# Integrated Expression-Invariant Face Recognition with Constrained Optical Flow

Chao-Kuei Hsieh[1], Shang-Hong Lai[2], and Yung-Chang Chen[1]

[1] Department of Electrical Engineering, National Tsing Hua University, Taiwan
[2] Department of Computer Science, National Tsing Hua University, Taiwan
d903915@oz.nthu.edu.tw, lai@cs.nthu.edu.tw,
ycchen@ee.nthu.edu.tw

**Abstract.** Face recognition is one of the most intensively studied topics in computer vision and pattern recognition. A constrained optical flow algorithm, which combines the advantages of the unambiguous correspondence of feature point labeling and the flexible representation of optical flow computation, has been proposed in our pervious work for face recognition from expressional face images. In this paper, we propose an integrated face recognition system that is robust against facial expressions by combining information from the computed intra-person optical flow and the synthesized face image in a probabilistic framework. Our experimental results show that the proposed system improves the accuracy of face recognition from expressional face images.

**Keywords:** Face recognition, expression recognition, constrained optical flow, expression normalization.

## 1 Introduction

Face recognition has been studied for the past few decades. Even though the 2D face recognition methods have been actively studied in the past, there are still inherent disadvantages and drawbacks. It was shown that the recognition rate can drop dramatically when the head pose and illumination variations are too large, or when there is expression on the face image. Pose, illumination, and expression variations are three essential issues to be dealt with in the research of face recognition.

Some authors have proposed different approaches to deal with such expression variations. One way [1] is to compute the optical flow between the testing and training face image. Another way [2] used a mask or a morphable model for the image registration in a face recognition system. In our previous work, we combined the advantages of the above two approaches: the unambiguous correspondence of feature point labeling and the flexible representation of optical flow computation. A constrained optical flow algorithm was proposed, which can deal with position movements and intensity changes at the same time when handling the corresponding feature points. We have then applied the algorithm not only to the application of face recognition from expression normalization [3], but also on the inter- and intra-person

optical flow analysis [4], which can be used for further face and expression recognition. Both methods can improve the accuracy of the face recognition from expressional face images, even though different information is utilized in these two algorithms. In this paper, we propose to exploit two different types of information, i.e. the computed optical flow and the synthesized image, to improve the accuracy of face recognition. Experimental validation is given to show the improved performance of the proposed face recognition system.

The remaining of this paper is organized as follows. We briefly review the constrained optical flow computational technique and our previous works on expression normalization and expression optical flow analysis in section 2 and 3, respectively. The proposed face recognition system is presented in section 4. Section 5 gives some experimental results and section 6 concludes this paper.

## 2 Constrained Optical Flow Computation

The computational algorithms of traditional optical flow cannot guarantee that the computed optical flow corresponds to the exact pixels in different images, since the intensity variations due to expression may mislead the computation of optical flow. Teng et al. [5] proposed to minimize the following discrete energy function to compute the optical flow, which used an adaptive smoothness adjustment scheme considered both flow components ($u_i$ and $v_i$) and all the brightness variation multiplier and offset factors ($m_i$ and $c_i$)

$$
\begin{aligned}
f(\mathbf{u}) = &\sum_{i \in D} w_i \left( \frac{I_{x,i} u_i + I_{y,i} v_i + I_{t,i} + m_i I_i + c_i}{\sqrt{I_{x,i}^2 + I_{y,i}^2 + I_i^2 + 1}} \right)^2 \\
&+ \lambda \sum_{i \in D} \left( \alpha_{x,i} u_{x,i}^2 + \alpha_{y,i} u_{y,i}^2 + \beta_{x,i} v_{x,i}^2 + \beta_{y,i} v_{y,i}^2 \right) \\
&+ \mu \sum_{i \in D} \left( \gamma_{x,i} m_{x,i}^2 + \gamma_{y,i} m_{y,i}^2 + \delta_{x,i} c_{x,i}^2 + \delta_{y,i} c_{y,i}^2 \right),
\end{aligned}
\tag{1}
$$

Furthermore, it can be rewritten in a matrix-vector form and efficiently solved by the incomplete Cholesky preconditioned conjugate gradient (ICPCG) algorithm [5].

