

Comparing Feature Point Tracking with Dense Flow Tracking for Facial Expression Recognition

José V. Ruiz, Belén Moreno, Juan José Pantrigo, Ángel Sánchez

Departamento de Ciencias de la Computación
Universidad Rey Juan Carlos, C/Tulipán, s/n,
28933 Móstoles, Madrid, Spain

`jvruiz@alumnos.urjc.es`, `belen.moreno@urjc.es`, `juanjose.pantrigo@urjc.es`,
`angel.sanchez@urjc.es`

Abstract. This work describes a research which compares the facial expression recognition results of two point-based tracking approaches along the sequence of frames describing a facial expression: feature point tracking and holistic face dense flow tracking. Experiments were carried out using the Cohn-Kanade database for the six types of prototypic facial expressions under two different spatial resolutions of the frames (the original one and the images reduced to a 40% of its original size). Our experimental results showed that the dense flow tracking method provided in average for the considered types of expressions a better recognition rate (95.45% of success) than feature point flow tracking (91.41%) for the whole test set of facial expression sequences.

1 Introduction

Automatic Facial Expression Analysis (AFEA) is becoming increasingly important research field due to its many applications: human-computer intelligent interfaces (HCII), human emotion analysis, talking heads, among others [8]. The AFEA systems typically deal with the recognition and classification of facial expression data which are given by one of these two kinds of patterns: Facial Action Units and Emotion Expressions.

- Facial Action Units (AUs): correspond to subtle changes in local facial features related to specific facial muscles (i.e. lip corner depressor, inner brow raiser, ...). They form the Facial Action Coding System (FACS) which is a classification of the possible facial movements or deformations without being associated to specific emotions. Descriptions of the AUs were presented in [9] and they can appear individually or in combination with other AUs.
- Emotion Expressions: are facial configurations of the six prototypic basic emotions (disgust, fear, joy, surprise, sadness and anger) which are universal along races and cultures. Each emotion has a correspondence with the given prototypic facial expression [9].

Many papers in the AFEA literature deal with the analysis or recognition of expressions by considering both types of patterns. Examples of works related to the recognition of some AUs are [1][13]. Examples of works which deal with the emotion expressions analysis are [11][15]. There are also papers which consider the classification of facial patterns expression using both AUs and emotion expressions [10]. Our work only consider the emotion expression patterns for recognition.

The development of robust algorithms with respect to the individual differences in expressions need from large databases containing subject of different races, ages, gender, etc. in order to train and evaluate these systems. Two examples of relevant Facial Expression databases are: Cohn-Kanade facial expression database [6] and the Japanese woman facial expression database (JAFFE) [14]. Some survey papers [8][2] offer overviews to describe the facial expression analysis algorithms. The different approaches in this area consider three main stages in an AFEA system: (a) face detection and normalization, (b) feature extraction and (c) expression classification.

Next, we only refer some of the papers related to the feature extraction stage related with our work. With respect to the extraction of facial expression features that model the facial changes, they can be classified [2] according to their nature in: (a) deformation features and (b) movement features. Deformation features do not have into account the information of the pixel movement, and they can be obtained from static images. Movement features are centred in the facial movements and they are applied to video sequences. The more relevant techniques which use these features are: (a) the use of movement models, (b) difference images, (c) marker tracking, (d) feature point tracking and (e) dense optical flow. The last two types of feature extraction methods have been extensively experimented and compared in our work. The integration of optical-flow with movement models increases their stability and improves the facial movement interpretation and related acial expression analysis.

Dense optical flow computed in regions (windows) was used in [12] to estimate the activity of twelve facial muscles. Among the feature point tracking based methods a representative work is [13] where lip, eye, eyebrows and cheeks models were proposed and the feature point tracking was performed for matching the model contours to the facial features. A spatial-temporal description which integrated dense optical flow, feature point tracking and high gradient component analysis in the same hybrid system, was proposed in [1], were HMM were used for the recognition of 15 AUs. A manual initialization of points in the neutral face of the first frame was performed.

