Facial Expression Understanding in Image Sequences Using Dynamic and Active Visual Information Fusion

Yongmian Zhang
Dept. of Computer Science
University of Nevada, Reno
Reno, NV 89557, USA

Qiang Ji
Dept. of Electrical, Computer and System Engineering
Rensselaer Polytechnic Institute
Troy, NY 12180, USA

Abstract

This paper explores the use of multisensory information fusion technique with Dynamic Bayesian networks (DBNs) for modeling and understanding the temporal behaviors of facial expressions in image sequences. Our approach to the facial expression understanding lies in a probabilistic framework by integrating the DBNs with the facial action units (AUs) from psychological view. The DBNs provide a coherent and unified hierarchical probabilistic framework to represent spatial and temporal information related to facial expressions, and to actively select the most informative visual cues from the available information to minimize the ambiguity in recognition. The recognition of facial expressions is accomplished by fusing not only from the current visual observations, but also from the previous visual evidences. Consequently, the recognition becomes more robust and accurate through modeling the temporal behavior of facial expressions. Experimental results demonstrate that our approach is more admissible for facial expression analysis in image sequences.

1. Introduction

A facial expression is indeed the human behavior. It often reveals not only the nature of the deformation of facial features, but also the relative timing of facial actions as well as their temporal evolution. It is clearly of interest for human-computer interactions and human behavior analysis that an automated facial expression recognition system can recognize the facial actions, yet modeling their temporal behavior so that various stages of the development of a human emotion can be visually analyzed and dynamically interpreted by machine. More importantly, it is often the temporal changes that provide critical information about what we try to infer and understand human emotions that possibly link to the facial expressions. One interesting aspect of this work is to model such dynamic behaviors and momentary intensities of facial expressions.

1.1. Previous Work

Approaches in facial expression analysis are generally distinguished as spatial analysis and spatio-temporal analysis. The major works on spatial analysis for facial expression recognition have focused on using Neural Networks (NNs). More recently, Gabor wavelet [1], and rule-based [2] have also been attempted. A common limitation of these works is that the recognition bases on static cues from still face images. We focus on how to model the temporal behaviors of facial expressions from dynamic appearances in an image sequence. Consequently, this sets our work significantly apart from the above approaches.

There have been several attempts to track and recognize facial expressions over time [3]-[7]. The current works on the spatio-temporal analysis for facial expression understanding, in our view, have the following shortcomings: **1**) facial motion information is obtained mostly by computing dense flow between successive image frames. Flow estimates are easily disturbed by the variation of lighting and non-rigid motion, and also sensitive to the inaccuracy of image registration and motion discontinuities; **2**) facial temporal information usually takes from three discrete expression states in an expression sequence: the beginning, the peak and the end of the expression. The facial movement itself is not measured. Hence, this can not reflect the temporal evolution and the momentary intensity of an observed expression, which are indeed more informative in human behavior analysis; **3**) besides HMM, most of proposed methods lack the sufficient expressive power to capture the temporal behaviors exhibited by facial expressions. HMM can model uncertainties and time series, on the other hand, it lacks the ability to represent induced and non-transitive dependencies. However, a facial expression is induced not only by its temporal information, but also by a great number of AU combinations and transient cues, which are non-transitive; **4**) the issue of the occlusion in facial expressions has not been directly addressed by the existing works.

1.2. Overview of Our Approach

The following constitutes the framework based on which our approach is developed: *1) Facial Motion Measurement*: the measurement of facial motion is through tracking of facial features by simultaneously using an active Infra-Red (IR) illumination and Kalman Filtering. The pupil positions detected by using IR illumination are used to constrain the detection and tracking of other feature positions so that the facial motion can be accurately measured under large varia-

tions of head motion; *2) Facial Expression Representation*: facial expression representation integrates the DBNs with the facial AUs from psychological views. The DBNs provide a coherent and unified hierarchical probabilistic framework to represent not only the causal relations of facial expressions to the complex combinations of facial AUs, but also temporal behaviors of facial expressions; *3) Facial Expression Recognition*: facial expression recognition lies in a framework of dynamic and active multisensory information fusion. This facilitates dynamically modeling temporal evolution of facial expressions, and increasing the robustness in handling occluded expressions or the uncertainty of feature measurement in dynamic imagery.

2. Facial Feature Tracking

Our approach to facial feature tracking relies on an active IR illumination as shown in Fig. 1(a), so that it can detect pupils under large variation in lighting and head orientations. To assist pupil detection, the person's face is illuminated with an IR illuminator, which produces the dark/bright pupil effect [8], [9], as shown in Fig. 1(b)(c). The detected pupils are subsequently tracked using a technique based on combining Kalman filtering with mean shift tracking. The pupil positions are then used to constrain the possible positions of other facial features. The simultaneous use of IR illumination and Kalman Filtering makes our approach more robust and insensitive to the variations in lighting, head motion, and non-rigid expression change. Details on pupil detection and tracking may be found in [9].

