Underdetermined Approach to Real-time Face Tracking and Recognition

Hisayoshi Chugan Yuki Oka Takeshi Shakunaga
Okayama University
1-1-1, Tsushima-Naka, Kita-ku, Okayama 700-8530, Japan
{chugan,oka,shaku}@chino.cs.okayama-u.ac.jp

Abstract

In a real-time face tracking and recognition system proposed by Oka and Shakunaga, an optimum weighted average of registered images are estimated and the weights are used for face identification and shape inference.Although their method works well even when a target face changes pose in photometric changes, both the person identification and expression recognition could not be robustly solved at the same time because of a capacity problem. This paper proposes underdetermined approach to solve the capacity problem. Although a single underdetermined system often results in some performance reduction, parallel implementation can remarkably improve the performance. Experimental results showed that the proposed method successfully worked for 10-person discrimination when 10 expressions were registered for each person even when an image sequence included many face motion, expression, photometric change and occasional occlusions.

1 Introduction

In the area of face tracking and recognition, many techniques have been developed to cover face appearance changes during tracking. In the statistical approach [3, 10, 1, 7], appearance changes, analyzed with manifolds or other statistic models are used. While these methods seem to provide feasible answers in practical situations, their performance directly depends on the training set. In active appearance model(AAM) approach, a mixed eigenspace of appearances and shapes is used, where shape is represented by the 2D positions of feature points [2, 4, 8]. In the AAM tracker [4], a sophisticated image alignment technique is developed for covering dynamic shape changes. On the other hand, photometric effects have not been efficiently considered as well as shape changes.

In recent work, several kinds of appearance changes have been discussed. Xu et al. [9] showed that a local multilinear model is useful for face tracking in gradual deformation, pose changes and photometric changes, although the proposed system did not work in real-time. Among them, Oka and Shakunaga [5, 6] proposed an efficient method for a real-time tracking and recognition to cover pose and photometric changes. However, the number of face shapes was at most 25 in their real-time implementation, and scalability seems a severe problem since their method needs to solve linear equations. This paper proposes a feasible solution to increase the number of shapes to 100 or more.

2 Weight Equations in Tracking and Recognition

2.1 Sparse 3D eigentracker

The real-time tracking and recognition method proposed in [5, 6] is composed of two eigen-based methods – sparse 3D eigentracker and augmented eigenface. The sparse 3D eigentracker is implemented by a particle filter in 6D pose space and high-dimensional eigenface to track a rigid face with taking photometric effects into account.

The augmented eigenface is the eigenface augmented by an associative mapping to the 3D face shape that is specified by a set of volumetric face models. An associative mapping is generalized to subspace-to-one mappings to cover the photometric image changes of a fixed shape. This technique, called photometric adjustment, is introduced and combined with associative mapping.

However, to keep weight equations stably solvable in the overdetermined system, the number of registered persons should be sufficiently less than the dimensionality of the eigenface. In this paper, we propose an underdetermined approach to refrain the unstability problem even when the number of registered persons is not sufficiently less than the dimensionality.

2.2 Universal and individual eigenfaces

Let $\mathbf{V}_{kl}$ denote an n-dimensional intensity vector of the k-th person under the l-th lighting condition. Let K and L indicate the number of persons, and the number of lighting conditions, respectively. The universal and individual eigenfaces are constructed and used as follows.

When a set of intensity vectors, $\{\mathbf{v}_{kl}\}$, are calculated by $\mathbf{v}_{kl} = \mathbf{V}_{kl}/\mathbf{1}^\top\mathbf{V}_{kl}$, the universal eigenface is constructed by average vector $\overline{\mathbf{v}}$ and m-principal eigenvectors $\mathbf{\Phi}_m$. Let this be described as $\langle\overline{\mathbf{v}}, \mathbf{\Phi}_m\rangle$.

