

Facial Action Unit Tracking and Facial Activity Recognition Based On Dynamic Bayesian Network

M. Shanmuga Priya¹

¹PG Student, Park College of Engineering and Technology
Email : priyamanimozhi1991@gmail.com

Abstract-The tracking and recognition of facial activities from images or videos have attracted great attention in computer vision field. Facial activities are characterized by three levels. First, in the bottom level, facial feature points around each facial component, i.e., eyebrow, mouth, etc., which describe the detailed of face shape information are captured. Second, in the middle level, facial action units, defined in the facial action coding system that represents the contraction of a specific set of facial muscles, i.e., lid tightened, eyebrow raiser, etc are identified. Finally, in the top level, six prototypical facial expressions representing the global facial muscle movement which are commonly used to describe the human emotion states are measured. In existing approaches focus on the three levels of facial activities, and track them separately. But proposed work introduces a unified probabilistic framework based on the Dynamic Bayesian Network to simultaneously and efficiently represent the facial movements in different levels which improves the efficiency of the system when compared with existing systems.

Index Terms— Bayesian network, expression recognition, facial action unit

recognition, facial feature tracking, simultaneous tracking and recognition.

I INTRODUCTION

The recovery of facial activities in image sequence is an important and challenging problem. In recent years, plenty of computer vision techniques have been developed to track or recognize facial activities in three levels. First, in the bottom level, facial feature tracking, which usually detects and tracks prominent facial feature points (i.e., the facial landmarks) surrounding facial components (i.e., mouth, eyebrow, etc.), captures the detailed face shape information. Second, facial actions recognition, i.e., recognize facial Action Units (AUs) defined in the Facial Action Coding System (FACS) [1], try to recognize some meaningful facial activities (i.e., lid tightened, eyebrow raiser, etc.). In the top level, facial expression analysis attempts to recognize facial expressions that represent the human emotional states. The facial feature tracking, AU recognition and represent the facial activities in three levels from local to global, and they are interdependent problems. For example, facial feature tracking can be used

In the feature extraction stage in expression/AUs recognition and expression/AUs recognition results can provide a prior distribution for facial feature points. However, most current methods only track or recognize the facial activities in one or two levels, and track them separately, either ignoring their interactions or limiting the interaction to one way. In addition, the estimates obtained by image-based methods in each level are always uncertain and ambiguous because of noise, occlusion and the imperfect nature of the vision algorithm. The proposed facial activity recognition system consists of two main stages: offline facial activity model construction and online facial motion measurement and inference. Specifically, using training data and subjective domain knowledge, the facial activity model is constructed offline. During the online recognition, as shown in Fig. 1, various computer vision techniques are used to track the facial feature points, and to get the measurements of facial motions, i.e., AUs. These measurements are then used as evidence to infer the true states of the three level facial activities simultaneously.

II LITERATURE REVIEW

1) Active Shape Models-Their Training and Application

Model-based vision is firmly established as a robust approach to recognizing and locating known rigid objects in the presence of noise, clutter, and occlusion. It is more problematic to apply model based methods to images of objects whose appearance can vary, though a number of approaches based on the use of flexible templates have been proposed. The problem with existing methods is that they sacrifice model specificity in order to

accommodate variability, thereby compromising robustness during image interpretation. They argue that a model should only be able to deform in ways characteristic of the class of objects it represents. It describes a method for building models by learning patterns of variability from a training set of correctly annotated images. These models can be used for image search in an iterative refinement algorithm analogous to that employed by Active Contour Models. The key difference is that our Active Shape Model (ASM) can only deform to fit the data in ways consistent with the training set.

They have described Point Distribution Models (PDMs)-statistical models of shape which can be constructed from training sets of correctly labeled images. A PDM represents an object as a set of labeled points, giving their mean positions and a small set of modes of variation which describe how the object's shape can change. Applying limits to the parameters of the model enforces global shape constraints ensuring that any new examples generated are similar to those in the training set. Given a set of shape parameters, an instance of the model can be calculated rapidly. The models are compact and are well suited to generate-and-test image search strategies. Active Shape Models (ASMs) exploit the linear formulation of PDMs in an iterative search procedure capable of rapidly locating the modeled structures in noisy, cluttered images even if they are partially occluded. Object identification and location are robust because the models are specific in the sense that instances are constrained to be similar to those in the training set.