(a)   (b)
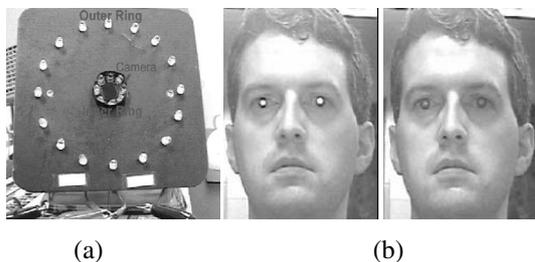
Fig. 1. (a) IR camera setting; (b) dark/bright pupil effect

3. Facial Information Extraction

3.1. Linguistic Descriptions

It is generally believed that the six facial expressions including happiness, sadness, anger, disgust, fear, surprise, can use culture and ethnically independent AUs to linguistically describe them. Such AUs were developed by Ekman and Friesen in their FACS [10]. We adapt the AU-coded descriptions of facial expressions in the FACS to describe the six emotional expressions. To facilitate our introduction to facial expression modeling, we illustrate the facial AUs that are pertinent to the six expressions in Fig. 2, which are directly adapted from [10].

A facial expression is indeed the combination of AUs. We group AUs of facial expressions as primary AUs and auxiliary AUs. By the primary AUs, we mean that the AUs or AU combinations can be clearly classified as or strongly pertinent to one of the six expressions without ambiguities. In contrast, an auxiliary AU is the one that can only be additively combine with primary AUs to provide complementary support to the facial expression classification. As such, a facial expression contains primary AUs and auxiliary AUs. For example, an AU9 can be directly classified as a type of a disgust; while it is ambiguous to classify a single AU17 to be a disgust. When an AU9 and an AU17 appear simultaneously, the facial appearance is more to be the expression of a disgust. Table I gives a summary of primary AUs and auxiliary AUs which are associated with six expression categories. By combining the primary AUs within the same category, the belief of classification to that category increases, as shown in Fig. 3(a). However, the combination of the primary AUs across different categories may result in: 1) a primary AU combination belonging to other categories, e.g., the combination of AU1 (a primary AU of the "sadness") and AU5 (a primary AU of the "surprise") generates a primary AU combination of the "fear" as illustrated in Fig. 3(b); 2) increasing ambiguity, e.g., when AU26 (a primary AU of the "surprise") combines with AU1 (a primary AU of the "sadness"), the belief of "surprise" is reduced and the ambiguity of classification may be increased as shown in Fig. 3(c). These characters are systematically represented by a probabilistic framework presented in Section 4.

(a)   (b)   (c)

Fig. 3. Examples of AU combination; (a) AU12+AU6 enhances the happiness; (b) AU1+AU5 becomes the fear; and (c) AU26+AU1 increases ambiguity between the surprise and fear

3.2. Facial Feature Extraction

In order to automatically extract the activation of the muscles from a face image, we have to quantitatively code AUs into facial feature movements that can be extracted directly from the face image. Fig. 4 presents facial geometrical relationships and furrow regions. In order to adapt the FACS description for machine recognition of facial expressions, there is a need to establish a simple and unique relation between changes of the feature points and the corresponding AUs. Table II gives quantitative description of AUs based

Fig. 2. Face action units for the components of six basic facial expressions (adapted from Ekman and Friesen [10])

Emotional Category	Primary Visual Cue					Auxiliary Visual Cue					
	AU(s)	AU(s)	AU(s)	AU(s)	AU(s)	AU	AU	AU	AU	AU	Transient Feature(s)
Happiness	6	12				25	26	16			wrinkles on eye outer corners; presence of nasolabial furrow
Sadness	1	15	17			4	7	25	26		
Disgust	9	10				17	25	26			
Surprise	5	26	27	1+2							furrows on the forehead
Anger	2	4	7	23	24	17	25	26	16		vertical furrows between brows
Fear	20	1+5	5+7			4	5	7	25	26	

on their linguistic descriptions. Consequently, the facial visual changes are automatically measurable on imagery. Additionally, the symmetric property of human face allows us to generate the redundant visual information for some feature points (e.g., feature points surrounding eyes) which can be used to reduce the information uncertainty possibly resulted from missing feature, inaccurate tracking or partial occlusions. Some feature points, such as eye outer corners, are vulnerable to be occluded by head rotation, and thus, we use more than one way to measure the feature displacements related to these points. Failure of one way is therefore supplemented by another way. Take AU4 for example, we measure not only $\angle HFI$ (see Fig. 4), but also the movement about the vertex of the upper eyelids as shown in Fig. 5.

The regions of facial wrinkles and furrows as shown in Fig. 4 are located by the facial feature positions through feature tracking. The presence of furrows and wrinkles on an observed face image can be determined by edge feature analysis in the areas that transient features appear. Fig. 6 shows the examples of transient feature detection. The contraction and extension of facial muscles may deform the initial nasolabial fold to a particular shape as depicted in Fig. 7(a). After a nasolabial fold is detected, we approximate the shape of a nasolabial fold as a quadratic form by fitting a set of the detected edge points in a least-square sense. The coefficients signifies the curvature of the nasolabial fold exhibited by some facial AUs.

