Journal of Mobile Multimedia, Vol. 10, No. 3&4 (2014) 294–308

FUSION OF VISIBLE IMAGES AND THERMAL IMAGE SEQUENCES FOR AUTOMATED FACIAL EMOTION ESTIMATION

HUNG NGUYEN FAN CHEN KAZUNORI KOTANI
Japan Advanced Institute of Science and Technology
1-1 Asahidai, Nomi, Ishikawa, Japan
nvhung@jaist.ac.jp, chen-fan@jaist.ac.jp, ikko@jaist.ac.jp

BAC LE
University of Science, VNU - HCMC, Vietnam
227 Nguyen Van Cu, Ho Chi Minh city, Vietnam
lhbac@hcmuns.edu.vn

The visible image-based approach has long been considered the most powerful approach to facial emotion estimation. However it is illumination dependency. Under uncontrolled operating conditions, estimation accuracy degrades significantly. In this paper, we focus on integrating visible images with thermal image sequences for facial emotion estimation. First, to address limitations of thermal infrared (IR) images, such as being opaque to eyeglasses, we apply thermal Regions of Interest (t-ROIs) to sequences of thermal images. Then, wavelet transform is applied to visible images. Second, features are selected and fused from visible features and thermal features. Third, fusion decision using conventional methods, Principal Component Analysis (PCA) and Eigen-space Method based on class-features (EMC), and our proposed methods, thermal Principal Component Analysis (t-PCA) and norm Eigen-space Method based on class-features (n-EMC), is applied. Applying our suggested methods, experiments on the Kotani Thermal Facial Emotion (KTFE) database show significant improvement, proving its effectiveness.

Keywords: facial emotions, thermal images, emotion estimation, feature fusion, decision fusion, thermal image sequences, t-PCA, n-EMC, KTFE database.

1 Introduction

The detection and estimation of human emotions is a challenging task. In the last decade, automated estimation of human emotions has attracted the interest of many researchers, because such systems will have numerous applications in security, medicine, and especially human-computer interaction. Many previous works [1], [2], [3] proposed have been inclined towards developing facial expression estimation. Nevertheless, there is a lack of accurate and robust facial expression estimation methods to be deployed in uncontrolled environments. When the lighting is dim or when it does not uniformly illuminate the face, the accuracy decreases considerably. Moreover, human emotions estimation based on only the visible spectrum has proved to be difficult in cases where there are emotion changes that expressions do not show. Using thermal infrared (IR) imagery, which is not sensitive to light conditions, is a new and innovative way to fill the gap in the human emotions estimation field. Besides, human emotions could be manifested by changing temperature of face skin which is obtained by an IR camera. Consequently, thermal infrared imagery gives us more information to help us robustly estimate human emotions. Although there are many significant advantages when we use IR imagery, it has several drawbacks. Firstly, thermal data are subjected to change together with body temperature

caused by variable ambient temperatures. Secondly, presence of eyeglasses may result in loss of useful information around the eyes. Glass is opaque to IR, and object made of glass act as temperature screen, completely occluding the parts located behind them. Hence, the sensitivity of IR imagery is decreased by facial occlusions. Thirdly, there are some facial regions not receptive to the emotion changes. To eliminate the effects of these challenging problems above, we propose fusion of visible images and sequences of thermal images. To estimate seven emotions, we use the fusion of conventional methods, PCA and EMC, and our proposed methods, t-PCA and n-EMC, over obtained the fusion features.

2 Related work

In the recent years, a number of studies have demonstrated that thermal infrared imagery offers a promising alternative to visible imagery in facial emotion estimation problems by better handling the visible illumination changes. Jarlier et al. [4] extracted the features as representative temperature maps of nine action units (AUs) and used K-nearest neighbor to classify seven expressions. The database for testing has four persons and the accuracy rate is 56.4%. Khan et al. [5] suggested using Facial Thermal Feature Points (FTFPs), which are defined as facial points that undergo significant thermal changes in presenting an expression, and used Linear Discriminant Analysis (LDA) to classify intentional facial expressions based on Thermal Intensity Values (TIVs) recorded at the Facial Thermal Feature Points (FTFPs). The database has sixteen persons with five expressions and the accuracy rate ranges from 66.3% to 83.8%. Trujillo et al. [6] proposed using a local and global automatic feature localization procedure to perform facial expression in thermal images. They used PCA to reduce the dimension and interest point clustering to estimate facial feature localization and Support Vector Machine (SVM) to classify three expressions. B.Hernández et al. [7] used SVM to classify the expressions "surprise", "happy", "neutral" from two inputs. The first input consists of selections of a set of suitable regions where the feature extraction is performed, second input is the Gray Level Co-occurrence Matrix used to compute region descriptors of the IR images. Nhan et al. [8] extracted time, frequency and time-frequency features from thermal infrared data to classify the natural responses in terms of subject-indicated levels of arousal and valence stimulated by the International Affective Picture System. Yoshitomi et al. [9] used two dimensional detection of temperature distribution on the face using infrared rays. Based on studies in the field of psychology, several blocks on the face are chosen for measuring the local temperature difference. With Back Propagation Neutral Network, the facial expression is recognized. The recognition accuracy reaches 90% with "neutral", "happy", "surprising" and "sad" expressions. However, the testing database is obtained from only one female frontal view. Yoshimomi generated feature vectors by using a two-dimensional Discrete Cosine Transformation (2D-DCT) to transform the grayscale values of each block in the facial area of an image into their frequency components, and used them to recognize five expressions, including "angry", "happy", "neutral", "sad", and "surprise". The mean expression accuracy is 80% with four test subjects [10]. Koda et al. used the idea from [10] and added a proposed method for efficiently updating of training data, by only updating the training data with "happy" and "neutral" facial expression after an interval [11]. The expression accuracy increased from 80% to 87% with this new approach. All these studies with thermal infrared imagery have shown that the facial temperature changing is useful for estimating the human emotions.

