NAIST-IS-MT9451019

修士論文

口唇の動き特徴量を用いた個人識別

大湊 吉行

1996 年 3 月 15 日

奈良先端科学技術大学院大学
情報科学研究科 情報処理学専攻

本論文は奈良先端科学技術大学院大学情報科学研究科に
修士 (工学) 授与の要件として提出した修士論文である。

大湊 吉行

指導教官 ： 千原 國宏 教授
横矢 直和 教授
佐藤 宏介 助教授

口唇の動き特徴量を用いた個人識別*

大湊 吉行

内容梗概

本論文はアクセスコントロールの新たなアプローチとして、口唇の動き特徴量を利用した個人識別について述べたものである。

顔画像を用いた従来の個人識別は静止した顔の幾何学的特徴を用いたものがほとんどであり，人間の動きの癖といった動的特徴を用いた個人識別はほとんどなされていない。そこで本稿では個人の動的な特徴として口唇の動きに注目し、口唇の動き特徴量を用いた個人識別方法を提案した。利用者は本人固有のパスワードを無声発話する。発話を記録したカラー動画像からフレーム毎に口唇色領域を色彩により切り出し、口唇の外接長方形を決定した。つぎにフレーム毎に開口度を計測し、整列することで時系列データが得られ、これを口唇の動き特徴量と定義した。得られた口唇の動き特徴量とあらかじめ登録されている特徴量の間で DP マッチングを行い、個人識別を行った。実験を行った結果、発話時間の長いものなら個人識別が十分可能であり、visual password として使用に耐えれることが確かめられた。

キーワード

口唇の動き特徴量、 開口度、 visual password、 DP マッチング、口唇の色領域

*奈良先端科学技術大学院大学 情報科学研究科 情報処理学専攻 修士論文, NAIST-IS-MT9451019, 1996 年 3 月 15 日.

Visual Person Identification Using Lip Motion Features*

Yoshiyuki Ominato

Abstract

This paper proposes a new approach of person identification using lip motion features for access control. This approach focuses on the information of the lip motion, whereas conventional approaches utilize still images which is containing the whole human face and are based on methods like template matching. The testee is asked to make a specific utterance, and the image sequence containing the mouth motion is taken by a video camera. For each frame of the image sequence, lip color area is extracted by HSL transform, and the circumrectangle of the mouth is determined. As a result of the above process, the opening ratio of the mouth (width versus height of the mouth) is estimated from the obtained bounding box of the mouth for each frame, and the sequence of the opening ratio of the mouth is registered as visual password. A person is identified by comparing the acquired sequence of the opening ratio of the mouth with the visual passwords in a database by DP matching. The experimental result shows that this method realizes good person identification when length of utterance is not short.

Keywords:

lip motion features, opening ratio, visual password, DP matching, lip color area

*Master's Thesis, Department of Information Processing, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-MT9451019, March 15, 1996.

目 次

1 はじめに **1**

2 個人識別 **3**

2.1 セキュリティ・サービス 3

2.1.1 アクセスコントロールにおける個人識別 4

2.1.2 侵入監視 4

2.1.3 防災監視 4

2.1.4 無人店舗などの運用 5

2.2 個人識別技術 6

2.2.1 個人識別手法 6

2.3 従来の顔画像を用いた個人識別方法 8

2.3.1 正面顔画像による個人識別 8

2.3.2 横顔による個人識別 10

2.3.3 三次元画像処理技術を用いた顔画像識別 11

3 口唇特徴を用いた個人識別 **12**

3.1 動的特徴 12

3.2 無声発話パスワード 13

3.3 口唇特徴パラメータ 15

3.4 識別処理の流れ 17

3.5 動画像取り込み 19

3.6 口唇特徴量抽出処理 22

3.6.1 口唇の色領域抽出 23

3.6.2 口唇領域の抽出 27

3.7 口唇存在予想領域の設定 29

3.8 マッチング処理 31

3.8.1 DP マッチング 32

4 実験・結果 **35**
4.1 実験装置および実験対象画像 35
4.1.1 実験装置 35
4.1.2 実験対象画像 38
4.2 実験 39
4.2.1 実験に対する制限 39
4.2.2 口唇領域抽出実験 39
4.2.3 考察 40
4.2.4 個人識別実験 41
4.2.5 実験結果 41
4.2.6 考察 47

5 結論 **50**

謝辞 **51**

参考文献 **52**

図 目 次

2.1 構造解析的手法の特徴量の一例 9
2.2 濃度値のマッチングにおける特徴量例 9
2.3 横顔輪郭線と特徴点の一例 10
2.4 ３次元顔画像計測一例 11

3.1 概念図 14
3.2 口形パラメータ 15
3.3 処理のながれ 18
3.4 動画像取り込みシステム構成 20
3.5 DFM メモリ内の画像データ構造 21
3.6 DFM の構成 21
3.7 口唇特徴抽出処理 22
3.8 明度 ・色相 ・彩度の関係 24
3.9 口唇の部分画像例 25
3.10 口唇色領域の色相 ・彩度 ・明度のヒストグラム一例 26
3.11 彩度 ・色相図からのカラー領域分割図 26
3.12 ラベリング走査領域 27
3.13 処理画像結果 28
3.14 口唇存在予想領域 29
3.15 開口度測定結果一例 31
3.16 A、B 2 つの時系列の時間軸の対応づけ 33

4.1 システム構成 36
4.2 DFM (フレームグラバ) 36
4.3 DFM 外部クロックアダプタ 37
4.4 indigo2 外観 37
4.5 フレームメモリユニット MPU-F100A 37
4.6 入力画像例 (口唇周辺画像) 38
4.7 口唇領域抽出結果 39
4.8 異なる発話時におけるコスト関数値 42

4.9 異なる発話時における認識結果 42
4.10 無声（閉口状態）時におけるコスト関数値 44
4.11 無声（閉口状態）における認識結果 44
4.12 『あ』発話時におけるコスト関数値 45
4.13 『あ』発話時における認識結果 45
4.14 『千原研究室』発話時におけるコスト関数値 46
4.15 『千原研究室』発話時における認識結果 46

表 目 次

2.1 個人識別技術 6

3.1 DFM の仕様 20
3.2 口唇の動き特徴量計測時間の比較 30

4.1 個人識別結果 43

第 1 章 はじめに

現在は情報化社会といわれ、情報の確保、機密保持がより重要視されてきている。さらにコンピュータの高速化とそれに伴う画像処理能力の向上、および社会への急速な普及により、従来は人間もしくは半自動で行っていたセキュリティ・サービスの自動化 ・インテリジェンス化が試みられている。セキュリティ ・サービスは、アクセスコントロール、侵入監視など多岐にわたっている。なかでも、情報化が進展する現代社会において、情報のアクセス管理に用いる個人識別の重要性はより高まってきている。

現在、個人識別は主にパスワードや ID カードなどを用いて行っているが、常に盗難や本人詐称の危険性をはらんでおり、その改善が求められている。そこで最も詐称が困難であり , 生来我々人間の誰もが持つ特徴である指紋や顔などの肉体的属性を利用した個人識別が注目されている。

なかでも顔による個人識別は人間が日常行っているようにもっとも自然な個人識別法であり、指紋などと違いプライバシーに関わる抵抗感の問題も少ない。また非接触での識別が可能であるため近年のマンマシンインターフェース向上への要求と相俟って、数多くの個人識別の報告がなされている [2][7]。

しかしながら従来の顔画像を用いた個人識別は、顔の部位の位置的情報や形状情報などのいわば顔の静的情報を用いたものが大半であり、静的情報を用いたアプローチはほとんど出尽くしてきたと思われる。

この顔画像認識手法の限界を打破するために、本研究では顔の動的情報、すなわち人間の動き、しぐさの個人差を用いた個人識別を提案する [3]。動きの特徴量としては、人間の顔の部分でも最もその動きの変化に富み、その動きに特徴がある口唇に焦点をあて特徴量として用いることにする。

利用者には本人特有のパスワードを発話してもらい、これらの発話から得られ

た口形変化より時系列特徴量を抽出する。この際利用者は実際に発声を行わず、口形のみにより個人の識別を行う、いわゆる Visual Password [1] を用いた個人識別手法の構築を目的とする。

第 2 章では研究の背景となるセキュリティ・サービスの概要、個人識別技術の紹介を行う。

第 3 章では本研究で行う動き特徴パラメータに対する説明と口唇の切り出しから口唇の動き特徴量なる口唇の時系列特徴量の計測までの一連の処理について述べる。

第 4 章では本研究で提案した口唇の動き特徴量を用いた個人識別の有用性を調べるための実験と結果について述べ、それに伴う考察を行う。

最後に結論で本研究の内容についてまとめ、今後の展望について述べる。

第 2 章 個人識別

本章では、本研究の背景としてセキュリティとその要素技術の 1 つである個人識別技術の紹介をした後，なかでも顔情報を用いた個人識別の従来手法について詳しく述べる。

2.1 セキュリティ・サービス

セキュリティとはもともと日常生活の安全や安心を供給するサービスである [16]。セキュリティ・サービスは侵入・盗難などの防犯、火災・設備異常などの防災はもとより道路・鉄道などの危険防止など、多岐にわたっている。当初は監視人による巡回等で安全確認を行うといった労働集約的サービスが主であったが、現在はセンサ技術の発達や監視システムの統合に伴う集中管理的な機械警備の体制が主流となっている。

しかしながらセンサ技術等の発達した現在でも最終的な判断は人間による監視に頼っており、そのために疲労や不注意による単純ミスの危険が常に伴う。

このような人為的ミスを防ぐために画像処理技術を活用し、サービスの自動化・省力化・均質化を行う動きが活発である。以下にその適用分野について概観する。

2.1.1 アクセスコントロールにおける個人識別

施設や情報処理システムへのアクセスを管理することをアクセスコントロールと呼んでいる。部屋や施設への出入りを管理する最も原始的な方法は、鍵の所持の有無によるか守衛か門衛による目視確認によるかである。鍵については物理的な鍵の他にIDカードや暗証番号入力などの電子的な鍵もある。一方、情報システムへのアクセスにはパスワード、IDカード、暗証番号などで本人の確認を行っている。いずれにしても、鍵の紛失・盗難、暗証番号の盗用に対する本人詐称を防ぐには不十分である。

このような問題に応えるのが、本人固有の生体情報、例えば指紋、掌形、眼底血管パターンあるいは顔を用いて個人識別を行う方式で一部は実用化されている。

2.1.2 侵入監視

侵入監視は敷地や建物、あるいは侵入禁止区域への故意の侵入、あるいは不注意による侵入などを監視して財産の保全や危険を未然に防止するセキュリティ・サービスの一つである。

このサービスは主に監視人による定点監視や巡回監視、あるいはTVカメラによる遠隔監視により行っている。これらはいずれも人手に頼るものであるために、疲労・不注意などによる人為的ミスがつきまとう。そこでカメラにより監視し、それを人手で監視するのではなく画像処理技術を活用することによる自動監視システムを実現することが望まれている。

2.1.3 防災監視

都市火災や森林火災の発見、気象条件による延焼予測、消火計画、避難誘導などは災害を最小限に抑える上で重要なセキュリティ・サービスの一つである。この分野の監視においても侵入監視の項目で述べたようなＴＶカメラによる定点監視あるいは無線操縦ヘリコプターによる移動監視がサービスの主流となっている。