2) Facial Action Coding System(FACS)

The Facial Expression Coding System (FACES) was developed as a less

time consuming alternative to measuring facial expression that is aligned with dimensional models of emotion. The system provides information about the frequency, intensity, valence, and duration of facial expressions. The selection of the variables included in the system was based on theory and previous empirical studies. Adopting the descriptive style of Ekman and similar to the work of Notarius and Levenson (1979), an expression is defined as any change in the face from a neutral display (i.e., no expression) to a non-neutral display and back to a neutral display.

When this activity occurs, a frequency count of expressions is initiated. Next, coders rate the valence (positive or negative) and the intensity of each expression detected. Notice that this is quite different from assigning an emotion term to each expression. While FACES requires coders to decide whether an expression is positive or negative, it does not require the application of specific labels. That is, judgments about emotion, in this case whether an expression is positive or negative, are made by persons who are considered to be familiar with emotion in a particular culture. In addition to valence and intensity, coders also record the duration of the expression.

They have developed Facial Action Coding System for facial expressions based on discrete emotions specifically designed to measure human muscle movements. They first measure only two facial expressions such as anger or happiness. Most of the discrete emotions are designed to measure only basic emotions. They achieved how to identify human facial expression with the FACS model. But, this approach fails to expose all human facial movements. This approach is inconsistent to measure the facial dimensions. In future they would like to

present all human face expressions and rigid head movements.

3) Robust Facial Feature Tracking Under Varying Face Pose and Facial Expression

Hierarchical multi-state pose-dependent approach for facial feature detection and tracking under varying facial expression and face pose. For effective and efficient representation of feature points, a hybrid representation that integrates Gabor wavelets and gray-level profiles is proposed. To model the spatial relations among feature points, a hierarchical statistical face shape model is proposed to characterize both the global shape of human face and the local structural details of each facial component.

Furthermore, multi-state local shape models are introduced to deal with shape variations of some facial components under different facial expressions. During detection and tracking, both facial component states and feature point positions, constrained by the hierarchical face shape model, are dynamically estimated using a Switching Hypothesized Measurements (SHM) model. Experimental results demonstrate that the proposed method accurately and robustly tracks facial features in real time under different facial expressions and face poses.

They presented a posed expression for human face. Because of facial expression recognition is very challenging due to posed expressions. They developed multi-state pose-dependent hierarchical shape model for tracking varying face pose and face expression. They proposed posed expression method for facial recognition and achieved small pose variation in human face. This approach will lose to identify large pose variation. It improves the accuracy and robustness of facial feature tracking. Need

lot of training dataset. In future they plan to combine the relationship between different facial expressions.

III EXISTING SYSTEM

Facial feature points encode critical information about face shape and face shape deformation. Accurate location and tracking of facial feature points are important in the applications such as animation, computer graphics, etc. Model free approaches are general purpose point trackers without the prior knowledge of the object. Each feature point is usually detected and tracked individually by performing a local search for the best matching position.

Facial expression recognition systems usually try to recognize either six expressions or the AUs. Image-based approaches, which focus on recognizing facial actions by observing the representative facial appearance changes, usually try to classify expression or AUs independently and statically.

The idea of combining tracking with recognition has been attempted before, such as simultaneous facial feature tracking and expression recognition and integrating face tracking with video coding. However, the model free methods are susceptible to the inevitable tracking errors due to the aperture problem; noise, and occlusion, Simple parallel mechanism may not be adequate to describe the interactions among facial feature points.

Discrete states still cannot describe the details of each facial component movement, i.e., only three discrete states are not sufficient to describe all mouth movement

IV PROPOSED SYSTEM

In proposed a hierarchical framework based on Dynamic Bayesian Network (DBN) for simultaneous facial feature tracking and facial expression recognition.

By systematically representing and modeling inter relationships among different levels of facial activities, as well as the temporal evolution information, the proposed model achieved significant improvement for both facial feature tracking and AU recognition, compared to state of the art methods.