Recently, a little attention has been paid to facial emotion estimation by using fusion information from visible images and thermal information. Wang et al. [12] proposed both decision-level and feature-level fusion methods using visible and IR imagery. In feature-level, they used tools for the

Active Appearance Model (AAM) to extract features and extracted three features of head motion for visible feature and calculated several statistical parameters including mean, standard deviation, minimum and maximum as IR features. To select the feature, they used F-test statistic. They also used Bayesians networks (BNs) and SVMs to obtain the feature fusion. In decision-level, BNs and SVMs are used to classify three emotions, happiness, fear and disgust. The results show that their methods improved about 1.35% accuracy compare with only using visible features. Yoshitomi et al. [13] proposed decision-level fusion of voices, visual and IR imagery to recognize the affective states. DCT is used to extract the visible and IR features, then two neutral networks are trained for obtained visible and IR features, respectively. For voice recognition, Hidden Markov Models (HMMs) are used. To decide the results, simple weighted voting is used. Following the related work, there are a few researches using fusion of visible and thermal imagery or these approaches that use the extracted features from a single infrared thermal image may lose some useful information which could be contained in the sequences. Therefore, we consider two methods of facial emotion estimation by fusing visible images and sequence of thermal imagery at decision-level and feature-level respectively.

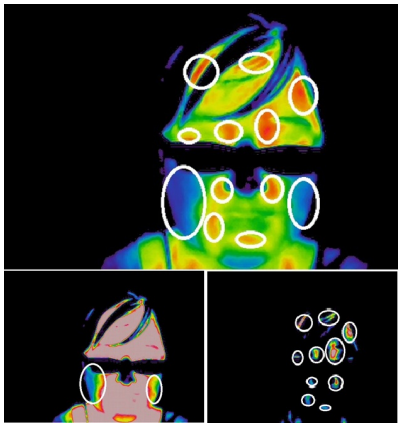

Fig. 1. t-ROIs.

3 Methods

In this section, we propose a feature fusion method to integrate visible images and sequence of thermal images by delicate selection of representative features (i.e. t-ROI) in Section 3.1, two classification methods (t-PCA and n-EMC) and a decision-level fusion method which automatically explores the best fusion weights of features in Section 3.2.

3.1 Feature-level fusion

Before selecting features, we perform some preprocessings such as face normalization, noise deduction.

First, with sequences of thermal images, we find the regions of interest based on t-ROIs.
In our definition, interest regions are regions in which temperature increases or decreases significantly when human emotions change. We use the two regions which are the hottest and coldest regions of

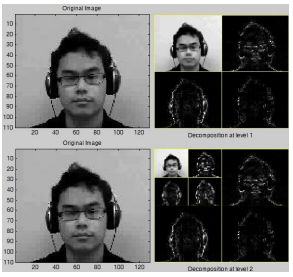

Fig. 2. Wavelet decomposition at level 1 and 2.

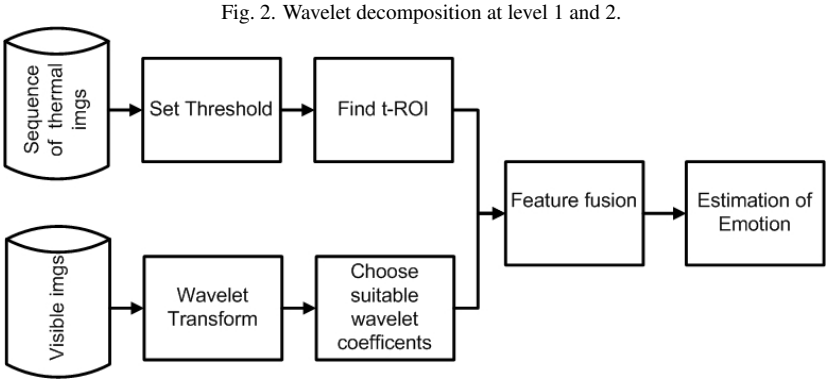

Fig. 3. Feature fusion of visible images and sequences of thermal images.

the face, except the eyeglasses, usually the forehead, eyeholes, and cheek-bone regions, as our interest regions. Before finding the t-ROIs, to avoid any ambient temperature change from frame to frame, we update the temperature of each point of each frame based on the difference between mean of ambient temperature and mean of the first m frame ambient temperature.
Let h be a map from face ($Fa \subset R^2$) to temperature ($T \subset R$) space

$$h : Fa \to T$$

$$(i, j) \mapsto h(i, j)$$

We obtain the t-ROIs by using the following equations:

$$\Delta T_{Fa} = T^{Fa}_{Max} - T^{Fa}_{Min}; \delta T_{Fa} = \Delta T_{Fa}/5$$

$$L^{Fa}_{k,idx} = \{(i, j) \in Fa | T^{Fa}_{Min} + \delta T_{Fa} * (idx - 1) \leq h(i, j) < T^{Fa}_{Max} - \delta T_{Fa} * (5 - idx)\} \quad (1)$$

where $T^{Fa}_{Max}, T^{Fa}_{Min}$ are maximum and minimum of temperature of each human face at frame k, respectively; $idx \in \{2, 5\}$.

After obtaining the t-ROIs for each frame, we find the dominant levels of frames. A frame a is more dominant than frame b if only if temperature change of all t-ROIs between frame a and frame a-1 is bigger than temperature change of all t-ROIs between frame b and frame b-1. Based on the obtained dominant level of each frame, we can automatically put the weight for each frame.

Second, with visible images, to eliminate of effects of noise and to omit unnecessary details, we use multiresolution analysis with Antonini filter bank. After two level of analysis with 7/9 Antonini filter bank, we keep coefficents of LL part wavelet transform [14]. Figure 2 shows wavelet decomposition at level 1 and 2 over our visible data.

