Facial Expression Recognition Using Game Theory

Kaushik Roy and Mohamed S. Kamel

Centre for Pattern Analysis and Machine Intelligence
University of Waterloo, ON, Canada
kaushik.roy@uwaterloo.ca

Abstract. Accurate detection of lip contour is important in many application areas, including biometric authentication, human computer interaction, and facial expression recognition. In this paper, we propose a new lip boundary localization scheme based on Game Theory (GT) to improve the facial expression detection performance. In addition, we use GT for selecting the proper set of facial features. We apply the Extended Contribution-Selection Algorithm (ECSA) for the dimensionality reduction of the facial features using a coalitional GT-based framework. We have conducted several sets of experiments to evaluate the proposed approach. The results show that the proposed approach has achieved recognition rates of 93.1% and 92.7% on the JAFFE and CK+ datasets, respectively.

Keywords: Facial expression recognition, coalitional game theory, extended contribution selection algorithm.

1 Introduction

Automatic lip boundary detection plays an important role in numerous application areas, including biometric authentication, human computer interaction, and facial expression recognition (FER) [1]. However, extracting the lip boundary from the mouth region is quite complicated and difficult due to large variations emerged from different speakers, illumination conditions, poor texture of lips, weak contrast between lip and skin, high deformability of lip, wrinkles, and occlusions by beard and moustache. Various methodologies for lip contour detection have been published during the last few years [2]. The state-of-the-art lip contour detection methods can be divided into three broad categories: 1) model-based, 2) color/gray level analysis-based, and 3) level set-based approaches. The model based approaches may not provide satisfactory performance due to poor contrast between lip and skin color. The color/gray level analysis-based approaches are computationally efficient. However, these methods result in large color noise and are sensitive to color contrast. The basic idea behind the level set-based approaches is to represent the lip contours as the zero level set of an implicit function defined in a higher dimension, usually referred as the level set function, and to evolve the level set function according to a partial differential equation (PDE). The level set-based methods mainly depends on the image gradients and thus, are highly sensitive to the presence of noise and poor image

N. Mana, F. Schwenker, and E. Trentin (Eds.): ANNPR 2012, LNAI 7477, pp. 139–150, 2012.

contrast. In this paper, we propose a lip localization approach based on Game Theory (GT) to track the lip boundary from the mouth region. We propose to apply a parallel game-theoretic decision making procedure by modifying Chakraborty and Duncan's algorithm [3], which combines (1) the region-based and gradient-based boundary finding methods, and (2) integrates the complementary strengths of each of these individual methods. This GT-based scheme is robust to noise and poor localization and performs well against weak lip/skin boundaries. Previous work on FER has focused mainly on the issues of feature extraction and facial pattern classification. However, less effort has been given to the critical issue of feature selection. In this paper, we propose a two-stage feature selection scheme, which focuses on feature ranking and redundancy reduction. First, we apply Information Gain (IG) to rank the best textural features [4]. Though IG has been regarded as one of the most efficient measures of feature ranking in classification problem, it has a serious weakness of ignoring the redundancy among higher ranked features. Therefore, finally, we propose a feature reduction scheme in the context of GT [5]. The GT-based feature selection scheme is generally quite effective in a rapid global search of large, non-linear and poorly understood spaces. It also suggests a particularly attractive approach in solving the problem occurred due to highly dimensionality of facial features with small sample size. The game-theoretic scheme takes into account the correlation of features while reducing the dimensionality. An iterative algorithm for feature ranking, called the contribution-selection algorithm [5], is enhanced, called Enhanced Contribution-Selection Algorithm (ECSA) hereafter, and used to select the optimal feature subset. This algorithm depends on the Multi-perturbation Shapley analysis (MSA), a framework that is based on the coalitional GT, to estimate the usefulness of features and rank them accordingly.