Let $\mathbf{PV}$ denote a part of an image, where $\mathbf{P}$ is an $n \times n$ diagonal matrix having diagonal elements that are either 1 or 0. The projection $\mathbf{s}$ of $\mathbf{PV}$ is calculated by

$$\widetilde{\mathbf{s}} = (\mathbf{P}\widetilde{\mathbf{\Phi}}_m)^+(\mathbf{PV}), \tag{1}$$

when $\widetilde{\mathbf{\Phi}}_m = [\mathbf{\Phi}_m \ \overline{\mathbf{v}}]$ and $\widetilde{\mathbf{s}} = \left[\alpha\mathbf{s}^\top \ \alpha\right]^\top$, and $(\mathbf{P}\widetilde{\mathbf{\Phi}}_m)^+$ denotes the Moore-Penrose pseudo inverse of $\mathbf{P}\widetilde{\mathbf{\Phi}}_m$. Once $\widetilde{\mathbf{s}}$ is calculated from a given part of image $\mathbf{PV}$, the normalized projection of $\widetilde{\mathbf{s}}$ is given by $\widehat{\mathbf{s}} = \left[\mathbf{s}^\top \ \ 1\right]^\top$.

For each person k, a set of $\mathbf{s}$-representations $S_k = \{\mathbf{s}_{kl} \mid l = 1, \cdots, L\}$ is calculated by projection of a set of intensity vectors $\{\mathbf{V}_{kl} \mid l = 1, \cdots, L\}$ to universal eigenface. The k-th individual eigenface $\langle \overline{\mathbf{s}}_k, \mathbf{\Theta}_k \rangle$ is constructed from S_k in the $\mathbf{s}$-domain, where $\overline{\mathbf{s}}_k$ and $\mathbf{\Theta}_k$ denote the average and the k-th individual eigenspace.

2.3 Weight equations

In Oka and Shakunaga [5, 6], linear equations are solved for both person identification and shape inference. Let these be called the weight equations. The definition and the solution of the weight equations are summarized as follows.

(1) Projection to universal eigenface
When an image vector $\mathbf{V}$ is selected and $\mathbf{P}$ is designed, the projection $\mathbf{s}$ of $\mathbf{PV}$ to the universal eigenface is calculated using Eq. (1). (We can set $\mathbf{P} = \mathbf{I}$, when full projection is necessary.)

(2) Photometric adjustment
In the $\mathbf{s}$-domain, for each k, a projection of $\mathbf{s}$ to the k-th individual eigenface is calculated by

$$\mathbf{s}_k = \mathbf{\Theta}_k {\mathbf{\Theta}_k}^\top (\mathbf{s} - \overline{\mathbf{s}}_k) + \overline{\mathbf{s}}_k. \qquad (2)$$

(3) Solving the weight equations
After all $\mathbf{s}_k$ are calculated from $\mathbf{s}$, the following linear equations, named the weight equations, are described as

$$\widehat{\mathbf{S}}_K \mathbf{w} = \begin{bmatrix} \mathbf{s}_1 & \cdots & \mathbf{s}_K \\ 1 & \cdots & 1 \end{bmatrix} \mathbf{w} = \widehat{\mathbf{s}}, \qquad (3)$$

where $\widehat{\mathbf{S}}_K = [\widehat{\mathbf{s}}_1 \ \cdots \ \widehat{\mathbf{s}}_K]$ and $\mathbf{w} = [w_1 \ \cdots \ w_K]^\top$. The optimum solution of Eq. (3) is given by $\mathbf{w} = \widehat{\mathbf{S}}_K^+ \widehat{\mathbf{s}}$ and indicates the weights of individual person. Person identification is accomplished by selecting

$$k_{max} = \text{argmax } w_k. \qquad (4)$$

2.4 Problems with weight equations

In Oka and Shakunaga [5, 6], the weight equations serve an essential role in the tracking and recognition framework.

Let K and M denote the number of persons and $m + 1$, where m is the dimensionality of the eigenface. Then, the computational cost to solve the weight equations is $O(K^3)$ in the overdetermined system. For implementing a real-time tracking and recognition of human faces, there is no fatal problem in the computational cost for the weight equations when $K \approx 100$.