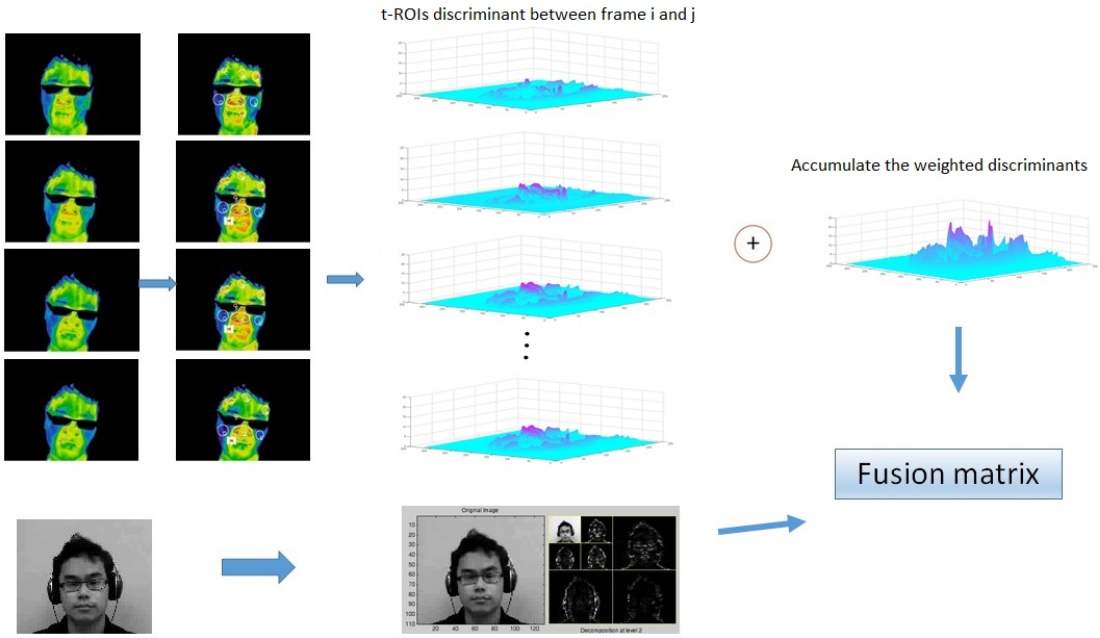

Fig. 4. A example procedure for fusion of visible images and sequences of thermal images.

To select the feature between visible feature and thermal feature, we perform feature-level fusion of visible and thermal image by using t-ROIs and PCA.

Step 1. Find t-ROIs over sequence of thermal images.

Step 2. Apply Wavelet transform over visible facial images and keep LL.

Step 3. Apply PCA over accumulate the weighted discriminant t-ROI.

Step 4. Build matrix from feature vectors obtained from step 2 and 3.

Step 5. Using PCA, EMC, PCA-EMC, t-PCA and n-EMC to classify emotions.

Figure 4 shows a example procedure for fusion of visible images and sequences of thermal images.

3.2 Decision-level fusion

To estimate human emotions, we use a decision fusion method of three conventional methods (PCA, EMC and PCA-EMC) and our proposed methods (thermal-Principal Component Analysis (t-PCA) and norm-Eigenspace method based on class feature (n-EMC)).

With PCA, the aim is to build a face space, including the basis vectors called principal components, which better describes the face images [15]. The difference between PCA and EMC is that PCA finds the eigenvector to maximize the total variance of the projection to line, while EMC is obtained eigenvector to maximize the difference between the within-class and between-class variance [16]. The Figure 5 shows the advantage of EMC over PCA.

Figure 6 shows the procedure of estimating human emotions using PCA. Figure 7 shows the procedure of estimating human emotions using EMC.

To analyze the human emotion using thermal infrared images, we propose two methods, thermal-Principal Component Analysis (t-PCA) and norm-Eigenspace method based on class feature (n-EMC).

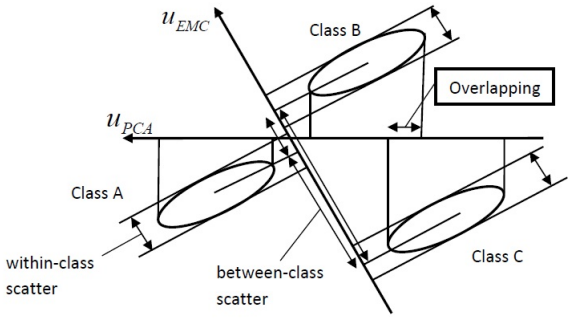

Fig. 5. Examples of a eigenvector of PCA and EMC [17].

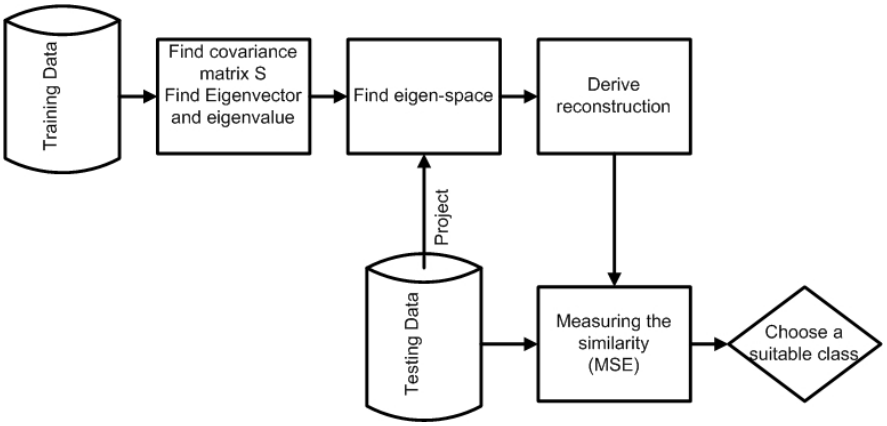

Fig. 6. Estimation of emotion using PCA.

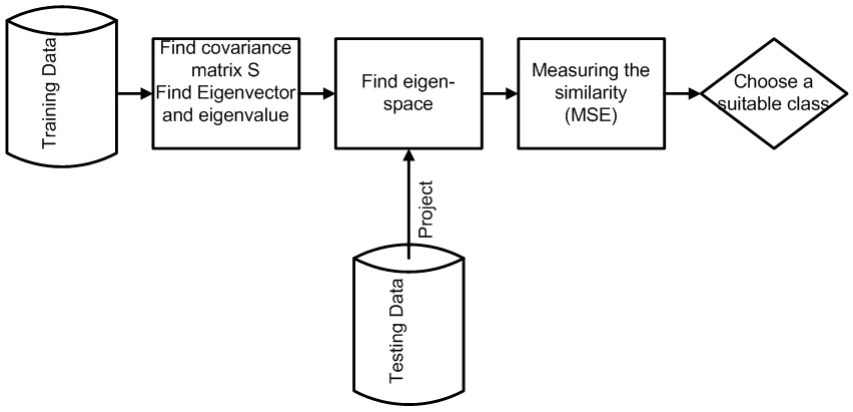

Fig. 7. Estimation of emotion using EMC.

To illustrate the feasibility of using eigen-space to fulfill facial emotion estimation task, thermal PCA (t-PCA) is modified from the PCA reconstruction method and evaluated over thermal data [15]. With PCA, the aim is to build a face space, including the basis vectors called principal components, which better describe the face images. PCA has several advantages over other face recognition schemes in its speed and simplicity [15]. We modified PCA to estimate facial emotion from thermal data.

Let F be a set of classes to be analyzed. Here, F is a set of all emotion classes. Assume that M_m^f thermal frames of training data are given as the facial temperature pattern for each class $f \in F$ where $F = \{anger, disgust, fear, happiness, neutral, sadness, surprise\}$. Let Γ_i^f be the $i-th$ facial temperature pattern where $i = \overline{1, M_f}$; the dimension of Γ_i^f, $n \times m$, is equal to the number of pixels in a thermal frame, and each element of Γ_i^f indicates the temperature of each pixel.