In order to guarantee the computed optical flow to be consistent to the motion vectors at these corresponding feature points, we modify the unconstrained optimization problem in the original formulation of the optical flow estimation to a constrained optimization problem [3] given as follows:

$$
minimize \ f(\mathbf{u}) = \mathbf{u}^T K \mathbf{u} - 2\mathbf{u}^T \mathbf{b} + c,
$$
$$
subject \ to \ u(x_i, y_i) = \bar{u}_i, \ and \ v(x_i, y_i) = \bar{v}_i, \forall (x_i, y_i) \in S
$$

where $S$ is the set of feature points and $(\bar{u}_i, \bar{v}_i)$ is the specified optical flow vector at the $i^{th}$ feature point. We applied a modified ICPCG procedure to solve this constrained optimization problem and the details are referred to [3].

## 3   Previous Works

In this section, we briefly review expression normalization [3] and inter- and intra-person optical flow analysis [4] in our previous works.

### 3.1   Expression Normalization

The face recognition problem can be considered as to determine the class $c$ that minimizes the difference between the reference image $\boldsymbol{R}_c$ and the synthesized neutral image from the testing facial expression image $\boldsymbol{T}$ by using $\boldsymbol{R}_c$ as the reference. After the image alignment and normalization of the input testing image $\boldsymbol{T}$ and reference image $\boldsymbol{R}_c$, where $c = 1,2,\ldots,C$, and $C$ is the total number of subjects in the face database, we can formulate the face recognition problem as follows:

$$\arg\min_c \left\| \boldsymbol{R}_c - Syn\big(\boldsymbol{T};OF\left(\boldsymbol{T};\boldsymbol{R}_c\right)\big) \right\|  \tag{2}$$

To further improve the computational efficiency, we modify the face recognition problem as follows:

$$\arg\min_c \left\| Syn\big(\boldsymbol{R}_c;OF\left(\boldsymbol{R}_c;\boldsymbol{NE}_0\right)\big) - Syn\big(\boldsymbol{T};OF\left(\boldsymbol{T};\boldsymbol{NE}_0\right)\big) \right\|  \tag{3}$$

where $NE_0$ is a universal neutral face image. To be more specific, instead of transforming the input image to the neutral image for each class, we now transform all images to a universal coordinate as $NE_0$.

   We define the operation, $Syn\big(\boldsymbol{T};OF\left(\boldsymbol{T};\boldsymbol{R}_c\right)\big)$, as the *OF-Syn* operator. The modified system flow chart is shown in Fig. 1. Although there are $C+1$ OF-Syn operators in total, the $C$ OF-Syn operations among them can be performed in advance, thus only one such operation is needed in the testing or recognition phase.
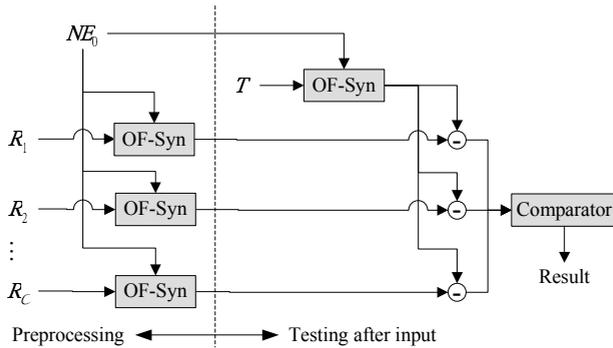


**Fig. 1.** The proposed expression-invariant face recognition flow chart according to Eq. (3)

### 3.2   Expression Optical Flow Analysis

The traditional expressive optical flow is computed from a neutral face image $NE_i$ of person $i$ to an expression image $EX_{i,k}$ with expression k of the same subject. However,

the computed optical flows are generally not in the same coordinate, since the geometry of neutral faces is different for different persons. Some research only considered motion vectors at certain feature points to overcome this problem, but only limited information about facial movement is used in this case.
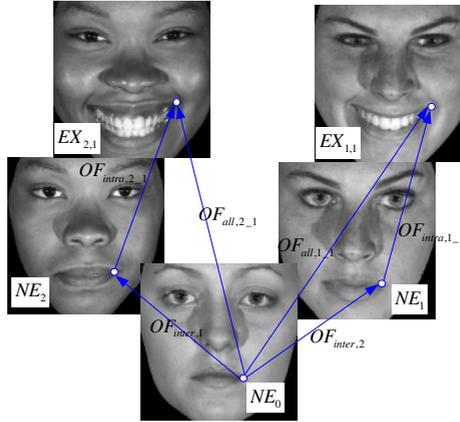