2 Proposed facial expression recognition system

This section outlines our facial expression recognition system from video sequences. We implemented two point-tracking strategies based on optical flow methods: the first one is feature-based and considers the movement of 15 feature points (i.e. mouth corners) and the second one is holistic and uses the displace-

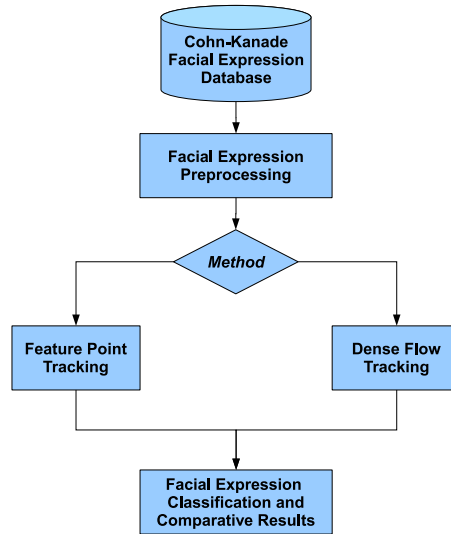


Fig. 1. Facial expression system architecture.

ment of the facial points which are densely and uniformly placed on a grid centered on the central face region. Our system is composed by four subsystems or modules: pre-processing, feature point tracking, dense flow point tracking and facial expression classification. Figure 1 represents the components of our facial expression recognition system. The following subsections detail the involved stages in each module.

2.1 Pre-processing

The pre-processing stage will be different for any of the two facial point tracking methods. In the case of feature point tracking method, we first manually select the two inner eye corners points (one from each eye) which will be used to normalize the global feature point displacements in each expression along the sequence of frames (avoiding scaling problems among the different faces in the database) and to compute the face normalization angle (avoiding that faces have different orientations).

For the dense flow tracking, normalization requires the following steps: locate manually five considered facial points placed near the face symmetry axis, computing the face angle normalization, obtaining a rectangle containing the central facial using a heuristic procedure and splitting this rectangle into three regions whose size in the vertical direction is readjusted by considering the standard facial proportions. The dense flow tracking pre-processing procedure is illustrated by the Figure 2 (where the A and B points are also used for the pre-processing stage in the considered feature point tracking method).

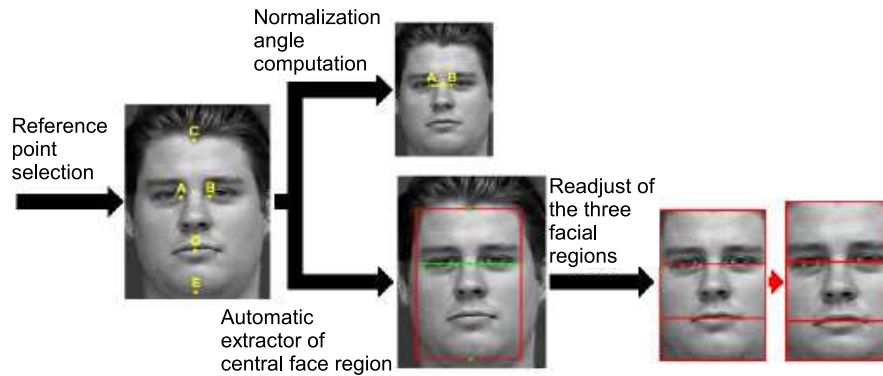


Fig. 2. Dense flow tracking pre-processing

2.2 Feature Point Tracking

Expressions are recognized (after applying the previous pre-processing to the first frame of each video sequence) by computing the sum of displacement vectors for each of the considered facial feature points from each frame to the next one in the expression video-sequence. This task can be decomposed into two substages: feature point location and optical flow application to feature points.

Feature point location In our approach, the considered feature points are manually extracted from the first frame of the sequence and no face region segmentation stage is required. The set of 15 considered feature points is represented in Figure 3. These points are located in the following facial regions: 4 on the eyes (two points for eye which are placed on the upper and lower eyelids), 4 on the eyebrows (two points for eyebrow which are the innermost and outermost ones), 4 on the mouth (which correspond to the corners and the upper and lower middle points), and 3 other detectable points (one is placed on the chin, and the other two are symmetrically placed one on each cheek). Similar subsets of points were also used by other authors [8].

These points are selected using the computer mouse in the first frame of the expression and then they are automatically tracked along the rest of frames in the video sequence describing the expression using Lucas-Kanade optical flow algorithm [3].

Since there are some differences with respect to pose and size between the images of individuals in the database, it is common to previously normalize all the facial images in the video sequences to guarantee that point displacement measures are correctly computed. For this aim, we have used the vector defined by the two inner eye corners to normalize the considered facial point displacements of faces in scale and orientation.