Compute the mean of training data $\Psi^f = \frac{1}{M^f}\sum_{i=1}^{M_f}\Gamma_i^f$ and let the normalized vector be $\Phi_i^f = \Gamma_i^f - \Psi^f$. We seek a set of M orthonormal vectors, u_k^f, which best describes the distribution of the training data. The kth vector, u_k^f is chosen by

$$\lambda_k^f = \frac{1}{M_f}\sum_{i=1}^{M_f}((u_k^f)^\tau \Phi_i^f)^2. \tag{2}$$

is a maximum, subject to $(u_l^f)^\tau u_k^f = \begin{cases} 1, & \text{if } l = k. \\ 0, & \text{if } otherwise. \end{cases}$

The eigenvectors and eigenvalues are the vector u_k^f and scalar λ_k^f of covariance matrix $C^f = \frac{1}{M_f}\sum_{i=1}^{M_f}(\Phi_i^f(\Phi_i^f)^\tau = A^f(A^f)^\tau$ where $A^f = [\Phi_1^f, \Phi_2^f, ..., \Phi_{M_f}^f]$

$(2) \Rightarrow \lambda_k^f = \frac{1}{M_f}\sum_{i=1}^{M_f}((u_k^f)^\tau\Phi_i^f)((u_k^f)^\tau\Phi_i^f)^\tau)$

$\Leftrightarrow \lambda_k^f = (u_k^f)^\tau \frac{1}{M_f}\sum_{i=1}^{M_f}(\Phi_i^f)((\Phi_i^f)^\tau((u_k^f)^\tau)^\tau)$

$\Leftrightarrow \lambda_k^f = (u_k^f)^\tau(\frac{1}{M_f}\sum_{i=1}^{M_f}(\Phi_i^f(\Phi_i^f)^\tau)u_k^f$

$\Leftrightarrow \lambda_k^f(u_k^f)^\tau = (u_k^f)^\tau(\frac{1}{M_f}\sum_{i=1}^{M_f}(\Phi_i^f(\Phi_i^f)^\tau)u_k^f(u_k^f)^\tau$

$\Leftrightarrow \lambda_k^f(u_k^f)^\tau = (u_k^f)^\tau(\frac{1}{M_f}\sum_{i=1}^{M_f}(\Phi_i^f(\Phi_i^f)^\tau)u_k^f(u_k^f)^\tau$

$\Leftrightarrow \lambda_k^f(u_k^f)^\tau = (u_k^f)^\tau C^f$

$\Leftrightarrow (\lambda_k^f(u_k^f)^\tau)^\tau = ((u_k^f)^\tau C^f)^\tau$

$\Leftrightarrow \lambda_k^f(u_k^f) = (C^f)^\tau((u_k^f)^\tau)^\tau \Leftrightarrow \lambda_k^f(u_k^f) = (C^f)^\tau u_k^f$

$\Leftrightarrow \lambda_k^f(u_k^f) = (A^f(A^f)^\tau)^\tau u_k^f \Leftrightarrow \lambda_k^f(u_k^f) = A^f(A^f)^\tau u_k^f$

$$\Leftrightarrow \lambda_k^f(u_k^f) = C^f u_k^f \tag{3}$$

The size of covariance matrix C^f, $nm \times nm$, is too large to determine $n \times m$ eigenvectors and eigenvalues. Turk et al. [18] suggested a computationally feasible method to find these eigenvectors. Let a matrix $H^f = (A^f)^\tau A^f$, then size of H^f is size $M \times M \ll nm \times nm$; let ν_i^f denote the eigenvectors of H^f.

We have $(A^f)^\tau A^f \nu_i^f = \mu_i^f \nu_i^f \Leftrightarrow A^f(A^f)^\tau A^f \nu_i^f = A^f \mu_i^f \nu_i^f \Leftrightarrow A^f(A^f)^\tau A^f \nu_i^f = A^f \mu_i^f \nu_i^f$

$$\Leftrightarrow C^f A^f \nu_i^f = \mu_i^f A^f \nu_i^f \tag{4}$$

From Equation(4), $A^f\nu_i^f$ is the eigenvector of $C^f = A^f(A^f)^\tau$ [18]. Therefore we can obtain ρ^f eigenvector of C^f by calculating the $\rho^f(\rho^f \ll nm)$ eigenvectors (ν_i^f) of matrix H^f and multiplying

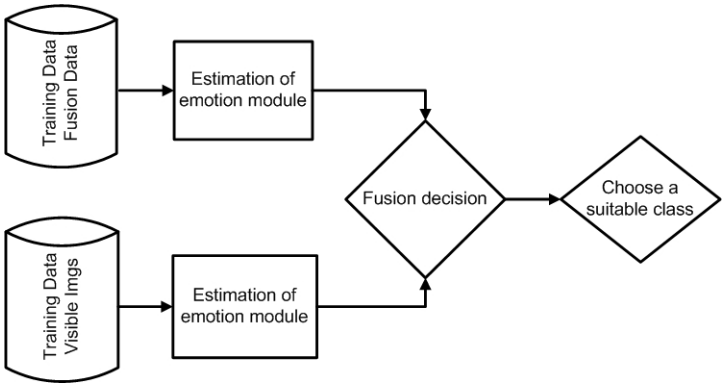

Fig. 8. Estimation of emotion using decision fusion.

A^f to ν_i^f. After obtaining eigenfaces from the training data of each emotion, we map facial thermal training data to feature spaces by $\omega_i^f(train) = (u_i^f)^\tau(\Gamma^f - \Psi^f), i = \overline{1, \rho^f}$.

We use the idea that if the input frame is much similar to some emotion training set, the reconstructed data will have less distortion than the data reconstructed from other eigenvectors of training emotions [15]. For each testing facial thermal frame Γ_{test}, firstly we project it onto the eigenfaces of each class.

$\omega_{test}^f = (U^f)^\tau(\Gamma_{test} - \Psi^f)$, where $U^f = (u_i^f), i = \overline{1, \rho^f}$.

Secondly, for each emotion, we find the feature which is most similar to the testing projected vector by calculate the angle between vector of training feature space and testing projected vector.

$$\beta^f = \underset{i}{\operatorname{argmax}} \frac{\omega_{test}^f \omega_i^f(train)}{\|\omega_{test}^f\| \|\omega_i^f(train)\|}; i = \overline{1, \rho^f}. \tag{5}$$

Thirdly, we find reconstruction of the testing data by the obtained feature in each class.