2 Lip Localization

In this section, we mainly focus on GT-based lip boundary detection approach. First, we apply Viola-Jones method [6] to detect the presence of a human face. An Active Shape Model with Local Features (ASMLF) [7] is applied to find a set of 60 facial feature points from the detected face image. The selected set of feature points is used to delineate the regions of interest. We apply the H-minima transform on the complemented input image to reduce all minima in the detected mouth image whose depth is less than a threshold. To enhance the quality of the image and to suppress the effect of noise, a 2D Wiener filter is deployed. Watershed transformation is used to divide the filtered image into several catchment basins, which consist of its own regional minimum. The pixels of final outcome are labelled according to a specific catchment basin number. [2]. This process is shown in Figure 1. Finally, we apply a parallel game-theoretic decision making procedure by modifying the Chakraborty and Duncan's algorithm [3], which combines the region-based segmentation and the boundary finding methods for the optimal estimation of lip boundary. The game is usually played out by a set of decision makers (or "players"), which in our case, corresponds to the two segmentation schemes, namely, the region-based and the

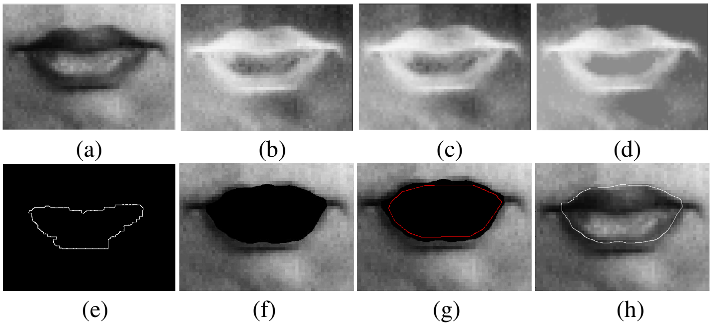


Fig. 1. Lip contour detection using GT. (a) Delineated mouth region from an original face image. (b) Complemented image. (c) Image after H-minima transform. (d) Enhancement using 2D Weiner filter. (e) Image after watershed segmentation. (f) Segmented area filled-up with black pixels. (g) Contour initialization. (h) Final contour.

gradient-based segmentation methods [3]. The lip segmentation problem can be formulated as a two-player game. If p^1 represents the set of strategies of the Player 1, and p^2 denotes the set of strategies of the Player 2, then each player tries to minimize the payoff function, $F^i(p^1, p^2)$. The main objective is to find the *Nash Equilibrium* (NE) of the system $(\bar{p}^1, \bar{p}^2)$, such that:

$$F^1(\bar{p}^1, \bar{p}^2) \leq F^1(p^1, \bar{p}^2), F^2(\bar{p}^1, \bar{p}^2) \leq F^2(\bar{p}^1, p^2) \quad (1)$$

We move toward the NE iteratively by taking t as the time index and can formulate the game as follows:

$$p^1_{t+1} = \underset{p^1 \in P^1}{\arg min} F^1(p^1, p^2_t);\ p^2_{t+1} = \underset{p^2 \in P^2}{\operatorname{argmin}} F^2(p^1_t, p^2) \quad (2)$$

In [3], Chakraborty and Duncan proved that there is always an existing NE solution if F^1 and F^2 are of the following form [3]:

$$F^1(p^1, p^2) = f_1(p^1) + \alpha f_{21}(p^1, p^2) \quad (3)$$

$$F^2(p^1, p^2) = f_2(p^2) + \beta f_{12}(p^1, p^2) \quad (4)$$

where α and β are scaling constants, F^i is bounded in $p^i \in P^i$, F^i is continuously second-order differentiable in $p^i \in P^i$, and there is an existing closed neighborhood of $u^i \subseteq p^i$ such that F^i is strongly convex in u^i. In the region-based method, the image is partitioned into connected regions by grouping the neighbouring pixels of similar intensity levels. The adjacent regions are then merged under some criteria involving the homogeneity or sharpness of the region boundaries. Now, if $y_{i,j}$ is the intensity of a pixel at (i, j) of the original image and $x_{i,j}$ is the intensity of a pixel at (i, j) of the segmented image, then, a common approach is to minimize an objective function of the form [3]:

$$E = \sum_{i,j}(y_{i,j} - x_{i,j})^2 + \lambda^2(\sum_{i,j}\sum_{i_s,j_s}(x_{i,j} - x_{i_s,j_s})^2) \quad (5)$$

where, i_s and j_s are indices in the neighborhood of pixel $x_{i,j,}$ and λ is a constant. In the above equation, the first term on the right-hand side represents the data fidelity term, and the second term on the right-hand side enforces the smoothness. To find the lip contour, the objective functions are described as follows:

For the region-based module (Player 1):

$$F^1(p^1, p^2) = \min_x \left[\sum_{i,j}(y_{i,j} - x_{i,j})^2 + \lambda^2 \left(\sum_{i,j}(x_{i,j} - x_{i-1,j})^2 + \sum_{i,j}(x_{i,j} - x_{i,j-1})^2\right)\right] + \alpha\left[\sum_{i,j\in A_{\vec{p}}}(x_{i,j} - u)^2 + \sum_{i,j\in \bar{A}_{\vec{p}}}(x_{i,j} - v)^2\right] \quad (6)$$

where, $y_{i,j}$ is the intensity of the original image, $x_{i,j}$ is the intensity of the segmented image given by p^1 as mentioned earlier, u is the intensity inside the contour given by p^2, and v is the intensity outside the contour given by p^2. $A_{\vec{p}}$ corresponds to the points that lie inside the contour, and $\bar{A}_{\vec{p}}$ represents those points that lie outside the contour. The first term on the right-hand side of (6) minimizes the difference between the pixel intensity values and the obtained region, as well as enforces continuity. The second term tries to match the region and the contour. In the region growing approach, we select an initial area within the region of interest for the lip boundary detection. At each iteration, the neighbouring pixels are observed and the value of E is measured from (5). The pixels, for which the value of E is less than a predefined threshold, are accepted into the region. The objective function of the Player 2 (i.e., the boundary finding module) is as follows:

Fig. 2. Lip contour detection using GT on CK+ database

$$F^2(p^1, p^2) = \operatorname*{argmax}_{\vec{p}} \left[M_{gradient}(I_g, \vec{p}) + \beta M_{region}(I_r, \vec{p})\right] \quad (7)$$

where, $\vec{p}$ denotes the parameterization of the contour given by p^2, I_g is the gradient image, I_r is the region segmented image, and β is a constant. In this paper, we apply a Variational Level Set (VLS)-based active contour model to parameterize and represent the lip contour data during the game-theoretic propagation [8]. Figure 2 shows the lip segmentation results. For feature extraction, we apply a discrete set of 1D log-Gabor kernels, which contains 4 spatial frequencies and 6 different orientations from 0° to 180°, differing in 30 steps that makes a filter bank of 24 different Gabor filters. These log-Gabor filters are deployed to each of the images and filter responses are obtained only at the selected fiducial points [9]. Therefore, the facial expressions in an input image are represented by a feature vector of length 1440 elements (60 fiducial points, 24 filter responses per point).

3 Feature Ranking and Reduction

We propose a two-stage feature selection strategy, including featuring ranking and removing the redundant terms. The IG has been regarded as one of the most efficient measures of feature ranking in classification problem [4]. Therefore, in the first stage, we apply a feature ranking measure based on IG to select the most informative and relevant features. The IG can be calculated as [4]:

$$InG(x_j) = -\sum_{k=1}^{C} P(y_k).logP(y_k) + P(x_j)\sum_{k=1}^{C} P(y_k|x_j).\log P(y_k|x_j) + P(\overline{x_j})\sum_{k=1}^{C} P(y_k|\overline{x_j}).\log P(y_k|\overline{x_j}) \quad (8)$$

where $P(y_k)$ is the probability of a sample feature vector belonging to the class y_k, $P(x_j)$ is the probability of a sample feature vector containing the feature x_j, $P(y_k|x_j)$ is the conditional probability of y_k given the feature x_j. The number of class is denoted by C. In our application, the extracted feature set of 1440 elements from each facial image is ranked by IG. In the second stage, we apply a redundancy reduction algorithm in the context of cooperative or coalitional games, a notion from GT [5]. The algorithm is based on the MSA [5], a framework which relies on GT to estimate the effectiveness of features. This algorithm iteratively computes the usefulness of features and selects them accordingly using forward selection process.