いずれも人手による常時監視もしくは長時間の監視が前提であるため機械化・省力化に画像処理技術の活用が期待されている。

2.1.4 無人店舗などの運用

CD/ATM ブースをはじめ自動販売機コーナなどの無人店舗は今後ますます増加の傾向にある。この分野において開店時の不正防止、機器の障害チェック、客のモニタリング、また閉店時の客のシャッタ閉じ込め防止など画像処理技術への期待は大きい。

上記の他にも鉄道関連での軌道内の障害物のチェックや駅ホームの安全チェック、道路の安全確保など様々な分野に画像処理技術の適用が期待されている。

これまでセキュリティ・サービスにおける画像処理技術の適用分野について概観してきたが、なかでも特に施設・建物への入出管理を正しく行ったり、あるいはコンピュータシステム内の情報資源への不当なアクセスを防ぐための技術である個人識別技術が近年の情報化社会とコンピュータの社会への浸透と相俟って注目され、その自動化・インテリジェンス化が望まれている。2.2節では個人識別技術の紹介を行う。

2.2 個人識別技術

2.2.1 個人識別手法

一般的に考えられる主な個人識別手法 [10] について考えると、おおよそ表 2.1 のような分類がなされる。

表 2.1 一般的な個人識別手法

	識別対象	具体例
0	目視確認	守衛，門番
1	所持品	鍵，印鑑，ＩＤカード等
2	固有の記憶情報	パスワード，暗証番号等
3	習慣化された動作による生成物	サイン，筆跡等
4	肉体的な属性	指紋，顔，網膜等

これらは (0) → (4) の順に詐称や偽造による不正使用が困難になるものと考えられる。近年のコンピュータ社会における自動化・省力化へのニーズを考えると (0) を用いる方法から、(1) ~ (4) の方法へ移行している。

現在，コンピュータによる個人識別においては、(1) のなかでも機械的な認識が容易な ID カード等の磁気記録を用いたものや、(2) の暗証番号を用いた個人識別システムが主に実用化されている。しかしながら現在実用化されているような ((1),(2) に分類される) ものは、盗難・紛失等による本人詐称が避けられないものがある。

このような問題に対して表 2.1の (4) に示してあるような肉体的な属性を用いて個人識別を行う試みが数多くなされている。その中でも最も実用化が進んでいるものは指紋を用いたものである。

しかしながら、指紋による個人識別はその使用用途が犯罪を連想させることから心理的抵抗が大きく利用者に嫌悪感を抱かせる可能性があることと、指紋を採取することによるプライバシー侵害の問題という欠点が存在する。この点が指紋を個人識別に適用する際の大きなネックとなっている。また網膜パターンを用い

る個人識別法は装置に顔を密着させなければならないなどの煩雑さが存在する。

これらの肉体的な属性を用いた個人識別法に対して，顔を用いた個人識別は次に示すような優位性を有している。

- 提示が容易
- 識別装置と非接触で識別可能
- 詐称が困難
- プライバシーによる抵抗感が少ない

上記に示した利点から顔画像を用いた個人識別手法に対する研究が数多く行われている。2.3節では顔画像を用いた個人識別方法に関する紹介を行う。

2.3 従来の顔画像を用いた個人識別方法

本節では従来の顔画像を用いた個人識別方法についての紹介を行う。顔による個人識別は人間が日常行っている最も自然な個人識別であり、顔は生まれた時から常に世の中に公開しているので、指紋・掌形と異なりプライバシーによる抵抗感の問題が少ないと言われている。顔を用いた個人識別手法の代表例を以下に示す [15]。

2.3.1 正面顔画像による個人識別

正面顔画像を用いた個人識別は主に以下の様に大別することが出来る。

1. 構造解析的手法

   眼・鼻・口・耳などの顔の構成要素と顔自体の輪郭線を抽出し、それらの形状・大きさ・相対的位置関係等を解析して図 2.1に示すようなパラメータ表現して特徴量とする手法である。この手法は特徴量が直観的で扱いやすいという利点があるものの、各パラメータを正確に抽出するための複雑な処理が必要であり、精度の高い自動抽出が困難である。

2. 濃淡値のマッチングによる手法 [12]

   顔画像の濃淡値を直接、特徴量と見なして画像どうしの重ね合わせによる類似度等を用いてマッチングを行う手法である。この方法は原理は簡単であるが、撮影するときの照明条件や重ね合わせる際の大きさと位置の正規化の精度に識別精度が大きく影響されるという欠点がある (図 2.2)。

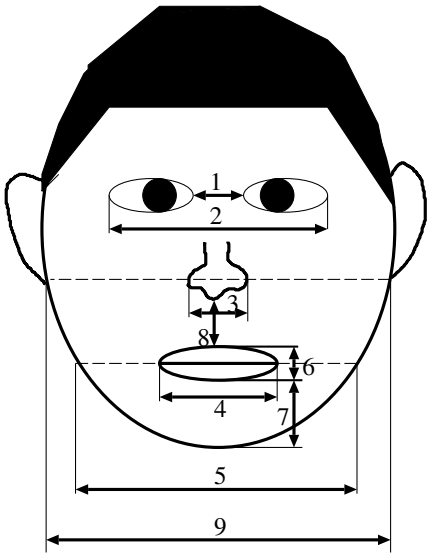

図 2.1 構造解析的手法の特徴量の一例

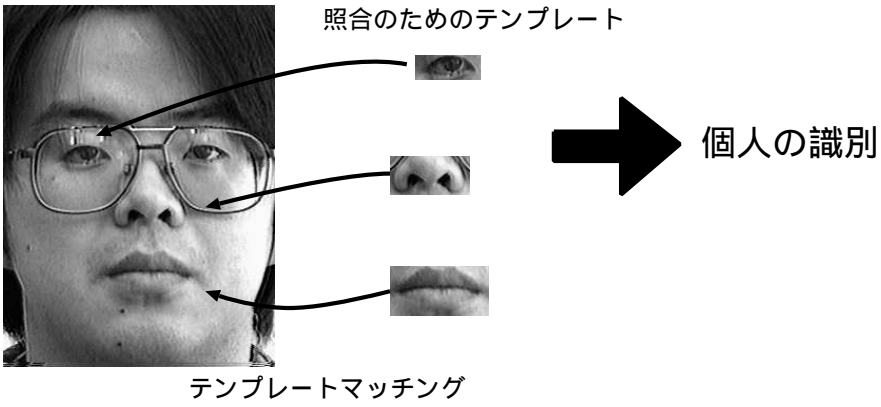

図 2.2 濃度値のマッチングにおける特徴量例

2.3.2 横顔による個人識別

人物横顔像より輪郭線を求め、輪郭線上からいくつかの特徴点を抽出し、その特徴点の座標を用いて個人を識別するのが主な手法である。

Hermon らは図 2.3に示すように横顔から 9 個の特徴点を抽出し、これらの特徴点より距離 ・面積 ・曲率などを示す 11 個の特徴量を求めて、特徴ベクトルを作成し個人の識別を行っている。この手法で 121 名より 3 枚づつ、すなわち 363 枚の横顔画像を用いて識別実験を行った結果、86 ~ 100% の識別率を得ている。しかしながら輪郭線は人手により作成されており、これを計算機に入れて特徴点を抽出しているので、完全自動化方式とはいい難い。

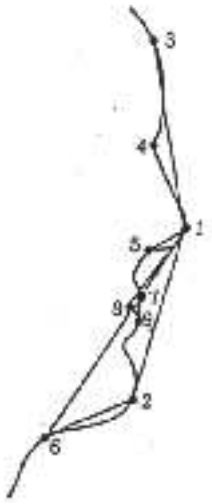

図 2.3 横顔輪郭線と特徴点の一例

2.3.3 三次元画像処理技術を用いた顔画像識別

上述したような 2 次元顔画像を用いて特徴点を抽出し、個人識別を行う手法の難点は前髪 ・ヒゲ ・眼鏡等の影響により、特徴点が正確に抽出できないことである。 3 次元画像処理はこれらの難点を克服する手段として期待されている。

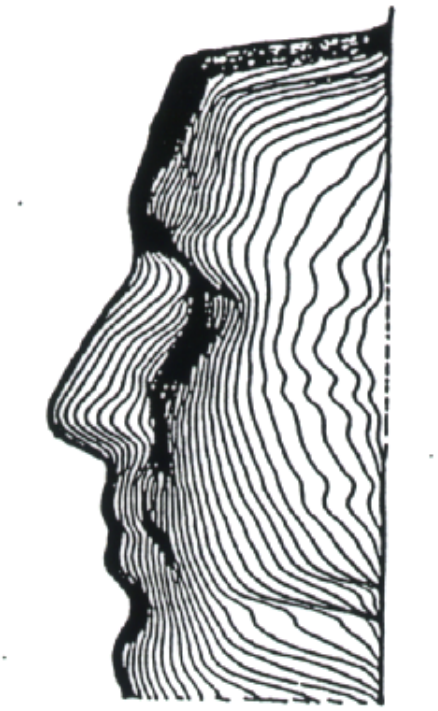

図 2.4　3 次元顔画像計測一例

しかしながらこの手法も 3 次元データの位置合わせや、確実かつ高速な 3 次元データ獲得法が困難であるという欠点 (技術的問題) が存在する。

その他の研究としてはニューラルネットワークを用いて顔画像識別を行うものや、顔画像を KL 展開して展開係数の比較によって識別する手法などが提案されている。いずれの手法もマッチング領域の位置ずれなどの変動に弱いという欠点が存在する。

第 3 章 口唇特徴を用いた個人識別

本章ではまず口唇の動的特徴について述べ、それを用いることによる利点について述べる。つぎに口唇特徴量について述べ、口唇特徴量の抽出 ・マッチングなど一連の処理のながれについて述べる。

3.1 動的特徴

通常人間は瞬時に相手が自分の知人であるか否か，そして知人であればその人が誰なのかを簡単に判断することが可能である。また実際に会わなくとも写真やテレビ画面を通して、人物を判断することが可能である。これから顔画像は個人識別を行う上で重要な情報を有していることがわかる。しかしながら 2.3節で述べたような顔画像認識の基本となる顔の位置的情報、形状情報のみに頼る個人識別手法は限界に近づきつつある。

その一例として、一般に人間は一見して区別が困難な一卵性双生児を識別する場合、その相手を知ることによって二人の差異を認識することがある。これは人間が顔情報のみならず，人間のしぐさ、ふるまいからも相手を判断しているからである。すなわち人間は後天的ではあるがそのしぐさ，ふるまいにも個人識別の手がかりとなるような特徴を有しており、それ故たとえ容姿が似ている一卵性双生児でも識別することが可能となるのである。

この例にならうと、一卵性双生児などの場合、静止した顔画像を用いて個人識別を行うのは困難である。そこでこの静的といえる従来顔画像認識手法に対し、我々は顔の動画像を用いた個人識別手法を提案する。

3.2 無声発話パスワード

本研究は顔の部位の中でも特に動きの変化に富み、もっとも個人の特徴を有していると考えられる口唇に注目し、口唇の動き特徴量を定義し、それを用いた個人識別法の構築を目的とする。

話者は本人固有の発話を行う。発話の際の画像情報から得られる口唇特徴量の時系列データを口唇の動き特徴量とし、個人の識別を行う。なお、この時音声信号は記録しない。なぜなら、本研究の目的が動き情報のみでどれほど効果のある個人識別結果が得られるかを評価することにあるからである。

また、口唇の動き特徴量を用いた個人識別には通常の静止顔画像を用いた個人識別と比較して以下に示すような 2 重のセキュリティが得られるといった利点がある。

- パスワード (本人固有の発話) ... 非公開情報
- 顔 ... 公開情報

一般にパスワードは個人の性格より類推され、盗用される恐れがある。しかしながらたとえパスワードが盗用されたとしても、口唇の動き特徴量から行う個人識別が存在することによりパスワードの盗用のみでは本人詐称は不可能である。