**Fig. 2.** Illustration of decomposing input optical flow ($OF_{all}$) to inter-person ($OF_{inter}$) and intra-person ($OF_{intra}$) parts

We proposed a different solution for optical flow normalization, as shown in Fig. 3. Instead of computing the intra-person optical flow $OF_{intra,i\_k}$ directly from the neutral face to an expressive face image for each person, we start from a global neutral face $NE_0$ to obtain the inter-person optical flow $OF_{inter,i} = OF(NE_0; NE_i)$ and overall optical $OF_{all,i\_k} = OF(NE_0; EX_{i,k})$. The intra-person optical flow can then be generated by pixel-wise differencing as follows:

$$OF_{intra,i\_k} = OF_{all,i\_k} - OF_{inter,i}. \qquad (4)$$

There are two advantages of doing it this way: (1) all expressive face images of every subject have the same number of motion vectors; (2) all optical flows are computed on the same geometry of $NE_0$.

After obtaining the normalized optical flows from different expressions, we consider it as a problem of subspace modeling. In other words, we can extract $K$ optical flow bases $OB_{intra,k}$ to describe intra-person optical flows. Moreover, when there is an input image, we can recognize the face by determining the person $i$, whose inter-person optical flow $OF_{inter,i}$ makes the intra-person optical flow be best spanned by the trained optical flow bases. This can be formulated as the following optimization problem:

$$\underset{i,\, b_{k,k=1,\dots,K}}{\arg\min} \left\| \left( OF_{input} - OF_{inter,i} \right) - \sum_{k=1}^{K} b_k OB_{intra,k} \right\| \qquad (5)$$

Furthermore, the spanning coefficients $b_k$ of each basis $OB_{intra,k}$ may be used for expression recognition. A negative coefficient does not make sense in physical

expression motion, thus we proposed a modified eigenvector algorithm to enforce non-negative projection coefficients.

## 4   Proposed Face Recognition System

There are two types of information generated by the constrained optical flow algorithm: the optical flow and the synthesized neural face image by image warping with the computed optical flow. As discussed in the previous section, the two face recognition methods based on the expression normalized images and the computed optical flow are described in section 3.1 and 3.2, respectively. In the first method, the optical flows for the input expression-variant images are different from different subjects in terms of geometry and dimensionality, thus the optical flow information was not used for comparison in the first method. On the other hand, in the second method, we compute the optical flow in the opposite direction, i.e. from the global neutral face to an input expression variant face, to preserve the same geometry for the computed optical flow, but it cannot be used directly to synthesize the corresponding neutral image for comparison. Only partial information was exploited in each of these two methods.

In this paper, we intend to integrate these two methods into the proposed expression-invariant face recognition system to fully exploit the information of the optical flow and the synthesized neutral image for face recognition. To do this, we formulate the problem as follows:

$$\max_{\mathbf{N}_i,E} P(\mathbf{N}_i, E \mid \mathbf{I}), i = 1, 2, ..., N \tag{6}$$

where $\mathbf{I}$ is the input image, $\mathbf{N}_i$ is a neutral face image in training data set, and $E$ denotes the expression optical flow vector between $\mathbf{I}$ and $\mathbf{N}_i$. Based on the posterior probability, equation (6) can be rewritten as

$$\max_{\mathbf{N}_i,E} P(\mathbf{N}_i) P(E) P(\mathbf{I} \mid \mathbf{N}_i, E) \cdot \tag{7}$$

Furthermore, the occurrence probability of each candidate is assumed equally probable, i.e. $P(\mathbf{N}_i)$ is a constant for all $i$. The formulation can be simplified as

$$\max_{\mathbf{N}_i,E} P(E) P(\mathbf{I} \mid \mathbf{N}_i, E) \cdot \tag{8}$$

There are two parts in equation (8), i.e. the probability of the expression movement $P(E)$, and the probability of the input image under the condition of the subject $\mathbf{N}_i$ with the expression $E$. As discussed before, a single type of optical flow cannot keep the uniformity of dimensionality and geometry in both circumstances. Thus. we define $P(E)$ and $P(\mathbf{I} \mid \mathbf{N}_i, E)$ separately.