TABLE II

THE DESCRIPTIONS OF FACIAL VISUAL CUES

AUs	Facial Visual Cues	Visual Channel(s)
AU1	$\angle FHJ$, $\overline{JF}$ increased OR $\overline{JF}$ increased, $l8$ non-increased	Brow
AU2	$l8$ increased and $\overline{JF}$ non-increased furrow in $\Box Z$ increased	Brow, Wrinkler
AU4	$l8$, $\overline{FJ}$, $\overline{JJ'}$ decreased, $\angle HFI$ increased and wrinkle in $\Box Y$ increased	Brow, Wrinkler
AU5	$l6$, $\overline{JF}$ and $\overline{JJ'}$ increased	Lid
AU6	nasolabial furrow presence and wrinkle in $\Box V$ increased	Nasolabial, Wrinkler
AU7	$\angle HFI$ non-increased and $\angle HGF$ increased	Lid
AU9	wrinkle increased in $\Box X$ nasolabial furrow presence OR $\overline{PF}$, $\overline{FJ}$ decreased	Wrinkler, Nasolabial
AU10	$l4$ decreased and $|\overline{FC} - \overline{F'C'}|$ increased, nasolabial presence OR $\overline{OD}$ decreased, $\overline{DB}$, $\overline{C'C}$ increased	Lip, Nasolabial
AU12	$\overline{FC}$, $\overline{F'C'}$ decreased, $\overline{CC'}$ increased, $\overline{GI}$ non-increased	LipCorner
AU15	$\overline{FC}$, $\overline{F'C'}$, $\overline{CC'}$ increased	LipCorner
AU16	$\overline{OD}$ non-change, $\overline{DB}$ decreased	Lip
AU17	$\overline{OB}$ decreased and wrinkle in $\Box U$ presence	Chin, Wrinkler
AU20	$\overline{CC'}$ increased and $\overline{FC}$, $\overline{F'C'}$ non-change	LipCorner
AU23	$\overline{DB}$, $\overline{CC'}$ decreased	Lip
AU24	$\overline{DB}$ decreased, $\overline{CC'}$ non-change	Lip
AU25	$\overline{DB}$ increased, $\overline{DB} < T_1$, $\overline{CC'}$ non-increased	Mouth
AU26	$T_1 < \overline{DB} < T_2$, $\overline{CC'}$ non-increased	Mouth
AU27	$\overline{DB} > T_2$, $\overline{CC'}$ non-increased	Mouth

COMPUTER SOCIETY
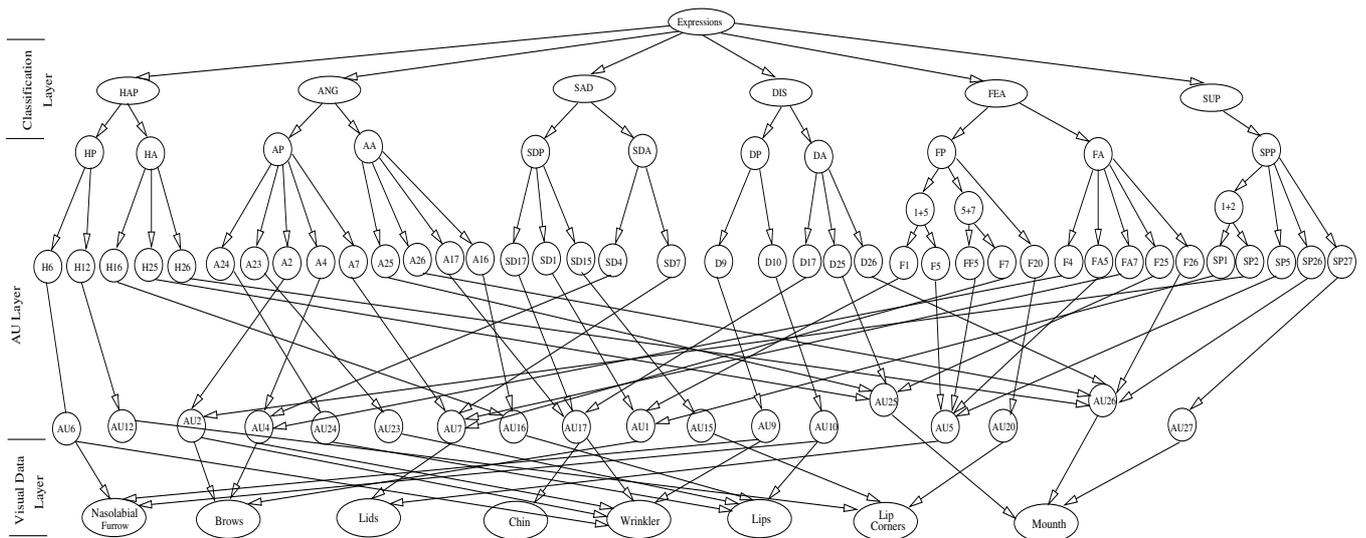
Fig. 8. BN causal model of six basic emotional expressions Note: HAP - Happiness; SAD - Sadness; ANG - Anger; SUP - Surprise; DIS - Disgust; FEA - Fear. Node HP means primary AUs for Happiness; while HA means auxiliary AUs for happiness, and so forth
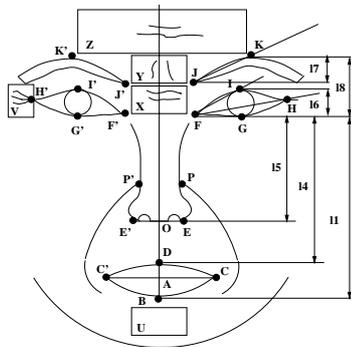
Fig. 4. The geometrical relationships of facial feature points, where rectangles represent the regions of furrows and wrinkles.
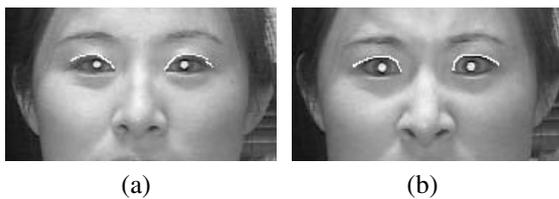
Fig. 5. The vertex of the upper eyelid usually shifts to eye inner corner as a subject performs anger; (a) neutral state; (b) anger