However, since $\widehat{\mathbf{s}}$ and $\widehat{\mathbf{S}}_K$ are generated during tracking, they often include many noise. In order to solve the weight equations stably, K should be sufficiently less than M. On the other hand, M could not be so large because the dimensionality of the eigenface affects the other aspects of the real-time tracker. Therefore, another feasible solution should be found for the weight equations.

3 Parallel Underdetermined Approach

3.1 Solution in underdetermined system and parallel approach

A feasible solution is provided by parallel underdetermined systems as follows: Let m' denote the dimensionality of each independent subspace in the m-dimensional eigenspace. When $m = Jm'$ holds, we can easily implement J-parallel underdetermined systems in the entire eigenspace. That is, the weight equations defined in Eq.(3) are rewritten to

$$\begin{bmatrix} \mathbf{S}_K^{(1)} \\ \vdots \\ \mathbf{S}_K^{(J)} \\ \mathbf{1}^\top \end{bmatrix} \mathbf{w} = \begin{bmatrix} \mathbf{s}^{(1)} \\ \vdots \\ \mathbf{s}^{(J)} \\ 1 \end{bmatrix}, \qquad (5)$$

where $\mathbf{S}_K^{(j)}$ and $\mathbf{s}^{(j)}$ denote the j-th m'-row submatrix and subvector of $\mathbf{S}_K$ and $\mathbf{s}$, respectively. Then, the j-th weight equations are represented as

$$\begin{bmatrix} \mathbf{S}_K^{(j)} \\ \mathbf{1}^\top \end{bmatrix} \mathbf{w}^{(j)} = \begin{bmatrix} \mathbf{s}^{(j)} \\ 1 \end{bmatrix}, \qquad (6)$$

where $\mathbf{w}^{(j)}$ is the optimum solution of the j-th weight equations.

After solving all the equations, the average of all the optimum solutions is represented by

$$\mathbf{w} = \frac{1}{J} \sum_{j=1}^{J} \mathbf{w}^{(j)}. \qquad (7)$$

The final weight vector is used for person identification and shape inference. (When partial projections are combined with weight equations, partial subimages are more precisely approximated by weighted averages of dictionary images. In the cases, similarities are calculated in each subimage in each underdetermined systems.)

When $m' + 1 < K$ holds, Eq. (6) is underdetermined, and there is $(K - m' - 1)$-dimensional solution space of $\mathbf{w}^{(j)}$. In the underdetermined system, the pseudo inverse solution provides $\mathbf{w}^{(j)}$ so as to minimize ${\mathbf{w}^{(j)}}^\top \mathbf{w}^{(j)}$ in the solution space.

In the SVD implementation, since the computational cost of the linear equations is $O((m' + 1)^3)$ in each underdetermined system, the total computational cost for the parallel underdetermined systems is $O(J(m'+1)^3)$. This cost is lower than the computational cost $O(K^3)$ for solving Eq. (3). Although the final optimum vector (Eq. (7)) is different from the solution of Eq. (3), it can provide a reliable approximation.

3.2 Biased weight equations

When the weight equations are underdetermined, all the weights$(w_1 \cdots w_K)$ are estimated so as to minimize ${\mathbf{w}^{(j)}}^\top \mathbf{w}^{(j)}$ in the solution space without considering image similarities between the unknown image $\mathbf{v}$ and each $\mathbf{v}_k$. If appropriate image similarities are considered in the optimization, better weighted average is expected to be generated by the weight equations.

Let us assume that the image similarity between $\mathbf{s}$ and $\mathbf{s}_k$ is given by inverse distance of them.Let us introduce a $K \times K$ diagonal matrix $\mathbf{B}$ to rebalance the weights with considering inverse distances.

$$\mathbf{B} = diag \left(\ d^{-1}(\mathbf{s}, \mathbf{s}_1) \quad \cdots \quad d^{-1}(\mathbf{s}, \mathbf{s}_K) \ \right) \qquad (8)$$

$$\text{where} \qquad d(\mathbf{s}, \mathbf{s}_k) = \sqrt{(\mathbf{s} - \mathbf{s}_k)^\top (\mathbf{s} - \mathbf{s}_k)}. \qquad (9)$$

In $\mathbf{B}$, diagonal terms indicate inverse distances from $\mathbf{s}$ to each $\mathbf{s}_k$. When $\mathbf{s}_k = \mathbf{s}$, a large number should be used for the k-th diagonal term instead of ∞.

