Evaluation of Spatio-Temporal Regional Features For 3D Face Analysis

Yi Sun and Lijun Yin
Department of Computer Science
State University of New York at Binghamton, Binghamton, NY.

Abstract

3D facial representations have been widely used for face recognition and facial expression recognition. Both local and global features can be extracted from either static or dynamic models in both spatial and temporal domains. 3D local features are referred to the features in regional facial areas while 3D global features are referred to the features from the entire facial region. In this paper, we address the issue of performance assessment of facial analysis in terms of global features versus local features, static models versus dynamic models, and spatial domain versus temporal domain. Based on the existing work of using 3D spatio-temporal HMM for facial analysis, we propose to extend it to a local-temporal HMM in order to provide an explicit comparison of global features and local features. A dynamic 3D facial expression database and a static facial expression database are used for experiments. The performance for six prototypic facial expression classification and face identification is analyzed and reported.

1. Introduction

The development of 3D imaging technologies allows the facial behavior analysis to be advanced from the traditional two-dimensional domain to the three-dimensional domain [3, 17, 19, 10]. There are some successful approaches using 3D models for face recognition [6, 13, 2, 4, 12, 16, 23, 18, 7, 5, 24, 20] and facial expression recognition [28, 25, 8, 22, 21]. However, unlike the well-studied 2D-based face analysis [11, 29, 15, 9, 1], there are few studies to evaluate the 3D face analysis systematically with respect to various feature representations (e.g., global versus local, static versus dynamic, and spatial versus temporal.) This paper intends to address this issue based on the 3D static/dynamic facial expression databases [27, 26], with the goal of improving various facial analysis performances, including facial expression classification and face identification. This paper focuses on the performance evaluation based on following observations or arguments:

(1) *3D static versus 3D dynamic*: The face models can be captured by a single frame (i.e., static) or a sequential frames (i.e., dynamic). Since the 3D facial surface is changed along with the expression change, the facial behavior can be better characterized in a 3D dynamic domain.

(2) *3D global versus 3D local*: The global features are selected from an entire facial surface, which is a good reflection of an individual facial identity. However, local features from local expressive regions could be more informative for local facial actions, leading to a more effective identification of facial expressions.

(3) *3D spatial versus 3D temporal*: Temporal dynamics of a face refers to the change of the facial surface along with time, while spatial dynamics refers to the relationship of different facial regions due to the action of facial expressions. Temporal dynamics of a 3D face could follow a certain pattern related to a certain expression, which is crucial for 3D facial expression classification.

Given the above observation and motivated by the existing work using 3D surface primitive labeling features and spatio-temporal Hidden Markov Model (HMM) for facial expression recognition [21, 20], we propose a local temporal HMM for 3D dynamic model analysis. We will further investigate the performance of facial expression recognition and face identification by comparing the results using various 3D feature representations: global vs. local, spatial vs. temporal, and static vs. dynamic.

Section 2 gives an overview of the existing system for face model processing and analysis. Section 3 describes the 3D local region-based temporal HMM model and its combination with a spatial HMM model. Experimental results and evaluations of 3D facial features under various global/local and spatio/temporal domains will be reported in Section 4, followed by a final discussion and conclusion in Section 5.

2. Overview Of Existing 3D Dynamic Facial Analysis System

The framework of the existing 3D dynamic face analysis system has been developed in our previous work [20, 21]. It includes face model pre-processing, 3D face sequence learning and recognition. Here is a brief overview: facial surface features are represented by a primitive label map.

978-1-4244-3993-5/09/$25.00 ©2009 IEEE

A geometric surface labeling approach is used to represent each vertex on a face model. Given 3D face model sequences, a generic model is used to adapt each frame of the 3D model sequence. To perform the model adaptation, a set of predefined 83 control points are defined and tracked. As a result, each range model in the sequence can be represented by a feature vector composed of all the vertex labels in the face region. Due to the high dimensionality of the feature vector and in order to better separate different expressions, we use Linear Discriminative Analysis (LDA) approach to project the labeled face models into an optimized feature space (denoted as OPT). Finally, the HMM model is trained for classifying six prototypic facial expressions.

