1. Problem

In this paper, we address the problem of facial feature tracking under varying facial expressions and poses by proposing a face shape prior model that is constructed from the Restricted Boltzmann Machine (RBM) and its variants.

Figure 1 Facial feature tracking when face is under facial expression and head pose change.

2. Motivation

Observations:
1. There exist patterns of face shape.
2. The face shape depends on facial expressions and head poses.

Motivation: To increase the accuracy and robustness of facial feature tracking algorithm, a face shape prior model that captures the face shape pattern with varying facial expressions and poses should be utilized.

3. Model

- A generative model based on Restricted Boltzmann Machines that captures the face shape patterns. It decomposes the shape variations into expression related and pose related parts.
- Part I: Variations of frontal face shapes with expression is modeled using Deep Belief Networks.
- Part II: Transferring frontal face shape to shape with different head poses is modeled with three-way RBM.

Figure 2 Face shape prior model based on Restricted Boltzmann Machines. X: feature point locations in frontal view. Y: feature point locations in non-frontal view. H: hidden nodes.

4. Model Training

- Model training is based on Contrastive Divergence algorithm [CD] [Hinton, 2002] [Mohamed et al., 2011] for part I and II separately.
- For part I, the Deep Belief Network is trained in a layer-wise manner so that training is relatively efficient [Mohamed et al., 2011].
- For part II, the three-way RBM model is trained by maximizing the joint likelihood \( p(x, y) \) using CD algorithm.

Specifically, the derivative [Memisevic et al., 2010] of the log-likelihood \( L(x, y; \theta) \) can be written as

\[
\frac{\partial L}{\partial \theta} = \frac{\partial}{\partial \theta} \log p(x, y; \theta) = \frac{\partial}{\partial \theta} \log p(x, y; \theta) - \frac{\partial}{\partial \theta} \log p(h; \theta)
\]

Using CD algorithm, equation (7) can be approximated with Gibbs sampling with the following equations:

\[
p(h_k = 1|x, y) = \sigma \left( \sum_i w_{ki} \sum_j \phi(x_{ij}|\theta) + w_k \right)
\]

\[
p(x, y) = \frac{1}{e^{-E(x, y)}}
\]

\[
p(x, y, h) = \sum_j \left( \sum_i \phi(x_{ij}|\theta) \sum_h \phi(h_i|\phi) \right)
\]

5. Facial feature tracking with shape model

- Estimation of the true facial feature point locations \( Y^* \) with facial expressions and pose base on measurements \( Y_m \) and prior model in a probabilistic formulation:

\[
Y^* = \arg \max_Y P(Y_m|H)^P(Y)
\]

- \( P(Y) \) can be estimated from the samples generated by the prior model shown in figure 3 using Kernel Density Estimation method or with a Gaussian Assumption.

To generate the sample, \( Y \) is initialized with \( Y_m \). Then, the algorithm updates \( X, H, Y \) sequentially using Gibbs sampling within part II. Furthermore, the algorithm updates \( X \) by sampling within part I. Finally, the output sample of \( Y \) is generated by the updated \( X \) and previously estimated \( H_2 \).

6. Experimental Results

- Table 1 Experimental results on CK+ database
- Table 2 Experimental results on MMI database
- Table 3 Experimental results on ISL database

(a) Sample sequence from American Sign Language database

(b) Sample sequence from ISL database

Figure 3 Performance of the proposed model based on synthetic data. (a) face with outlier (left eyebrow tip); (b) correction of (a); (c) face with corrupted points on left half face; (d) correction of (c). First row: FrontalRBM (Part I); Second row: PoseRBM (Part I and II).

Figure 4 Facial feature tracking on sample sequences.