An iterative algorithm for the feature selection, called ECSA, is used to optimize the performance of the classifier on unseen data [5]. The ECSA algorithm combines both the filter and wrapper approaches. However, unlike the filter methods, the features are ranked at each step by using the classifier as a black box. The ranking is based on the *Shapley value* [5], a well known concept from GT, to estimate the importance of each feature for the task at hand by taking into account the interactions between the features. Formally, we can define the coalitional game theory by the pair *(N, u),* where $N = \{1, 2, \dots\dots, n\}$ is the set of all players and $u(F)$, for every $F \subseteq N$, denotes a real number associating a value with the coalition *F.* GT represents the contribution of each player to the game by constructing a certain value function, which assigns a real-value to each player and the values correspond to the contribution of the players in achieving an optimal payoff. The calculation of the contribution value is based on the Shapley value [5]. Essentially, the Shapley value of a player is a weighted mean of its marginal value, averaged over all the possible subsets of players. If we transform the concept of GT into the arena of facial feature subset selection, in which the contribution of each feature is estimated to generate a classifier, the players *N* are mapped to the features of a dataset and the payoff is denoted by a real valued function *u(F).* We can calculate the contribution values from the sampled permutations of the whole set of players, with *d* being the bound on the permutation-size:

$$\theta_i(u) = \frac{1}{|\Pi_d|}\sum_{\pi\in\Pi_d} \Delta_i(F_i(\pi)) \quad (9)$$

where Π_d denotes the set of sampled permutations on subsets of size *d*. The ECSA is iterative in nature, and can either adopt a forward selection or backward elimination approach. In this paper, we consider only the Forward Feature Selection (FFS) approach for the GT-based framework. Based on the contribution value, ECSA ranks each feature and then selects features with the lowest contribution. The algorithm continues to calculate the contribution values of the remaining features, given those that have already been selected, and selects the new features, until the contribution values of all the candidate features fall below a contribution threshold. The algorithm can be regarded as a generalization of filter methods. However, the main idea of the algorithm is that the contribution value is calculated for each feature according to its assistance in improving the classifier's performance, which is generated using a specific induction algorithm, and in conjunction with other features. We propose the following payoff function that minimizes the within-class distance and maximizes the between-class distance:

$$min\big(u(F)\big) = min(W_l) + min(1/(B_l + 1)) \tag{10}$$

In (10), W_l denotes the within-class distance and can be defined as

$$W_l = \frac{1}{n_l} \sum_{k=1}^{n_l} (X_k^l - m_l)^T (X_k^l - m_l) \tag{11}$$

where $l = 1, 2, 3, \dots\dots, c$ and m_l is the mean vector of class l. B_l in (10) denotes the between-class distance and can be defined as

$$B_l = \sum_{l=1}^{c} P_l \times (m_l - m)^T (m_l - m) \tag{12}$$

where m is the mean vector of all samples. The ECSA algorithm in its FFS version is depicted as follows:

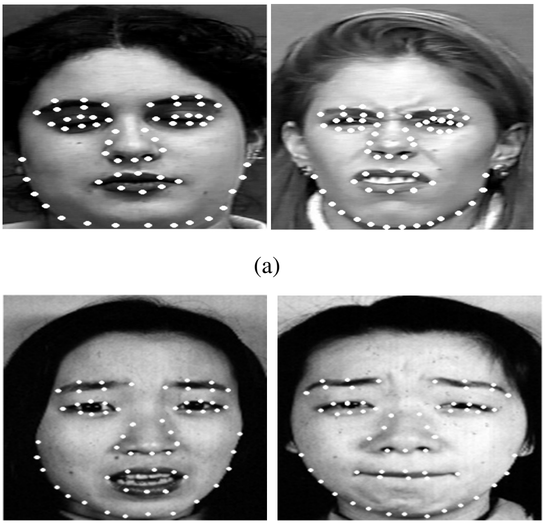
$Extended\ Contribution\ Selection\ Algorithm\ (P, T_d, d, f\ )$:
P is the set of input features, T_d *denotes the contribution threshold,* d *is the maximal permutation size for estimating contribution values, and f is the selected feature subset size in each phase. This algorithm calculates the contribution value of feature, p using the payoff function,* $u(F)$ *described above, and selects at most f features with the lowest contribution values that fall below* T_d.