Fig. 3. The 15 selected facial feature points.

Optical flow application to feature points The optical flow tracking is applied between each pair of consecutive frames in the video sequence. Lucas-Kanade method [3] for computing the optical flow has been applied to estimate the displacement of points. Lucas-Kanade algorithm is one of the most popular gradient-based (or correlation) methods for motion estimation computing in a video sequence. This method tries to obtain the motion between two image frames which are taken at times t and $t + \delta t$ at every pixel position assuming a brightness constancy.

We computed the global displacement vector for each considered facial point in any expression by applying the Lucas-Kanade algorithm between each pair of consecutive frames. These corresponding inter-frame displacement vectors are then added to obtain the global displacement vector corresponding to each point along the expression.

Once applied this algorithm, two values are computed for each feature point displacement vector along the sequence of frames for any facial expression: its normalized module (in pixels) and the normalized angle of the displacement vector. A feature vector v is created for each facial expression sequence containing the pair of displacement features for each of the considered $N = 15$ facial points and a natural number T which codifies the type of facial expression:

$$v = [|\mathbf{p}_1|, \theta_{\mathbf{p}_1}, |\mathbf{p}_2|, \theta_{\mathbf{p}_2}, \dots, |\mathbf{p}_N|, \theta_{\mathbf{p}_N}, T] \quad (1)$$

where $|\mathbf{p}_i|$ represents the module of the displacement vector corresponding to feature point \mathbf{p}_i and $\theta_{\mathbf{p}_i}$ the angle of this vector. The whole set of these feature vectors corresponding to the considered facial video sequences is properly partitioned in two independent files (training and test files) used by the SVM algorithm to classify the considered types of expressions.

2.3 Dense Flow Point Tracking

Due to the difficulty of a precise extraction of the considered feature points (even by manually marking the points in the first frame, since these points

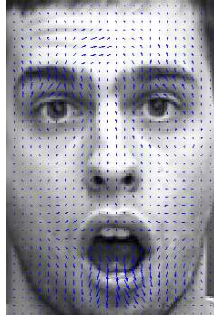


Fig. 4. Result of applying dense optical flow by reducing the spatial resolution for a surprise expression.

usually correspond to a region of pixels), we have also considered the tracking of a grid of uniformly spaced points of the central facial region. This region is automatically extracted by the method explained in subsection 2.1. Since the facial frames in the considered database have a 640×480 spatial resolution, and neighbour points in the face (along the consecutive frames) present a high correlation, it becomes computationally expensive to apply the Lucas-Kanade algorithm to each point contained in the considered facial region. Therefore, we applied a two-level Gaussian pyramid (that is equivalent to a low-pass filter) to decrease $1/16$ the number of points to which the optical flow computation is applied. In this way, the optical flow algorithm is now computed on 3,750 points instead of the around 60,000 points contained in the considered central facial region. Moreover, the facial movement vectors between frames now become more smooth and continuous (see Figure 4).

2.4 Facial expression classification

We used a Support Vector Machine (SVM) for our facial expression recognition experiments. A SVM is a classifier derived from statistical learning theory that has interesting advantages: (1) ability to work with high-dimensional data and (2) high generalization performance without the need to add a-priori knowledge, even when the dimension of the input space is very high. The problem that SVMs try to solve is to find an optimal hyperplane that correctly classifies data points by separating the points of two classes as much as possible. SVMs have also been generalized to find the set of optimal separating hyperplanes for a multiclass problem. Excellent introductions to SVM can be found in [4][5].

We used the SVMTool [7] tool for our facial classification experiments: It requires from a training and a testing stage. During the training stage, a set of the SVM parameters are adjusted (i.e. those related with the type of kernel used by the classifier). Once the SVM has been trained, we use the test set of facial expression sequences to compare the performance of both considered facial expression recognition approach: feature point and dense flow tracking.



Fig. 5. Software tool visual interface.

3 Experimental results

3.1 Cohn-Kanade facial expression database

For our facial recognition experiments we used the Cohn-Kanade facial expression database [6]. Image data consist of approximately 500 frame sequences from about 100 different subjects. The included subjects range in age from 18 to 30 years, 65 percent of them were female; 15 percent were African-American and 3 percent Asian or Latino.