4. Modeling Facial Expressions
4.1. BN Model of Facial Expression

According to the causal relations between AUs and the six expression categories as linguistically described in Table I, we build the BN model as shown in Fig. 8. Our BN model of facial expression consists of three primary layers: classification layer, facial AU layer and sensory information layer. The classification layer consists of a class (hypoth-

Fig. 6. Transient feature detection; (a) horizontal wrinkles between eyes; (b) horizontal wrinkles on the forehead

esis) variable C including six states $c_1, c_2, \cdots, c_6$, which are in the correspondence of $happiness, sadness, disgust, surprise, anger$ and $fear$, and a set of attribute variables denoted as HAP, ANG, SAD, DIS, SUP and FEA. The goal of this level of abstraction is to find the probability of class state c_i, which represents the chance of class state c_i given facial observations. The AU layer is analogous to linguistic description of the relation between AUs and facial expressions as described in Table I. Each expression category consists of primary AUs and auxiliary AUs. A primary AU contributes stronger visual cue to the understanding of the facial expression than an auxiliary AU does. Hence, the likelihood of primary AUs to the facial expression is higher than that of auxiliary AUs. The lowest level of layer in the model is the sensory information layer containing visual information variables, such as *Brows, Lips, Lip Corners, Eyelids, Cheeks, Chin, Mouth, Nasolabial Furrow*, and *Wrinkles*. All variables in this layer are observable.

4.2. Dynamic Modeling

Facial expressions can be said to express emotions and they vary according to subject-environment interaction. As illustrated in Fig. 9, an expression sequence, in many cases, sequentially contains multiple expressions with a different

(a)

(b)
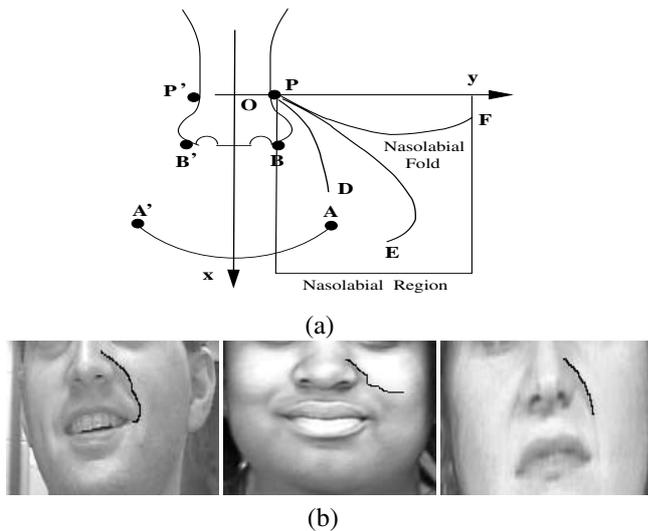
Fig. 7. (a) The possible shapes of a nasolabial fold due to facial activities; (b) the example results of nasolabial fold detection

intensity of facial action due to the change of object in environment or in thought, and evolves over time until the visual information about the underlying the emotional expression can be classified. Modeling such temporal behaviors of facial expressions can lead to better understand the human emotion at each stage of its development.
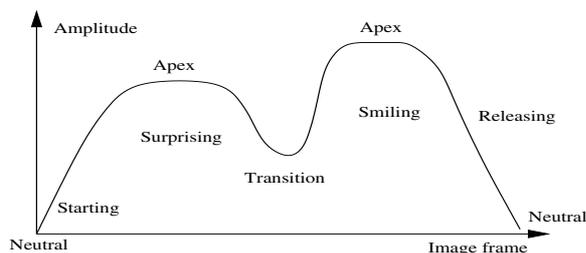
Fig. 9. An illustration that an expression sequence contains two emotional expressions, a surprise followed by a smile

