分散協調による顔認識のための能動的登録

加藤 文和　沖 昌並　向川 康博　尺長 健

岡山大学 情報工学科

〒700-8530 岡山市津島中 3-1-1

kato@chino.it.okayama-u.ac.jp

あらまし
自然な環境下における顔認識のためには，対象人物の検出，追跡，登録，認識の4つのタスクが必要不可欠である．本稿では，分散協調視覚の枠組で，これらを統一して扱うシステムを提案する．不特定人物を対象とするためには，人物の追跡と同時に顔画像を登録する必要があるが，照明変動に対してロバストな認識を実現するためには，顔画像の登録が最も重要となる．そこで，本研究では，複数のカメラを用いて対象人物を適切に選択する能動的登録によって，効率良く辞書を登録する方法について提案する．

Active Registration for Face Recognition by Cooperative Distributed Vision

Takekazu KATO, Shohei OKI, Yasuhiro MUKAIGAWA, Takeshi SHAKUNAGA
Department of Information Technology, Okayama University
3-1-1 Tsushima-naka, Okayama, 700-8530, JAPAN
kato@chino.it.okayama-u.ac.jp

Abstract
The face recognition in natural environments requires four tasks which are detection, tracking, registration and recognition. This paper presents a system which integrates these tasks by Cooperative Distributed Vision. In order to cope with the unknown person, face images should be registered while a face is tracked. For the robust recognition against lighting changes, the effective registration is the most important. For this purpose, we propose an Active Registration method which realizes the effective selection of the target persons by using many distributed cameras.

1　はじめに

個人認識技術は，セキュリティシステムや人間とコンピュータとのインタフェースなど広い分野に応用が可能である．安定な個人認識のためには，対象人物やその位置，シーンの照明条件などに制限がなく，またそれらが動的に変化するような自然な環境下での認識技術が不可欠である．

計算機による個人認識法の一つに顔画像を用いる方法があり，今までに多くの手法が提案されてきた [1, 2, 3, 7, 8, 10]．しかし，これらの手法は，認識対象となる人物の顔画像を管理された環境下において登録しておく必要があり，自然な環境下での登録，認識を行うことは困難である．

不特定の人物を対象とする場合，あらかじめ辞書を用意しておくことができないため，移動する対象を検出，追跡しながら同時に辞書の登録を行わなければならない．また，照明条件や顔の向きの変化などに対して安定に認識を行うためには，様々な条件に対応できる辞書を構築しなければならない．つまり，自然な環境下での安定な個人認識のためには，効率の良い辞書の登録が最も重要な課題となる．そこで本研究では，複数の分散カメラを使い，それらを互いに協調させることによって，登録対象やその条件を能動的に選択し，効率の良い登録を行う．

本稿では，分散協調視覚 [4, 5] の枠組にもとづいて，検出，追跡，登録，認識を統合したシステムを提案し，さらに複数カメラを使った能動的登録によって，安定な個人認識に必要な辞書を効率良く構築する方法について述べる．

2　システムの概要

本章では，分散協調による人物の検出，追跡，登録，認識システムの概要について述べる．

2.1　システムの構成要素

本システムは複数のエージェントの集合と二つのユニットから構成される．

2.1.1　エージェント

エージェントは，パン・チルトをコントロール可能なアクティブカメラを備えた計算機 (観測ステーション) に対応するモデルであり，それぞれ対象人物の検出，追跡，顔画像の抽出等の機能を持つ．また，各エージェントはネットワークによって接続され，相互に通信することができる．

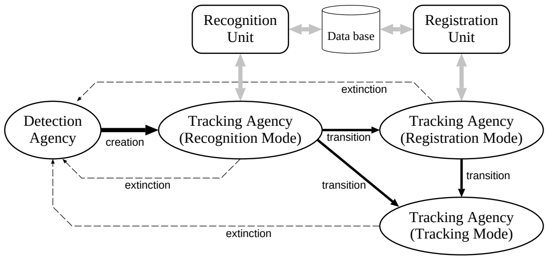


図 1: システムの概要

2.1.2　検出 Agency と追跡 Agency

システム内で同じタスクを持つエージェントの集合を Agency と呼び，Agency に属するエージェントをメンバと呼ぶ．各 Agency には，Agency マスターと呼ぶ特別な役割りを持つエージェントが必ず一つ存在し，Agency のメンバは Agency マスターによって管理される．システム内の全てのエージェントは，必ずいずれかの Agency に属し，Agency マスターによってタスクを与えられる．

