Human Emotion Estimation Using Wavelet Transform and t-ROIs for Fusion of Visible Images and Thermal Image Sequences

Conference Paper · July 2014

DOI: 10.1007/978-3-319-09153-2_17

4 authors:

Hung Nguyen
Japan Advanced Institute of Science and Technology
8 PUBLICATIONS **21** CITATIONS

SEE PROFILE

Fan Chen
Japan Advanced Institute of Science and Technology
37 PUBLICATIONS **247** CITATIONS

SEE PROFILE

Kazunori Kotani
Japan Advanced Institute of Science and Technology
69 PUBLICATIONS **779** CITATIONS

SEE PROFILE

Bac Le
Ho Chi Minh City University of Science
107 PUBLICATIONS **815** CITATIONS

SEE PROFILE

Human Emotion Estimation using Wavelet Transform and t-ROIs for Fusion of Visible Images and Thermal Image Sequences

Hung Nguyen[1], Fan Chen[1], Kazunori Kotani[1], and Bac Le[2]

[1] Japan Advanced Institute of Science and Technology,
1-1 Asahidai, Nomi, Ishikawa, Japan
{nvhung,chen-fan,ikko}@jaist.ac.jp
[2] University of Science, Ho Chi Minh city,
227 Nguyen Van Cu, Ho Chi Minh city, Vietnam
lhbac@hcmuns.edu.vn

Abstract. Most studies in human emotion estimation focus on visible image-based analysis which is sensitive to illumination changes. Under uncontrolled operating conditions, estimation accuracy degrades significantly. In this paper, we integrate both visible images and thermal image sequences. 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 Principal Component Analysis (PCA), Eigen-space Method based on class-features (EMC), PCA-EMC is applied. Experiments on the Kotani Thermal Facial Emotion (KTFE) database show the effectiveness of proposed methods.

Keywords: Human emotions, thermal images, emotion estimation, feature fusion, decision fusion, thermal image sequences, KTFE database.

1 Introduction

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] 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 sequence of thermal images. To estimate five emotions, we use the fusion of conventional methods Principal Component Analysis (PCA), Eigenspace Method based on class-features (EMC), and PCA-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. Sophie Jarlier et al. [2] 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%. M.M.Khan et al. [3] 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%. L.Trujillo et al. [4] 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.Hernandez et al. [5] 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. B.R.Nhan et al. [6] 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. Y.Yoshitomi et al. [7] 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. Y. 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 [8]. Y.Koda et al. used the idea from [8] 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 [9]. 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. [10] 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. [11] 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 human emotion estimation by fusing visible images and sequence of thermal imagery at decision-level and feature-level respectively.

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 and a decision-level fusion which explores the best fusion weights of features in Section 3.2.

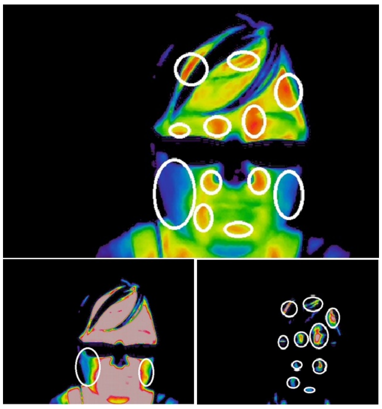

Fig. 1. An example of t-ROIs.

3.1 Feature-level fusion

Before selecting features, we perform some preprocessings. 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 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 f be a map from face ($F \subset R^2$) to temperature ($T \subset R$) space

$$f : F \to T$$

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

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

$$\Delta T_F = T^F_{Max} - T^F_{Min}; \delta T_F = \Delta T_F/5$$

$$L^F_{k,idx} = \{(i, j) \in F | T^F_{Min} + \delta T_F * (idx - 1) \leq f(i, j) < T^F_{Max} - \delta T_F * (5 - idx)\} \quad (1)$$

where T^F_{Max}, T^F_{Min} are maximum and minimum of temperature of each human face at frame k, respectively; $idx \in \{2, 5\}$.
Second, with visible images, to eliminate of effects of non-uniform illumination and to omit unnecessary details, we use wavelet transform with Antonini filter bank [12].

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.