$\Gamma_{reconst}^f = U^f \omega_{\beta f}^f + \Psi^f$

Finally, we choose an emotion in which the reconstructed testing data and the original data are the most similar.

$$\gamma = \underset{f}{\operatorname{argmax}} \frac{\Gamma_{test} \Gamma_{reconst}^f}{\|\Gamma_{test}\| \|\Gamma_{reconst}^f\|}; f = \overline{1, 7} \tag{6}$$

The second valuation to estimate human emotion uses n-EMC over thermal data. n-EMC is modified from EMC [19]. The difference between EMC and n-EMC is formulation to calculate the difference between the within-class and between-class variance.

In mathematics, with n-EMC, instead of finding the eigenvectors, u_k^f and eigenvalues λ_k^f of covariance matrix $C^f = \frac{1}{M_f} \sum_{i=1}^{M_f} (\Phi_i^f (\Phi_i^f)^\tau = A^f (A^f)^\tau$ where $A^f = [\Phi_1^f, \Phi_2^f, ..., \Phi_{M_f}^f]$, we find eigenvectors, u_k^f and eigenvalues λ_k^f of matrix $S = \|S_B - S_W\|_2$ where

$$M = \sum_{f \in F} M_f \tag{7}$$

$$\Psi^f = \frac{1}{M^f} \sum_{i=1}^{M_f} \Gamma_i^f; \Psi = \frac{1}{M} \sum_{f \in F} \sum_{i=1}^{M_f} \Gamma_i^f \tag{8}$$

$$S_B = \frac{1}{M}\sum_{f\in F} M_f \|\Psi_f - \Psi\|_2 \|\Psi_f - \Psi\|_2^{\tau}. \tag{9}$$

$$S_W = \frac{1}{M}\sum_{f\in F}\sum_{i=1}^{M_f} \|\Gamma_i^f - \Psi_f\|_2 \|\Gamma_i^f - \Psi_f\|_2^{\tau}. \tag{10}$$

For each testing facial thermal frame Γ_{test}, firstly we project it onto the eigenfaces of each class. $\omega_{test}^f = (U^f)^{\tau}(\Gamma_{test} - \Psi^f)$, where $U^f = (u_i^f), i = \overline{1,\rho^f}$.
Secondly, for each emotion, we find the feature which is most similar to the testing projected vector by calculate the angle between vector of training feature space and testing projected vector.

$$\beta^f = \underset{i}{\operatorname{argmax}} \frac{\omega_{test}^f \omega_i^f(train)}{\|\omega_{test}^f\| \|\omega_i^f(train)\|}; i = \overline{1,\rho^f}. \tag{11}$$

Finally, we choose an emotion which has maximum of β^f

$$\gamma = \underset{f}{\operatorname{argmax}}\, \beta^f; f = \overline{1,7} \tag{12}$$

To estimate human emotions, we use decision fusion method of PCA, t-PCA, EMC and n-EMC. Figure 8 shows the general procedure to estimate human emotions using decision fusion.

When using decision fusion of PCA (t-PCA), we used the estimation of emotion module as described in figure 6. To determine the best class of emotions, after using PCA (t-PCA), the voting method with weights is used. The weights are set to fusion data and visible image, respectively. We determine the emotion class f of input image by choosing j satisfied minimum of following equation:

$$f = argmin\left(w_1 * MSE_j^{VI} + w_2 * MSE_j^{FU}\right) \tag{13}$$

where MSE_j^{VI} and MSE_j^{FU} are mean square errors calculated at class j of visible image and fusion data. We set $w_1 = \frac{4}{3}$ and $w_2 = \frac{2}{3}$ in experiment.

To estimate human emotion using decision fusion of EMC (n-EMC), we used the estimation of emotion module as described in figure 7. To determine the best class of emotions, after using EMC (n-EMC), the voting method with weights is used. The weights are set to fusion data and visible image, respectively.

We determine the emotion class f of input image by choosing j satisfied maximum of following equation:

$$k = max_i \frac{f_h^{VI} * F_i^{VI}}{\|f_h^{VI}\| * \|F_i^{VI}\|}, i = \overline{1,n} \tag{14}$$

$$g = max_i \frac{f_h^{FU} * F_i^{FU}}{\|f_h^{FU}\| * \|F_i^{FU}\|}, i = \overline{1,n} \tag{15}$$

$$f = argmax\left(w_1 * k + w_2 * g\right), \tag{16}$$

where n is a number of the training images of class j; f_h^{VI} and f_h^{FU} are testing image h of visible image and fusion data, respectively; F_i^{VI} and F_i^{FU} are vector i of eigenface of visible image and fusion data, respectively. We set $w_1 = \frac{2}{3}$ and $w_2 = \frac{4}{3}$ in experiment.