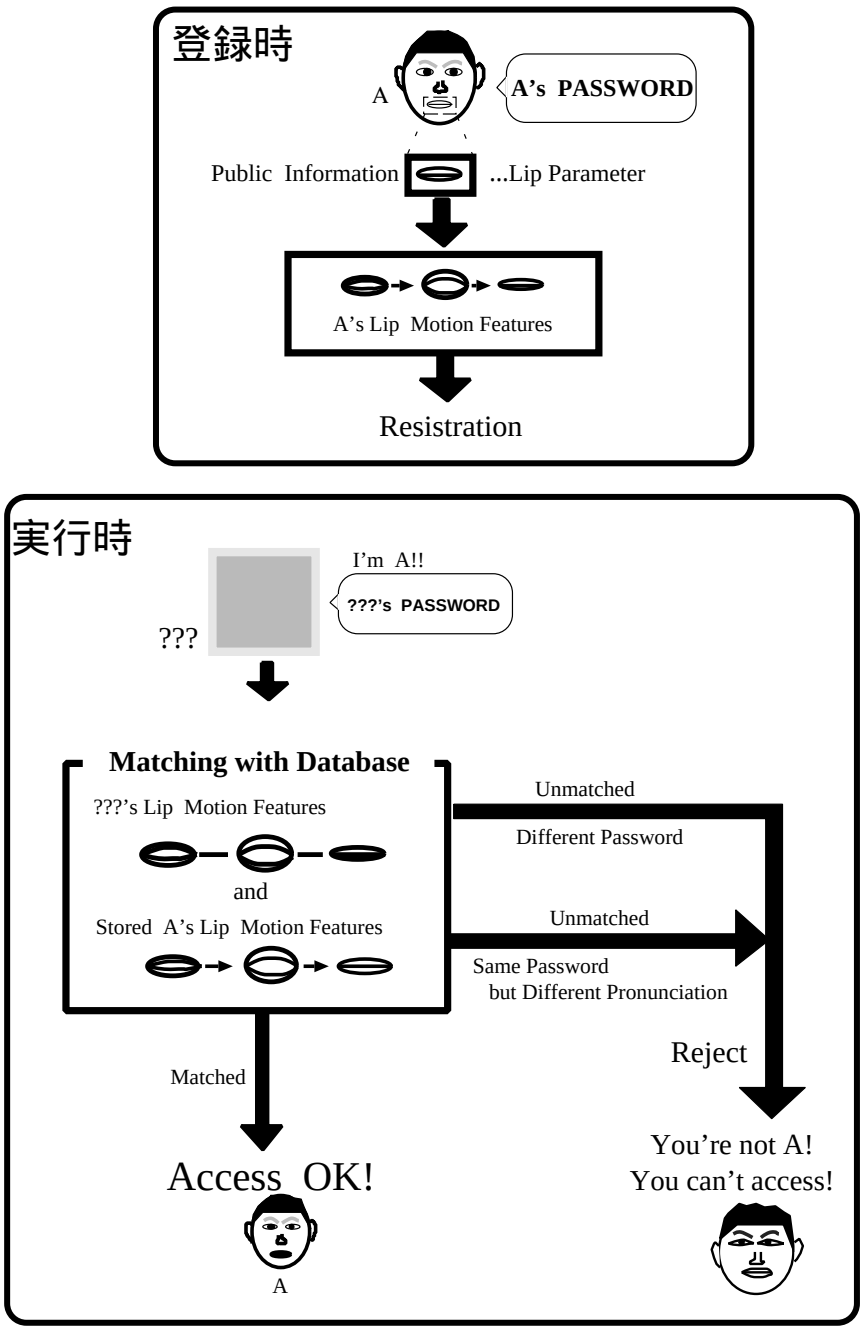
登録時
A
A's PASSWORD
Public Information
...Lip Parameter
A's Lip Motion Features
Resistration
実行時
I'm A!!
???
???'s PASSWORD
Matching with Database
???'s Lip Motion Features
and
Stored A's Lip Motion Features
Unmatched
Different Password
Unmatched
Same Password
but Different Pronunciation
Reject
Matched
Access OK!
A
You're not A!
You can't access!

図 3.1 概念図

3.3 口唇特徴パラメータ

口の動きは主に次に示す 7 つのパラメータによって表現できると考えられている（図 3.2）[21], [20]。

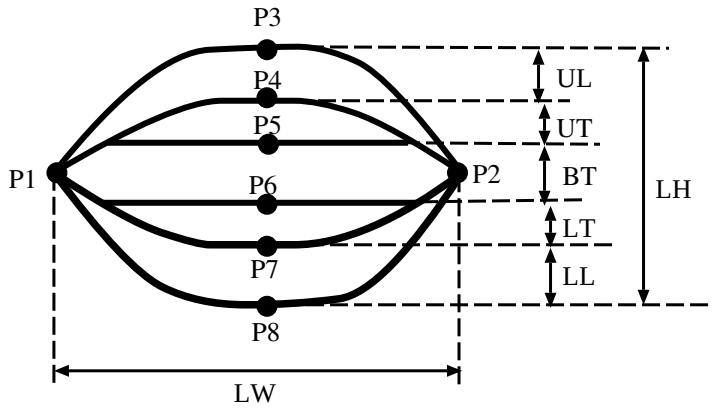

図 3.2 口形パラメータ

- 口の横幅 LW
  唇の左端と右端の間の距離
- 口の開き LH
  上唇上端と下唇下端の間の距離
- 上歯幅 UT
  上歯の見えている部分の上端と下端の間の距離
- 下歯幅 LT
  下歯の見えている部分の上端と下端の間の距離
- 歯間幅 BT
  上歯下端と下歯上端の間の距離
- 上唇幅 UL
  上唇の上端と下端の間の距離

- 下唇幅 LL
  下唇の上端と下端の間の距離

この中でも特に発話に際し、変化が大きいと思われる口の横幅 LW と口の開き LH を用いた開口度 OR を口形パラメータとして定義する。開口度 OR は (3.1) 式のように表現される。

$$OR = \frac{LH(\text{上唇上端と下唇下端の間の距離})}{LW(\text{唇の左端と右端の間の距離})} \tag{3.1}$$

3.4 識別処理の流れ

口唇の動き特徴量を用いた個人識別を行う際の処理のながれを図3.3に示す。カラー撮像した口唇周辺動画像を動画像メモリに取り込んだ後、計算機に1フレーム毎にデータを取り込み、口唇の特徴量の抽出処理を行う。入力動画像全てのフレームに対して口唇の特徴量を抽出し、得られる時系列データを口唇の動き特徴量と定義する。このデータに対して登録データとのマッチングをし、その結果から個人の識別を行う。

なおフレームレートは$30(frame/sec)$である。発話に当たってはまず口の閉じた状態を初期状態とし、話者はなるべく自然な状態で発話するようにする。また発話終了後にも口を閉じてもらうようにした。以下、各々のサブブロックの処理について説明する。

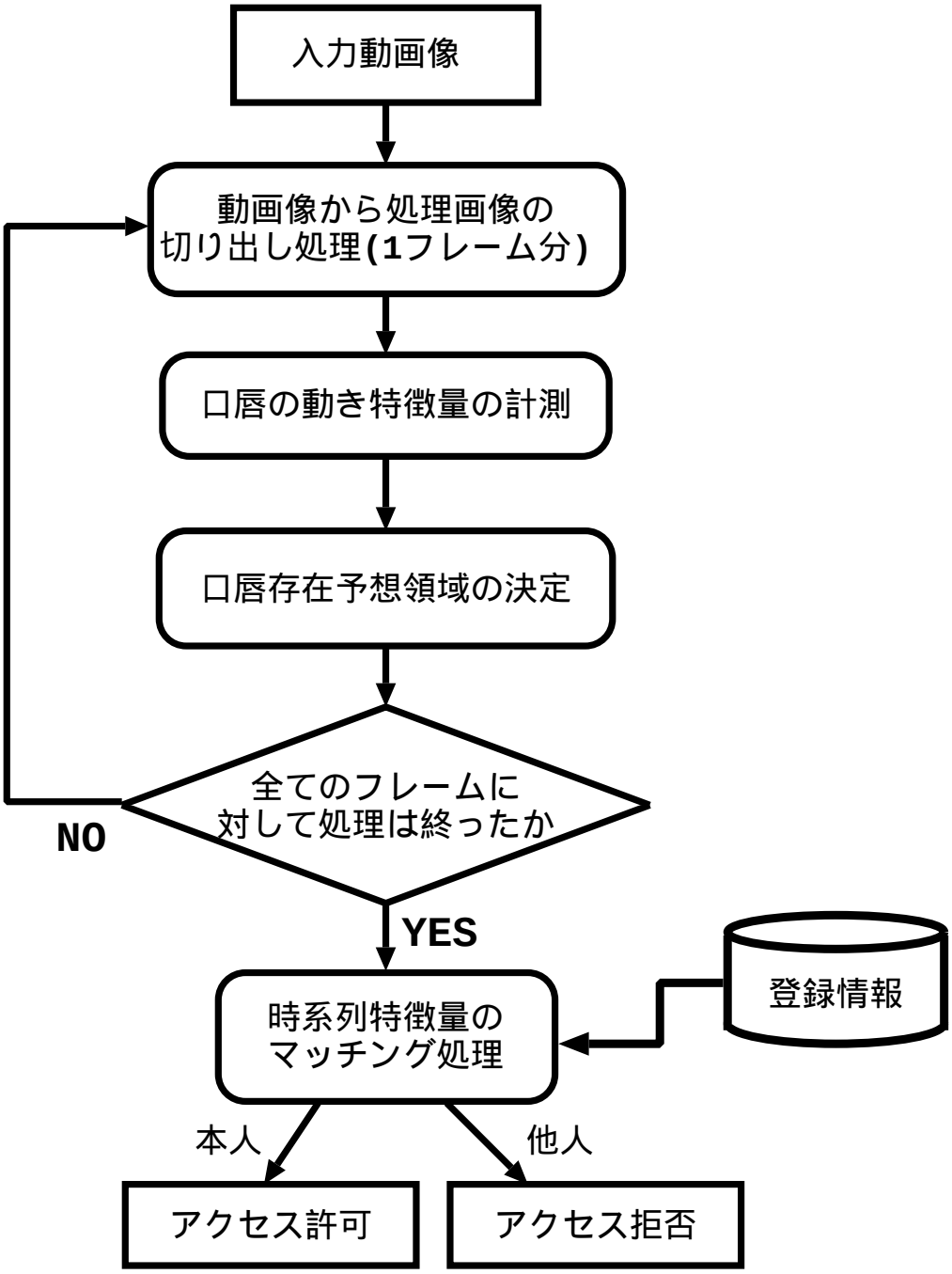
入力動画像
動画像から処理画像の
切り出し処理(1フレーム分)
口唇の動き特徴量の計測
口唇存在予想領域の決定
全てのフレームに
対して処理は終ったか
NO
YES
登録情報
時系列特徴量の
マッチング処理
本人
他人
アクセス許可
アクセス拒否

図 3.3 処理のながれ

3.5 動画像取り込み

動画像の取り込みは動画像取り込み装置DFM (Dynamic video Frame Memory、計測技術研究所製) を用いて以下の手順で行われる。動画像取り込みのためのシステム構成を図 3.4に示す。