1. $SelectedFeatures := \phi$
2. $For\ each\ \ p \in P \setminus SelectedFeatures$
   i) $ECNT_p = Contributions\ (p, SelectedFeatures, d)$
3. $if\ min\ ECNT_p < T_d$
   i) $SelectedFeatures :=$
      $SelectedFeatures\ \cup Selection\ \big(\{ECNT_p\}; f, T_d\big)$
   ii) $Goto\ Step\ 2$
   else
   iii) $return\ SelectedFeatures$

The case $SelectedFeatures := \phi$ is handled by returning the fraction of majority class instances. In our feature selection algorithm, we use the Shapley value heuristically to estimate the contribution value of a feature. The decision tree is used for the feature selection, and the SVM is deployed to perform the actual prediction on the selected features based on the performance of the classifier [10].

4 Performance Evaluation

Extensive experiments were conducted on the following two databases, namely, JAFFE [11], and Cohn-Kanade Version 2 or Ck+ [12]. Figure 3 exhibits the results of facial feature tracking using the ASMLF with the game-theoretic approach. From the Figure 3, we can find that the applied game-theoretic approach with the local features accurately finds the feature points on both of the databases. The GT-based lip contour detection process, further, enhances the feature tracking performance. An extensive set of experiments was conducted on all the datasets, and the coupling coefficients, α and β were set to 0.27 when the game-theoretic integration module was used. To obtain the contour data of the lip boundary during game-theoretic evolution, the selected parameter values using the VLS algorithm were set to $\mu = 0.001, v = 2.0$, $\lambda = 5.0$ and time step $\tau = 3.0$. The number of models in the shape models was 21 and 35 for the JAFFE and CK+ databases, respectively. To show the effectiveness of the lip contour detection process, we compared the proposed GT-based lip contour detection scheme with Geodesic Active Contour (GAC) [13], ASM proposed by Cootes et al. [14] and ASMLF [7] as shown in Figure 4. The GT-based scheme substantially improves the accuracy of lip detection, especially when mouths are open. The reason is that our proposed scheme uses the region-based information



(a)

(b)

Fig. 3. Facial point detection results. (a) Cohn Kanade database. (b) JAFFE database.

as well as the gradient data with the game-theoretic fusion method for lip localization. In the first stage of feature selection, we applied IG to rank the feature from the extracted set of 1440 log-Gabor elements. Fig. 5 shows the accuracy of the selected feature subsets with a different number of top-ranked features obtained using IG on two datasets. Fig.5 (a) shows that IG achieves a reasonable accuracy of 93.80% when the number of selected features is 640. Therefore, a selected set of 640 top-ranked feature elements is used as input to the GT-based feature selection criteria for the further improvement of FER. Similarly, we can find from Fig. 5(b) that an accuracy accuracy of 93.25% is reached when the length of the feature vector is 800. After obtaining 800 top-ranked features from IG, we input them to the GT for further reduction of feature size. The proposed cooperative GT-based feature selection approach is used to reduce the feature dimension without compromising the recognition accuracy. The ECSA is prone to overfitting on the validation set. The

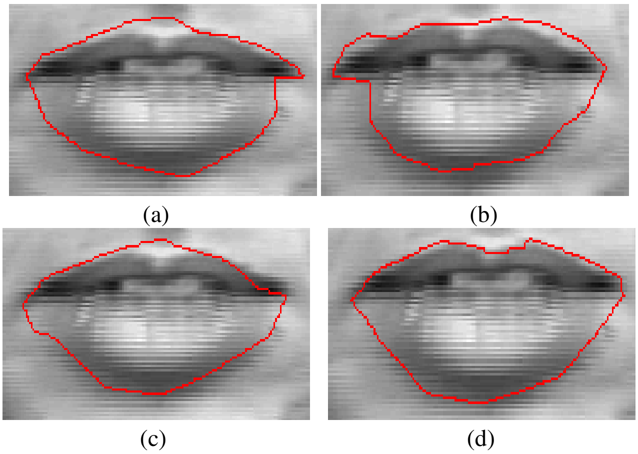


Fig. 4. Comparison of different lip contour detection methods. (a) GAC, (b) ASM, (c) ACMLF, and (d) ASMLF with GT.