Substituting $\mathbf{w}^{(j)} = \mathbf{B}\mathbf{w}'^{(j)}$, the following linear equations, named the biased weight equations, are described as,

$$\widehat{\mathbf{S}}_K^{(j)}\mathbf{B}\mathbf{w}'^{(j)} = \widehat{\mathbf{s}}^{(j)}, \tag{10}$$

where $\widehat{\mathbf{S}}_K^{(j)} = [\widehat{\mathbf{s}}_1^{(j)} \ \cdots \ \widehat{\mathbf{s}}_K^{(j)}]$. The optimal solution of Eq. (10) is given by

$$\mathbf{w}^{(j)} \quad = \mathbf{B}[\widehat{\mathbf{S}}_K^{(j)}\mathbf{B}]^{+}\widehat{\mathbf{s}}^{(j)}. \tag{11}$$

It should be noted that Eq. (11) still provides a solution of the original weight equations. Since $\mathbf{w}'^{(j)\top}\mathbf{w}'^{(j)}$ is minimized in the biased weight equations, $\mathbf{w}^{(j)}$ is optimized with considering distances between $\mathbf{s}$ and each $\mathbf{s}_k$.

In the parallel implementation, each underdetermined system could be transformed to the biased weight equations with using the same distances measured in the universal eigenface.

Relation to nearest neighbor criterion
There is a simple relation between the biased weight equations and the nearest neighbor criterion. In our formulation, the weight equations are specified in the m-dimensional eigenspace, and distances are also measured in the same dimensional eigenspace. However, the dimensionalities can be different from each other.

Suppose that the weight equations are specified in a very low dimensional space with keeping the distances defined in m-dimensional space. When the weight equations get specified in 0-dimensional space, $\widehat{\mathbf{S}}_K = \mathbf{1}^\top$ and $\widehat{\mathbf{s}} = 1$. Therefore, Eq. (11) becomes

$$\mathbf{w} = \mathbf{B}[\mathbf{1}^\top\mathbf{B}]^{+} = \frac{1}{\sum_{k=1}^{K} d^{-2}(\mathbf{s},\mathbf{s}_k)} \begin{bmatrix} d^{-2}(\mathbf{s},\mathbf{s}_1) \\ \vdots \\ d^{-2}(\mathbf{s},\mathbf{s}_K) \end{bmatrix}. \tag{12}$$

In this case, the heaviest person indicated by Eq. (4) becomes equivalent to the nearest neighbor person.

4 Experiments

4.1 Training set and universal and individual eigenfaces

The training set consisting of 10 faces in 10 expressions, called Data-10x10, is used in this paper. For each combination of person and expression, a face shape was taken by a range finder, and 24 images were taken by a camera under different lighting conditions. Therefore, Data-10x10 consists of 2400 images and 100 shapes.

A 140D universal eigenface, called EF10x10, was constructed from 2400 images in Data-10x10. The dimensionality of EF10x10 was determined by the smallest dimension where the cumulative contribution rate reached 95%. The average and the principal components of $\tilde{\mathbf{\Phi}}_m$ of EF10x10 are shown in Fig. 1. The augmented eigenface, called AEF10x10, was also constructed from EF10x10 and associative mapping to 3D shape. In EF10x10, individual photometric eigenfaces were also constructed from 24 images for each person in each expression. The individual eigenfaces were used for photometric adjustment.

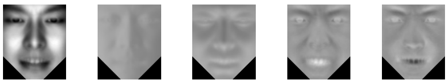
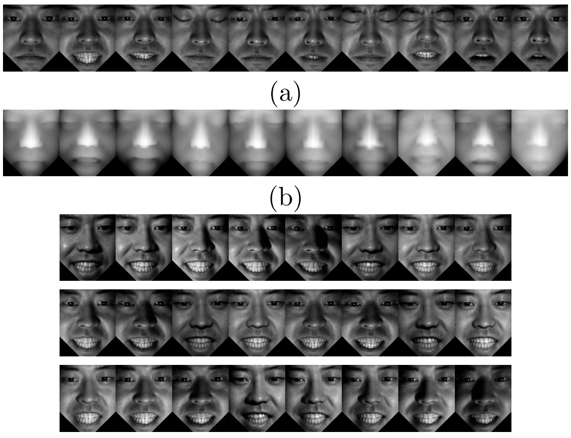


Figure 1. $\tilde{\mathbf{\Phi}}_m$ of EF10x10.