3. Spatio-Temporal HMMs and Their Variation

A 3D facial sequence provides dynamics in both single frame (across a face model surface) and multiple frames (across time). The spatial configuration provides dynamic information of different regions of a face during expressions. It is interesting to see that given a certain expression, how one part of a facial surface could change along with the other parts of the facial surface, and how they encode the spatial relationship and function as a whole. The temporal change of a facial surface provides dynamic information of a facial surface varied along with time. It is also interesting to see how much variations could be across frames with respect to different subjects. Another issue to be investigated is whether the local features or the global features provide more crucial cues for face behavior classification.

To study these performances, we designed five HMM representations with different variations, namely Global Temporal HMM, Local Spatial HMM, Local Temporal HMM, Global Combined HMM, and Local Combined HMM.

3.1. Global Temporal HMM (T_G-HMM)

The global feature based temporal HMM takes each frame (whole facial region) of a face sequence as an observation and explores the temporal dynamics of the entire facial surface along the time axis. As described in the previous section, given a sequence of facial models, we transform each frame (one 3D face model) into an optimized feature space (OPT). Then we apply the training process to learn basic HMMs on the OPT space in order to distinguish facial expressions (or face identities.) Since this method tracks the temporal changes of an entire face, we call it Global Temporal HMMs (T_G-HMM) (see Figure 1). The decision made by this method is denoted as $Decision^{T_G}$ [21].

During the training stage, the statistical information and the temporal dynamics of the trained data are learned by HMMs [14]. For facial expression classification, *each prototypic expression* is modeled as an N-state HMM, e.g., $N = 6$. For face identification, *each subject* is modeled as an HMM.

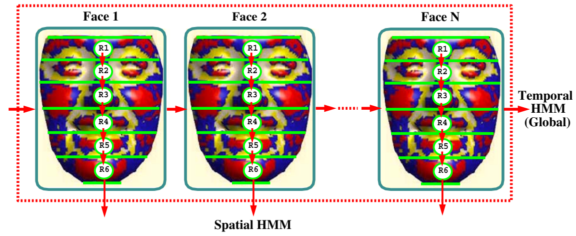

Figure 1. Structure of the *global* 3D dynamic HMM, which is applied to T_G-HMM, S_L-HMM, and C_G-HMM.

3.2. Local Spatial HMM (S_L-HMM)

Considering the facial configuration, different regions of a face perform correlatively with certain relations during an expression action. Unlike the global temporal HMM method, we segment the facial surface into six regions from top to bottom [20] ($R1 - R6$ as shown in Figure 1) for each face model based on the tracked 83 feature points. We then transform each region of a face from the labeled surface space to an optimized compact space separately. Therefore, we derive six optimized feature vectors. If we take the six regions as six states of an HMM, every single facial frame model has a sequence of 6-states with an observation of the optimized feature vectors (shown as a vertical line of each facial model in Figure 1.) In the case of facial expression classification, for example, there will be an output as a decision for a certain expression for each single frame. Since each 3D frame model will get their likelihood for one distinct expression, we use the Bayesian decision rule to decide which one of the six prototypic expressions that the single frame belongs to.

Since we take N frames as a subsequence, a majority voting strategy is applied on all N frames after the classification of each frame. The subsequence is classified to an expression if the majority of the N frames are classified to this expression. Because this method tracks the spatial dynamics of a 3D face model, we call it Local Spatial HMM (S_L-HMM), and the decision made by this method is denoted as $Decision^{S_L}$.

3.3. Local Temporal HMM (T_L-HMM)

Different regions of a face may perform differently with different expressions. For example, when smiling, mouth and cheek regions usually show larger movements than the other regions. Given this observation, we partition a facial surface into six regions $R1$, $R2$, ..., $R6$, including eyes, eye-

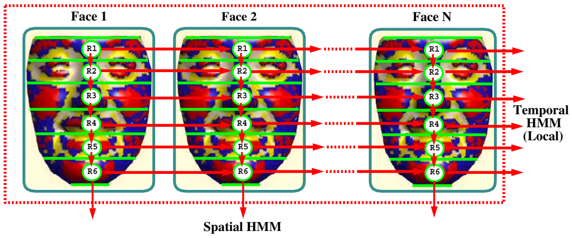

Figure 2. Structure of the *local* 3D dynamic HMM, which is applied to T_L-HMM and C_L-HMM.

brows, noses, and mouth, which are the same segmentations as used in the spatial HMM method. When observing the state change of a local region across a sequence, we are able to use the facial features of the local region to train a temporal HMM. In other words, each local region of a facial surface will learn an HMM for each distinct expression separately (as illustrated in Figure 2). Given six local regions and six prototypic facial expressions, a total of 36 T-HMMs are established for an entire facial surface.