図 1 に示すように，Agency には検出 Agency と追跡 Agency の二種類がある．検出 Agency は，決まった範囲内に出現する人物の検出を行い，追跡 Agency は特定の人物を追跡し，追跡対象に関する認識や登録を行う．追跡 Agency はタスクに応じて以下の 3 つのモードを持ち，それらのモード間を遷移することによってその役割を変化させる．

[T1] **認識モード:** 対象人物の顔画像を抽出し，その向きを判別する．そして，得られた顔画像と方向の情報を認識ユニットに送り，既に登録されている人物との同定を依頼する．この結果，対象人物が登録されていない人物である場合や，登録状態が不十分である場合，登録モード ([T2]) に遷移し，そうでなければ追跡モード ([T3]) に遷移する．

[T2] **登録モード:** 対象人物の顔画像を抽出し，その向きを判別する．そして，抽出された顔画像を方向毎に仮登録し，登録された画像集合が照明変動を十分に許容するものであるかどうかの判定を行う．全ての方向に関する登録が十分になれば，追跡モード ([T3]) に遷移する．

[T3] **追跡モード:** 人物の追跡のみを行う．

2.1.3 認識ユニットと登録ユニット

ユニットはシステム内で特定の処理を行うものであり，認識ユニットと登録ユニットの二つから成る．登録ユニットは登録モードの追跡Agencyによって仮登録された顔画像を受け取り，認識のための辞書を構築する．また，認識ユニットは認識モードの追跡Agencyからの依頼を受け，登録ユニットによって構築された辞書に基づいて個人同定を行う．ユニットはエージェントと同様に観測ステーション上に実装することができ，またエージェントとは別にサーバーとして実装することもできる．

2.2 システムの協調動作

分散協調視覚システムでは，限られた範囲のみを把握できる複数の分散エージェントが互いに協調しあうことによって，シーン全体を把握する．このとき対象の追跡といったタスクでは実時間性が要求され，エージェント間での密な協調を必要とする．しかしシステム全体で密な協調を行う場合，システムの規模が大きくなると通信量が膨大になりシステムは破綻してしまう．そこで本システムでは，同じタスクを持つエージェントの集合，つまりAgencyを一つの機能単位とし，Agency内での密な協調とAgency間の粗な協調の2段階の協調動作を考えることによって，柔軟なシステムを構築する．

2.2.1 Agency内の協調動作

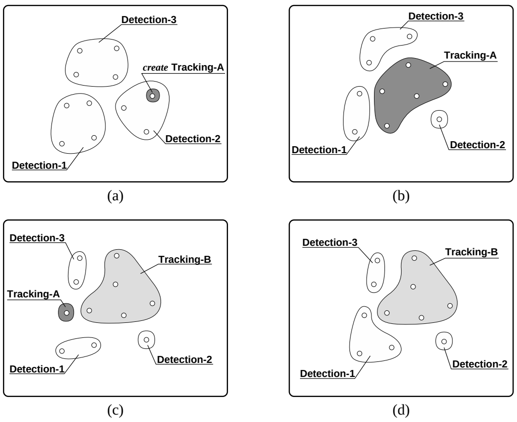
Agency内のエージェントが密に協調動作することによって，Agencyごとのタスクを実現する．Agencyが行うタスクはAgencyの種類やそのモードによって異なる．

検出Agencyの協調動作

検出Agencyは決まった観測範囲内を各エージェントで分担して観測する．観測範囲の分担はAgencyマスターが決め，各エージェントはそれに従って観測を続ける．

追跡Agencyの協調動作

追跡Agencyは，Agency内のエージェントの間で協調しながら，特定の対象の追跡を行う．この追跡Agency内の協調は基本的に松山らによって提案された協調的追跡のプロトコル[6]に従う．



(a) (b) (c) (d)

図2: システムの典型的な変化の例

また，認識モードでは，追跡対象の顔画像から個人同定を行い，登録モードでは，追跡対象の顔画像の登録を行う．これらのタスクに応じて，追跡Agency内の各エージェントは顔画像の抽出を行い，それをAgencyマスターに集約する．