Static BN modeling of facial expression works with visual evidences and beliefs from a single instant in time, and it lacks the ability to express temporal dependencies between the consecutive occurrences in image sequences. To overcome this limitation, we use DBNs for modeling the dynamic aspect of a facial expression. Our DBN for facial expressions is made up of interconnected time slices of SBNs exactly described above, and the relationships between two neighboring time slices are modeled by the first order Hidden Markov model. The relative timing of facial actions during the emotional evolution is described by moving a time frame in accordance with the frame motion of a video sequence, so that the visual information at the previous time provides diagnostic support for current visual evidences. Eventually, the belief of the current hypothesis

of expression is inferred relying on the combined information of current visual cues through causal dependencies in the current time slice, as well as the preceding evidences through temporal dependencies. Fig. 10 shows the temporal dependencies by linking the top level of SBN in Fig. 8. Consequently, the visual evidences from the preceding time
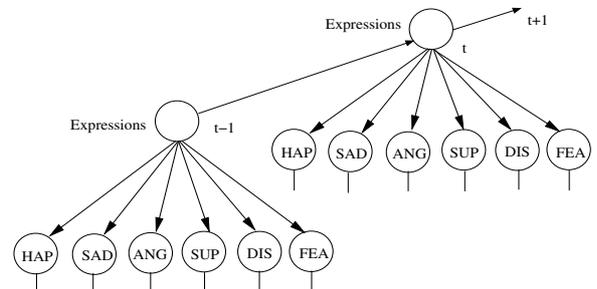
Fig. 10. The temporal links of DBN for modeling facial expression (two time slices are shown). Node notations are given in Fig. 8

slice serves as new prior information, and the prior information is integrated with current data to produce a posterior estimate of current facial expression.

4.3. Active Fusion of Visual Cues

A facial expression involves simultaneous changes of facial features on multiple feature regions. Some of such facial deformations extracted from face images possibly result from the errors in facial feature detection and tracking due to the limitation of tracking accuracy. For facial activities at a given time, there is a subset of visual information that is the most informative for the current goal to reduce the ambiguity of classification the most. If we can actively and purposively choose such visual cues for fusion, we can achieve a desirable result in a timely manner while reducing the ambiguity of classification to a minimum.

From the multisensory information fusion point of view, we have n hypothesis of emotional states, $H = \{h_1, \cdots, h_n\}$. The visual observations, $\mathbf{E} = \{E_1, \cdots, E_m\}$, which is obtained from m diverse visual sources, forms an information vector. The information fusion is to estimate a posterior probability that $H = h_i$ is true given $\mathbf{E}$, i.e., $P(H = h_i | \mathbf{E})$. By the most informative sources, we mean the sensory data from a subset, $E \subset \mathbf{E}$, that after integrating with the existing data, can maximize the certainty of H, given E, i.e., it can lead the probability of hypothesis, $P(H = h_i | E \subset \mathbf{E})$, closest to either 1 or 0, and the ambiguity of the hypothesis can be reduced to a minimum. In our approach, a fusion system is cast in a DBN framework as shown in Fig. 11. The DBNs provide a dynamic knowledge representation and control structure that allows sensory information to be combined according to the rules of probability theory. The active sensor controller (see Fig. 11) serves as sensor cueing that allows the

recognition system to actively select a subset of facial features to produce the visual information that is the most relevant to the current emotional state. Which subset of facial feature regions needs to be sensed is determined by evaluating the uncertainty reducing potential of possible consequences resulted by sensing various facial feature regions (visual channels).
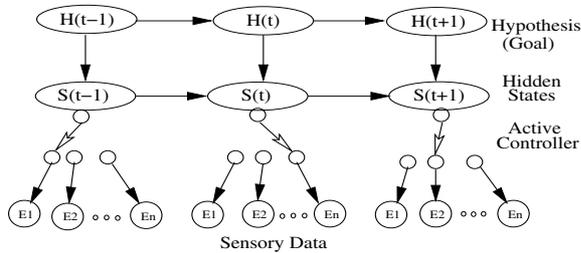
Fig. 11. A conceptual framework of DBN based active information fusion. The system consists of a Goal, Hidden States, an Active Sensor Controller, and Sensory Data

The uncertainty reducing potential is formulated in the framework of mutual information theory. Extending Shannon's measure of entropy, we can obtain the uncertainty reducing potential $I(H|E_1, E_2)$ if two sensors, E_1 and E_2, are fused, i.e.,

$$
\begin{aligned}
I(H|E_1, E_2) &= ENT(H) - ENT(H|E_1, E_2) \\
&= -\sum_{h \in H} p(h) \log(p(h)) + \sum_{h \in H} \sum_{e_1 \in E_1} \sum_{e_2 \in E_2} \Big\{ \\
&\quad p(e_1, e_2) p(h|e_1, e_2) \log p(h|e_1, e_2) \Big\},
\end{aligned}
\quad (1)
$$

where h and e_i are the possible outcomes of H and E_i; $p(h|e_1, e_2)$ is estimated by applying Bayes' theorem under the assumption that sensor $\{E_1, \cdots, E_n\}$ are conditionally independent each other:

$$
p(h|e_1, e_2) = \frac{p(h|e_1) p(h|e_2)}{p(h) \sum_{h \in H} \frac{p(h|e_1) p(h|e_2)}{p(h)}},
\quad (2)
$$

In the above equations, $p(h|e_1)$ and $p(h|e_2)$ are directly obtained by using BN probability inference. The prior probability of hypothesis $p(h)$ is revised based on evidential observations from the preceding time slice through the state evolution probability. Eq. (1), (2) can be extended for applying on fusing any number of sensors. In the interest of space, we will describe this in the publications elsewhere.