Since the features extracted from the six local regions could generate six different classification results, we take the majority voting strategy to determine the expression type of the subsequence. If more than two regions are classified as a same expression, such expression is taken as the recognized expression for this subsequence. If there is no majority expression to be recognized among the six regions ($R_1, ..., R_6$), the expression with the maximum likelihood (probability) of the region will be chosen as the recognized expression of this subsequence. This procedure is formulated as the following equation:

$$R_c = \arg\max_{R_k}[\frac{P(\lambda_{c*}^{k}|O^{R_k})}{\sum_{i=1}^{C} P(\lambda_i|O^{R_k})}], k = 1, 2, ..., 6 \quad (1)$$

where λ_{c*}^{k} is the expression type determined by the region R_k. As a result, the expression of the region R_c with the maximum likelihood is selected as the recognized expression of the subsequence.

Since the regional features of a facial surface are used to learn their temporal changes, and the classified expression is determined by either the majority voting or the maximum probability of observations of local regions, we call the HMM a Local Temporal HMM (T_L-HMM). The decision resulted from this method is denoted as $Decision^{T_L}$.

3.4. Global Combined-HMM (C_G-HMM)

To take advantage of both spatial dynamics and temporal dynamics of facial surface movements, we combine the spatial HMM and the temporal HMM to construct the pseudo-2D HMM, called Global combined HMM (C_G-HMM) as illustrated in Figure 1.

We take the results of the global T_L-HMM ($Decision^{T_G}$) and the local S-HMM ($Decision^{S_L}$), and make a final decision ($Decision^{C_G}$) by fusing the two outputs. Here is the procedure to derive $Decision^{C_G}$ of the combined HMM:

IF $Decision^{T_G} == Decision^{S_L}$
  $Decision^{C_G} = Decision^{T_G}$
ELSE
  IF $Confidence^{S_L}$ is more than a threshold
    $Decision^{C_G} = Decision^{S_L}$
  ELSE
    $Decision^{C_G} = Decision^{T_G}$
END

We define $Confidence^{S_L}$ as the ratio of the majority votes versus the number of frames of the query model subsequence. In our experiment, we take six frames as a subsequence which corresponds to 6 states of the T_G-HMM, and we take an empirical value 0.67 as the threshold. Till this end, if the majority of frames (4 or more) of a query sequence are recognized as an expression A by the S_L-HMM, the query subsequence is classified as the expression A. Otherwise, the expression is classified to the $Decision^{T_G}$. Essentially, the combined HMM uses the learned facial temporal characteristics to compensate for the learned facial spatial characteristics.

3.5. Local Combined-HMM (C_L-HMM)

Similar to the combined global spatio-temporal HMM as defined above (C_G-HMM), we construct a local combined HMM as illustrated in Figure 2. Rather than using T_G-HMM, we use the local temporal HMM (T_L-HMM) to integrate with the S_L-HMM, resulting a combined local spatial-temporal HMM, as denoted C_L-HMM. The final decision is based on the local temporal HMM and the local spatial HMM. Here is the procedure for $Decision^{C_L}$ of the combined HMM:

IF $Decision^{T_L} == Decision^{S_L}$
  $Decision^{C_G} = Decision^{T_L}$
ELSE
  IF $Confidence^{S_L}$ is more than a threshold
    $Decision^{C_G} = Decision^{S_L}$
  ELSE
    $Decision^{C_G} = Decision^{T_L}$
END

Similarly, the same $Confidence^{S_L}$ as defined in the previous subsection is used for the condition check. This local combined-HMM integrates the local temporal dynamics and the local spatial dynamics to improve the 3D facial sequence classification.

4. Experiment and Evaluation

4.1. Methods and Data

We take our 3D dynamic facial expression database (4DFE) [26] for experiments.

