCS '99), pp. 207–212, 1999.
- [5] Kathryn Elliot, Carman Neustaedter, and Saul Greenberg. Stickyspots: Using location to embed technology in the social practices of the home. In *Proceedings of the 1st international conference on Tangible and embedded interaction*, pp. 79–86, 2007.
- [6] 松原靖子, 小林一郎. 働く女性のための家庭内情報集約システムの開発. 人工知能学会全国大会論文集, pp. 179–182, 2008.
- [7] Wellner P. Interacting with paper on the digitaldesk. *Communications of the ACM*, pp. 87–96, 1993.
- [8] J Rekimoto and M Saitoh. Augmented surfaces: a spatially continuous work space for hybrid computing environments. *ACM Transactions Computer-Human Interaction*, pp. 307–322, 1999.
- [9] H Koike, Y Sato, and Y Kobayashi. Integrating paper and digital information on enhanced-desk: a method for realtime fingertracking on an augmented desk system. *ACM Transactions Computer-Human Interaction*, pp. 307–322, 2001.
- [10] Chih-Sung (Andy) Wu, Susan J Robinson, and Alexandra Mazalek. Wikitui: leaving digital traces in physical books. In *ACE '07: Proceedings of the international conference on Advances in computer entertainment technology*, pp. 264–265, 2007.

- [11] Willow Garage. Opencv wiki. <http://opencv.willowgarage.com/wiki/>.
- [12] NVIDIA. Nvidia cuda zone. http://www.nvidia.co.jp/object/cuda_home_jp.html.
- [13] Kazutaka Yasuda, Takeshi Naemura, and Hiroshi Harashima. Thermo-key: Human region segmentation from video. *IEEE Computer Graphics and Applications*, Vol. 24, No. 1, pp. 26–30, 2004.
- [14] Yoichi Sato, Yoshinori Kobayashi, and Hideki Koike. Fast tracking of hands and fingertips in infrared images for augmented desk interface. In *Proc. 2000 IEEE International Conference on Automatic Face and Gesture Recognition (FGR 2000)*, pp. 462–467, 2000.
- [15] Desney S. Tan and Randy Pausch. Pre-emptive shadows: Eliminating the blinding light from projectors. In *CHI '02: CHI '02 extended abstracts on Human factors in computing systems*, pp. 682–683, 2002.
- [16] Anthony Tang, Carman Neustaedter, and Saul Greenberg. Videoarms: Embodiments for mixed presence groupware. *People and Computers XX* Engage, pp. 85–102, 2007.
- [17] Philip Tuddenham and Peter Robinson. Distributed tabletops: Supporting remote and mixed-presence tabletop collaboration. In *Proceedings Second Annual IEEE International Workshop on Horizontal Interactive Human-Computer Systems TABLETOP '07*, pp. 19–26, 2007.
- [18] Sherrah Jamie and Gong Shaogang. Skin colour analysis. http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/GONG1/cvOnline-skinColourAnalysis.html.
- [19] 松橋聡, 藤本研司, 中村納, 南敏. 顔領域抽出に有効な修正 hsv 表色系の提案. *テレビジョン学会誌*, Vol. 49, No. 6, pp. 787–797, 1995.
- [20] 岡兼司, 陳欣蕾, 中西泰人, 佐藤洋一, 小池英樹. 拡張机型インタフェースのための複数指先の追跡とその応用. *情報処理学会論文誌: コンピュータビジョンとイメージメディア*, Vol. 44, pp. 73–84, 2003.
- [21] S. Iwabuchi, B. Shizuki, K. Misue, and J. Tanaka. Natural storage in human body. In *Proceedings of the 9th International Conference on Knowledge-Based Intelligent Information and Engineering Systems (KES2005)*, No. 4, pp. 430–436, Sep. 14-16, 2005.