To further define $P(E)$ with preservation of identical geometry and dimensionality for each $\mathbf{N}_i$, we use the same strategy in method 2, i.e. the intra-person optical flow. With equation (4), the motion information $E$ in $P(E)$ is defined as

$$\boldsymbol{u}(x, y)_{@NE_0} = \boldsymbol{v}(x, y)_{@NE_0} - \boldsymbol{w}(x, y)_{@NE_0}, \tag{9}$$

where $v(x, y)$ is the overall optical flow from global neutral face $NE_0$ to input image $\mathbf{I}$, $w(x, y)$ is the inter-person optical flow from $NE_0$ to the guessed neutral face N, and $u(x, y)$ is the intra-person optical flow from $\mathbf{N}_i$ to $\mathbf{I}$. Moreover, and the symbol '$@NE_0$' denotes the optical flow represented with the geometry of $NE_0$, even though the intra-person optical flow is defined as the pixel-wised movement from $\mathbf{N}_i$ to $\mathbf{I}$.

The probability of $P(\mathbf{I} \mid \mathbf{N}_i, E)$ is then defined as the similarity between input image $\mathbf{I}$ and the synthesized image from neutral face $\mathbf{N}_i$ and the computed optical flow movement, i.e.

$$P\left(\mathbf{I} \mid \mathbf{N}_i, E\right) \propto \exp\left\{-\left\|I - Syn\left(N_i; \boldsymbol{u}_{@N_i}\right)\right\|^2 \middle/ \sigma^2\right\}. \tag{10}$$

Since the optical flow used for synthesizing neutral face $\mathbf{N}_i$ to a certain expression must be represented with the same geometry of $\mathbf{N}_i$, the intra-person optical flow $u(x, y)_{@NE_0}$ in equation (9) is not appropriate in this circumstance. An estimated intra-person optical flow under the geometry of $\mathbf{N}_i$ is needed in equation (10). The MAP optimization problem is now rewritten as

$$\max_{\mathbf{N}_i, E} P\left(E\right) P\left(\mathbf{I} \mid \mathbf{N}_i, E\right) = \max_{\mathbf{N}_i, E} P\left(\boldsymbol{u}_{@NE_0}\right) \exp\left\{-\left\|I - Syn\left(N_i; \boldsymbol{u}_{@N_i}\right)\right\|^2 \middle/ \sigma^2\right\}, \tag{11}$$

Thus two components are needed to compute in the above formulation: one is the intra-person optical flow at the geometry of $NE_0$ ($Intra@NE_0$), and the other is the intra-person optical flow at the geometry of $NE_i$ ($Intra@\mathbf{N}_i$). We can first compute the optical flow $Intra@NE_0$ directly and warp it to obtain $Intra@NE_i$ (Fig. 1, defined as procedure 1), or oppositely, we can compute $Intra@NE_i$ first and then warp to $Intra@NE_0$ (Fig. 1, defined as procedure 2). In the first flow, we compute $Intra@NE_0$ as the diagram shown in Fig. 3, i.e.

$$\begin{aligned} Intra @ NE_0 &\triangleq Input @ NE_0 - Inter @ NE_0 \\ &= OF\left(NE_0; Input\right) - OF\left(NE_0; NE_i\right). \end{aligned} \tag{12}$$

The optical flow $Intra@NE_i$ can be simply obtained by computing $OF(NE_i; Input)$. After collecting one type of intra-person optical flow, we can further obtain the other one by nonlinear warping with the inter-person optical flow as shown in Fig. 4. Take $Intra@NE_i$ to $Intra@NE_0$ for example. The movement of each pixel has been obtained in $Intra@NE_i$ calculation. The corresponding position in $NE_0$ of each pixel in $NE_i$ can be determined easily through inter-person optical flow $Inter@NE_i$. In most cases, the corresponding position is not on integer grid. We can estimate the motion of each non-integer pixel by bilinear interpolation.