Sequences of frontal images representing a prototype facial expression always start with the neutral face and finish with the expression at its higher intensity (the corresponding frames are in this way incrementally numbered). These sequences were captured with a video camera, digitized into 640 by 480 pixel arrays with 8-bit precision for grayscale values, and stored in jpeg format.

For many of the subjects in this database, the six basic facial expression sequences were captured: joy, surprise, anger, fear, disgust and sadness. The number of frames per sequence is variable and its average value is 18. A subset of 138 subjects of the database (all of them containing the sequences corresponding to the six types of expressions) was used in our experiments.

3.2 Experiments: description and results

Most of the components of the proposed expression recognition system were programmed in MATLAB. Figure 5 presents the interface of the tool for a sample face when the feature point tracking method is applied. A PC Pentium 4 at 2.2 GHz with 1GB of RAM memory was used for the algorithm development and tests.

Experiments were organized in four methods by considering the two compared optical flow point tracking methods and two different spatial image resolutions:

- feature point tracking using the original 640×480 frame spatial resolution in the Cohn-Kanade database (FPT 1:1),
- feature point tracking by reducing to the 40% original resolution (FPT 1:0.4),
- dense flow tracking using the original resolution (DFT 1:1), and
- dense flow tracking by reducing to the 40% original resolution (DFT 1:0.4).

For the experiments, a total of 246 image sequences of facial expressions were used. The training set was composed by 180 sequences (30 for each type of basic facial expression) and the test set used the resting 66 sequences (11 for each type of expression). SVM Torch classifier was trained using three types of kernels (polynomial, Gaussian and sigmoid, respectively) and manually adjusting their corresponding parameters to improve the classification results. Best recognition results in average were always obtained using the Gaussian kernel.

Next, we compare the best recognition results for the four approaches (FPT 1:1, FPT 1:0.4, DFT 1:1 and DFT 1:0.4, respectively) using the test set of 66 expression sequences. Table 1 presents the best recognition rates achieved for each type of basic facial expression using the four considered approaches. We also show in the last row of this table the best average recognition result for the six types of facial expressions using the four compared methods. The best values of SVM Torch parameters: *std* (standard deviation for the Gaussian kernel) and *c* (trade-off value between training error and margin) are also shown for each method.

Best average facial expression recognition results were achieved with the dense flow tracking method at the original frame resolution (95.45% of success rate). The difference of recognition results using this same method but reducing the frame resolution to a 40% of its original size is negligible (95.20% of correct recognition). However, the application of Lucas-Kanade algorithm to this second method reduces its computation time an 85% in average (209.29 seconds for DFT 1:1 and 31.55 seconds for DFT 1:0.4, respectively). Using the feature point tracking method, the average success recognition rate at the original resolution is 91.41% and 88.38% by reducing the frame resolution to a 40%, respectively. In this second case, by reducing the spatial resolution the average time of applying of Lucas-Kanade is reduced about a 60% (124.4 seconds for FPT 1:1 and 49.77 seconds for DFT 1:0.4, respectively). The best recognized expression for the DFT 1:1 method is sadness with a 100% of success rate and the worst recognized one is fear with a 90.91% of success using our test set.

We also show in Table 2 the corresponding confusion matrix relating the six types of expressions for the DFT 1:1 method.

It is difficult to compare the results of the presented facial expression recognition methods with other works considering a similar approach than the presented in this work approach. Lien et al [1] also used feature and dense flow tracking but their recognition approach is based on the Facial Action Coding System (FACS) to recognize action units (describing the expressions) but considering only for experiments the point displacements of the upper face region (above both eye brows). The recognition is performed using Hidden Markov Models (HMM) as

Table 1. Best recognition results obtained by the four methods for each type of expression.

Facial Expression	FPT 1:1 (<i>std</i> =300, <i>c</i> =10)	FPT 1:0.4 (<i>std</i> =700, <i>c</i> =100)	DFT 1:1 (<i>std</i> =7000, <i>c</i> =100)	DFT 1:0.4 (<i>std</i> =2500, <i>c</i> =100)
Joy	90.91	96.97	93.94	95.45
Surprise	98.48	92.42	96.97	98.48
Sadness	90.91	87.88	100.00	92.42
Anger	86.36	78.79	95.45	93.94
Disgust	92.42	90.91	95.45	96.97
Fear	89.39	83.33	90.91	93.94
Average	91.41	88.38	95.45	95.20

Table 2. Confusion matrix for the DTF 1:1 method.