(a)

(b)

(c)

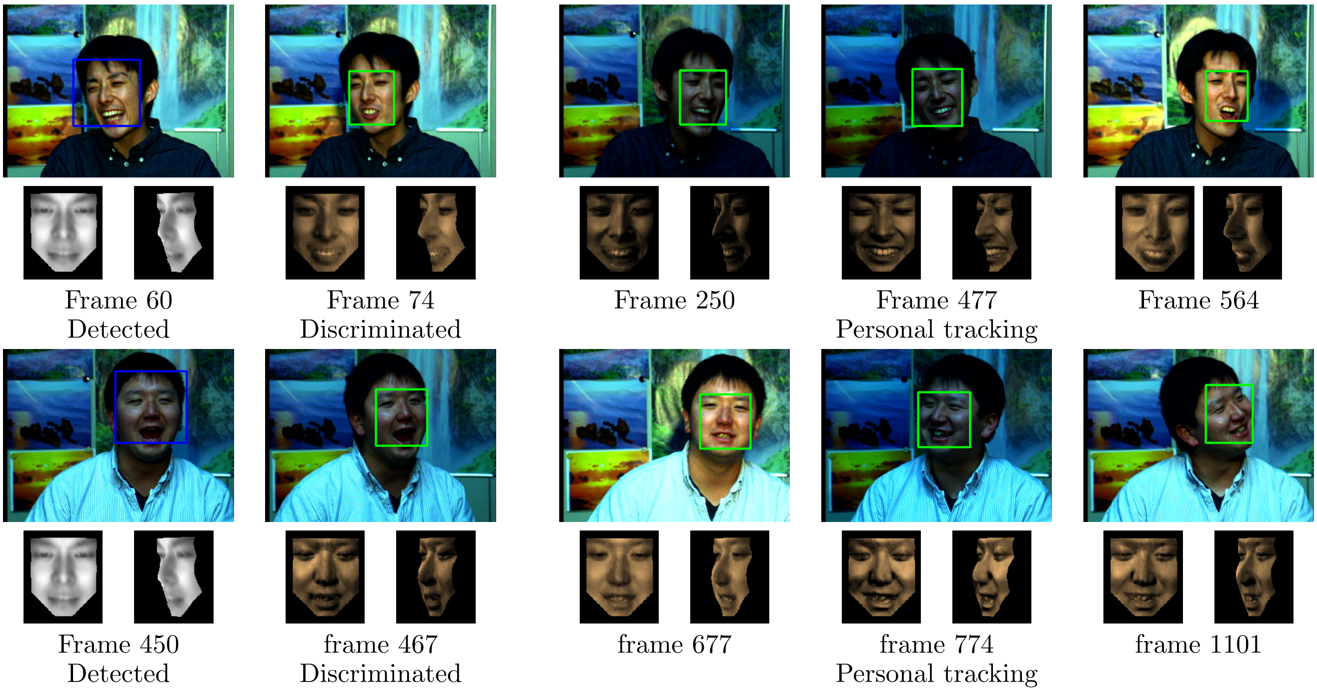
Figure 2. Examples of Data-10x10

4.2 10-person discrimination in expressional and photometric changes

We evaluated the performance of tracking and recognition when augmented eigenface was constructed for 10 persons with 10 expressions for each person. In this experiment, individual eigenfaces were generated for each combination of person and expression from 24 images in each combination. Therefore, 10 expression subspaces were prepared for each person. For each expression of each person, 24 images were continuously taken by switching 24 light sources automatically. A range image was also taken separately by a range sensor. Since a 24-image set and the range image were taken separately, they were automatically calibrated to suppress positional noises. Figure 2 indicates a set of images and shapes: (a) shows 10 expressions of a person taken under a particular lighting condition, (b) indicates 10 shape data of (a), and (c) shows 24 images of a particular expression taken under different lighting condition.