5. Experimental Results

In the first experiment, we create a short image sequence in which a subject performs a smile followed by a surprise, as shown in Fig. 12(a). We can see visually that, the temporally resolution of the expression varies over time as it seems to be in spontaneous interaction. Fig. 12(b) provides

the analysis result by using our facial expression model. This plot naturally profiles the momentary emotional intensity and the dynamic behavior of facial expression that the magnitude of the facial expression gradually evolves over time. Such a dynamic aspect of facial expression modeling can more realistically reflect the evolution of a spontaneous expression starting from a neutral state to the apex, and then gradually releasing. Since there are inter-personal variations with respect to the amplitudes of facial actions, it is often difficult to determine the absolute emotional intensity of a given subject through machine extraction. In this approach, as we can observe from the result, the relative change of the emotional magnitude can be well modeled at each stage of the emotional development.

(a)

(b)
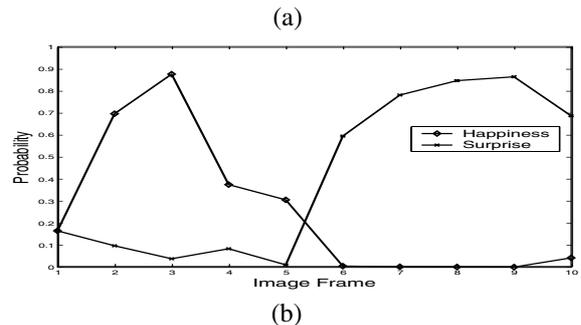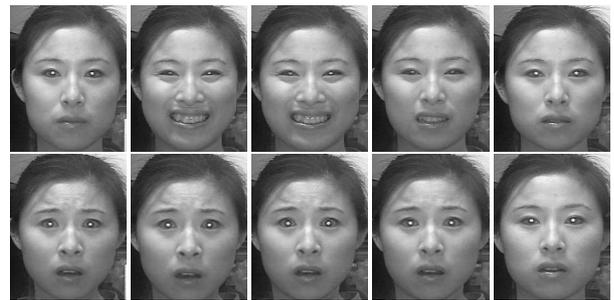
Fig. 12. (a) An image sequence shows a subject performing smiling followed by surprising; (b) the probability distributions of facial expressions, where only the distributions of "happiness" and "surprise" are shown

The benefits of our approach can be best shown when an image sequence presents the facial features which are in missed-detection or missed-tracking due to occlusions and image noises. Fig. 13(a) gives such an image sequence in which the facial features in some frames are assumed to be in fully missed-tracking. Fig. 13(b) depicts the result by our facial expression modeling, in which we can readily observe that, though the image sequence has facial features fully mis-tracked in some of frames, the facial expression can still be assessed correctly by reasoning over time. The inability of current facial expression recognition systems to correlate and reason about facial temporal information over time is an impediment to provide a coherent overview of the dynamic behavior of facial expressions in an image sequence since it is often the temporal changes that provide

critical information about what we try to infer and understand the human emotional expressions. By integrating the current multiple visual cues and the preceding evidences, our approach tends to be more robust in handling partial occluded facial expressions. An example is given in Fig. 14.

(a)

(b)
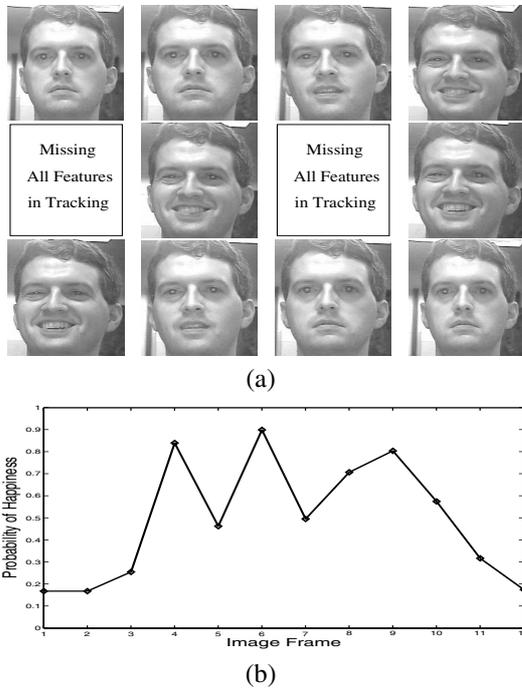
Fig. 13. (a) An image sequence with some image frames in fully missed-tracking; (b) result from our facial expression model