	Joy	Surprise	Sadness	Anger	Disgust	Fear
Joy	11	0	0	0	0	0
Surprise	0	11	0	0	0	0
Sadness	0	0	11	0	0	0
Anger	0	1	0	8	2	0
Disgust	0	0	0	0	11	0
Fear	4	1	0	0	1	5
Total						

classifier. They only reported the average expression recognition rate for the feature point tracking (85%) and for the dense flow tracking method (93%).

4 Conclusion and future work

We have implemented a semi-automatic facial expression recognition system using sequences of frames describing the expression. Two approaches based on optical flow of facial points have been compared (feature point tracking and dense flow tracking, respectively) at two different frame resolution (original one and reducing to a 40% the spatial resolution of the frames). Experiments were performed with the Cohn-Kanade database of facial expressions. We can conclude from our tests that dense optical flow method (using SVM Torch [7] as classification tool with a properly-tuned Gaussian kernel parameters) provided better recognition results (95.45%) than the equivalent feature point tracking approach (91.41%). The dense flow tracking method also offers two additional advantages: similar recognition results for the two considered spatial frame resolutions and a smaller number of points need to be located (5 points in the preprocessing shown in Fig. 2 instead the 17 points required in the feature facial tracking: 2 for preprocessing and 15 to be tracked). However, as a disadvantage

dense flow tracking presents a much higher processing time specially working at the original frame resolutions.

As future work, a first improvement for our system is the automatic search of considered preprocessing and feature points in the first frame of the sequence. It is also desirable to adapt the system to recognize several degrees of intensities in each basic expression. A more complete fair comparison of our results with other related works using the same expression database is also needed.

Acknowledgements

This research has been partially supported by the Spanish projects TIN2008-06890-C02-02 and URJC-CM-2008-CET-3625.

References

1. J.J. Lien, T. Kanade, J.F. Cohn and C.C. Li, Automated Facial Expression recognition Based on FACS Action Units, Proc. of the Third IEEE Intl. Conf. on Face and Gesture Recognition, 390-395, 1998
2. Y. Tian, T. Kanade and J.F. Cohn, Facial Expression Analysis, in: S.Z Li and A.K. Jain (editors): Handbook of Face Recognition, Springer, 247-275, 2004
3. B.D. Lucas and T. Kanade, An iterative image registration technique with an application to stereo vision, Proc. 7th Int. Joint Conf. on Artificial Intelligence, 674-679, 1981
4. Cristianini, N., Shawe-Taylor, J. An Introduction to Support Vector Machines. Cambridge University Press, 2000
5. Vapnik, V. The Nature of Statistical Learning Theory. Springer, New York, 1995
6. Kanade, T., Cohn, J. F., Tian, Y. Comprehensive database for facial expression analysis. Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition (FG'00), Grenoble, France, 46-53, 2000
7. R. Collobert and S. Bengio, SVMtorch: Support Vector Machines for Large-Scale Regression Problems, Journal of Machine Learning Research (1):143-160, 2001
8. B. Fasel and J. Luttin, Automatic Facial Expression Analysis: Survey, Pattern Recognition, 36(1):259-275, 2003
9. P. Ekman and W. Friesen, The Facial Action Coding System: A Technique for the Measurement of Facial Movement, Consulting Psychologists Press, 1978
10. G. Littlewort, M. S. Bartlett, I. Fasel, J. Susskind, J. Movellan, Dynamics of facial expression extracted automatically from video, Image and Vision Computing 24:615-625, 2006
11. M. Lyons, J. Budynek, S. Akamatsu, Automatic classification of single facial images, IEEE Trans. Pattern Analysis and Machine Intelligence 21(12), 1999
12. K. Mase, Recognition of facial expression from optical flow, IEICE Transactions, E. 74(10):3474-3483, 1991.
13. Y.L. Tian, T. Kanade, and J. Cohn, Recognizing action units for facial expression analysis, IEEE Trans. on Pattern Analysis and Machine Intelligence, 23(2):1-19, 2001.
14. <http://www.kasrl.org/jaffe.html>
15. Z. Wen and T. Huang, Capturing subtle facial motions in 3d face tracking, Proc. ICCV, 2003.