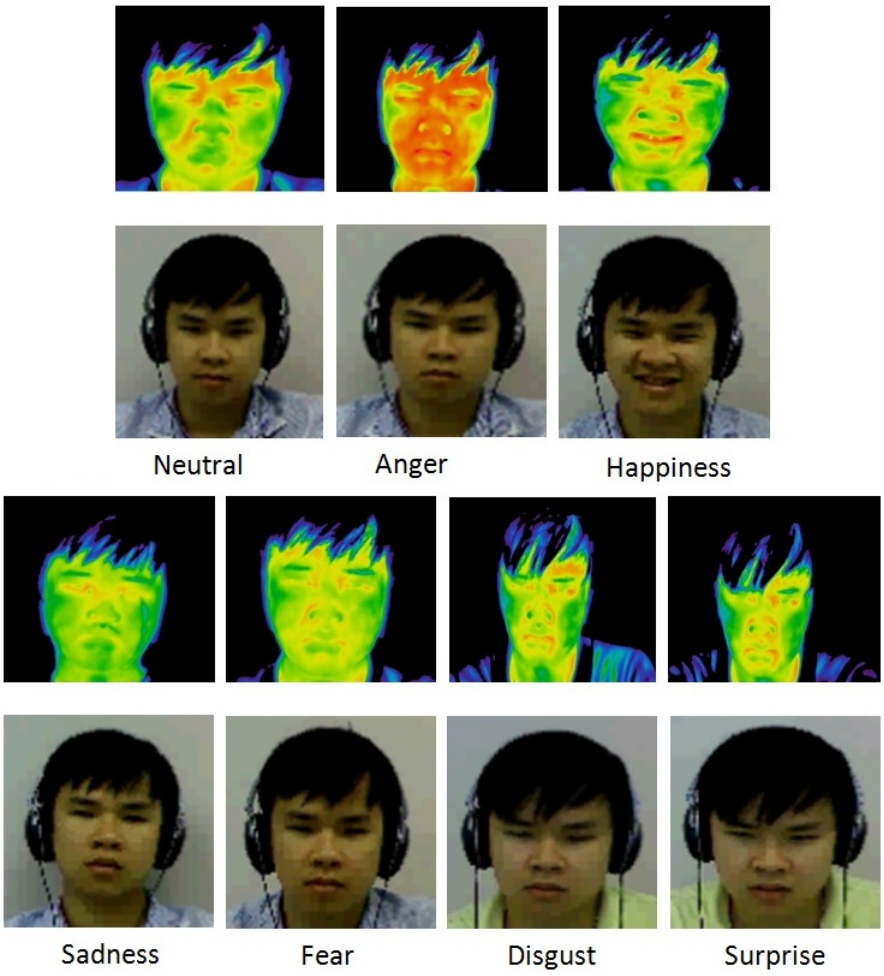

Fig. 9. Sample thermal and visible images of seven emotions.

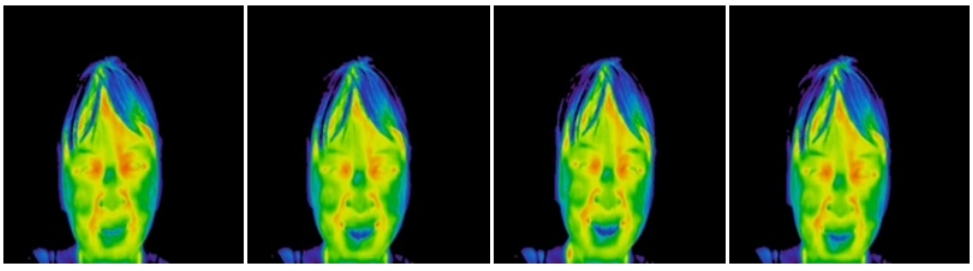

Fig. 10. Sample sequence of thermal images.

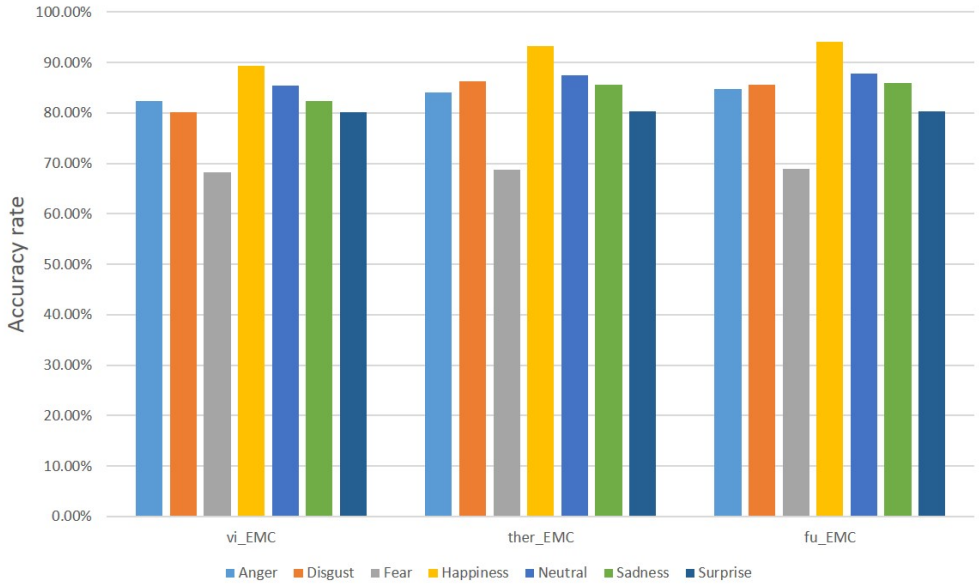

Fig. 11. Facial emotion estimation results using EMC.

4 Experiments and analysis

4.1 Database

The KTFE database [17] includes 152GB visible and thermal facial emotion videos, visible facial expression image database and thermal facial expression image database. This database contains 30 subjects who are Vietnamese, Japanese, Thai and Chinese from 11 year-old to 32 year-old with seven emotions. The example of visible and thermal images is shown in Figure9.

From draw data of KTFE database, we extract manually visible images and sequences of thermal images based on self-reports of participants, expressions and changing of facial temperatures. Causing the time-lag phenomenon, the sequence of thermal images are designed from a frame which we extracted the visible image to a frame which is after the participant emotion is neutral. Figure 10 shows a sample sequence of thermal images.

4.2 Experimental results

In our experiments, we separate the training and testing data as 60% and 40% of total visible images, sequence of thermal images, and fusion of visible image and thermal image sequence.

Figure 11 shows the results of facial emotion estimation of EMC with visible images (vi_EMC), sequence of thermal images (ther_EMC) and fusion of visible images and sequences of thermal images (fu_EMC). Following [20], accuracy of estimating human emotion using thermal images is lower than using visible images. Because emotions of thermal images are always not clearer than emotions of visible images. Therefore, with EMC methods, good for classification, the results using visible images are better than results using thermal images. However, with our new results, accuracy of estimating human emotion using sequences of thermal images is better than using visible images. When using fusion information, the average accuracy increases of 2.77% in compare with using only visible features, especially for happiness. In general, average accuracy of each emotion increases

