Exemplar-Based Face Parsing

Brandon M. Smith 1 Li Zhang 1
1University of Wisconsin - Madison

Jonathan Brandt 2 Zhe Lin 2 Jianchao Yang 2
2Adobe Research

Motivation
A common task in face image analysis is parsing an input face image into facial parts, e.g., left eye and upper lip. Most previous methods accomplish this task by marking a few landmarks or contours on the input face image. Instead, we seek to mark each pixel on the face with its semantic part label, that is, our algorithm parses a face image into its constituent facial parts.

Previous Landmarks, Contours Ours

Pros

- Vectorized representation
- Encodes ambiguity
- Generalizes to hair, teeth, ears, etc. across datasets

Cons

- Ambiguous localization
- Inconsistent definitions across datasets
- Not vectorized, but can be combined with landmarks and contours

Step 1: Nonrigid Exemplar Alignment
For each keypoint in each top exemplar, perform a local search in the input image to find the best match; record the matching score. Warp the label map of each exemplar nonrigidly using a displacement field interpolated from the match location offsets.

Step 2: Exemplar Label Aggregation
Aggregate warped label maps using weights derived from the keypoint matching scores in Step 1. The weights are spatially varying and favor exemplar pixels near good keypoint matches.

Step 3: Pixel-Wise Label Selection
Produce a label probability vector at each pixel by attenuating each channel in the aggregated label map. The attenuating weights are trained offline to correct for label population biases.

Runtime Pre-Processing
Extract dense SIFT descriptors in the input image. Search for a subset of top exemplar faces, each associated with a similarity transformation that aligns the exemplar face to the input face.

Quantitative Results

Our method

- Vectorizes exemplar label maps
- Encodes ambiguity
- Generalizes to hair, teeth, ears, etc. across datasets

Ours, full pipeline

- Vectorizes exemplar label maps
- Encodes ambiguity
- Generalizes to hair, teeth, ears, etc. across datasets

Extensions of Our Approach

- Contour Estimation
- Hair Segmentation

We can synthesize the input face by replacing the exemplar label vectors with the color channels from the exemplar images.

Large segmentation errors occur infrequently, but when they do occur, errors are almost always localized to the mouth region. Unusual mouth expressions like those shown above are not represented well in the exemplar images, which results in poor label transfer from the top exemplars to the test image.

Qualitative Results

Our method

- Vectorizes exemplar label maps
- Encodes ambiguity
- Generalizes to hair, teeth, ears, etc. across datasets

Ours, omit Step 3

Comparison with a face parsing algorithm (Warrell & Prince), and three face alignment algorithms (segments were derived from the contours generated by these algorithms).

Failure Cases on Mouths Due to Insufficient Exemplars

Ours, omit Step 3

The results in (c) exemplifies the problem with the label weights used to maximize the diagonal of the confusion matrix. We instead show accuracy using the F-measure (harmonic mean of precision and recall) and we optimize label weights to maximize the F-measure.

F-Measures for LFW Images

Ours, full pipeline

- Vectorizes exemplar label maps
- Encodes ambiguity
- Generalizes to hair, teeth, ears, etc. across datasets

Comparison with a face parsing algorithm (Warrell & Prince), and three face alignment algorithms (segments were derived from the contours generated by these algorithms).

F-Measures for Helen Images

Ours

- Vectorizes exemplar label maps
- Encodes ambiguity
- Generalizes to hair, teeth, ears, etc. across datasets

Liu et al. is a nonparametric scene parsing algorithm. The only area where Liu et al.'s system is more accurate than ours is on the face skin. The difference is primarily due to our algorithm incorrectly hallucinating skin in hair regions, while Liu et al.'s system does not.

Adobe Systems Incorporated, and an NSF Graduate Research Fellowship.