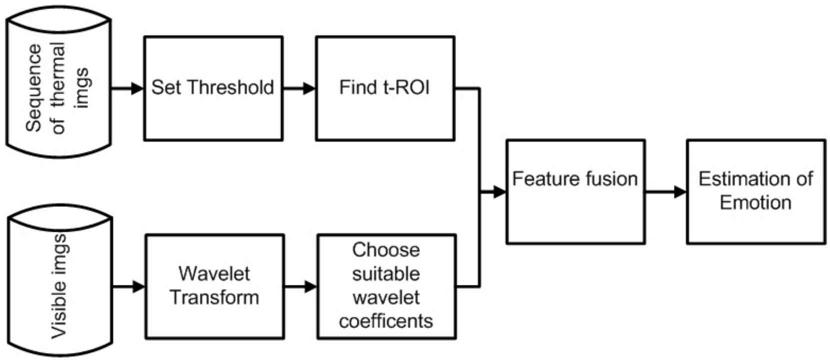

Fig. 2. Feature fusion of visible and sequence thermal image.

Step 2. Apply Wavelet transform over visible facial images and keep LL.
Step 3. Apply PCA over each t-ROI.
Step 4. Build matrix from feature vectors obtained from step 2 and 3.
Step 5. Using PCA, EMC to classify emotions

3.2 Decision-level fusion

To estimate human emotions, we use decision fusion method of PCA, EMC and PCA-EMC.
With PCA, the aim is to build a face space, including the basis vectors called principal components, which better describes the face images [13]. The difference between PCA and EMC is that PCA finds the eigenvector to maximize the total variance of the projection to line, while EMC [14] obtains eigenvectors to maximize the difference between the within-class and between-class variance.

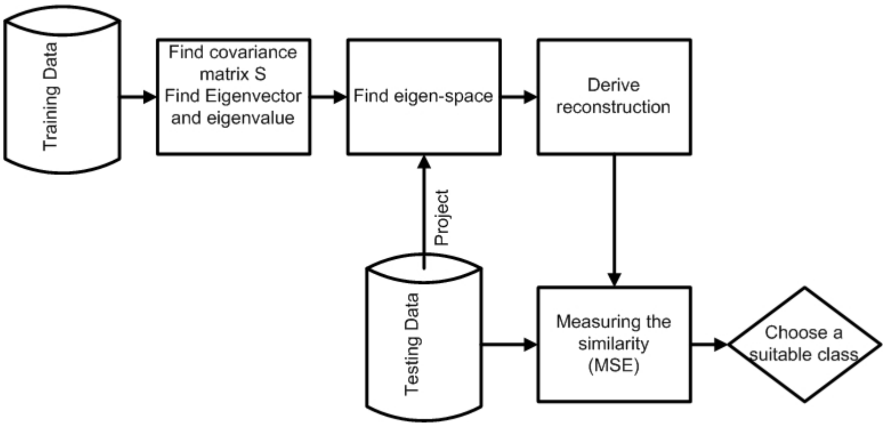

Fig. 3. Estimation of emotion using PCA.

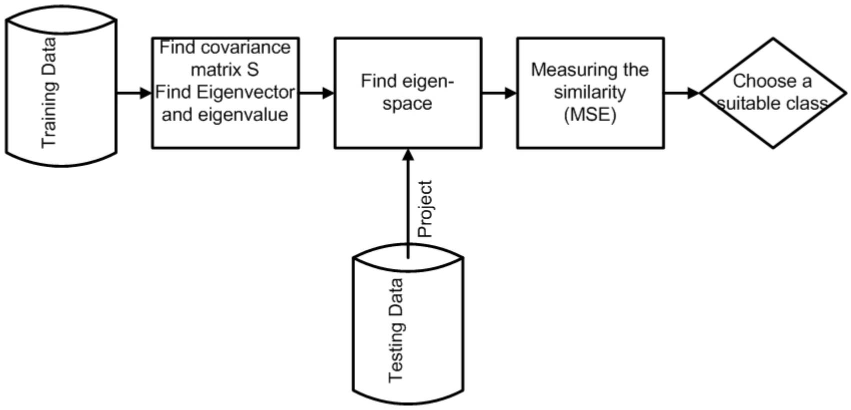

Fig. 4. Estimation of emotion using EMC.