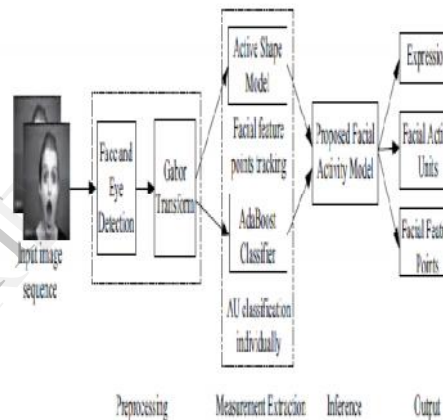


Figure 5.1 Block diagram of facial expression identification

V MODULES DESCRIPTION

1) Preprocessing & Segmentation:

The first step is to adjust the gray level of a flame image according to its statistical distribution. This segmentation process to be handled according to the image size the segmentation blocks will vary like 8X16, 4X16.
















Segmentation is a process of extracting and representing information from an image is to

group pixels together into regions of similarity.

2) Dataset preparing:

The overall possible datasets are collected from all the sources to identify the particular human face expressions. Feature extracting based on the datasets. So while collecting datasets, must add all the possible expressions from the social people.

The proposed model is evaluated on two databases, i.e., the CK+ database and MMI facial expression database. The advantage of using this database is that it contains a large number of videos that display facial expressions.

 AU1 Inner brow raiser	 AU2 Outer brow raiser	 AU4 Brow lowerer	 AU5 Upper lid raiser
 AU6 Cheek raiser	 AU7 Lid tightener	 AU9 Nose wrinkler	 AU12 Lip corner puller
 AU15 Lip corner depressor	 AU17 Chin raiser	 AU23 Lip tightener	 AU24 Lip pressor
 AU25 Lip part	 AU26 Jaw drop	 AU27 Mouth stretch	

List of AUs and their relationships

3) Measurement Extraction

a) Facial Feature Tracking

Facial feature, which usually detects and tracks important facial feature points (i.e. the facial landmarks) and facial components (i.e. mouth, eyes) capture the details of face shape information. Facial feature tracking can be used to extract the face expression, and this result can provide a prior distribution for facial feature points. DBN helps to improve

the facial feature tracking performance. This facial feature tracking could be classified into two categories: Model free and Model based method. Model free method usually detected matching position. Model based method such as ASM (Active Shape Model).

Tracked the facial feature point measurements through an Active Shape Model (ASM) based approach [1], which first searches each, point locally and then constrains the feature points based on the ASM model, so that the feature points can only deform in specific ways found in the training data. The ASM model is trained using 500 key frames selected from the training data, which are 8-bit gray images with 640×480 image resolution. All the 26 facial feature point positions are manually labeled in each.

ASM to improve the robustness and accuracy of feature point tracking. For example, for mouth, they used three ASMs to represent the three states of mouth, i.e., widely open, open and closed.

b) AU classification:

Model-based methods overcome this weakness by making use of the relationships among AUs, and recognize the AUs simultaneously [6]. Facial expression recognition systems usually try to recognize either six expressions or the AUs. Over the past decades, there has been extensive research on facial expression analysis. Current methods in this area can be grouped into two categories: image-based methods and model-based methods.

Image-based approaches, which focus on recognizing facial actions by observing the representative facial appearance changes, usually try to classify expression or AUs independently and statically. This kind of

method usually consists of two key stages. First, various facial features, such as optical flow, explicit feature measurement (e.g., length of wrinkles and degree of eye opening), Local Binary Patterns (LBP), Independent Component Analysis (ICA), etc., are extracted to represent the facial expressions or facial movements. Given the extracted facial features, the expression/AUs are identified by recognition engines, such as Support Vector Machines (SVM).

Model based methods overcome this weakness by making use of the relationships among AUs, and recognize the AUs simultaneously.

c) Facial expression identification:

Given the facial action model [9] and image observations, all three levels of facial activities are estimated simultaneously through a probabilistic inference by systematically integrating visual measurements with the proposed model.

Compared to the previous related works, proposed work has the following features.

1) First, build a DBN model to explicitly model the two-way interactions between different levels of facial activities. In this way, not only the expression and AUs recognition can benefit from the facial feature tracking results, but also the expression recognition can help improve the facial feature tracking performance.

2) Second, recognize all three levels of facial activities simultaneously. Given the facial action model and image observations, all three levels of facial activities are estimated simultaneously through a probabilistic inference by systematically

integrating visual measurements with the proposed model.