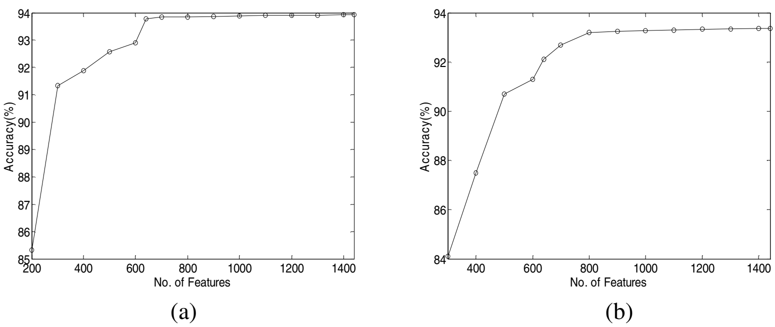
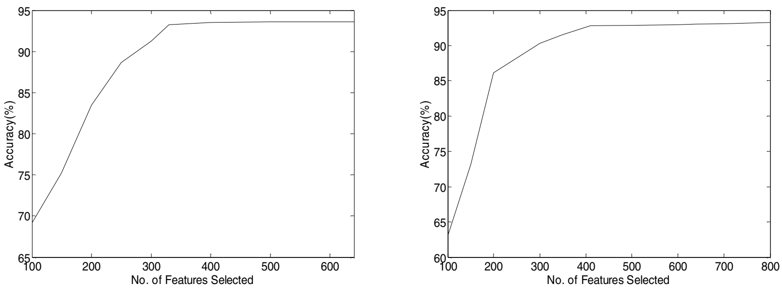


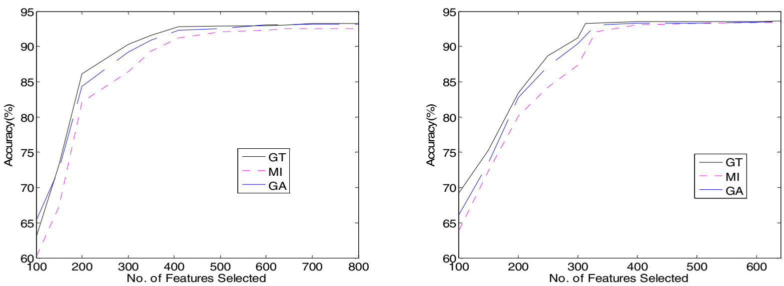
Fig. 5. Feature ranking using information gain. (a) JAFFE database. (b) CK+ database.

"curse of dimensionality" appears, and the irrelevant features are selected if the classifier's performance is evaluated on a small validation set. Since the number of samples from most facial expression research is limited, the cross-validation procedure is commonly used to evaluate the performance of a classifier. For the JAFFE and CK+ datasets, we use a 4-fold cross validation to obtain the validation accuracy. Figs. 6 shows the cross-validation accuracies of the selected feature subsets for the FFS scheme on two datasets. From Fig. 6, we can see that the reasonable accuracy is obtained with the FFS scheme when the number of selected features is (a) 313 in JAFFE dataset and (b) 410 in CK+ datasets. In order to show the effectiveness of the proposed feature selection scheme, we compared our GT-based method with Genetic Algorithm (GA) [15] and Mutual Information (MI) [4]. For GA, we deployed the same fitness function used in the current study. Fig. 7 clearly demonstrates that our GT-based scheme outperforms the other two methods in terms of dimensionality reduction, especially when the feature vector length is comparatively small. We can see from Fig. 7(a) that GT achieves an accuracy of 92.81% when the length of feature vector is 410, while GA reaches a similar accuracy level when the feature vector