The difference between the within-class and between-class variance is calculated as following:

$$S = S_B - S_W. \tag{2}$$

$$S_B = \frac{1}{M}\sum_{f \in F} M_f(\overline{x}_f - \overline{x})(\overline{x}_f - \overline{x})^{\tau}. \tag{3}$$

$$S_W = \frac{1}{M}\sum_{f \in F}\sum_{f \in F}^{M_f} M_f(\overline{x}_{fm} - \overline{x}_f)(\overline{x}_{fm} - \overline{x}_f)^{\tau}. \tag{4}$$

$$\overline{x}_f = \frac{1}{M}\sum_{m=1}^{M_f} x_{fm}; \overline{x} = \frac{1}{M}\sum_{f \in F}\sum_{m=1}^{M_f} x_{fm}. \tag{5}$$

where F is a set of expression classes, M_f facial-patterns are given for each class $f \in F$ and x_{fm} is an N-dimension vector of the $m-th$ facial patterns,$m = \overline{1, M_f}$

Figure 3 shows the procedure of estimating human emotions using PCA. Figure 4 shows the procedure of estimating human emotions using EMC.
To estimate human emotions using PCA-EMC, first, we use PCA to reduce the dimension and apply EMC to the obtained eigentspace.

Figure 5 shows the general procedure to estimate human emotions using decision fusion. When using decision fusion of PCA, we used the estimation of emotion module as described in figure 3.
To determine the best class of emotions, after using PCA, the voting method with weights is used. The weights, 2/3 and 4/3 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:

$$f = argmin\left(w_1 * MSE_j^{VI} + w_2 * MSE_j^{FU}\right) \tag{6}$$

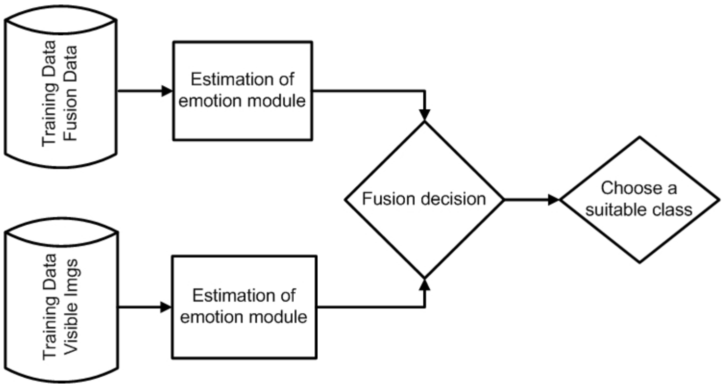

Fig. 5. Estimation of emotion using decision fusion.

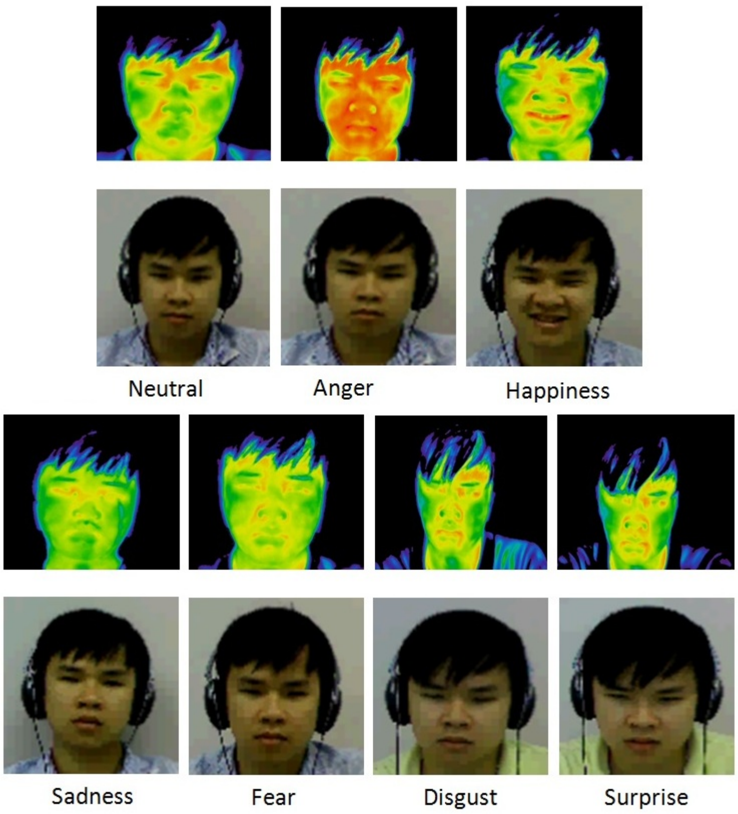

Fig. 6. Sample thermal and visible images of seven emotions.