Test image sequences
For each of 10 persons, a 1200-frame image sequence was taken in a room under an artificially controlled lighting condition while the person kept changing expressions and face poses. Therefore, the test image sequences include a lot of photometric and expressional changes along with pose changes.

Three quality levels of cropped images
Some pose estimation errors are inevitable during face tracking because of photometric and pose changes.


Frame 60
Detected

Frame 74
Discriminated

Frame 250

Frame 477
Personal tracking

Frame 564

Frame 450
Detected

frame 467
Discriminated

frame 677

frame 774
Personal tracking

frame 1101

Figure 4. Examples of tracking and recognition.

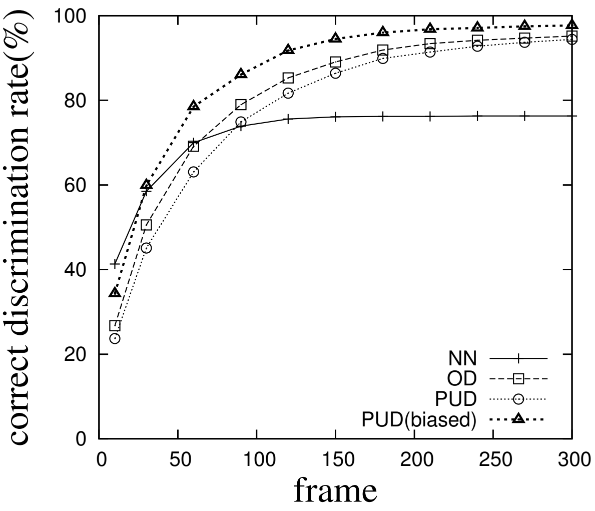


Figure 3. Change of person discrimination rate.

They often affect the quality of projection to the eigenface, and result in erroneous weight estimation. Furthermore, the tracker sometimes fails to track in severe conditions. Taking these situations into account, we classify a cropped image into three classes in each frame and design robust and fast discrimination rules in each quality level.

The three classes are defined by the following three parameters; rotation from estimated pose and frontal face (r), normalized correlation between input image and projected image (c), and maximum weight (w_{max}).These parameters show different aspects of the tracking result. When r is large, the image is likely to be unreliable. Parameter c indicates how well an input image is represented by the universal eigenface, and w_{max} indicates how well the weight equations are solved. By using these parameters, we can define the following three classes:

1. **Good frames** $r \leq 30deg.$, $c \geq 0.995$, $w_{max} \geq \theta_1$.
2. **Effective frames** Not good frames, $r \leq 30deg.$, $c \geq 0.992$, and $w_{max} \geq \theta_2$.
3. **Ineffective frames** All other frames.

θ_1 and θ_2 are tuned to 0.7 and 0.65 for the overdetermined system, and 0.35 and 0.3 for the parallel underdetermined system. Note that the parallel underdetermined system provides milder peaks than does the overdetermined system. Therefore, θ_1 and θ_2 were set to smaller values.

Discrimination rules for image sequence

Robust and fast face discrimination rules are defined for good and effective frames as follows:

1. If the current frame is good, the image is discriminated as a person whose weight is maximum.
2. If the current frame is effective, and if a person whose weight is maximum is the same as the person in the previous and the second previous effective frames, the image is discriminated as the person.

Compared methods

The following methods are compared:

1. NN: Nearest neighbor discrimination in 140D EF10x10.
2. OD: Overdetermined system of the original weight equations: $\mathbf{s}$ and $\mathbf{S}_K$ are coded in 140D EF10x10.

3. PUD: 5-parallel underdetermined systems of the weight equations: $\mathbf{s}^{(j)}$ and $\mathbf{S}_K^{(j)}$ are coded in the j-th 28D subspace for EF10x10.