2.2.2 Agency間の協調動作

Agency間で協調することによって，システム全体の動作をコントロールする．Agency間の協調動作は，Agencyマスターが代表して他のAgencyマスターと通信し，その結果に応じてAgencyの動作を変えたり，エージェントのやりとりを行うことによって実現する．

図2にシステムの典型的な変化の例を示す．初期状態において，3個の検出Agencyが存在し，追跡Agencyは存在しない．検出Agencyはそれぞれの観測エリア内で人物の検出を行っている．検出Agency内のエージェントが人物を検出すると，追跡Agencyが生成される(図2(a))．追跡Agencyは，まず認識モードで対象人物の認識を行う．認識の結果，対象人物が登録されていない人物なら，このAgencyは登録モードに遷移する．

登録モードでは，可能な限りの登録が完了するまで，エージェントを統合しながら登録を行う(図2(b))．登録が完了すると，追跡Agencyは追跡モードに遷移する．

追跡モードでは，追跡Agencyは多数のエージェントを必要としない．もし，他に認識モードや登録モードの追跡Agencyが近くに存在すれば，エージェントはそれらのAgencyに移動する(図2(c))．最後に，対象人物が観測範囲から消えると，追跡Agencyは消滅する(図2(d))．

この例のように，本システムは，1)Agency の生成と消滅，2)Agency 間のエージェントの移動，の 2 種類のイベントによって，システムの状態を変化させる．以下でそれぞれについて述べる．

追跡 Agency の生成と消滅

検出 Agency 内のエージェントが人物を検出すると，もしシステム内に同じ人物を追跡する追跡 Agency が存在しなければ，対象を検出したエージェントのみから構成される新しい追跡 Agency を生成する．また，人物が観測範囲から出ると，追跡 Agency は消滅し，Agency 内の全てのエージェントは他の Agency に移動する．

エージェントの移動

Agency が現在のタスクにおいてさらに多くのエージェントを必要とする場合，Agency は他の Agency からエージェントを取り込む．また，Agency があまりエージェントを必要としない場合，エージェントを他のエージェントに移動させる．これらの移動は Agency 間で通信を行うことによって決定する．移動するエージェントを決定することを契約と呼び，契約が成立した場合のみエージェントを移動させる．実際には，Agency 間の契約は Agency マスターが代表して行う．そのため，契約に必要な情報は全て Agency マスターに集約する．

Agency マスター自身が他の Agency に移動する場合，Agency 内の他のエージェントにマスターを変更してから移動する．このときマスターが持つ全ての情報を新しいマスターに伝える．ただし，単一のエージェントのみから構成される Agency，つまり Agency マスターのみを持つ Agency の場合，マスターを他のエージェントに変更することができないため，他の Agency に移動することができない．そこで，他の Agency のエージェントと交換し，そのエージェントを新たなマスターとすることで，自分自身の移動を行う．

3 エージェントの機能

本システムでは，各エージェントは人物の検出や追跡，また顔画像の抽出などを行う．以下でこれらのタスクに必要な処理について述べる．

3.1 対象の検出と追跡

対象の検出と追跡には，あらかじめ人物の顔領域や頭部の色を求めておき，入力画像からこれらの色に近い領域を求め，その位置関係が正しいものを顔領域として検出する．また，検出された顔領域の重心が常に画像の中心にくるようにカメラを制御することで，対象の追跡を行う．

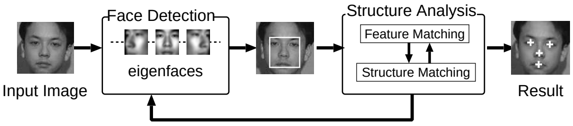


図 3: 顔画像の抽出

この方法は色の比較だけで実現可能なため，高速な処理が可能であり，背景差分に比べ背景に関する事前知識を必要としないため，カメラが移動したり背景が変動する場合にも適用可能である．さらに，色情報として色相を用いることで照明変動に対してロバストに検出を行うことができる．

3.2 顔画像の抽出

顔画像の抽出には eigenface [7, 10] と顔構造解析法 [8] を用いる (図 3)．eigenface は顔全体のパターンを比較する手法であるため，背景の変動や個人差に対して安定に顔画像を抽出することは困難である．また，顔構造解析法は，部分照合と構造照合を統合した手法であり，背景の変動や個人差に対してロバストであるという利点を持つが， 探索範囲が広い場合，部分照合の結果の組み合せが膨大になり処理に時間がかかるという欠点を持つ．そこで，eigenface で検出された顔画像に対して顔構造解析法を適用することで，探索範囲を限定し，高速かつ安定に顔画像の抽出を行う．