(1) *Person-independent*: We conducted person-independent experiments on 60 subjects selected from the database. To construct the training set and the testing set, we generated a set of 6-frame subsequences from each expression sequence. To do so, for each sequence of a subject, we chose the first six frames as the first subsequence. Then, we chose 6 consecutive frames starting from the second frame as the second subsequence. The process is repeated by shifting the starting index of the sequence every one frame till the end of the sequence. The rationale for this shifting is that a subject could come to the recognition system at any time, and thus the recognition process could start from any frame. As a result, $30780 (= 95 \times 6 \times 54)$ subsequences of 54 subjects were derived for training, and $3420 (= 95 \times 6)$ subsequences of the other 6 subjects were derived for testing. Following a ten-fold cross-validation procedure, we randomly select 54 subjects for training and the other 6 subjects for testing for ten trials. The average recognition rate of the ten trials is reported as the final result.

(2) *Person-dependent*: We conducted person-dependent experiments on the 60 subjects as well. We split each of the six video sequences into two parts: training set and test set. All six expressions of all subjects in the training set are used to learn the six expressional HMM models. To do so, we randomly select 90% of the 6-frame subsequences from all expression sequences of all subjects to construct a training set, and the remaining 10% of the subsequences to construct a testing set. Following the ten-fold cross-validation procedure, the experiment was conducted ten times with different training sets and testing sets.

For comparison, we also conducted the experiment on a 3D static facial expression database (3DFE) [27]. The 3DFE database contains 100 subjects with six prototypic facial expressions, and each expression has four static 3D models. This data set is only used to test the spatial HMM due to the lack of the temporal information.

In the following, we report the experimental results and their performances for the five types of HMMs respectively.

4.2. Performance of Facial Expression Classification

(1) T_G***-HMM***

We applied the global facial surface label map of the entire facial surface as the feature vector for temporal HMM training and testing. Prior to input, we applied LDA to transform the global feature vectors to the compact optimal feature vectors. Table 1 reports the recognition rates of six prototypic facial expression classification in both person-dependent case and person-independent case. As the results demonstrated, the global temporal HMM is able to learn and trace the temporal changes of the entire facial surface features. Unlike the person-dependent case, the person-independent case guarantees that the testing data has never appeared in the training set. Therefore, the recognition rate for the person-independent case is much lower (about 10%) than the person-dependent case.

Method	P-independent	P-dependent
T_G-HMM	79.81%	89.29%

Table 1. Expression recognition results using T_G-HMM with global 3D label features in both person-dependent case (*P-dependent*) and person-independent case (*P-independent*).

Methods	P-independent	P-dependent
S_L-HMM (1-frame)	57.07%	72.96%
S_L-HMM (6-frames)	70.08%	79.77%

Table 2. Facial expression recognition result using the spatial HMM with local label features with either single-frame or six-frames subsequences.

(2) S_L***-HMM***

Since the spatial dynamics within each frame model provides important cues for distinguishing various expressions, we split the facial surface to six local regions ($R1, ..., R6$). The six regions construct a local spatial domain for a 6-state HMM. Similar to the global feature label map, we construct the local feature label map for each region individually. Then, a LDA transformation is applied to transform the local feature vectors to compact optimal feature vectors for the six regions separately.

Two experiments have been conducted using the 6-state spatial HMM. First, we apply the spatial HMM to every single frame for expression classification. Second, the spatial HMM is applied to every six-frame subsequence, and a majority voting is carried out for the expression classification. The results are reported in Table 2 for both person-independent and person-dependent cases.

As compared to the results from the T_G-HMM method, the performance of the S_L-HMM method is significantly degraded due to the lack of temporal dynamic information of facial expressions. Table 2 shows that the 6-frame majority voting method can significantly improve the recognition performance against the single-frame based method for both person-dependent and person-independent cases.

(3) T_L***-HMM***

In contrast to the global temporal HMM, the local temporal HMM is based on the local regional features. To study whether the local features could outperform the global features, we build the temporal 6-state HMM for each of the lo-

Regions	P-independent	P-dependent
R1	74.85%	81.80%
R2	78.47%	82.39%
R3	60.12%	67.86%
R4	71.88%	78.67%
R5	69.35%	76.59%
R6	65.72%	78.08%
Combined (R1-R6) T_L-HMM	83.18%	89.94%

Table 3. Expression recognition rates using the temporal HMM with local surface label features on each local region individually. The last line shows the result of the combined local-region based approach (T_L-HMM).