The overall system flowcharts of procedure 1 and 2 are depicted in Fig. 5 and Fig. 6, respectively. In procedure 1, since the $OF$ block of $OF(NE_0; NE_i)$ can be pre-computed in the training process, only one optical flow calculation, the $Input@NE_0$ is needed in the testing process. However, the optical flow used for synthesis, which requires more precision, is obtained through a long computational procedure. This may damage the
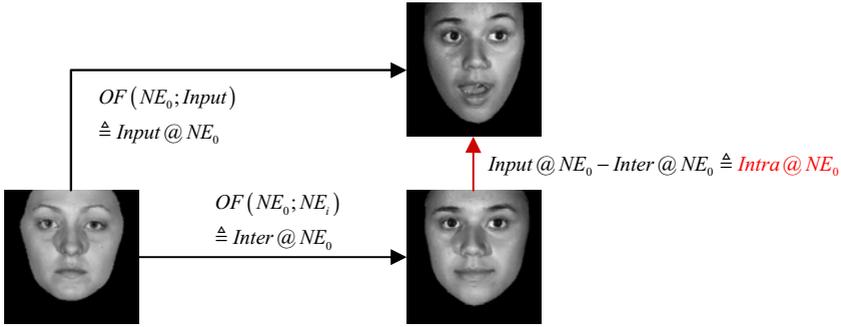
$OF(NE_0; Input)$
$\triangleq Input @ NE_0$

$Input @ NE_0 - Inter @ NE_0 \triangleq Intra @ NE_0$

$OF(NE_0; NE_i)$
$\triangleq Inter @ NE_0$

**Fig. 3.** Intra-person optical flow, $Intra @ NE_0$ calculation in procedure 1

$Intra @ NE_0$

$OF(NE_i; Input)$
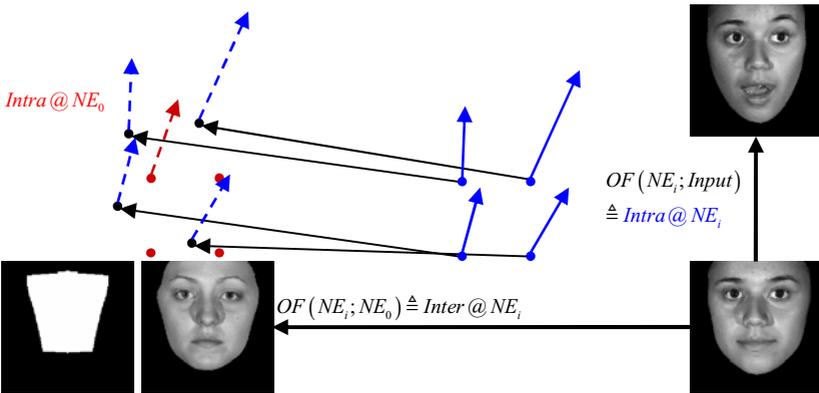$\triangleq Intra @ NE_i$

$OF(NE_i; NE_0) \triangleq Inter @ NE_i$

**Fig. 4.** Diagram of intra-person optical flow mapping from one person to another

quality of the synthesized result if any non-negligible inaccuracy is involved during the flow.

In procedure 2, on the other hand, even though the operation $OF(NE_0; NE_i)$ can be pre-computed in the training process as well, the *OF* block $OF(NE_i; Input)$ is required to compute for each guess. There are totally *C* times of *OF* computations in the testing procedure, which are directly followed by image synthesis block.

The objective function can be computed after the two types of intra-person optical flow are calculated. We considered the probability of intra-person optical flow as a mixture of Gaussian probability distribution with centers corresponding to the samples in the training optical flow dataset, that is,

$$P\left(u_{@ NE_0}\right) = \sum_{i=1}^{T} f_i(y) \propto \sum_{i=1}^{T} \exp\left\{-\frac{1}{2}(y-\overline{y}_i)^T \Sigma^{-1}(y-\overline{y}_i)\right\}, \qquad (13)$$