1. カラー CCD カメラより口唇周辺画像を撮像し、ビデオデープに記録
2. ビデオテープを VTR により再生し、再生された画像をフレームメモリユニットを介してDFM 外部クロックアダプタよりDFM に入力し、DFM 内の外部拡張メモリ (2GB) にマップ
3. 外部拡張メモリに取り込まれた動画像ファイルを 1 フレーム単位でホストコンピュータのメインメモリ上に伝送
4. 伝送された 1 フレーム分の口唇周辺画像データに対し以下に述べるサブブロックの処理を行う

また、DFM の仕様を表 3.1に、DFM メモリ内の画像データ構造を図 3.5に示す。

表 3.1 DFM の仕様

映像信号	ハイビジョン、NTSC,EDTV 等
画像データ	8bit/10bit、カラー/モノクロ切替え可能
メモリ	内蔵メモリ (16MB)、外部拡張メモリ (2GB)

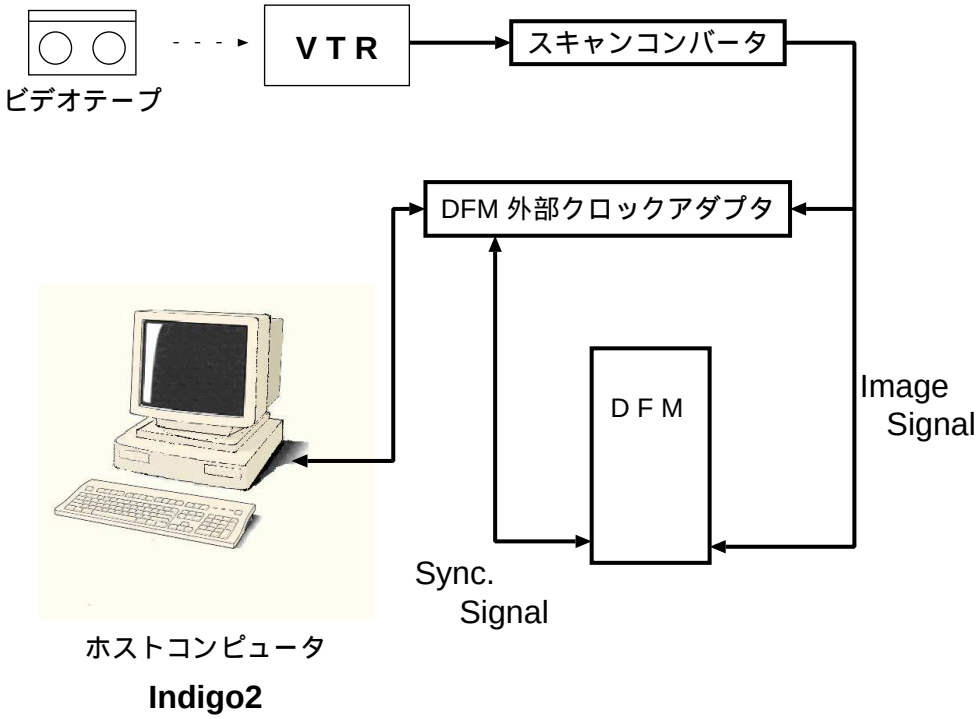

図 3.4 動画像取り込みシステム構成

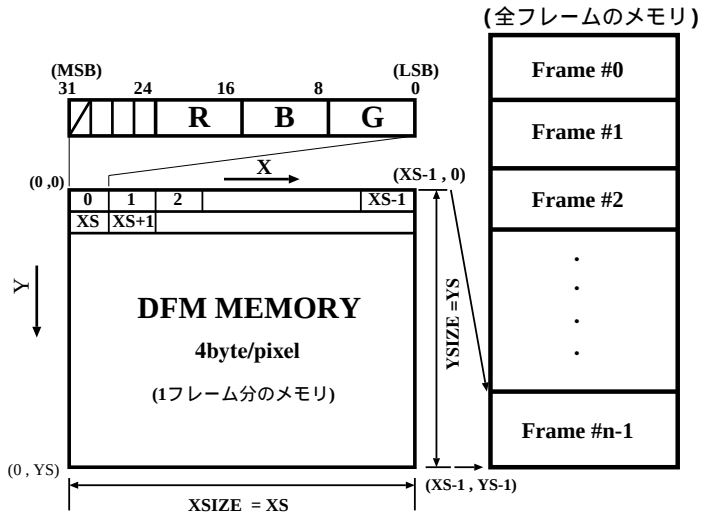

図 3.5 DFM メモリ内の画像データ構造

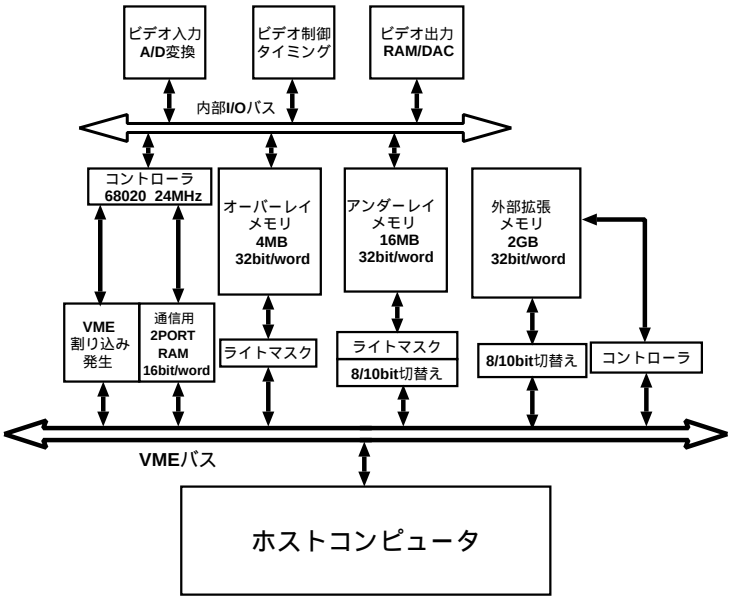

図 3.6 DFM の構成

3.6 口唇特徴量抽出処理

本節では図 3.3の中のサブブロックである口唇特徴量の計測処理について述べる。処理の流れを図 3.7に示す。口唇の抽出処理に関する研究は数多く存在している [8][9][5] が、本研究では、色情報を用いて口唇領域を切り出し、口唇特徴量の計測を行う。色情報を用いた口唇の切り出しは、環境に拘束される反面、口唇形状をモデル化することなく切り出しが可能であるという利点が存在する。

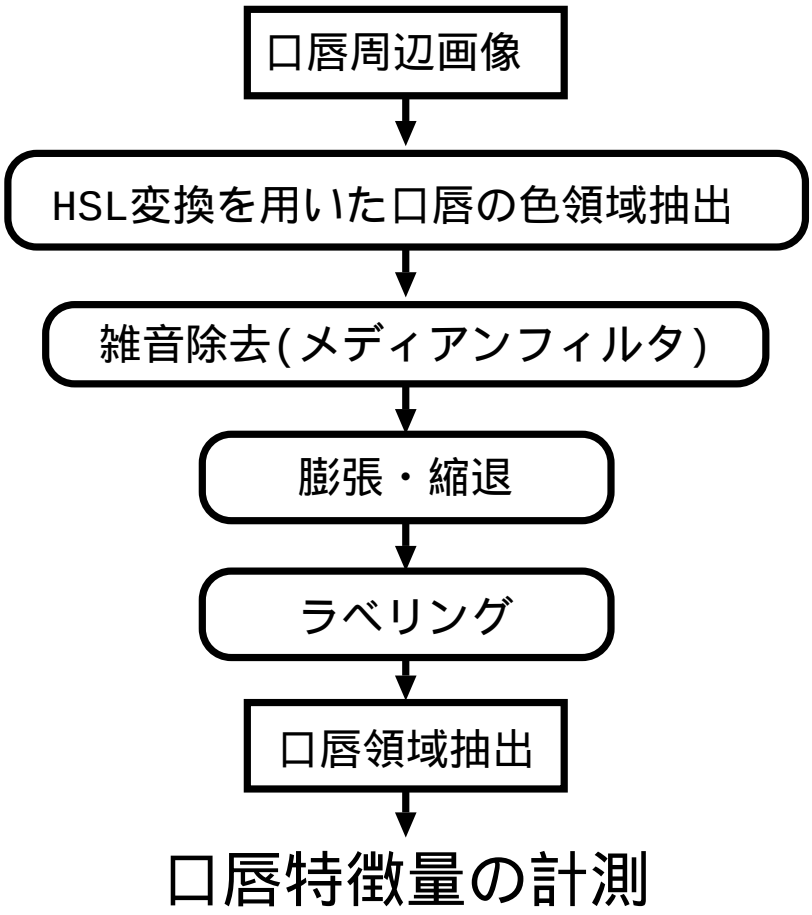

図 3.7 口唇特徴抽出処理

3.6.1 口唇の色領域抽出

口唇の切り出しには色情報を利用した口唇の色領域の切り出しを行う。通常、計算機上に取り込まれ、電子化されたカラー画像は、R,G,B 各々の輝度信号として分解され保存されている。このような色情報の表現方法は RGB 表色系といわれている。一般的に R,G,B 各々の信号間相互の相関が高く、また人間の心理的な色知覚とは異なっており、RGB 表色系は実際に色情報を扱った処理を行うのには不適切である [19]。

そこで本稿では各属性の相関が非常に低く、高速で変換が行え、かつ人間の色知覚に近い HSL 表色系を利用して、口唇の色領域の抽出を行う。以下、HSL 表色系 [13] について述べる。

HSL 表色系

R,G,B 各々の輝度信号に対し、色を明度 (Lightness)・彩度 (Saturation)・色相 (Hue) の 3 属性に変換する表色系である [17]。明度 L は物体表面の相対的な明暗に関する色の属性を示したものであり、R,G,B 各々の輝度値を用い (3.2) 式により変換される。

$$L = \frac{(R+G+B)}{3} \tag{3.2}$$

彩度 S は物体表面の色味の強さ、飽和度を示したものであり $X = Max(R,G,B)$、$Y = Min(R,G,B)$ とした時、(3.3) 式により変換される。

$$S = \frac{(X-Y)}{X} \times 255 \tag{3.3}$$

色相 H は赤・黄・緑・シアン・青・マゼンタなどの色知覚を特徴づける色の属性をいい、(3.4) 式のように場合分けし定義されている。

$$\begin{aligned} X = R\text{の時} & \\ H &= \left(\frac{G-B}{X-Y}\right) \times 60 \\ X = G\text{の時} & \\ H &= \left(\frac{B-R}{X-Y} + 2\right) \times 60 \end{aligned} \tag{3.4}$$