Methods	P-independent	P-dependent
C_G-HMM	82.19%	91.02%
C_L-HMM	86.41%	91.52%

Table 4. Expression recognition results using C_G-HMM and C_L-HMM.

cal regions ($R1, R2, ..., R6$) separately. Each subsequence contains 6 frames. The results of using the temporal HMM on individual regions are reported in Table 3. To study the performance of the combined individual regions, we take a majority voting approach as described in the previous section to generate a new result.

As we can see, although the local regions cover the significant expressive areas, a single stand-alone expressive region does not perform well to capture the difference between six prototypic expressions. Nevertheless, the integration of the classifications of local regions generates a much better result due to the account of relationship of regional expression information. Moreover, it has also shown that the combined local-region based approach (T_L-HMM) is better than the global feature based approach (T_G-HMM), especially for the person-independent case (Table 1).

In short, the above experiments show that the combined local surface regions (T_L-HMM) are better than the global surface region T_G-HMM in terms of facial feature representation and expression characterization. However, the global facial surface is better than any single local region.

(4)* C_G*-HMM

To combine the information learned from both global temporal dynamics and spatial dynamics, we implemented the C_G-HMM to classify the six prototypic expressions. As the results shown in Table 4, the recognition performances for both person-dependent and person-independent cases have been improved. This is mainly due to the fact that temporal changes of the global features can be compensated by the local spatial feature dynamics, and vise versa.

(5)* C_L*-HMM

Methods (3D Static)	P-independent	P-dependent
Global (PSFD)	64.12%	79.07%
Local (S_L-HMM)	67.34%	81.87%

Table 5. Summary of the experiments on 3D static database (3DFE).

Similar to the study of the combined global temporal-spatial HMM, we combine the information learned from both local temporal dynamics and spatial dynamics, and implemented the C_L-HMM to classify the six prototypic expressions. Table 4 shows that the C_L-HMM has achieved the best recognition performance among those experiments both person-independently and person-dependently. The result also shows that the fusion of local temporal dynamics and local spatial dynamics can perform better than using the individual regional features or the global features alone.

(6) 3D Surface Label Histogram

In order to compare the HMM based approaches with other approaches, we also implemented an approach using a global surface label histogram (called primitive surface feature distribution (PSFD) developed by Wang *et al.* [22]) to classify static facial expression models. We take every single frame of the 4DFE database as a static facial model, and applied same procedure as described in [22] to classify six expressions using the LDA classifier. The result is reported in Table 6 (the 1st row).

(7) Performance on 3D Static Facial Expression Database

In comparison, we applied the second database (3DFE [27]) for test. Due to the lack of temporal information of the 3D static models, we only use two approaches for comparison: Global PSFD approach and single-frame based spatial HMM approach. The performance is reported in Table 5. The single-frame based spatial S_L-HMM outperforms the single-frame PSFD based approach. However, both the static model based single-frame approaches (Table 5) can not compete with the 3D dynamic based multiple-frame approaches (Row 3 - Row 7 of Table 6).

In summary, the experiments on 3D dynamic facial expression database (4DFE) and 3D static facial expression database (3DFE) are reported in Table 6 and Table 5. As we can see, among all the experiments for 3D facial expression classification, the combined local temporal HMM (C_L-HMM) achieved the best performance. The results suggest that the fusion of local temporal dynamics and local spatial dynamics can significantly improve the performance of using either single frame models or regional/global features alone.

Methods (3D Dynamic)	P-independent	P-dep.
Global (1-frame PSFD)	53.24%	67.38%
Local (1-frame S_L-HMM)	57.07%	72.96%
Global (6-frame T_G-HMM)	79.18%	89.29%
Local (6-frame S_L-HMM)	70.08%	79.77%
Local (6-frame T_L-HMM)	83.18%	89.94%
Global (6-frame C_G-HMM)	82.19%	91.02%
Local (6-frame C_L-HMM)	86.41%	91.52%

Table 6. Summary of the experiments on 3D dynamic database (4DFE).

Methods	E-independent	E-dependent
Global (T_G-HMM)	91.14%	96.89%
Local (T_L-HMM)	78.39%	95.54%
Global (C_G-HMM)	95.65%	98.61%
Local (C_L-HMM)	92.49%	97.57%

Table 7. 3D Dynamic face recognition using Global temporal HMM (T_G-HMM), Local temporal HMM (T_L-HMM), Global combined HMM (C_G-HMM), and Local combined HMM (C_L-HMM) for both expression-independent and expression-dependent cases.