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

To estimate human emotion using decision fusion of EMC, we used the estimation of emotion module as described in figure 4. Figure 5 shows the procedure to estimate human emotions using decision fusion of EMC.
To determine the best class of emotions, after using EMC, the voting method with weights is used. The weights, 4/3 and 2/3 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:

$$k = max_i \frac{f_h^{VI} * F_i^{VI}}{\|f_h^{VI}\| * \|F_i^{VI}\|}, i = \overline{1, n} \tag{7}$$

$$g = max_i \frac{f_h^{FU} * F_i^{FU}}{\|f_h^{FU}\| * \|F_i^{FU}\|}, i = \overline{1, n} \tag{8}$$

$$f = argmax\left(w_1 * k + w_2 * g\right), \tag{9}$$

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; $w_1 = \frac{2}{3}$ and $w_2 = \frac{4}{3}$

4 Database

The KTFE database [15] includes 131GB visible and thermal facial emotion videos, visible facial expression image database and thermal facial expression image data-base. This database contains 26 subjects who are Vietnamese, Japanese, Thai from 11 year-old to 32 year-old with seven emotions. The example of visible and thermal images is shown in Fig.6.

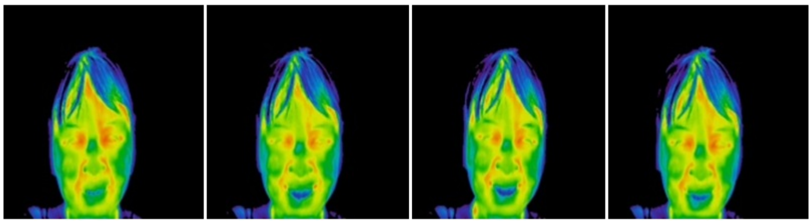

Fig. 7. Sample sequence of thermal images.

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.

5 Experimental results

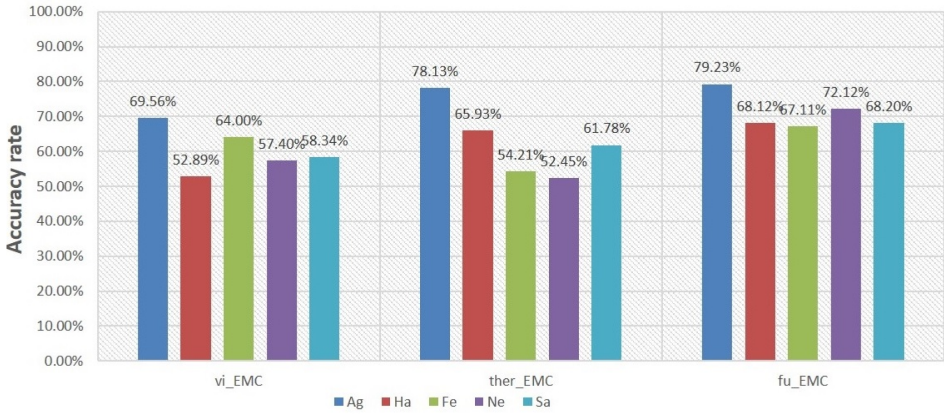

Fig. 8. Human emotion estimation results using EMC.

In our experiments, we separate the training and testing data as 70% and 30% of total visible images, thermal images, and fusion of visible and thermal image sequence. Fig.8 shows the results of emotion estimation of EMC with visible images (vi_EMC), thermal images (ther_EMC) and fusion of visible images and sequence of thermal images (fu_EMC). Accuracy of estimating human emotion using thermal images is lower than using visible images. 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. In general, average accuracy of each emotion increases when we use fusion information. The results prove the necessary of fusion information.