$X = B$の時

$$H = \left(\frac{R-G}{X-Y}+4\right)\times 60$$

但し ,$H < 0$ の時 $H = 360 + H$

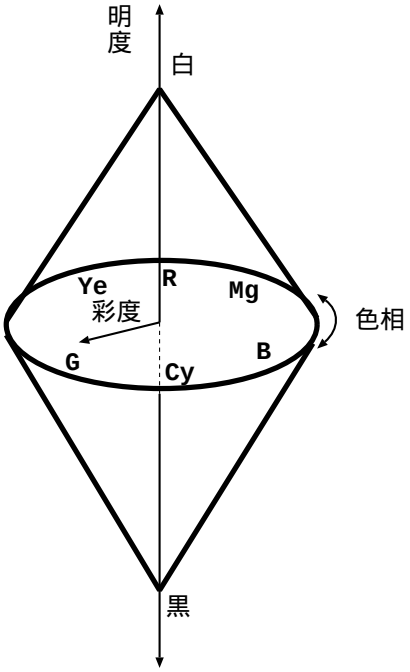

図 3.8 明度 ・色相 ・彩度の関係

口唇の色領域の決定

HSL 表色系に変換された口唇周辺画像に対して、口唇の色領域と思われる領域を切り出す。口唇の色領域は前もって手動で切り出しておいた口唇の部分画像 80 枚に対して色相 ・明度 ・彩度各々のヒストグラムを作成し、これを利用して口唇の色領域を決定した。口唇の部分画像の一例を図 3.9に、得られた色相 ・明度 ・彩度のヒストグラムを図 3.10に示す。口唇の色領域はヒストグラムを参照し、以下に示す色領域に存在すると考えた。

- 色相 H : 0 ~ 30 , 330 ~ 360
- 彩度 S : 120 ~ 240

なお、明度 L は照明の影響等、外乱の影響を受けやすい性質を持っているので、使うことを避ける。この色領域に存在する画素に 255、それ以外の画素に 0 の値を与える。

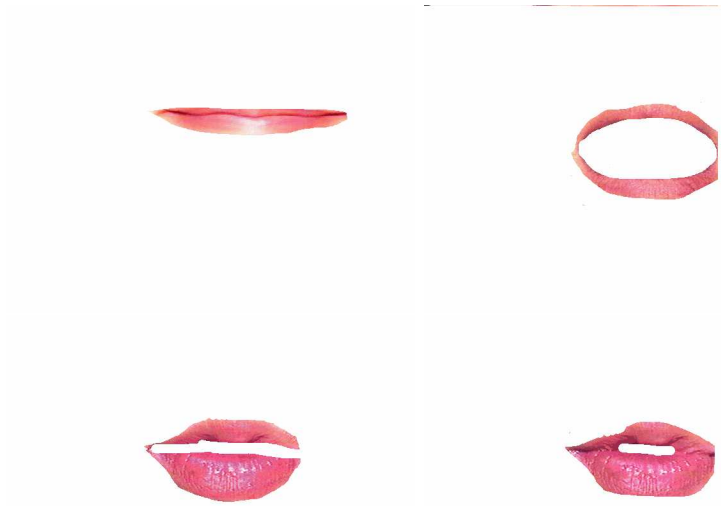

図 3.9 口唇の部分画像例

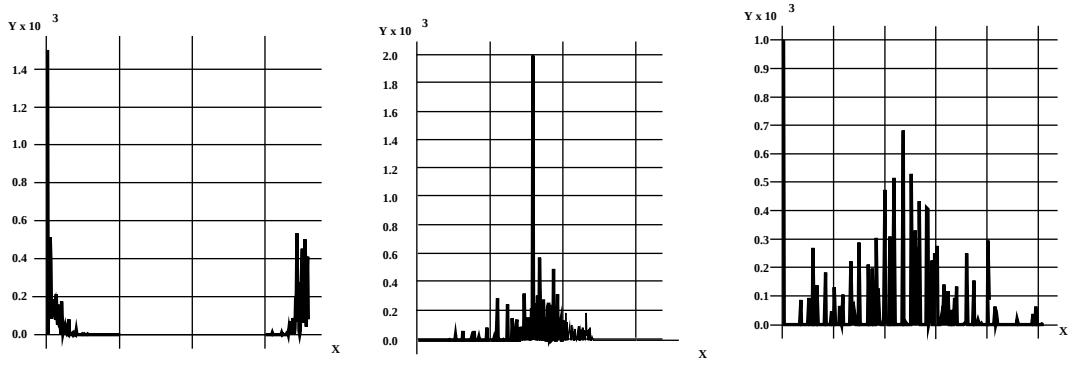

図 3.10　口唇色領域の色相 ・彩度 ・明度のヒストグラム一例

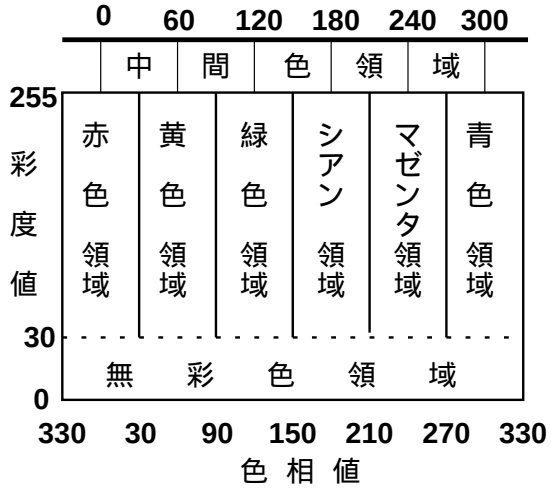

図 3.11　彩度 ・色相図からのカラー領域分割図

3.6.2 口唇領域の抽出

3.6.1節によって口唇の色領域が抽出されるが図3.13を見てもわかるように口唇領域と思われる場所以外にも口唇の色領域と同じ色領域をもつ画素やごましお雑音が存在している。以下の順に処理を行い、雑音の除去→ 口唇の輪郭抽出を行う[23],[11],[14]。

1. メディアンフィルタを用いて、ごましお雑音の除去を行う。
2. 口を大きく開けた時に時々生じる上唇部分と下唇部分の分離への対策として膨張 ・縮退処理を行う。
3. 得られた画像に対してラベリングを行い、口唇の外接長方形をとる。ここでラベリングを行う走査領域は第1のフレームでは図3.12の範囲、第2フレーム以降は、3.7節で決定される口唇存在予想領域とする。この処理によりラベリングされた画素のなかで存在する最も大きな連結成分を口唇とする。
4. 外接長方形をとり、口唇の特徴量である開口度を求める。

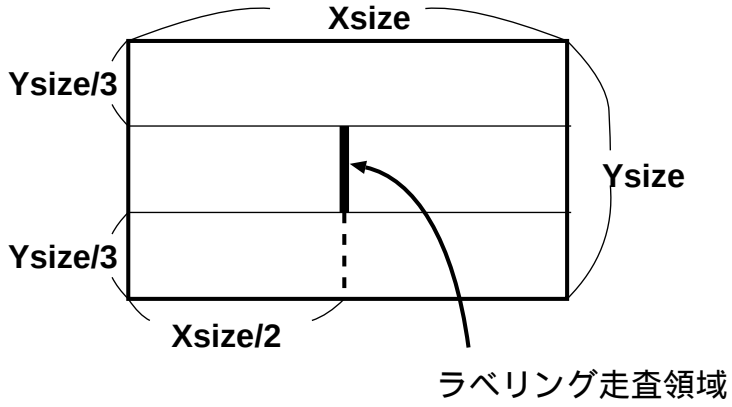

図 3.12 ラベリング走査領域

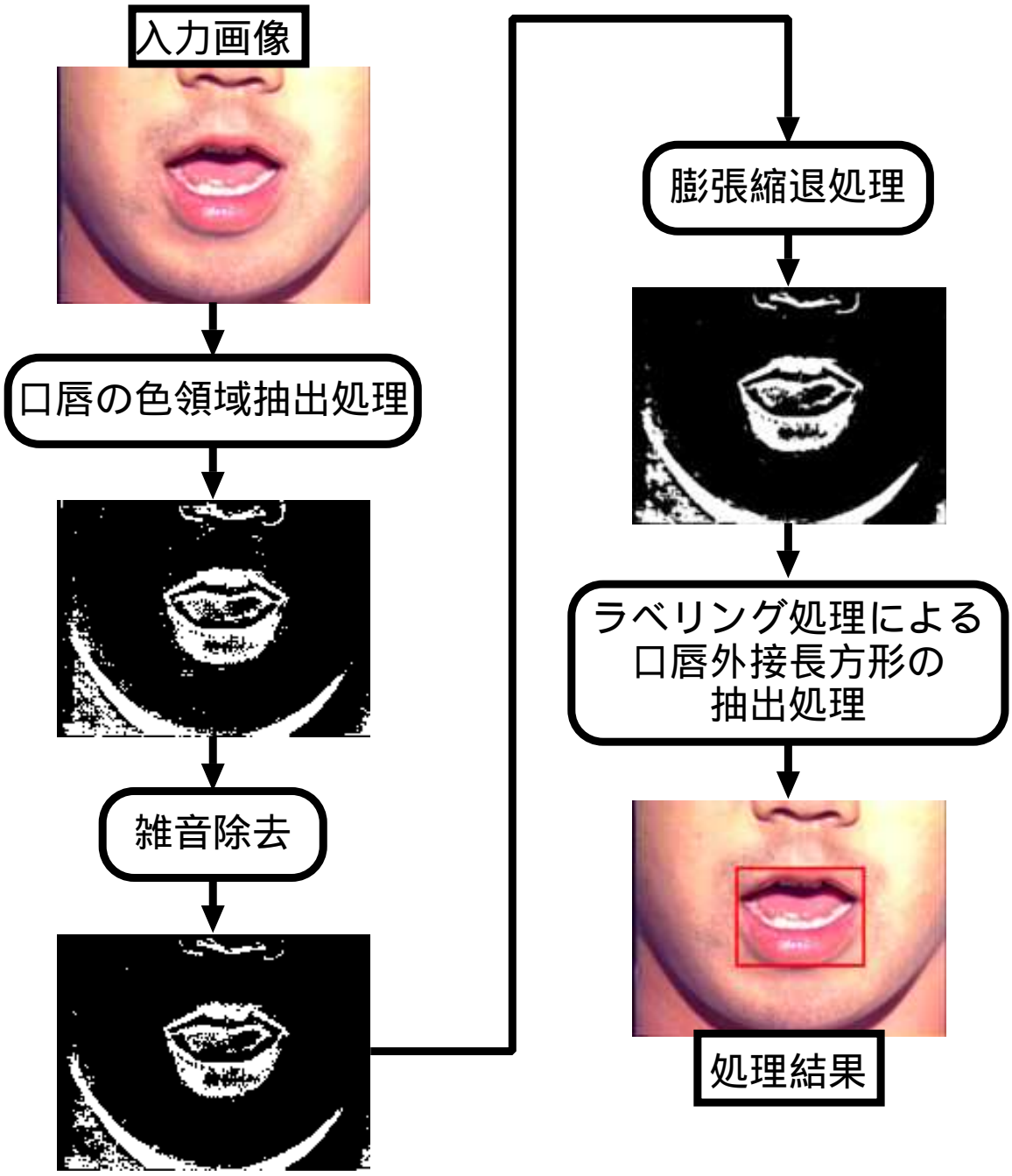
入力画像
口唇の色領域抽出処理
雑音除去
膨張縮退処理
ラベリング処理による
口唇外接長方形の
抽出処理
処理結果

図 3.13　処理画像結果

3.7 口唇存在予想領域の設定

上述した処理により口唇の外接長方形が得られる。この作業を全てのフレームに対して行うのだが、フレームの全ての領域に対して上記の処理を行うと膨大な時間が掛かる。そこで、2 フレーム以降からは口唇存在予想領域を設定し、その領域内で口唇特徴量抽出処理を行う。

口唇存在予想領域は最も口唇の動きの激しい時を想定し、以下の範囲内に存在すると考える (図 3.14)。

- 前フレームの外接長方形に対し、縦方向に+20%、横方向に+10%増大させた矩形領域

この口唇存在予想領域を設定することにより、口唇領域ではない誤った口唇の色領域を口唇領域と特定する危険性も減少する。口唇存在予想領域を設定して行う口唇抽出処理と、全領域に対して行う口唇抽出処理の手法を 70 枚分のフレームの動画像 10 組に対して行い、その処理時間を測定し、平均をとった。