顔の姿勢の変化に対応するため，顔の姿勢をいくつかの方向に分けて表現する．あらかじめ学習した各方向毎の eigenface を使って入力画像を探索する．また，探索時の部分画像のサイズを段階的に変化させることにより，顔のスケールの変化に対応する．探索時の相違度には DFFS(distance from feature space)[7] を用いる．

3.3 予備実験

まず，色情報を用いて顔領域を追跡する実験を行った．この結果を図 4 に示す．図中の実線の枠が顔領域，破線が頭部の領域を表している．この処理は SIMD 型並列画像処理プロセッサである IMAP-VISION を用いて，ビデオレート (30 fps) 以内で行うことができた．

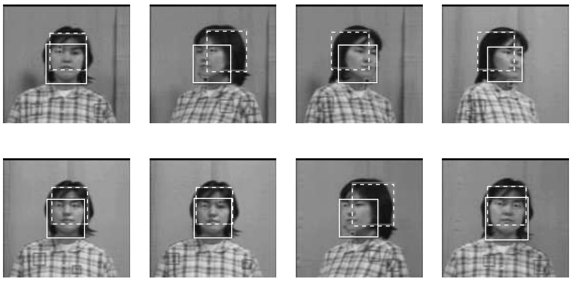


図 4: 追跡結果

図 5: Eigenface

図 6: 抽出結果

次に，顔画像の抽出と顔の向きの判別を行った．顔の向きは正面，右向き，左向きの 3 通りとした．抽出に用いた eigenface を図 5 に示す．図 4 の顔領域に対してそれぞれ顔画像の抽出と方向の判別を行った結果を図 6 に示す. この結果から正しく顔画像の抽出と方向の判別ができていることがわかる．また，この処理は Pentium-II(450MHz) を用いて 1 秒から 2 秒程度であった．

4　登録と認識

本章では，未知の照明条件において，不特定人物を対象とした場合の個人認識と，認識を安定に行うための辞書の登録方法について述べる．

4.1　認識

4.1.1　部分空間法による個人認識

追跡 Agency によって抽出された顔画像の認識には部分空間法を用いる．登録されている辞書から部分空間を作成し，抽出された顔画像の部分空間に対する DFFS と，DIFS(distance in feature space) 算出し，個人同定を行う．ただし，すべての feature space に対する DFFS の最小値が閾値を超えた場合には，未知の人物と見なす．

顔の方向や照明変動によって，顔の見え方は大きく変化するため，認識は難しくなる．顔の方向の変化に対応するために，顔画像は方向ごとに登録されているものとする．また，照明変動に対応するために，様々な照明条件下で撮影された登録画像を用いて，部分空間法によって認識を行う．つまり，各個人について，方向ごとの部分空間を作成する．

4.1.2　予備実験

まず，照明変動を十分に許容する辞書が獲得されていれば，未知の照明条件のもとでも安定な認識が可能であることを確認する実験を行った．6 人に対して照明条件を変えた 8 枚の顔画像を撮影した．照明条件は室内の蛍光灯のみの状態 (2 枚) と，ハロゲンランプを使って照明条件を 6 方向に変えた状態 (6 枚) である．切り出した顔の画像サイズは 50×50 である．実験では，各人物 8 枚中 7 枚を学習データとして固有空間を生成し，残りの 1 枚を認識用とした．認識に用いた固有空間の次元数は 3 であり，評価尺度として DFFS を用いた．図 7 に学習及び認識に用いた画像の例を示す．

顔認識結果を表 1 に示す．s1 ~ s6 が，各人物の 7 枚の学習データから生成した固有空間，f1 ~ f6 が認識用データを示しており，各番号が各人物に対応している．また，表中の値は DFFS を示している．この結果より，6 枚の認識用データ全てに対して認識が成功していることが分かる．

次に，登録されていない人物を正しく棄却できるかどうかを確認する実験を行った．6 人の画像集合それぞれに対して認識用 (1 枚) と学習用 (7 枚) の組合せを 8 つ構成し，この時生成される 48 個の固有空間を用いて実験を行った．認識に用いる DFFS の閾値を変化させたときの False Rejection Rate(FRR) と False Acceptance Rate(FAR) を図 8 に示す．この結果より適切な閾値を決めることによって，登録されていない人物は，未知の人物として判定できることが分かる．