4.3. Performance of Face Identification on 3D Dynamic Data

To evaluate the 3D spatio-temporal features (either global or local) for the face recognition task, we implemented the approach developed by Sun *et al.* [20], and applied the procedures as described in Section 3.1 - Section 3.5 for face classification using T_G-HMM, T_L-HMM, C_G-HMM, and C_L-HMM respectively. Unlike the facial expression classification where each expression is modeled as an HMM, the face classification requires each subject to be modeled as an HMM. We use a 12-state temporal HMM and a 6-state spatial HMM for each subject for training and testing. Same as the previous experiments, the 3D dynamic facial expression database (4DFE) [26] is employed for performance evaluation. We performed two types of experiments: expression-independent face recognition and expression-dependent face recognition. Table 7 shows the recognition results.

In contrast to the 3D facial expression recognition, the 3D face recognition shows that the global temporal HMM (T_G-HMM) is superior to the local temporal HMM (T_L-HMM). The combined global temporal HMM achieves the best performance in face classification.

This preliminary finding may suggest that the global feature representation could be advantageous over the local feature representation for 3D dynamic face classification, while the local features and their combinations could be more beneficial to the 3D dynamic facial expression classification.

5. Concluding Remarks

In this paper, we extend the existing work to evaluate the performance of 3D face model analysis in terms of spatial dynamics vs. temporal dynamics, global features vs. local features, and static models vs. dynamic models. The results show that distinguishing facial expressions and distinguishing face identities may require different preferences for facial feature representations (e.g., global/local and spatial/temporal). The preliminary experiments from the 3DFE and 4DFE databases show that in general, combined local temporal HMM (C_L-HMM) achieved the best performance for 3D facial expression classification, while the combined global temporal HMM (C_G-HMM) achieved the best performance for face classification. In other words, facial expressions are exhibited more likely by regional surface changes, while faces are identified more likely by an entire facial surface as a whole. Both tasks are benefited from the temporal dynamics of 3D model sequences.

The current work is only based on the HMM-based approaches and six predefined local regions with surface labels. Our future work is to verify the finding using different feature representations and a variety of other classification approaches. For example, in addition to the regional surface features, we will study the local action unit features. Another important extension of this work is to investigate a more effective approach for the facial region partition using anatomical knowledge or related psychological findings. We will further study the configuration of different expressive regions and their collaborative functions. For example, we may reorder the $R1 - R6$ in the HMM model to further study the relationship of the HMM states with respect to different facial behaviors. Finally, the integration of 3D dynamic models and 2D dynamic textures will also be one of our future work.

6. Acknowledgement

This work is based on the materials supported by the National Science Foundation under grants IIS-0541044, IIS-0414029, and the NYSTAR's James D. Watson Investigator Program.

References

[1] M. Bartlett, G. Littlewort, I. Fasel, J. Chenu, and J. Movellan. Fully automatic facial action recognition in spontaneous behavior. In *International Conference on Automatic Face and Gesture Recognition*, 2006. 1

[2] V. Blanz and T. Vetter. Face recognition based on fitting a 3d morphable model. *IEEE Trans. PAMI*, 25(9), Sept. 2003. 1

[3] K. Bowyer, K. Chang, and P. Flynn. A survey of approaches and challenges in 3d and multi-modal 3d+2d face recognition. *CVIU*, 101(1):1–15, 2006. 1

[4] A. Bronstein, M. Bronstein, and R. Kimmel. Three-dimensional face recognition. *International Journal of Computer Vision*, 64(1), 2005. 1

[5] A. Bronstein, M. Bronstein, and R. Kimmel. Expression-invariant representation of faces. *IEEE Trans. on IP*, 16(1), 2007. 1

[6] K. Chang, K. Bowyer, and P. Flynn. An evaluation of multimodal 2d+3d face biometrics. *IEEE Trans. PAMI*, 27(4), 2005. 1

[7] K. Chang, K. Bowyer, and P. Flynn. Multiple nose region matching for 3d face recognition under varying facial expression. *IEEE Trans. on PAMI*, 28(10), 2006. 1