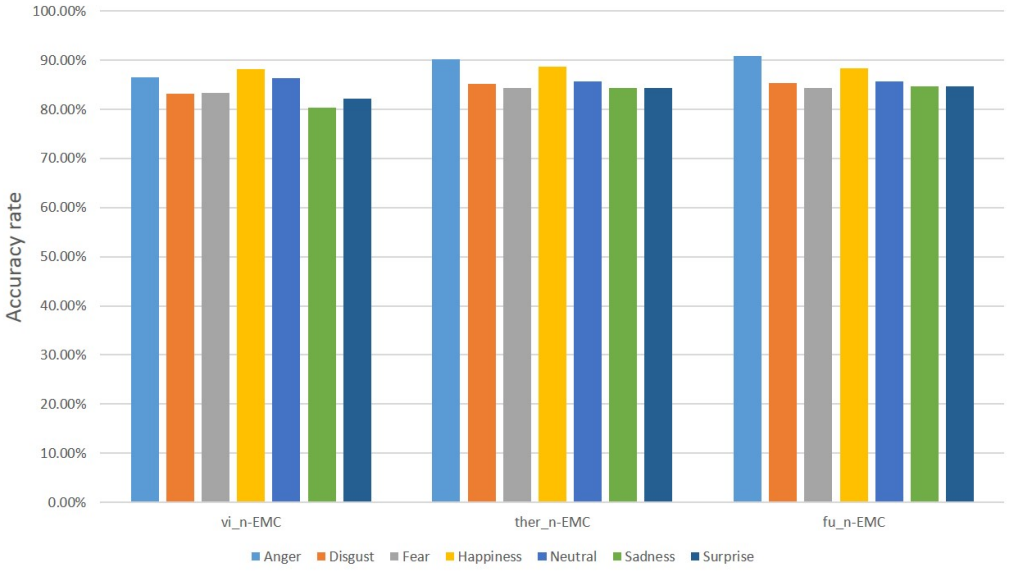

Fig. 12. Facial emotion estimation results using n-EMC.

when we use fusion information. The results prove the necessary of fusion information.

Figure 12 shows the results of facial emotion estimation of n-EMC with visible images (vi_n-EMC), sequence of thermal images (ther_n-EMC) and fusion of visible images and sequences of thermal images (fu_n-EMC). With n-EMC, the average accuracy using visible images, sequences of thermal images and fusion of visible images and sequences of thermal images increases 3.15%, 2.41%, 2.35%, respectively compared with EMC. Similar to the results using EMC, the accuracy using fusion data is better than using other data.

Fig.13 shows the results of facial emotion estimation of PCA with visible images (vi_PCA), sequences of thermal images (ther_PCA) and fusion of visible images and sequences of thermal images (fu_PCA). With PCA, accuracy using thermal images is better than accuracy using visible images. Although, emotions of thermal images are not clearer than emotion of visible image, PCA works better than EMC, which is good to classify each emotion. In general, with PCA, using fusion data gives the best results comparing using thermal images, visible images and sequence of thermal images.

Fig.14 shows the results of facial emotion estimation of t-PCA with visible images (vi_t-PCA), sequences of thermal images (ther_t-PCA) and fusion of visible images and sequences of thermal images (fu_t-PCA). With t-PCA, the average accuracy using visible images, sequences of thermal images and fusion of visible images and sequences of thermal images increases 2.03%, 1.31%, 0.48% respectively in compare with PCA. Especially, for disgust with visible images, accuracy improvement is 11.97%. Our method, t-PCA, yields an average improvement of 2.33% in performance of facial emotion estimation compared with using only visible features.

In conclusion, comparing the results of visible images, sequences of thermal images, and fusion data, the accuracy of estimating emotion using fusion data is better than the accuracy of estimation emotion using only visible images, thermal images and sequences of thermal images.

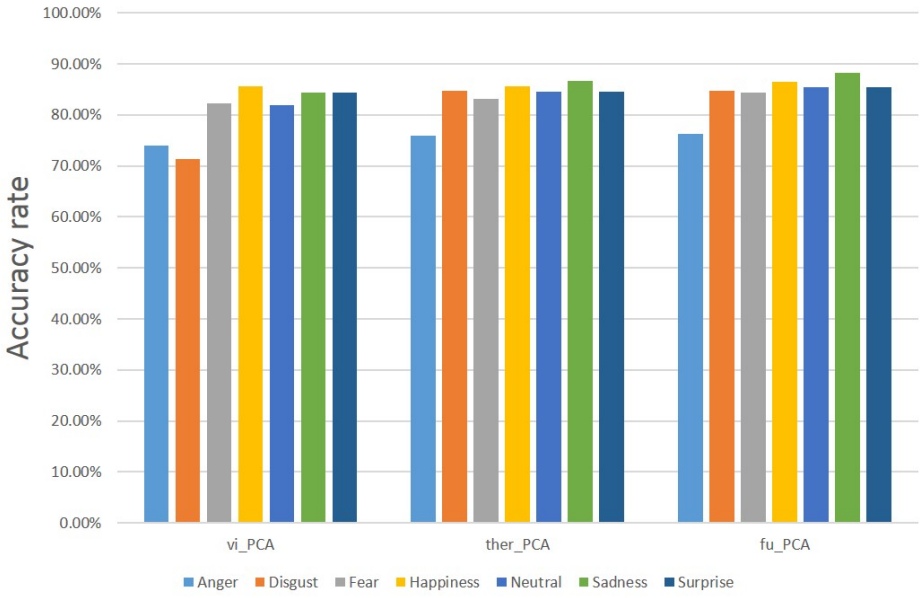

Fig. 13. Facial emotion estimation results using PCA.

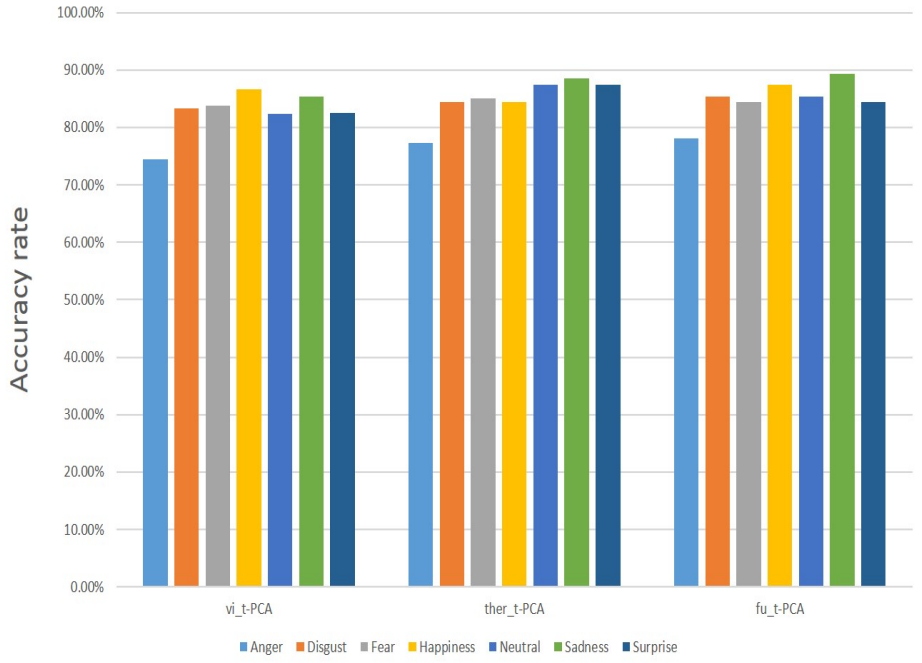

Fig. 14. Facial emotion estimation results using t-PCA.