As we have previously remarked, the facial deformations extracted from images may involve the feature tracking errors. If we fuse all current available visual cues at a time, the feature changes caused by tracking errors are also taken into account, and consequently, cause more ambiguities in classification. Additionally, if we go over all visual channels for the information of facial feature changes each time, this will cost unnecessary computations. We therefore choose actively and purposively the visual cues for fusion. In this example as given in Fig. 15(a), we let the maximum of two visual channels plus the visual evidence at the preceding time to be integrated at a time. The comparative result given in Fig. 15(b) shows that, in terms of uncertainty reduction, the active fusion is better than the passive fusion (fusing all available visual cues at a time for this case).

Finally, we present an example of facial expression understanding in a live image sequence. The original image sequence of facial expressions has 600 frames containing the six emotional expressions plus the neutral state among them. We provide selected images that will, hopefully, convey our results. Fig. 16 provides such a facial expression sequence of only 60 images resulted from sampling the origi-

(a)

(b)
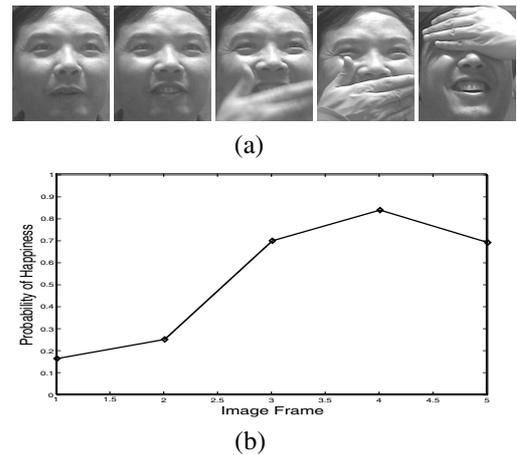
Fig. 14. (a) A posed image sequence performing occluded facial expression; (b) result from our facial expression model

(a)

(b)
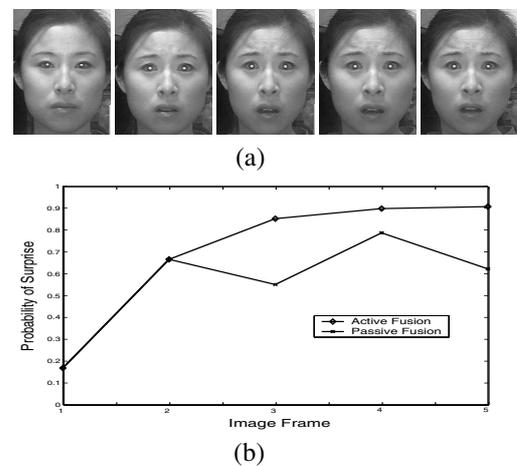
Fig. 15. (a) A segment of an image sequence; (b) comparative result between active and purposive fusion of visual cues and passive fusion of visual cues (fusing all available visual cues)

nal sequence in every 10 frames. Fig. 17 visually plots the probability distributions over six facial expressions for the image sequence as given in Fig. 16. The results consistently confirm that the dynamic aspect of our approach can lead to be more robust for facial expression analysis particularly in image sequences.