Fig.9 shows the results of emotion estimation of PCA with visible images (vi_PCA), thermal images (ther_PCA) and fusion of visible images and sequence 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 and visible images.

Fig.10 shows the results of emotion estimation of PCA-EMC with visible images (vi_PCA-EMC), thermal images (ther_PCA-EMC) and fusion of visible images and sequence of thermal images (fu_PCA-EMC). With PCA-EMC, the information of each data has been reduced. Therefore, the almost results of estimation using PCA-EMC are lower than using PCA or EMC. However, similar to the results using EMC and PCA, the accuracy using fusion data is better than using other data.

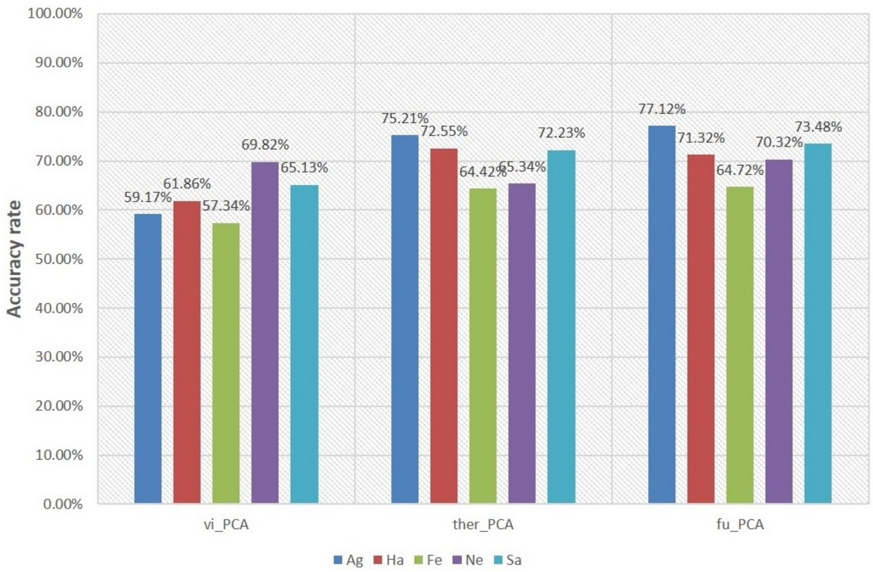

Fig. 9. Human emotion estimation results using PCA.

In conclusion, comparing the results of visible images, thermal images, and fusion data, the accuracies of estimating emotion using fusion data are better than the accuracies of estimation emotion using visible images and thermal images.

6 Conclusions

In this paper, we have proposed the fusion of visible features and thermal features for estimating human emotions. Our method has several advantaged points. First, to the best of our knowledge, this is one of the first methods using sequence of thermal images. Emotion is complex action of human. To understand it clearly, using a single image can not figure out the exact emotion. Besides, using thermal information with single frame can not give the right emotion. Therefore, it is necessary to use sequence of thermal images. Second, with t-ROIs, we fill the gaps of thermal image, eyeglass problem. Third, using wavelet transform for visible image gives several advantages such as to reduce the unnecessary coarse, so on. The fusion features, obtained from important visible features and necessary thermal feature, are better than only visible and thermal features. We also suggest decision fusion with weighted similarity measure for the conventional method PCA, EMC and PCA-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.