where $\overline{y}_i$ is the intra-person optical flow in the training data, *i* is the index and *T* is the number of training sample, and *y* denotes $u_{@ NE_0}$. As for the similarity between the input image and the synthesized images, we directly calculate the reciprocal of the
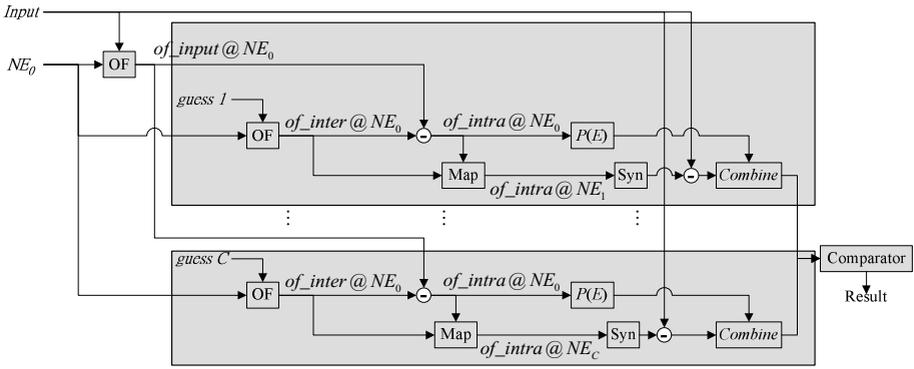
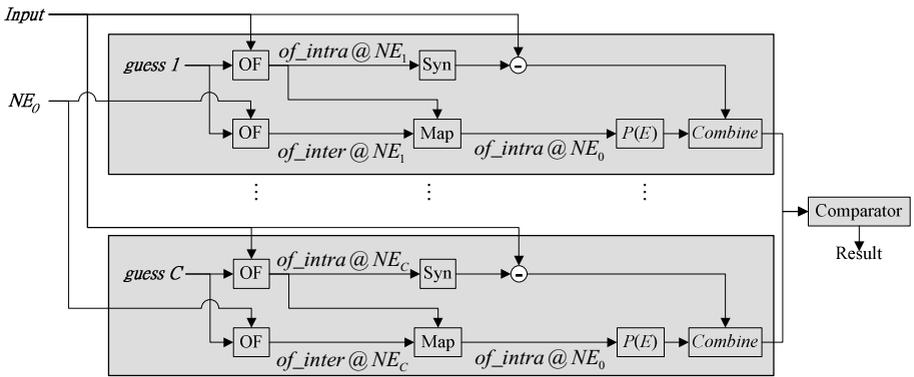**Fig. 5.** System flowchart of procedure 1



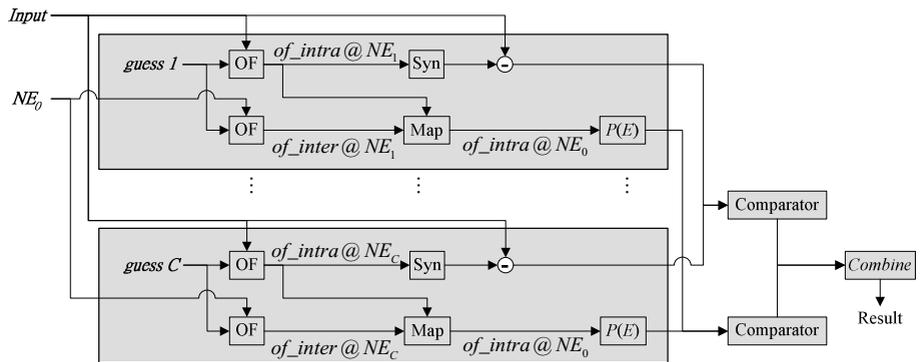**Fig. 6.** System flowchart of procedure 2



**Fig. 7.** Modified system flowchart of procedure 2

average of pixel-wise difference between them. According to the original definition, since the two values are under different scales, one will dominate the result of final probability. Instead of comparing the final result after combining the two values as

shown in Fig. 6, we modify the flow to Fig. 7. We compare and score the two values separately first, and then combine the scores for final decision as

$$S_{Final} = r \times S_{P(E)} + S_{P(\mathbf{I}|\mathbf{N}, E)}, \tag{14}$$

where $r$ is a weight determined empirically. $S_{P(E)}$ and $S_{P(\mathbf{I}|\mathbf{N}, E)}$ stand for the score of $P(E)$ and $P(\mathbf{I}|\mathbf{N}, E)$ respectively.

## 5    Experimental Results

Our experiments were performed on the Binghamton University 3D Face Expression (BU-3DFE) Database [6]. The BU-3DFE database contains the face images and 3D face models of 100 subjects (56 females and 44 males) each with a neutral face and 6 different expressions (angry, disgust, fear, happy, sad, and surprised) at different levels (from level 1 (weakest) to 4 (strongest)). Note that only the 2D face images were used in our experiments. Among them, 34 subjects are randomly selected for intra-person optical flow training, and the others are used as the testing set.

### 5.1    Pre-processing

We manually labeled 21 feature points, including 3 points for each eyebrow and 4 points for each eye, one at the nose tip and the other 6 around the mouth region. With the labeled points, the distance between the outer corners of both eyes is used as the reference to normalize face images.