6. Conclusion

Compared with the existing works on facial expression analysis, our approach enjoys several favorable properties: *First*, our approach has expressive power to capture the dependencies, uncertainties and temporal behaviors exhibited by facial expressions, so that dynamic behaviors of facial expressions can be well modeled. *Second*, our approach allows the recognition system to actively select a subset of the most relevant facial visual cues at a current time to correlate previous visual evidences, so that the uncertainty of
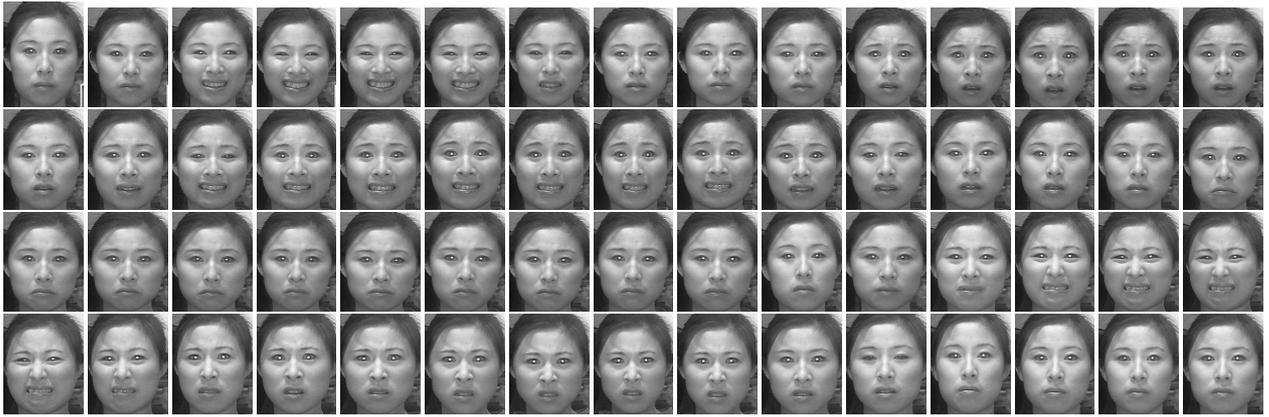
Fig. 16. An image sequence sampled from IR camera (from the left to the right and the top to the bottom)
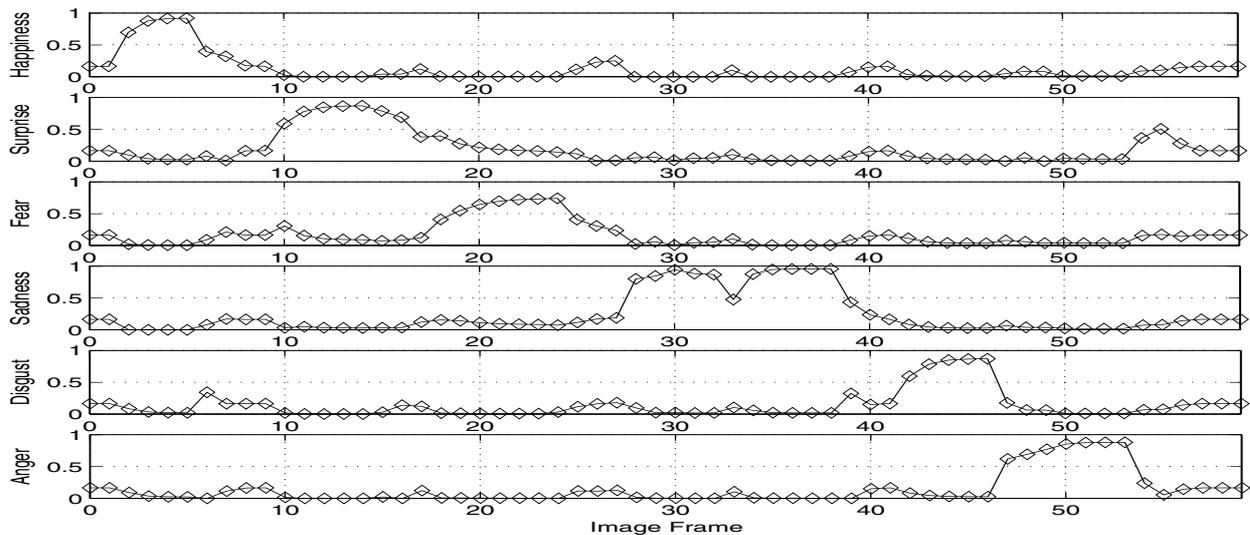
Fig. 17. Classification results by our facial expression model. The image sequence is given in Fig. 16

classification can be reduced to a minimum at each time. *Third*, taking advantages of DBNs and multisensory information fusion, our approach is more robust in facial expression analysis and facilities for handling occluded facial expressions. *Fourth*, our approach allows an image sequence have multiple expressions, and two consecutive expressions do not require to be temporally segmented by a neutral state, so that the facial expression analysis becomes more flexible in image sequences.

References

[1] M. J. Lyons, J. Budynek, and S. Akamatsu, "Automatic classification of single facial images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 21, no. 12, pp. 1357–1362, 1999.

[2] M. Pantic and L. Rothkrantz, "Expert system for automatic analysis of facial expression," *J. Image and Vision Computing*, vol. 18, no. 11, pp. 881–905, 2000.

[3] M. Rosenblum, Y. Yacoob, and L. S. Davis, "Human expression recognition from motion using a radial basis function network architecture," *IEEE Trans. NNs*, vol. 7, no. 5, pp. 1121–1137, 1996.

[4] M. J. Black and Y. Yacoob, "Recognizing facial expression in image sequences using local parameterized models of image motion," *Int'l J. Computer Vision*, vol. 25, no. 1, pp. 23–48, 1997.

[5] I. A. Essa and A. P. Pentland, "Coding, analysis, interpretation, and recognition of facial expressions," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, no. 7, pp. 757–763, 1997.

[6] N. Oliver, A. Pentland, and F. Bérard, "LAFTER: Lips and face real time tracker with facial expression recognition," in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, 1997.

[7] Y. Tian, T. Kanade, and J. F. Cohn, "Recognizing action units for facial expression analysis," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 23, no. 2, pp. 97–115, 2001.

[8] C. Morimoto, D. Koons, A. Amir, and M. Flicker, "Framerate pupil detector and gaze tracker," in *Proc. of the IEEE ICCV99 Frame-Rate Workshop*, (Kerkyra, Greece), Sep. 1999.

[9] 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 *Proc. Int'l Conf. Pattern Recognition*, Aug. 2002.

[10] P. Ekman and W. V. Friesen, *Facial Action Coding System (FACS): Manual*. Palo Alto, CA: Consulting Psychologists Press, 1978.