これらの結果から，各方向について，照明変動を十分に許容する顔画像が獲得できていれば，未知の照明条件のもとで，未登録の人物を含む場合でも安定な認識が可能であることが分かる．

4.2　能動的登録

前節で述べたように，照明変動を十分に許容する辞書が獲得できれば，安定な認識が可能である．

(a) Training Set　(b) Test Set

図 7: 顔画像の例

表 1: 認識結果 (DFFS)

	s1	s2	s3	s4	s5	s6
f1	**0.085**	0.350	0.338	0.344	0.437	0.563
f2	0.329	**0.127**	0.345	0.294	0.394	0.556
f3	0.289	0.278	**0.112**	0.283	0.276	0.365
f4	0.393	0.405	0.383	**0.070**	0.311	0.284
f5	0.315	0.254	0.300	0.215	**0.032**	0.273
f6	0.460	0.463	0.371	0.251	0.307	**0.058**

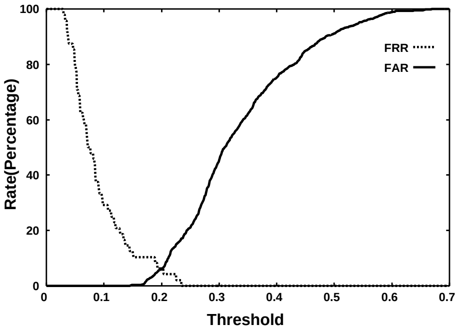


図 8: FRR,FAR と閾値の関係

このような辞書を獲得するためには，各カメラと対象人物の関係を固定せず，辞書の登録状況に応じて各カメラが対象とする人物を柔軟に交換する必要がある．

本研究では，複数のカメラを用いて効率良く辞書を登録するために，能動的登録を行う．能動的登録のためには，現在登録されている辞書が，どの程度照明変動を許容しているかを逐次評価し，各エージェントが適切な Agency に属するように動的に配置を変える必要がある．本節では，登録画像の評価方法と，適切なエージェントの配置について詳しく述べる．

4.2.1　登録画像の評価

各時刻において，現在登録されている顔画像が十分なものであるかを評価するための基準を定義する．理想的には，登録画像はあらゆる照明変動を許容しなければならない．つまり，登録画像から作成される部分空間と，あらゆる照明条件のもとで撮影された顔画像との距離が小さくなければならない．対象人物の全照明パターンは未知であるため，あらかじめ数人に対してあらゆる照明条件下での画像を撮影しておき，その平均画像を全照明平均パターンとして用いる．全照明平均パターンに含まれる各画像と登録画像から作成される部分空間の距離の最大値を登録画像の評価値として定義する．この評価値は値が小さいほど理想的な登録画像であると判断できる．

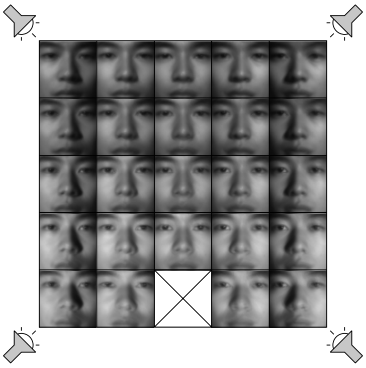


図 9: 全照明平均パターン

図 9 に今回評価に用いた全照明平均パターンを示す．左右に 5 段階，高さを 5 段階に変化させ，合計 25 通りの照明条件のもとで撮影を行った (実際には，カメラ自身の陰になる照明位置を除く 24 通りの画像を用いた)．

4.2.2　適切なエージェントの配置

適切なエージェントの配置は対象の位置や照明条件などによって時々刻々変化する．そこで，現在の状況に応じて Agency 間でエージェントを逐次移動させることによってシステムを動的に変化させ，適切な配置を実現する．

認識モードの追跡 Agency 内に，追跡対象の登録に貢献しないエージェントが存在するなら，そのエージェントが貢献できるタスクを他の Agency に要求する．逆に登録画像が不足している方向があり，その方向を登録できるエージェントがメンバに存在しない場合，他の Agency からその方向を登録できるエージェントを要求する．これらの要求が一致したとき，Agency 間でエージェントを移動させる．以下に移動するエージェントを決定する手順を示す．