結果を表 3.2に示す。

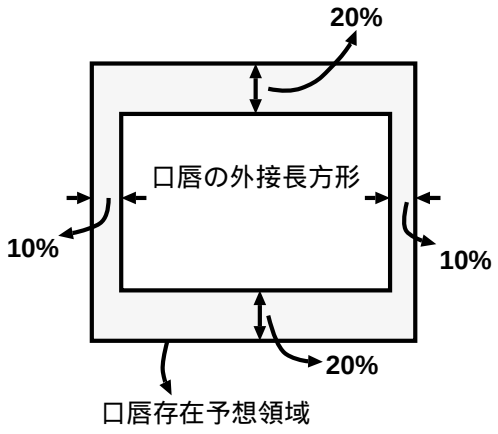

図 3.14 口唇存在予想領域

表 3.2　口唇の動き特徴量計測時間の比較

	INDIGO2
口唇存在予想領域を設定して行った口唇抽出処理	30 秒
全画面に対して行った口唇抽出処理	442 秒

口唇存在予想領域を設定し処理を行うことにより、全画面に対して処理を行うものに対して、およそ 13 倍処理速度が向上した。

3.8 マッチング処理

3.6節で示した処理によって図 3.15のような口唇特徴量の一次元時系列データが得られる。一般的に発話に際して同一人物が同じ単語を発声しても、その長さはその都度変わり、しかも非線形に伸縮する。そこで個人識別の為のマッチングを行う際に、発話時間長の変動を正規化（**DTW**:Dynamic time warping ）し、パターンを時間軸に対し非線形に整列化する必要がある。一次元時系列パターンを速やかに整列化し、それに基づいてパターン間の類似度を調べる方法として、本研究では DP マッチングを用いる [22],[4]。

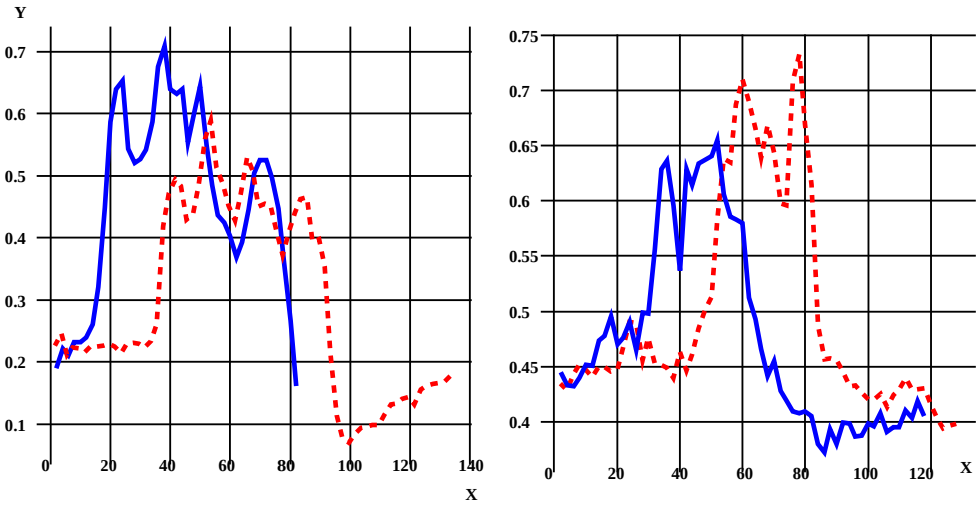

図 3.15 開口度測定結果一例

3.8.1 DP マッチング

DP マッチングはソ連の Slutsker らと日本の迫江らによってほぼ同時期に発表された手法であり、『2 つの異なる時系列パターンが最も良く一致するように時間軸を非線形に変換し、その後距離 (類似度) を求める手法』である [18]。この手法は現在音声認識の世界において代表的なマッチングアルゴリズムとなっている。

対応すべき 2 つの時系列を次のように表す。

$$\begin{cases} A = \mathbf{a_1}, \mathbf{a_2}, \mathbf{a_3}, \cdots, \mathbf{a_I} \\ B = \mathbf{b_1}, \mathbf{b_2}, \mathbf{b_3}, \cdots, \mathbf{b_J} \end{cases} \tag{3.5}$$

ここで、$\mathbf{a}_i$は一次元時系列パターン A の第 i フレームの特徴パラメータでベクトル表現とする。Iおよび Jは各々A、Bのフレーム総数で時間長に対応する。一般に Iと Jは時間軸が異なるのでそのまま単純には比較出来ない。そこで A、Bを時間軸に対して非線形に伸縮し、対応させる必要がある。ここで図 3.16に示すような A、Bからなる平面を考えると時間伸縮関数 (warping function)F、すなわち A、B両パターンの時間軸の対応づけはこの平面の格子点

$$F = c_1, c_2, \cdots, c_k, \cdots, c_K \tag{3.6}$$

で表現することができる。ただし、$c_k = (i_k, j_k)$ は i 軸では i_k、j軸では j_kの格子点を表している。2 つの特徴ベクトル $\mathbf{a_i}$と $\mathbf{b_i}$との距離を $d(c) = d(i, j)$ と表すと F に沿った距離 (経路) の総和は

$$D(F) = \frac{\sum_{k=1}^{K} d(c_k) w_k}{\sum_{k=1}^{K} w_k} \tag{3.7}$$

と表すことが出来る。本稿では (3.7) 式より得られた $D(F)$ の値をコスト関数値として定義する。3.7式において $D(F)$ の値が小さければ小さいほどパターン A、B間の対応づけは良い。すなわち、コスト関数値 $D(F)$ の値が小さければ類似度が高く、逆に値が大きければ、類似度は低い。ただし時間伸縮関数 F は以下のような条件を満たすように設定する。

- 単調性と連続性の条件

$$0 \leq i_k - i_{k-1} \leq 1, \quad 0 \leq j_k - j_{k-1} \leq 1, \tag{3.8}$$

- 境界条件

$$i_1 = j_1 = 1, \quad i_k = I, \quad j_k = J \tag{3.9}$$

- 傾斜制限条件

$$|i_k - j_k| \leq r \tag{3.10}$$

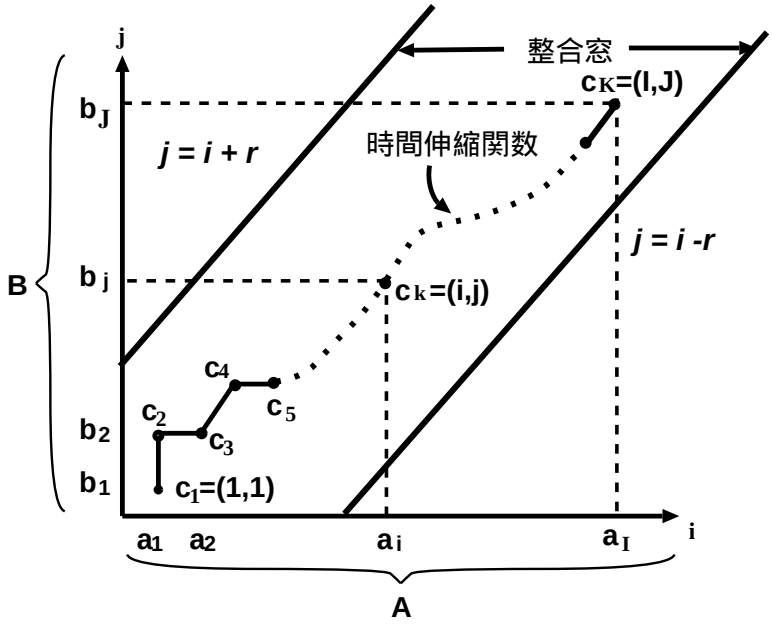

図 3.16　A、B 2 つの時系列の時間軸の対応づけ

今回は時系列データの対称性を保つためのアルゴリズムを採用した。これは以下の 3 式のなかで最も距離の短いものを経路として用いるという制限を上記の制限にさらに加えるものである。

$$i_k = i_{k-1} + 1 \quad \text{かつ} \quad j_k = j_{k-1} \tag{3.11}$$

$$i_k = i_{k-1} + 1 \quad \text{かつ} \quad j_k = j_{k-1} + 1 \tag{3.12}$$

$$i_k = i_{k-1} \quad \text{かつ} \quad j_k = j_{k-1} + 1 \tag{3.13}$$

またこの時、(3.7) 式の重み関数 w_kは $I+J$となる。さらに、(3.10) 式において $r=5$ と設定した。(3.7) 式は加法的なので動的計画法を用いて効率良く解くことが可能である。以下にアルゴリズムを示す。

1. 部分点列 $c_1, c_2, \cdots, c_k$ $(c_k(i,j))$ に対する部分和を $g(i,j)$ とし初期条件として $g(1,1)=d(1,1)$ と設定する。

2. (3.11) 式の経路制限を用いて以下の式を得る。

$$g(i,j)=\left\{\begin{array}{c} g(i-1,j) \\ g(i,j-1) \\ g(i-1,j-1) \end{array}\right\}+d(i,j) \tag{3.14}$$

   (3.14) 式を $i=1,2,\cdots,I \quad, j=1,2,\cdots,J,$ について計算する。

3. 2 の作業により得られた値 $g(I,J)$ より $D(F)$ は

$$D(F)=\frac{g(I,J)}{I+J} \tag{3.15}$$

   となる。このアルゴリズムにより口唇の動き特徴量のマッチングで行う。本稿では DP マッチングにおけるコスト関数 $D(F)$ の値に対して適当な閾値を設定することにより本人であるかそうでないかを決定する。

第 4 章 実験 ・結果

本章では第 3章で提案した口唇の動き特徴量を用いた個人識別手法についての評価実験を行う。

4.1 実験装置および実験対象画像

4.1.1 実験装置

実験に利用した装置を以下に示す。

- DFM (フレームグラバ) (図 4.2) [6]

  計測技術研究所製。動画像をメモリに取り込む装置である。

- DFM 外部クロックアダプタ (図 4.3)

  計測技術研究所製。本装置は 12.115MHz のクロック信号を DFM の外部クロックとして供給する装置である。DFM とは専用ケーブルで接続され、RS232C ケーブルによりフレームメモリユニット MPU-F100A と接続されている。

- フレームメモリユニット MPU-F100A(ソニー社製) (図 4.5)

- CCD カメラ (ソニー社製)

- INDIGO2(シリコングラフィックス社製 ワークステーション) (図 4.4)