5 Conclusions

In this paper, we have proposed the fusion of visible features and thermal features for estimating human emotions. Specially, two classification methods, t-PCA and n-EMC, are proposed to perform decision-level fusion. Our experiment on KTFE spontaneous database show that our methods yield and average improvement of 2.74% in performance of facial emotion estimation compared with using only visible features. Our method has several advantaged points. First, to the best of our knowledge, this is one of the first methods using sequences of thermal images. Emotion is complex action of human. To understand it clearly, using a single image cannot figure out the exact emotion. Besides, using thermal information with single frame cannot give the right emotion. Therefore, it is necessary to use sequences of thermal images. Second, with t-ROIs, we fill the gaps of thermal image, eyeglass problem. Third, using the weight discriminant features help our method decrease the running cost and increase accuracy rate. Because there are some frames more important than others, the weights set for frames are necessary. Fourth, using wavelet transform for visible image gives several advantages such as to reduce the unnecessary details, and so on. The fusion features, obtained from important visible features and necessary thermal feature, are better than only visible and thermal features. Two proposed classification methods, t-PCA and n-EMC, are successful to improve estimation accuracy. We also suggest decision fusion with weighted similarity measure for the conventional methods (PCA and EMC) and the our proposed methods (t-PCA, n-EMC) to increase the estimation accuracy. Experiments are tested in fusion database, specially designed from KTFE database. The results prove that the fusion of visible images and thermal image sequences performs better than either of the data. As a future work, we plant to improve the t-ROIs considerably, and also investigate more sophisticated fusion techniques for visible images and thermal image sequences and sequences of visible and thermal images.

References

1. Z.Zeng, M.Pantic, G.T.Roisman, and T.S.Huang (2009), *A survey of affect recognition methods: Audio, visual, and spontaneous expressions*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 31 (1), pp. 39-58.
2. B.Fasel and J.Luettin (2003), *Automatic facial expression analysis: a survey*, Pattern Recognition, Vol. 36(1), pp. 259-275.
3. M.Pantic, S.Member, L.J.M.Rothkrantz (2000) *Automatic analysis of facial expressions: The state of the art*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22, pp. 14241445.
4. S.Jarlier,D.Grandjean, S.Delplanque, K.N′Diaye, I.Cayeux, M.Velazco, D.Sander, P.Vuilleumier, and K.Schere (2011), *Thermal Analysis of Facial Muscles Contractions*, IEEE Transaction on Affective Computing, Vol. 2 (1), pp. 2-9.
5. M.M.Khan, R.D.Ward, and M.Ingleby (2009), *Classifying pretended and evoked facial expression of positive and negative affective states using infrared measurement of skin temperature*, ACM Transactions on Applied Perception, Vol. 6 (1), pp. 1-22.
6. L.Trujillo, G.Olague, R.Hammoud, and B.Hernandez (2005), *Automatic feature localization in thermal images for facial expression recognition*, IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops, pp. 14.
7. B.Hernández, G.Olague, R.Hammoud, L.Trujillo, and E.Romero (2007), *Visual learning of texture descriptors for facial expression recognition in thermal imagery*, Computer Vision and Image Understanding, Vol. 106, pp. 258 - 269.
8. B.R.Nhan, and T.Chau, T (2010), *Classifying affective states using thermal infrared imaging of the human face*, IEEE Transactions on Biomedical Engineering, Vol. 57, pp. 979-987.
9. Y.Yoshitomi, N.Miyawaki, S.Tomita, and S.Kimura (1997), *Facial expression recognition using thermal im-*

age processing and neural network, Robot and Human Communication, ROMAN '97 Proceedings , 6th IEEE International Workshop, pp. 380 - 385.
10. Y.Yoshitomi (2010), *Facial expression recognition for speaker using thermal image processing and speech recognition system*, Proceedings of the 10th WSEAS International Conference on Applied Computer Science, pp. 182-186.
11. Y.Koda, Y.Yoshitomi, M.Nakano, and M.Tabuse (2009),*A facial expression recognition for a speaker of a phoneme of vowel using thermal image processing and a speech recognition system*, The 18th IEEE International Symposium on Robot and Human Interactive Communication, ROMAN 2009, pp. 955-960.
12. S.Wang, S.He, Y.Wu, M.He, and Q.Ji (2014), *Fusion of visible and thermal images for facial expression recognition*, Frontiers of Computer Science, Vol 8(2), pp. 232-242.
13. Y.Yoshitomi, S.Kim, T.Kawano, and T.Kilazoe (2000), *Effect of sensor fusion for recognition of emotional states using voice, face image and thermal image of face*, Proceedings of the 9th IEEE International Workshop on Robot and Human Interactive Communication, pp.178 -183.
14. M.Antonini, M.Barlaud, P.Mathieu, and I.Daubechies (1992), *Image coding using wavelet transform*, IEEE Transactions on Image Processing, Vol. 1, pp.205 -220.
15. D.T.LIN (2006), *Facial expression classification using PCA and hierarchical radial basis function network*, Journal of Information Science and Engineering, Vol. 22, pp. 1033-1046.
16. T.Yabui, Y.Kenmochi, and K.Kotani, *Facial expression analysis from 3D range images; comparison with the analysis from 2D images and their integration*, 2003 International Conference on Image Processing, Vol.2, pp.879-882.
17. H.Nguyen, K.Kotani, F.Chen, and B.Le (2014), *A thermal facial emotion database and its analysis*, Image and Video Technology, Lecture Notes in Computer Science, Vol. 8333, pp. 397-408.
18. M. Turk and A. Pentland (1991), *Eigenfaces for recognition*, Journal of Cognitive Neuroscience, Vol.3, pp.71-86.
19. T.Kurozumi, Y.Shinza, Y.Kenmochi, and K.Kotani (1999), *Facial individuality and expression analysis by eigenspace method based on class features or multiple discriminant analysis*, 1999 International Conference on Image Processing, Vol.1, pp. 648-652.
20. H.Nguyen, F.Chen, K.Kotani, and B.Le (2014), *Human emotion estimation using wavelet transform and t-ROIs for fusion of visible images and thermal image sequences*, Computational Science and Its Applications ICCSA 2014, Lecture Notes in Computer Science, Vol.8584, pp. 224-235.