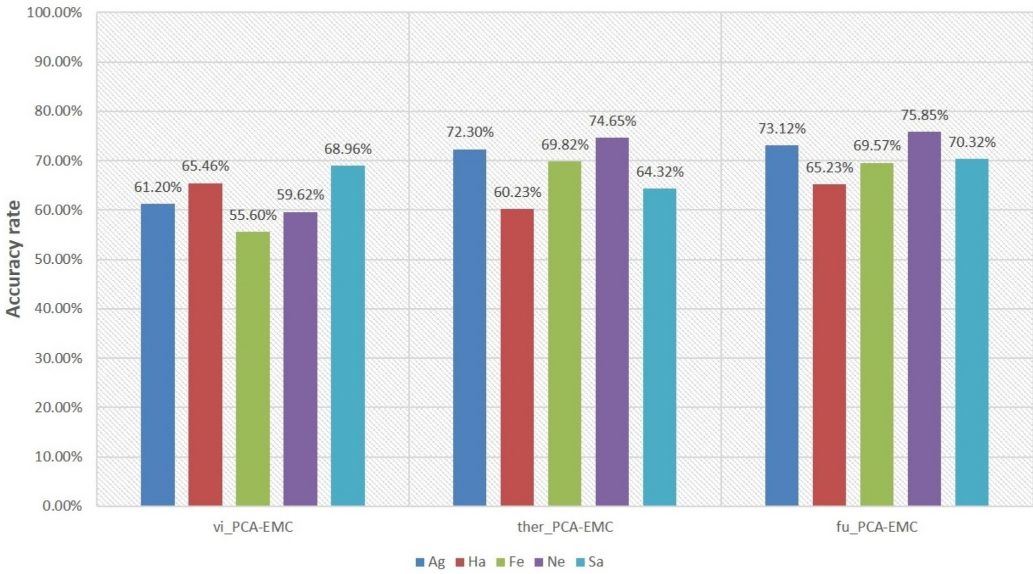

Fig. 10. Human emotion estimation results using PCA-EMC.

References

1. Zeng, Z., Pantic, M., Roisman, G.T., Huang, T. S.: A survey of affect recognition methods: Audio, visual, and spontaneous expressions, IEEE Trans. Pattern Anal. Mach. Intell, vol. 31, no. 1, pp. 39-58, (2009)
2. Jarlier,S., Grandjean,D., Delplanque, S., NDiaye, K., Cayeux, I., Velazco, M., Sander, D., Vuilleumier, P., Schere, K.: Automatic facial expression analysis: a survey, Pattern Recognition, vol. 36, pp. 259-275 (2003)
3. Khan, M.M., Ward, R.D., Ingleby, M.: Classifying pretended and evoked facial expression of positive and negative affective states using infrared measurement of skin temperature, Trans. Appl. Percept, vol. 6, no. 1, pp. 1-22 (2009)
4. Trujillo, L., Olague, G., Hammoud, R., Hernandez, B.:Automatic feature localization in thermal images for facial expression recognition, IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops, CVPR Workshops, pp. 14 (2005)
5. Hernández, B., Olague, G., Hammoud, R., Trujillo, L., Romero, E. : Visual learning of texture descriptors for facial expression recognition in thermal imagery, Computer Vision and Image Understanding, vol. 106, pp. 258 - 269 (2007)
6. Nhan, B.R., Chau, T.: Classifying affective states using thermal infrared imaging of the human face, IEEE Transactions on Biomedical Engineering, vol. 57, pp. 979-987, (2010)
7. Yoshitomi, Y., Miyawaki, N., Tomita, S., Kimura, S. : Facial expression recognition using thermal image processing and neural network, Robot and Human Communication, ROMAN '97 Proceedings , 6th IEEE International Workshop, pp. 380 - 385, (1997)

8. Yoshitomi, Y.: 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 (2010)
9. Koda, Y., Yoshitomi, Y., Nakano, M., Tabuse, M.: 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 (2009)
10. Wang, S., He, S., Wu, Y., He, M., Ji, Q.: Fusion of visible and thermal images for facial expression recognition, J Frontiers of Computer Science (2014)
11. Yoshitomi, Y., Kim, S., Kawano, T., Kilazoe, T.: 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 (2000)
12. Antonini, M., Barlaud, M., Mathieu, P., Daubechies, I.: Image coding using wavelet transform, IEEE Trans. Image Processing, vol. 1, pp.205 -220 (1992)
13. LIN, D.T.: Facial Expression Classification Using PCA and Hierarchical Radial Basis Function Network, Journal of Information Science and Engineering, vol. 22, pp. 1033-1046 (2006)
14. Kurozumi, T., Shinza, Y., Kenmochi, Y., Kotani, K.: Facial Individuality and Expression Analysis by Eigenspace Method Based on Class Features or Multiple Discriminant Analysis, ICIP (1999)
15. Nguyen, H., Kotani, K., Chen, F., Le, B.: A thermal facial emotion database and its analysis. In: Klette, R., Rivera, M., Satoh, S. (eds.) Image and Video Technology. LNCS, vol. 8333, pp. 397-408 . Springer, Heidelberg (2014)