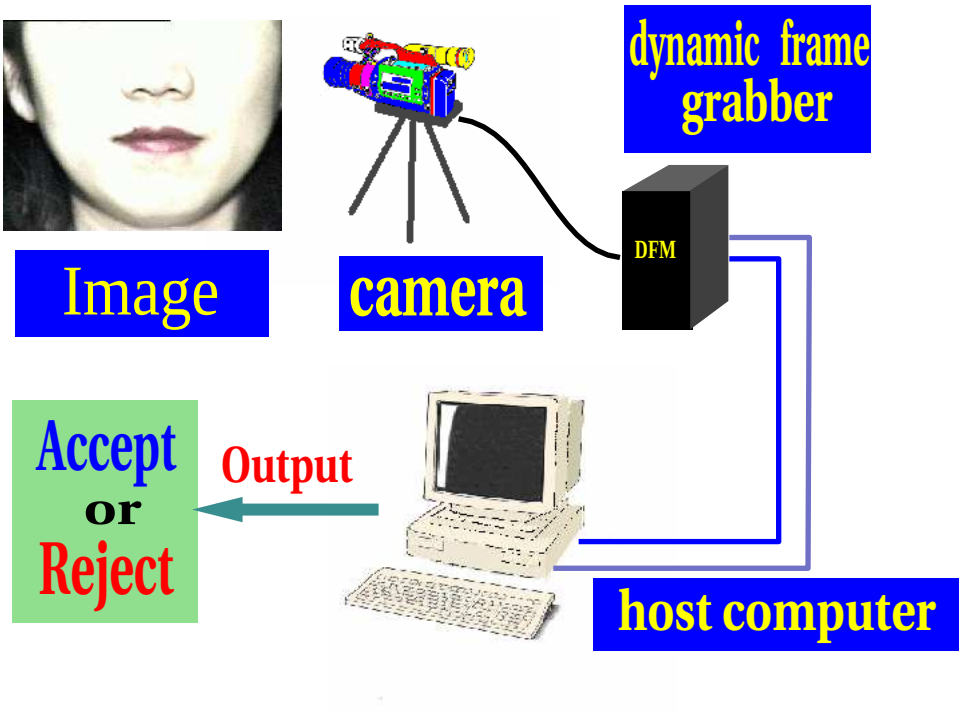

図 4.1　システム構成

図 4.2　DFM　(フレームグラバ)

図 4.3 DFM 外部クロックアダプタ

図 4.4 indigo2 外観

図 4.5 フレームメモリユニット MPU-F100A

4.1.2 実験対象画像

処理の対象となる画像は 720×480 画素のカラー画像である。(R,G,B 各々 8 bit/pixel)。撮像対象者はカメラの前に座り、照明条件が一定の下で口唇周辺が写るようにする。入力画像例を図 4.6に示す。

図 4.6 入力画像例 (口唇周辺画像)

4.2 実験

4.2.1 実験に対する制限

本実験を行うに当たって以下のように実験に対する条件・制限を設ける。

- 照明条件一定
- フレームレート $30(frames/sec)$
- 発話開始時・終了時には口を閉じる

4.2.2 口唇領域抽出実験

4.2.1節に示した条件下において、3.6節で述べた口唇の抽出処理がうまくいくかどうかの実験を行った。被験者は 10 人で、口の開閉を行ってもらった。図 4.7 は図 4.6で示した入力画像例の動画像系列に対する口唇領域抽出結果である。

図 4.7 口唇領域抽出結果

4.2.3 考察

4.2.1節で設定した条件下においては、口の開閉の如何に関わらず、口唇領域の抽出に成功した。

これは前述した予備実験により設定した口唇の色領域でわずかな照明の変化等の外乱に影響を受けやすい明度を用いないことにより、外乱による影響を受けにくくなったからだと考えられる。

さらに、ラベリング走査領域や口唇存在領域を設定することにより、口唇領域以外に存在する口唇色領域部分、例えば、顎の部分や鼻の部分等を効率的に除去した結果であると推測される。

しかし、本実験において一例だけではあるが口唇領域を正しく抽出できないものがあった。本手法は口唇の形状モデリングなどの煩雑な処理をしなくてよいために処理が簡単・高速である反面、口唇の周辺に大きなニキビ等、本来の口唇色領域内に存在していないものが混入していた場合、その部分をも含めて口唇領域と認識してしまうことがある。今回の失敗例もこの条件に当てはまるものであった (図 4.7の (d))。

4.2.4 個人識別実験

口唇の動き特徴量を用いた個人識別の有用性と、発話時の口唇の動き特徴量の個人性を調べるために評価実験を行った。発話パターンは以下のパターンを用いた。

- 無声 (閉口状態)
- 『あ』
- 『千原研究室』

ただし、ここでいう無声というのは口唇を閉じた状態のこと、いわゆる閉口状態のことを示している。

被験者 は10 人で上記の各発話を各々3 回ずつ発話する。また、この時被験者には日頃行っている発話と同じように発話するように要請した。

この各々の発話パターンに対して、3章で述べた処理を行い、各々の発話パターンに対して個人の識別率を調べた。

4.2.5 実験結果

異なる発話パターン時

本人が 『千原研究室』と発話し、本人が同じく発話した 『千原研究室』とマッチングを行った時のコスト関数値の累積度数と、他人が 『あ』と発話し、他人の『千原研究室』とマッチングを行った時のコスト関数値の累積度数を示したものを図 4.8に示す。また、コスト関数値に対し、様々な閾値を設定したときの本人受理率と他人棄却率を示したものを図 4.9に示す。

被験者が異なる発話をした場合、他人棄却率を 100%とした時、本人受理率が 95%となった。このことから、他人が異なる発話を行うと高い他人棄却率と本人受理率が得られることが証明され、パスワードによるセキュリティの存在が確かめられた。

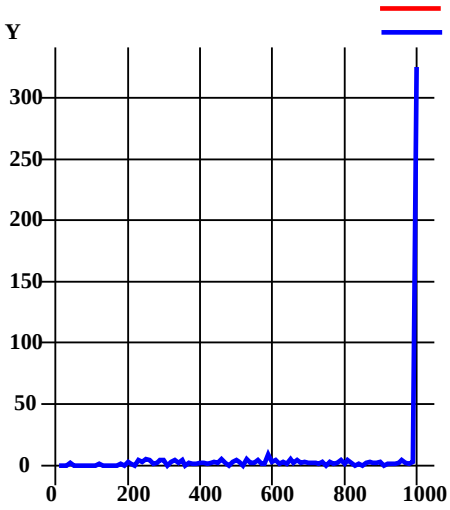

図 4.8　異なる発話時におけるコスト関数値

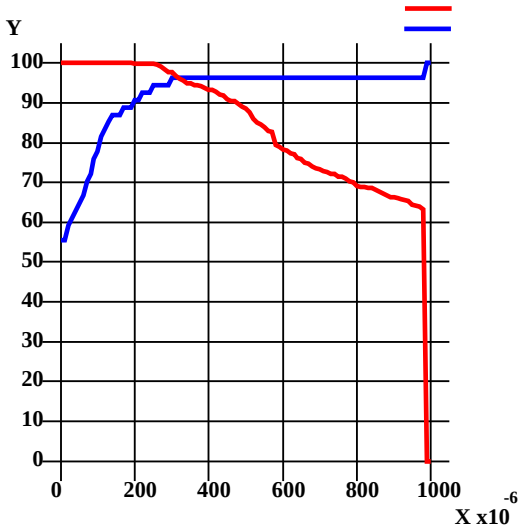

図 4.9　異なる発話時における認識結果

同一発話時

発話時の口唇の動き特徴量の個人性を調べるために、被験者が全員同じ発話を行った際の類似度の値を調べた。後述の考察のために本人受理率を全て 80% としたときの他人受理率を計算した。結果を 4.1に示す。

なお、無声 (閉口状態)、『あ』、『千原研究室』発話時のコスト関数値の累積度数を図 4.10、図 4.12、図 4.14に、そのときの認識結果を図 4.11、図 4.13、図 4.15に示す。