### 5.2    Face Recognition with Proposed System

As described in the previous section, we follow the modified flowchart shown in Fig. 8. We apply a mask (Fig. 8(b)) defined from the global neutral face $NE_0$ (Fig. 8(a)) to extract the region of interest. Moreover, the region inside the mouth is discarded, as illustrated in Fig. 8(e). Both the optical flow and the grayscales of the synthesized image within the mask will be used in face recognition process. Some experimental images are shown in Fig. 8. For an input image (Fig. 8(d)), we first position the corresponding mask (Fig. 8(e)) to obtain the masked image (Fig. 8(f)). After that, for each candidate in the database (Fig. 8(g) and 8(j)), the intra-person optical flow, i.e. $intra@NE_i$, is computed and used for virtual image synthesis (Fig. 8(h) and Fig. 8(k)). The masked images (Fig. 8(i) and 8(l)) can finally be applied for similarity comparison.

   Expression-invariant face recognition results are listed in Table 1-3. According to the results, the average face recognition rates based on the synthesized image or the intra-person optical flow individually are 85.86% and 82.39%, respectively. In addition to using the synthesized image and optical flow information separately, we also carry out the face recognition experiment by using the proposed integrated solution, i.e. based on Equ. (14). In this experiment, 10 to 1 points are given to the top ten candidates in each comparison, and we equally weight the information of synthesized image and intra-person optical flow, i.e. $r = 1$. The recognition rate is improved to 90.28% as shown in Table 3.
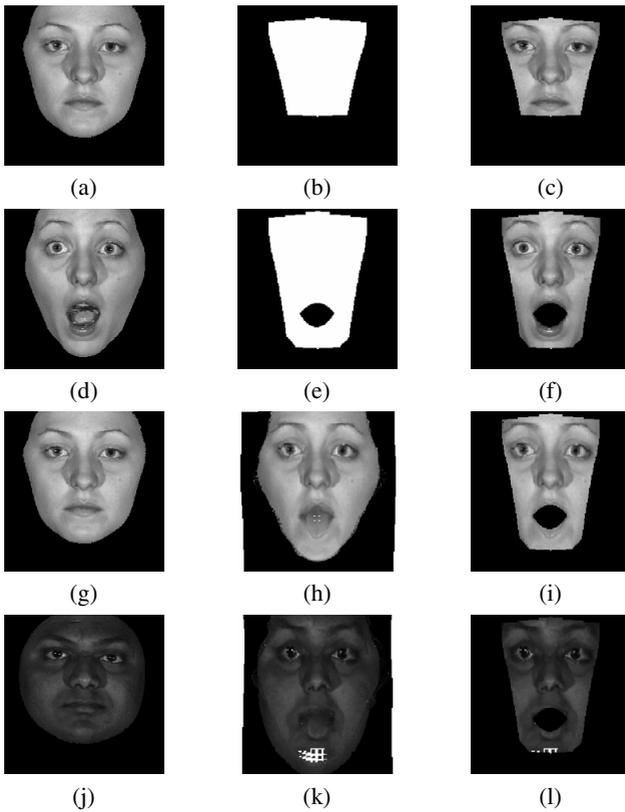
**Fig. 8.** Illustration of experimental images: (a) global neutral face, (b) mask image, (c) masked image of (a), (d) input image, (e) warped mask image, (f) masked input image, (g) guessed subject 1, (h) synthesized face from (g) to (d), (i) masked synthesized image (g) using mask image (e), (j) guessed subject 2, (k) synthesized face from (g) to (d), and (l) masked synthesized image (k) using mask image (e)

**Table 1.** Recognition result using the synthesized face images only

| Expression | Level 1(%) | Level 2(%) | Level 3(%) | Level 4(%) | Average |
|---|---|---|---|---|---|
| AN | 93.94 | 86.36 | 89.39 | 83.33 | |
| DI | 90.91 | 86.36 | 81.82 | 77.27 | |
| FE | 84.85 | 83.33 | 83.33 | 83.33 | |
| HA | 89.39 | 92.42 | 86.36 | 80.30 | |
| SA | 92.42 | 93.94 | 90.91 | 86.36 | |
| SU | 84.85 | 86.36 | 80.30 | 72.73 | 85.86% |

The impact of different weighting is also discussed in our experiment. We try different weightings ranging from 0.5 to 2 and the face recognition accuracies are depicted in Fig. 9. We can see that the proposed system can achieve the best accuracy at 93% when $r = 0.7$, which means the synthesized image is of higher significance.