1. 現在のモードが登録モードとなっている追跡 Agency について，追跡対象に関する全ての方向の登録画像の評価値を計算する．また，現在の登録画像に各エージェントから得られた

登録画像を追加したときの評価値の減少量を，そのエージェントの登録に対する貢献度として計算しておく．

2. Agency のメンバのうち，貢献度が閾値以下となるエージェントが存在するなら，他の全ての Agency にタスクを要求するメッセージを送る．このとき付加情報としてエージェントが動作する観測ステーションの位置や有効な観測範囲などを送る．
3. メッセージを受け取った Agency は，付加されている観測ステーションの 3 次元位置や観測範囲から，タスクを要求しているエージェントによって登録可能な顔画像の方向を計算する．現在その方向を観測しているエージェントがメンバに存在せず，かつその方向の登録画像の評価値が閾値以上である場合，メッセージを返す．このとき登録画像の評価値を付加情報として送る．
4. 最初の Agency は，返ってきたメッセージの中から付加されている評価値が最大のものを選択し，そのエージェントとの契約を成立させる．契約の成立した Agency には契約成立のメッセージを送り，それ以外の Agency には要求をキャンセルするメッセージを送る．
5. 契約の成立した Agency 間で，エージェントの移動を行う．

5　実験

複数のカメラを用いて能動的登録を行うシステムのプロトタイプを作成し，本手法の有効性を検証した．

5.1　試作システムの概要

本システムの各エージェントは以下の要素から構成される．

- Pan-Tilt-Zoom コントロール可能なカメラ (SONY EVI-G20)
- SIMD 型並列画像入力/処理用プロセッサ (IMAP-VISION)
- カメラコントロール，高次の画像処理，認識用の PC(PentiumII 300MHz)

また，各エージェント間は Fast Ethernet によって接続する．

以上のようなシステムを使って，複数カメラによる複数人物の登録実験を行った．今回は，原理的な有効性を確認することを目的とし，以下のような単純な条件のもとで実験を行った．

- カメラ 2 台，対象人物 2 人とし，各カメラの初期の観測対象は予め与えておく．また，新たな対象は登場しないものとする．このとき，カメラと対象人物が同数なので，単一のエージェントのみで Agency が構成され，互いに Agency マスターどうしの交換を行うことのみが可能である．
- 対象の 3 次元位置，姿勢は固定とし，予め与えておく．また，照明位置を適当な間隔で移動させ，対象の照明条件を変える．

比較のために，Agency 間で協調しない場合とする場合それぞれについて実験を行った．

5.2　協調しない場合

Agency 間で協調しない場合の登録の経過を図 10，11 に示す．それぞれ図 10 が登録された顔画像，図 11 がそのときの評価値の遷移を表している．登録画像を見ると，決まった方向のみの登録が行われ，他の方向について の登録は行われていないことがわかる．評価値についても正面に関しては低い値になっているが，他の方向には変化していない．また，時間がたつにつれ，評価値が変化しなくなり，無駄な登録を行っていることがわかる．このように，協調を行わない場合には Agency がエージェントを無駄に占有してしまい，有効な登録が行えない状況が生じる．

5.3　協調する場合

Agency 間で協調する場合の登録の経過を図 12，13 に示す．登録画像を見ると，適当なタイミングで対象の交換が行われ，2 人ともそれぞれのカメラで撮影可能な 2 つの方向について登録が行われていることがわかる．また，評価値の遷移を見ても，2 つの方向について十分な登録がされていることがわかる．Agency 間で協調を行うことによって，適切な登録が行えるように各エージェントを割当てることが可能になる．

6　おわりに

本稿では，分散協調視覚の枠組にもとづいて，複数の分散カメラの協調によって，人物の検出，追跡，登録，認識を統合して扱うシステムを提案した．

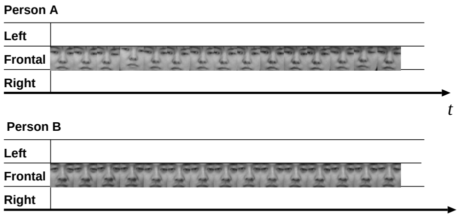


図 10: 登録画像 (協調しない場合)