表 4.1 個人識別結果

	本人受理率	対他人棄却率
無声 (閉口状態)	80%	71%
『あ』	80%	76%
『千原研究室』	80%	90%

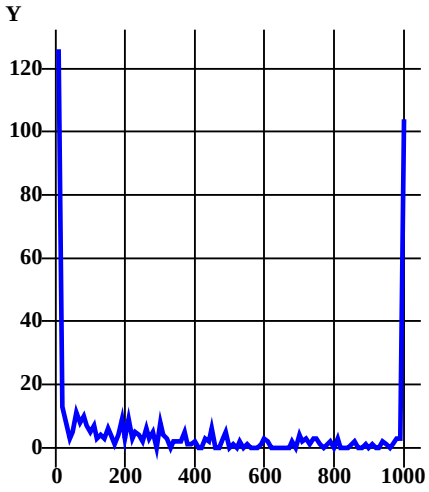

図 4.10　無声（閉口状態）時におけるコスト関数値

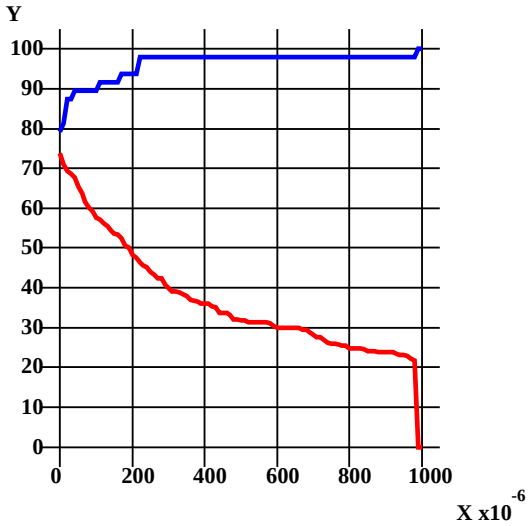

図 4.11　無声（閉口状態）における認識結果

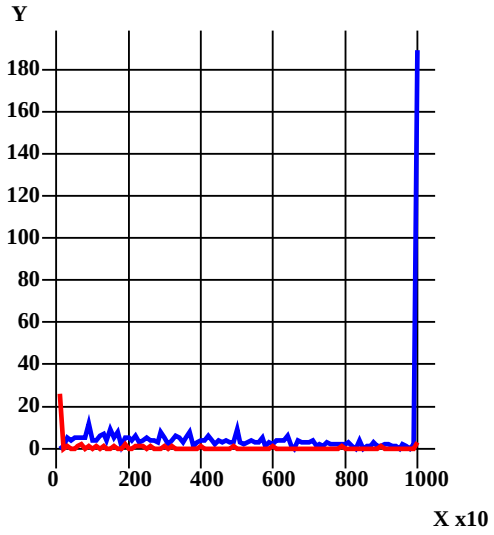

図 4.12　『あ』発話時におけるコスト関数値

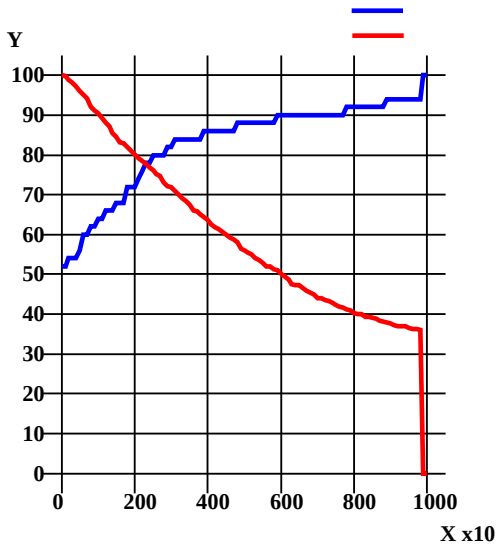

図 4.13　『あ』発話時における認識結果

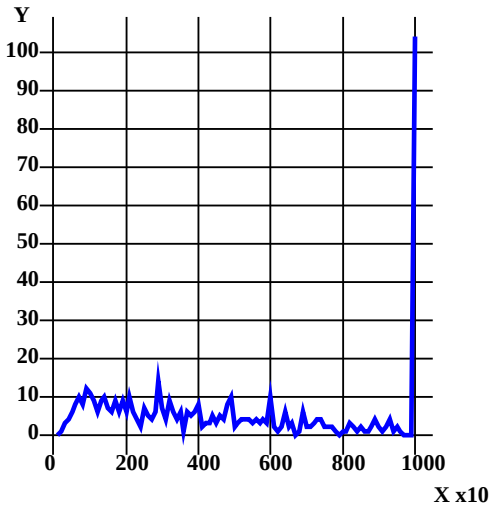

図 4.14　『千原研究室』発話時におけるコスト関数値

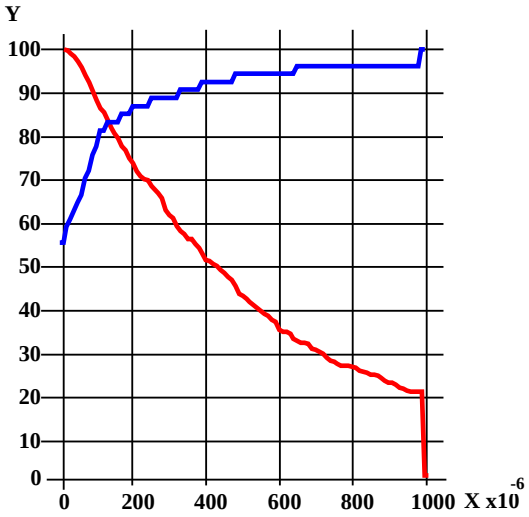

図 4.15　『千原研究室』発話時における認識結果

4.2.6 考察

4.2.5で発話パターンが無声、『あ』、『千原研研究室』という 3 つのパターンに対して個人識別実験を行った。

本実験において被験者には同一発話を行い、口唇の動き特徴量によりマッチングを行った。

この結果から発話パターンの長さによる口唇の動き特徴量の発現の度合と口唇の動き特徴量の個人性の有無を考える。

口唇の動き特徴量の個人性の有無

口唇の動き特徴量に個人性が存在するが否かを調べるために、無声時と発話パターン『千原研究室』に対してその認識率の違いについて考える。無声時と発話時の認識率を比べると、無声時よりも発話を行った際の方が高い個人識別率が得られている。

無声によるマッチングは実質的に顔の一部分である口唇の一次元形状情報のみ、すなわち静的な顔の形状情報のみでマッチングを行っており、動きに対する情報は有していない。一方、発話パターン『千原研究室』は口唇の動き特徴量を含んだパターンであるであることから、口唇の動き特徴量は個人性を有していることが確かめられる。

また無声時においては、本人同士のマッチングでコスト値が小さな結果が得られたとしても、他人とのマッチングの際にもまたコスト値が大きな値を持つという不確実性も有していた。それに対して、発話パターン『千原研究室』は口唇の一次元形状情報だけでなく、動きの特徴量も有しているため高い認識率が得られた。発話パターンの長さに対しては次項で述べる。

発話パターンの長さに対する認識率の違い

つぎに発話パターンの長さに関する認識率の違いであるが、今回実験で行った『あ』のような単音の場合、本人自身の発話のマッチングにも関わらずコスト値が大きな値を持つ場合や、逆に他人とのマッチング時においても、コスト値が低い値を持つ場合が多く存在した。

これは単音の場合は無声の時のような実質的に一次元形状情報だけでマッチングを行っている訳ではないが、それでも口唇の動きが単純であり、このような発話パターンにおいては、口唇の動き特徴量に個人性が現われにくいと考えられる。他人とのマッチング時においても、時にコスト値の低い結果を得ることとなった。

一方、『千原研究室』のような発話は口唇の開閉、すぼみ等が激しいことから、口唇の動きの特徴量が顕著に現れたと思われる。この発話パターンに対して本人同士のマッチングを行うと無声時や発話パターン『あ』における本人同士のマッチングにより算出される類似度の値が大きくばらけているのに比して、発話パターン『千原研究室』はコスト値が全て低い値を示した。

このことから、単音のような短かな発話においては口唇の動きに個人性が出にくいのではないかと思われる。

認識率について

今回口唇の動き特徴量を用いた個人識別を行ったが、本人対他人のマッチングに対してもコスト値が低くなる場合があった。この理由としては以下のようなものが考えられる。

- 開口度値に一意性がない

開口度とは口唇の縦横比のことを示しており、したがって、縦横比が同じになるようなパターン、例えば『う』のように口唇をすぼめた発話パターンと『お』のような口唇を大きく開けたようなパターンは同じ開口度を取りうる可能性を有している。これより口唇の動きが完全に開口度のみでは記述されていないのではないかと思われる。

この欠点を改善するための案としては、何か顔の中に基準となるような部分、例えば目頭間の距離のようなもの、を抽出し、その値を基準として口唇を正規化することにより、図 3.2 に示すようなパラメータを算出し、その距離の時系列データを用いてマッチングを行うといった方法がある。こ

の方法を用いることにより、口唇はより正確な記述が可能となるが、正規化するためには複雑な処理が必要になるというデメリットも存在する。

しかし、本研究において提案した手法は、口唇の記述に対して上述した欠点が存在してはいるが、その反面、正規化の必要性がなくまた口唇周辺領域のみの処理でよいので処理が軽く簡単であるというメリットがある。

第 5 章 結論

本論文では、口唇の動き特徴量を用いて個人識別を行う手法についての報告を行った。口唇の特徴量としては開口度なるものを提案し、全てのフレームに対して得られた開口度の時系列データを口唇の動き特徴量と定義した。得られた口唇の動き特徴量に対して DP マッチングを用いることによりマッチングを行った。

口唇の動き特徴量の有用性を調べるために無声、『あ』、『千原研究室』の 3 つの発話パターンに対して個人識別実験を行った結果、長い発話パターンでは口唇の動き特徴量に個人性が生じることを確認した。

本研究はセキュリティ・サービスの 1 つの分野であるアクセスコントロールのための研究と位置付けているが、その他にも犯罪捜査のための一技術としての応用や話者認識の一つの手がかりとしてなど、幅広い応用分野があると思われる。

今後は口唇を記述するパラメータ数を増やし、また処理速度を向上させることにより 'Visual Password' として実用的なシステムを構築することができる。

謝辞

本研究を進めるにあたり終始暖かい御指導を賜わるとともに、研究に取り組む姿勢や研究者としての心得などを諭して頂いた像情報処理学講座 千原 國宏教授に厚く御礼申し上げます。

副指導教官としてゼミ発表等で御助言頂いたソフトウェア基礎講座横矢 直和教授に感謝致します。

多くのミーティングは発表の場において研究の要点からご指導頂き、些細な質問にも丁寧な御回答を頂いた本講座 佐藤 宏介 助教授 に深く感謝いたします。

本研究にあたっての問題点や改良点の指摘や、計算機管理等に関する知識を与えて下さった先端科学技術研究調査センター 大城 理 助教授 に感謝いたします。

日頃より、信号処理の基礎や研究への取り組み方、プログラミング手法など多岐にわたり御助言を頂きました本講座 眞溪 歩 助手に感謝致します。

修論を書くにあたり数多くの適切な助言をして頂いた本講座 陳 謙助手 に感謝いたします。研究室における生活を快適なものにしてくださった秘書の米田 尚子嬢、教務技官の岸岡 了嬢に感謝致します。

本研究を進める上で必要となる基礎技術について御教授頂いた本講座博士後期課程土居元紀 氏に感謝いたします。

プログラミング技法において様々な助言を頂いた本講座修士 2 年の伴好弘氏並びに福本圭司氏、富田浩朗氏に感謝致します。また、本研究を行うにあたり、様々な助言を行ってくれた本講座認識グループの山瀬 弘之氏、楠瀬 賢也氏、柴田 豊彦氏、更科 渉氏に感謝します。

また、実験を行うにあたり、快く顔画像の撮像に御協力頂いた方々に深く感謝致します。

最後に、細井 聖氏、藤本 直登志氏、松尾 努氏、王 家寧氏、黒田智宏氏、永野眞己嬢、安室 喜弘氏を始めとする像情報処理学講座の諸氏には、本研究の様々な局面にて多大なる御協力いただきました。ここに感謝の意を表します。

参考文献

[1] K.Venkatesh Prasad. “visual passwords;spationtemporal cues for speaker recognition”. *ACCV‘95 Second Asian Conference on Computer Vision*, pp. I–146 – I150, Dec 1995.

[2] 佐々木 努 赤松 茂 深町映夫, 末永康仁. “正面顔画像の自動識別法の検討”. 信学技報, pp. 1–7, 1991.

[3] 大湊吉行, 佐藤宏介, 千原國宏. “口唇の動き特徴量を用いた個人識別”. 第 39 回システム制御情報学会研究発表講演会講演論文集, pp. 513–514, May 1995.

[4] 上坂吉則, 尾関和彦. “パターン認識と学習のアルゴリズム”. 文一総合出版, 1990.

[5] 内村圭一. “口唇周辺画像情報を用いた発話単語認識”. 画像ラボ, Jan. 1994.

[6] 計測技術研究所. *“DFM–Dynamic Video Frame Memory Version 2.0”*, June 1995.

[7] 土居元紀. “アクセスコントロールを目的とする顔画像照合システム”. Master’s thesis, 奈良先端科学技術大学院大学, Feb 1995.

[8] 森本広頼, 李亜兵, 亀井且有, 井上和夫. “画像処理による口唇領域の抽出”. 第 39 回システム制御情報学会研究発表講演会講演論文集, pp. 373–374, May 1995.

[9] 黒田英夫, 光田和臣, 池本雄一郎. “顔画像からの口部領域の自動抽出法”. 信学技報, Mar 1991.

[10] 塩野充, 真田英彦. “個人識別技術の最近の研究動向”. 信学技報, Sep 1992.

[11] 長谷川純一, 輿水大和, 中山晶, 横井茂樹. “画像処理の基本技法 <技術入門編>”. 技術評論社, 1986.

[12] 渋川勝久, 中辻康. “濃淡値から得られる特徴量を用いた顔画像の個人識別”. 画像ラボ, Vol. 5, No. 10, pp. 43–45, 1994.

[13] 松尾賢一, 上田勝彦, 梅田三千雄. “色彩情報を用いた情景画像からの文字列抽出”. 信学技報, pp. 49–56, Dec 1994.

[14] 八木伸行, 井上誠喜, 林正樹, 中須英輔, 三谷公二, 奥井誠人, 鈴木正一, 金次保明. “C 言語で学ぶ実践画像処理 “. オーム社, 1992.

[15] 南畝. “顔画像による個人識別の技術”. システム/制御/情報, Vol. 35, No. 7, pp. 415 –422, 1991.

[16] 南畝, 増田功. “セキュリティ技術における画像処理”. 画像電子学会誌, Vol. 22, No. 3, pp. 268 –278, 1993.

[17] 南畝, 中村納. “画像工学”. コロナ社, 1989.

[18] 中川聖一. “確率モデルによる音声認識”. 電子通信情報学会, 1988.

[19] 宮原誠, 吉田育弘. “色データ (r,g,b)–> (h,v,c) 数学的変換方法”. テレビジョン学会誌, Vol. 43, No. 10, pp. 1129–1136, 1989.

[20] 藤井智子, 黒川隆夫. “単語発話時の口形変化の分析と口話画像合成のための規則化”. *11th Symposium on Human Interface*, pp. 191–196, Oct. 1995.

[21] 藤井智子, 渡辺茂晃, 黒川隆夫. “発話時の口形変化と手話アニメーションへの口話の導入”. Human Interface 第 28 回研究会, Vol. 9, pp. 161–168, 1994.

[22] 古井貞熙. “デジタル音声処理”. 東海大学出版会, 1985.

[23] 田村秀行. “コンピュータ画像処理入門”. 総研出版, 1985.