(a) (b)

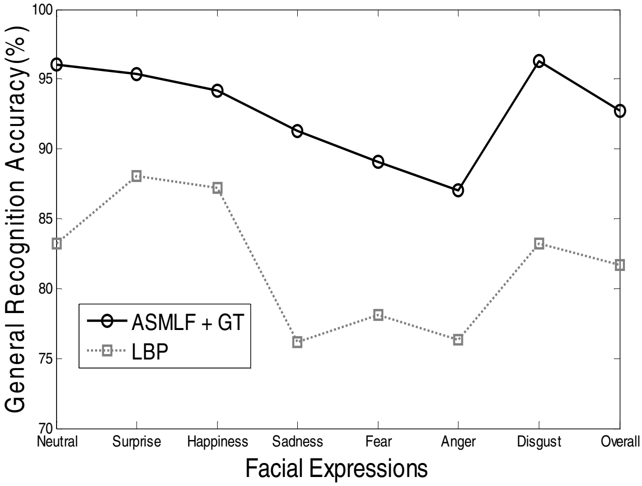
Fig. 6. Feature selection using GT with FFS. (a) JAFFE database. (b) CK+ database.



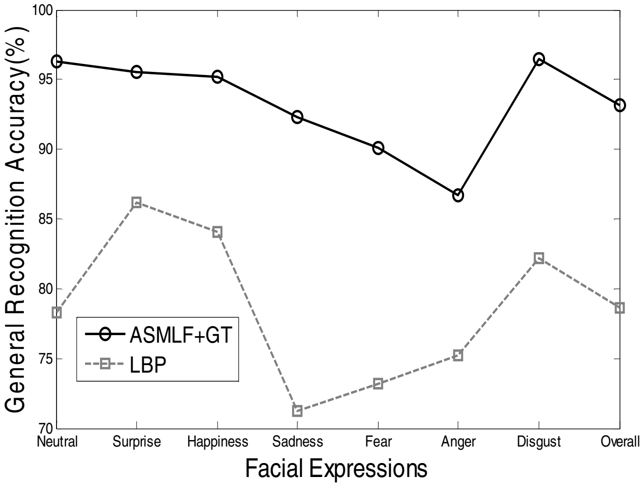
(a) (b)

Fig. 7. Comparison of different feature selection methods. (a) CK+ database. (b) JAFFE database.

length is 560. However, MI was outperformed by GT and GA in terms of both the accuracy and dimensionality. Similarly, in Fig. 7(b), GT shows an accuracy of 93.20% when the selected size of feature vector is 313, whereas the similar recognition accuracy can be reached by GA and MI when the feature vector sizes are 560 and 595, respectively. We used SVM to classify 7 categories of expressions, namely, neutral, happiness, sadness, surprise, anger, disgust, and fear. In order to analyze the performance of the classifier in recognizing individual expression, a training set of 70 images is produced from the JAFFE database. The training set



(a)



(b)

Fig. 8. Comparison of different facial shape localization methods. (a) Cohn-Kanade database. (b) JAFFE database.

contains 7 images for each of the 10 expresser, one image per expression. The other images are then used for testing. For CK+ database, we used a set of 120 facial images for training and the rest of the images were used for testing the overall accuracy of the proposed FER system. We also compared the performance of the proposed algorithm with other existing FER algorithms. Fig. 8 shows the comparison of the proposed algorithm with LBP [1]. We can find that the proposed shape guided approach with the GT-based lip contour detection process shows a better feature point tracking performance than the LBP scheme, since our proposed lip contour detection process enhances the feature point detection accuracy around the mouth region. We can find from Fig. 8 that our method outperforms the conventional LBP method with the accuracies of 93.1% and 92.7% on the JAFFE and CK+ datasets, respectively. Table 1 also demonstrates the comparison of our method with LBP and Boosted LBP reported in [1], and also with PCA-based approach proposed in [8] on JAFFE dataset. From Table 1, we can find that our classification rate on the JAFFE dataset outperforms the other techniques with an accuracy of 93.12%.