**Table 2.** Recognition result using intra-person optical flow only

| Expression | Level 1(%) | Level 2(%) | Level 3(%) | Level 4(%) | Average |
|:---:|:---:|:---:|:---:|:---:|:---:|
| AN | 87.88 | 87.88 | 81.82 | 66.67 | |
| DI | 83.33 | 75.76 | 68.18 | 60.61 | |
| FE | 84.85 | 81.82 | 77.27 | 75.76 | |
| HA | 93.94 | 92.42 | 83.33 | 74.24 | |
| SA | 96.97 | 93.94 | 90.91 | 87.88 | |
| SU | 90.91 | 86.36 | 81.82 | 72.73 | 82.39% |

**Table 3.** Recognition result using the integrated information, including the synthesized images and intra-person optical flow

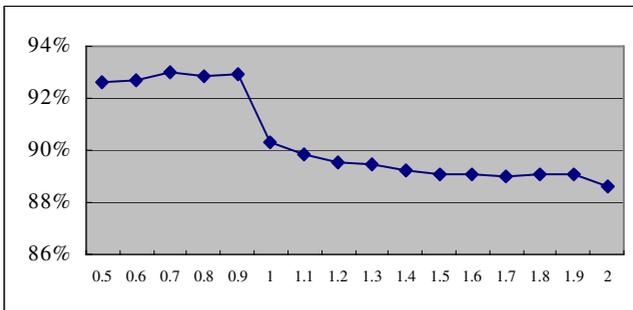| Expression | Level 1 (%) | Level 2(%) | Level 3(%) | Level 4(%) | Average |
|:---:|:---:|:---:|:---:|:---:|:---:|
| AN | 98.48 | 92.42 | 87.88 | 84.85 | |
| DI | 87.88 | 86.36 | 80.30 | 81.82 | |
| FE | 90.91 | 90.91 | 87.88 | 87.88 | |
| HA | 96.97 | 93.94 | 95.45 | 87.88 | |
| SA | 98.48 | 96.97 | 93.94 | 90.91 | |
| SU | 93.94 | 90.91 | 90.91 | 78.79 | 90.28% |



**Fig. 9.** Recognition result on the test data with different weighting parameters in eq.(14)

## 6   Conclusion

In this paper, we proposed an integrated expression-invariant face recognition system based on the constrained optical flow. In the previous works on optical flow based face recognition, either the synthesized face image or the intra-person flow is used individually for face and expression recognition. In this work, we proposed to integrate the information of synthesized images and the intra-person optical flow distribution probability to improve the face recognition accuracy. As the experimental results shows, the proposed system improves the accuracy of face and expression recognition on expressional face images. However, the proposed integrated system is more computationally costly compared to the previous works, since the optical flow computation, intra-person optical flow mapping and image synthesis are needed for all candidates in the database. This is the main research topic in our future study.

## References

1. Li, X., Mori, G., Zhang, H.: Expression-invariant face recognition with expression classification. In: Proc. 3rd Canadian Conf. on Computer and Robot Vision (June 2006)
2. Martinez, A.M.: Recognizing expression variant faces from a single sample image per class. In: Proc. IEEE Conf. Computer Vision Pattern Recognition (June 2003)
3. Hsieh, C.-K., Lai, S.-H., Chen, Y.-C.: Expression-invariant face recognition with accurate optical flow. In: Ip, H.H.-S., Au, O.C., Leung, H., Sun, M.-T., Ma, W.-Y., Hu, S.-M. (eds.) PCM 2007. LNCS, vol. 4810, pp. 78–87. Springer, Heidelberg (2007)
4. Hsieh, C.-K., Lai, S.-H., Chen, Y.-C.: Expressional face image analysis with constrained optical flow. In: Proc. of ICME, Hannover, Germany, June 23-26 (2008)
5. Teng, C.-H., Lai, S.-H., Chen, Y.-S., Hsu, W.-H.: Accurate optical flow computation under non-uniform brightness variations. In: Computer Vision and Image Understanding, vol. 97, pp. 315–346 (2005)
6. Yin, L., Wei, X., Sun, Y., Wang, J., Rosato, M.J.: A 3D facial expression database for facial behavior research. In: Proc. Intern. Conf. on Automatic Face and Gesture Recognition, pp. 211–216 (April 2006)