VI CONCLUSION

A hierarchical framework based on Dynamic Bayesian Network for simultaneous facial feature tracking and facial activity recognition systematically representing and modeling inter relationships among different levels of facial activities, as well as the temporal evolution information, the proposed model achieved significant improvement for both facial feature tracking and AU recognition, compared to state of the art methods.

Proposed method did not use any measurement specifically for expression, and the global expression is directly inferred from AU and measurements, with improved image-based computer vision technology.

Proposed system may achieve better results with little changes to the model. Facial feature point measurements and from their relationships.

The improvements for facial feature points and AUs come mainly from combining the facial action model with the image measurements.

Specifically, the erroneous facial feature measurements and the AU measurements can be compensated by the model's build-in relationships among different levels of facial activities, and the build-in temporal relationships.

This proposed model systematically captures and combines the prior knowledge with the image.

VII FUTURE ENHANCEMENT

In the future work, plan to introduce the rigid head movements, i.e., head pose, into the model to handle multi view faces. In addition, modeling the temporal phases of each AU, which is important for understanding the spontaneous expression, is another interesting direction to pursue. And also plan to introduce the image extraction from blur image, i.e., would like to convert original image by using Dynamic Bayesian Network with Navye Bayes Classifier. And also would like to understand more relationship between Facial Expression.

VIII REFERENCES

- 1) T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, “**Active Shape Models- Their Training and Application,**” Computer Vision and Image Understanding., vol.61, no.1, pp. 38-59, 1995.
- 2) B. A. Draper, K. Baek, M. S. Bartlett, and J. R. Beveridge, “**Recognizing Faces With PCA and ICA,**” Computer Vision and Image Understanding., vol.91, no.1-2, pp.115-137, 2003.
- 3) P. Ekman and W. V. Friesen, “**Facial Action Coding System(FACS),**” Manual. Palo Alto, CA, USA: Consulting Psychologists Press., vol.45, no.15, pp. 89-111, 1978.
- 4) X. W. Hou, S. Z. Li, H. J. Zhang, and Q. S. Cheng, “**Direct appearance models,**” in Processing IEEE Conference Computer Vision Pattern Recognition., vol. 1, pp. 828–833,2001.
- 5) Jacob Richard-Whitehill, “**Automatic Real-Time Facial Expression Recognition for Signed Language Translation**”, Department of Computer Science, University of the Western Cape., vol.32, no. 56, pp. 115-167, 2006.
- 6) S. J. McKenna, S. Gong, R. P. Würtz, J. Tanner, and D. Banin, “**Tracking facial feature points with Gabor wavelets and shape models,**” in Processing International Conference Audio- Video-Based Biometric Person Authentication., vol. 1206, pp. 35–42, 1997.
- 7) G. R. S. Murthy, R.S.Jadon “**Effectiveness of Eigen spaces for Facial Expressions Recognition**” International Journal of Computer Theory and Engineering, Vol. 1, no. 5, pp. 1793-8201, 2009.
- 8) Peng Yang, Qingshan Liu, Dimitris N. Metaxas, “**Boosting Coded Dynamic Features for Facial Action Units and Facial Expression Recognition**”, IEEE ., vol.51, no. 42 pp. 110-142, 2007.
- 9) C. Shan, S. Gong, and P. W. McOwan, “**Facial Expression Recognition Based On Local Binary Patterns: A Comprehensive Study,**” Image and Vision Computing., vol.27, no.6, pp. 803-816, 2009.
- 10) M. Valstar and M. Pantic, “**Fully Automated Recognition of the Temporal Phases of Facial Actions,**” IEEE Transactions on Systems Man and Cybernetics-Part B: Cybernetics., vol. 42, no. 1, pp. 28-43, 2012.
- 11) Yan Tong, Yang Wang, Zhiwei Zhu, and Qiang Ji, “**Robust Facial Feature Tracking Under Varying Face Pose and Facial Expression,**” Pattern Recognition., vol.40, no.11, pp. 3195-3208, 2007.
- 12) Z. Zhu, Q. Ji, K. Fujimura, and K. Lee, “**Combining Kalman filtering and mean shift for real time eye tracking under active IR illumination,**” in Processing IEEE International Conference Pattern Recognition., vol. 4 , pp. 318–321, 2002.