Table 1. Comparisons of different FER methods on JAFFE database

Methods	7-Class recognition (%)
LBP [1]	78.6
Boosted-LBP+SVM [1]	81.0
PCA [9]	87.51
Proposed	93.12

5 Conclusions

In this research effort, we have achieved two performance goals. First, a game-theoretic lip contour detection scheme is deployed to enhance the performance of lip localization method. The GT-based algorithm brings together the region-based and boundary-based methods and operates different probability spaces into a common information-sharing framework. The proposed algorithm localizes the lip boundaries from the delineated mouth regions that may be affected by low image intensities, poor acquisition process, opening of the mouth, variability in speaking style, teeth, wrinkles, and by the occlusions of beard and moustache. The localization scheme based on GT avoids the over-segmentation and performs well against the blurred outer lip/skin boundary. Experiments results show that the improved lip localization scheme can enhance the accuracy of the overall facial shape tracking process. Second, a two-stage feature selection approach based on IG and GT is used to find the subset of informative texture features. Further analysis of our results indicates that the proposed feature selection framework is capable of removing redundant and irrelevant features, outperforming other traditional approaches like GA and MI. We validate the proposed FER scheme on JAFFE and CK+ datasets with an encouraging performance.

References

1. Shan, C., Gong, S., McOwan, P.: Facial expression recognition based on Local Binary Patterns: A comprehensive study. Image and Vis. Comput. 27(6), 803–816 (2009)
2. Li, K., Wang, M., Liu, M., Zhao, A.: Improved level set method for lip contour detection. In: IEEE Intl. Conf. Image Process., pp. 673–676 (2010)
3. Chakraborty, A., Duncan, J.: Game-theoretic integration for image segmentation. IEEE Trans. Pattern Anal. and Machine Intell. 21(1), 12–30 (1999)
4. Makrehchi, M., Kamel, M.: Aggressive feature selection by feature ranking. In: Liu, H., Motoda, H. (eds.) Computational Methods of Feature Selection, pp. 313–330. Chapman and Hall/CRC Press (2007)
5. Cohen, S., Dror, D., Ruppin, E.: Feature selection via coalitional game theory. Neural Computa. 19, 1939–1961 (2007)
6. Viola, P., Jones, M.: Robust Real-Time Face Detection. Intl. J. Comp. Vis. 57(2), 137–154 (2004)
7. Ginneken, B., Frangi, A., Staal, J., Romeny, B., Viergever, M.: Active shape model segmentation with optimal features. IEEE Trans. Medical Imaging 21(8), 924–933 (2002)
8. Li, C., Xu, C., Gui, C., Fox, M.: Level set evolution without re-initialization: a new variational formulation. In: Proc. IEEE Intl. Conf. Comp. Vis. and Pattern Recog., pp. 430–436 (2005)
9. Bashyal, S., Venayagamoorthy, G.: Recognition of facial expressions using Gabor wavelets and learning vector quantization. Intl. J. Engg. App. of Artificial Intell. 21(7), 1–9 (2008)
10. Vapnik, V.: Statistical Learning Theory. John Wiley & Sons, New York (1998)
11. Lyons, M., Budynek, J., Akamatsu, S.: Automatic classification of single facial images. IEEE Trans. Pattern Anal. and Machine Intell. 21(12), 1357–1362 (1999)
12. Lucey, P., Cohn, J., Kanade, T., Saragih, J., Ambadar, Z.: The extended Cohn-Kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression. In: IEEE Intl. Conf. Computer Vis. and Pattern Recog. Workshop, pp. 94–101 (2010)
13. Paragios, N., Deriche, R.: Geodesic active contours and level sets for the detection and tracking of moving objects. IEEE Trans. Pattern Analysis and Machine Intelligence 22(3), 266–280 (2000)
14. Cootes, T., Taylor, C., Cooper, D., Graham, J.: Active shape models—their training and application. Computer Vis. Image Understand. 61(1), 38–59 (1995)
15. Goldberg, D.: Genetic algorithms in search, optimization, and machine learning. Addison-Wesley Professional (1989)