4. PUD(biased): 5-parallel underdetermined systems of the biased weight equations.

Experimental results
Figure 3 shows the correct discrimination rates vs. the number of frames from detection when 100 detection frames were randomly selected between 0 and 900 from each 1200-frame sequence. Therefore, 1000 test sequences in total were used for the experiment.

Among the compared methods, PUD(biased) showed the highest discrimination rate, and PUD also provided good result. Although OD and PUD(biased) provided no false discrimination (with 1.8% and 1.9% rejection, respectively), these curves showed that parallel underdetermined approach could perform person identififation more quickly than the original method.

As shown in Fig. 3, PUD reached a correct answer with probability about 60% within 30 frames (= 1 sec) and with probability 86% within 90 frames (= 3 sec). OD reached a correct answer with probability 26% within 30 frames and with probability 78% within 90 frames. These results showed that PUD could remarkably improve the performance.

Figure 4 shows how AEF10x10 worked with the parallel underdetermined method on test image sequences. When a face was detected, as shown in the leftmost image, the initial texture and shape were set to average ones and they were gradually updated by the augmented eigenface during tracking the face. Person identification was accomplished in the second image. After the identification, the identified person was more stably tracked and expression recognition was continued using 10 subspaces of the particular person, as shown in the rest 3 images. Under each input image, interpreted face expression is shown in frontal and oblique views. All the processes worked in real-time.

4.3 Comparison of computational time

As mentioned in 3.1, the total computational cost of the Eqs. (3) and (11) are $O(K^3)$ and $O(J(m'+1)^3)$, respectively.

In the current implementation, we compared processing times of the weight equations(Eq. (3)) and the biased weight equations(Eq. (11)). The processing times of Eqs.(3) and (11) were about 3.3 msec and 2.1 msec when 140D eigenface is applied for identification.

In another experiment, when the number of registered person increases 249, the processing times of Eq.(3) and Eq.(11) were about 6.1 msec and 2.1 msec. These results showed that parallel underdetermined approach can improve the computational cost of weight estimation.

5 Conclusions

This paper proposed the underdetermined approach to solving the weight equations proposed in Oka and Shakunaga [5, 6]. Although a single underdetermined system often results in some performance reduction, parallel implementation of underdetermined systems can remarkably improve the performance reduction. Experimental results show that the proposed method successfully works in real-time face tracking and recognition of 10 faces with 10 expressions.

Acknowledgment

This work has been supported in part by a Grant-In-Aid for Scientific Research (No.20300067) from the Ministry of Education, Science, Sports, and Culture of Japan.

References

[1] O. Arandjelovic, G. Shakhnarovich, J. Fisher, R. Cippola, and T. Darrel. Face recognition with image sets using manifold density divergence. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 581–588, 2005.

[2] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):681–685, 2001.

[3] K. C. Lee, J. Ho, M. H. Yang, and D. J. Kriegman. Video-based face recognition using probabilistic appearance manofolds. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 313–320, 2003.

[4] I. Matthews and S. Baker. Active appearance aodels revisited. *International Journal of Computer Vision*, 60(2):135–160, 2004.

[5] Y. Oka and T. Shakunaga. Sparse eigentracker augmented by associative mapping to 3d shape. In *Proceedings of IEEE Conference on Automatic Face and Gesture Recognition*, pages 649–656, 2011.

[6] Y. Oka and T. Shakunaga. Real-time face tracking and recognition by sparse eigentracker augmented by associative mapping to 3d shape. *Image and Vision Computing Journal*, 30(3):147–158, 2012.

[7] R. Wang, S. Shan, X. Chen, and W. Gao. Monifold-manifold distance with application to face recognition based on image set. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2008.

[8] H. Wu, X. Liu, and G. Doretto. Face alignment via boosted ranking model. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2008.

[9] Y. Xu and A. K. Roy-Chowdhury. A physics-based analysis of image appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33:1681–1688, 2011.

[10] S. K. Zhou, R. Chellappa, and B. Moghaddam. Visual tracking and recognition using appearance adaptive models in particle filter. *IEEE Transactions on Image Processing*, 13(11):1491–1506, 2004.