[8] Y. Chang, M. Vieira, M. Turk, and L. Velho. Automatic 3d facial expression analysis in videos. In *IEEE ICCV05 Workshop on Analysis and Modeling of Faces and Gestures*, Beijing, 2005. 1

[9] J. Cohn and T. Kanade. Use of automated facial image analysis for measurement of emotion expression. *The handbook of emotion elicitation and assessment. Oxford University Press Series in Affective Science*, J. Coan and J. Allen, Eds., New York, NY: Oxford., 2006. 1

[10] J. Kittler, A. Hilton, M. Hamouz, and J. Illingworth. 3d assisted face recognition: A survey of 3d imaging, modelling and recognition approaches. In *CVPR05 Workshop on A3DISS*, 2005. 1

[11] S. Li and A. Jain. *Handbook of face recognition*. Springer, New York, 2004. 1

[12] Y. Li, S. Gong, and H. Liddell. Constructing facial identity surfaces for recognition. *IJCV*, 53(1):71–92, 2003. 1

[13] X. Lu, A. Jain, and D. Colbry. Matching 2.5D face scans to 3D models. 28(1):31–43, Jan. 2006. 1

[14] A. Nefian. *A Hidden Markov Model based approach for face detection and recognition*. Ph.D. Thesis, Georgia Tech., 1999. 2

[15] M. Pantic and L. Rothkrantz. Automatic analysis of facial expressions: the state of the art. *IEEE Trans. PAMI*, 22(12), 2000. 1

[16] G. Passalis, I. Kakadiaris, and T. Theoharis. Intraclass retrieval of nonrigid 3d objects: Application to face recognition. *IEEE Trans. on PAMI*, 29(2), 2007. 1

[17] P. Phillips, P. Flynn, T. Scruggs, K. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of the face recognition grand challenge. In *IEEE Conf. on CVPR 2005, San Diego, CA*, 2005. 1

[18] C. Samir, A. Srivastava, and M. Daoudi. Three-dimensional face recognition using shapes of facial curves. *IEEE Trans. on PAMI*, 28(11), 2006. 1

[19] A. Scheenstra, A. Ruifrok, and R. Veltkamp. A survey of 3d face recognition method. In *AVBPA 2005*, pages 891–899, 2005. 1

[20] Y. Sun and L. Yin. 3d spatial-temporal face recognition using dynamic range model sequences. In *IEEE CVPR Workshop on Biometrics, 2008. Anchorage, Alaska.*, 2008. 1, 2, 6

[21] Y. Sun and L. Yin. Facial expression recognition based on 3d dynamic range model sequences. In *European Conference on Computer Vision (ECCV08)*, 2008. 1, 2

[22] J. Wang, L. Yin, X. Wei, and Y. Sun. 3D facial expression recognition based on prmitive surface feature distribution. In *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR06)*. 1, 5

[23] S. Wang, Y. Wang, M. Jin, X. Gu, and D. Samaras. 3d surface matching and recognition using conformal geometry. In *CVPR06*, 2006. 1

[24] S. Wang, Y. Wang, M. Jin, X. Gu, and D. Samaras. Conformal geometry and its applications on 3d shape matching, recognition and stitching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(7), 2007. 1

[25] Y. Wang, X. Huang, C. Lee, S. Zhang, Z. Li, D. Samaras, D. Metaxas, A. Elgammal, and P. Huang. High resolution acquisition, learning and transfer of dynamic 3d facial expressions. In *EUROGRAPHICS 2004*, 2004. 1

[26] L. Yin, X. Chen, Y. Sun, T. Worm, and M. Reale. A high-resolution 3d dynamic facial expression database. In *IEEE International Conference on Automatic Face and Gesture Recognition*, Amsterdam, the Netherlands, Sept. 2008. 1, 4, 6

[27] L. Yin, X. Wei, J.Wang, Y. Sun, and M. Rosato. A 3D facial expression database for facial behavior research. In *IEEE International Conference on Face and Gesture Recogntion (FGR06)*, 2006. 1, 4, 5

[28] Z. Zeng, M. pantic, G. Roisman, and T. Huang. A survey of affect recognition methods: audio, visual, and spontaneous expressions. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 31(1), 2009. 1

[29] W. Zhao, R. Chellappa, P. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Computing Surveys*, 35(4), Dec. 2003. 1