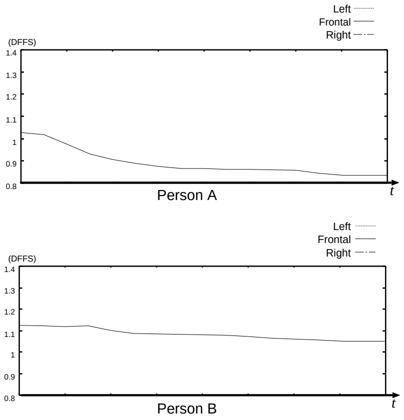


図 11: 評価値の遷移 (登録しない場合)

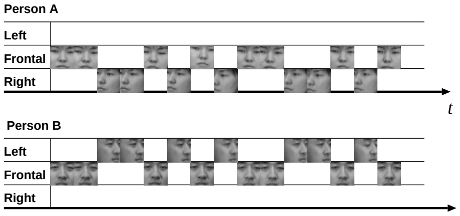


図 12: 登録画像 (協調する場合)

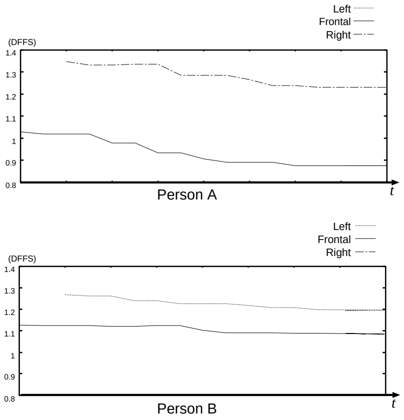


図 13: 評価値の遷移 (協調する場合)

さらに，本システムの具体的なタスクとして，対象を能動的に選択しながら登録を行う能動的登録を提案し，試作システムによる登録実験を行い，本手法によって安定な認識を行うために必要な辞書を効率良く登録できることを示した．

今後の課題としては，1) 顔画像の抽出や，登録画像の評価などの実時間性を要求する処理の高速化， 2) 対象やカメラが多い場合など，複雑な環境での安定な対象検出，追跡， 3) 大規模，かつ広範囲にわたる環境への適用などが挙げられる．

本研究は，日本学術振興会未来開拓学術推進事業 (JSPS-RFTF 96P00501) の補助を受けて行った．

参考文献

[1] Akamatsu, S., Sasaki, T., Fukamachi, H., and Suenaga, Y., “Automatic extraction of target images for face identification using the subspace classification method,” IEICE Trans. Inf. & Syst., vol. E76-D, no. 10, pp.1190-1198, 1993.

[2] Kanade, T., “Picture Processing by Computer Complex and Recognition of Human Faces,” Ph.D. Thesis, Kyoto University, 1973.

[3] Kirby, M. and Sirovich, L., “Application of the Karhunen-Loeve Procedure for the characterization of human faces”, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 12, no. 1, 1990.

[4] Matsuyama, T., “Cooperative Distributed Vision – Dynamic Integration of Visual Perception, Action, and Communication –,” Proc. of Image Understanding Workshop, Monterey CA 1998.

[5] 松山 隆司, “分散協調視覚 -視覚 行動 ·コミュニケーション機能の統合による知能の創発-”, MIRU’98, vol.1, pp.343–352, 1998.

[6] 松山 隆司, 和田 俊和, 丸山 昌之, “能動視覚エージェントによる移動対象の協調的追跡”, MIRU’98, vol.1, pp.365–370, 1998.

[7] Moghaddam, B. and Pentland, A., “Probabilistic Visual Learning for Object Representation,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol.19, no.7, pp.696–710, 1997.

[8] Shakunaga, T., Ogawa, K. and Oki, S., “Integration of Eigentemplate and Structure Matching for Automatic Facial Feature Detection,” Proc. FG’98, pp.94–99, 1998.

[9] Shakunaga, T., Mukaigawa, Y., Kato, T. and Oki, S., “Person Detection, Tracking, Registration and Recognition by Cooperative Distributed Vision,” Proc. of 2nd International Workshop on Cooperative Distributed Vision, pp.169–186, 1998.

[10] Turk, M. and Pentland, A., “Eigenfaces for Recognition,” Journal of Cognitive Neuroscience, vol.3, no.1, pp.71–86, 1991.