A Face Annotation Framework with Partial Clustering and Interactive Labeling

Problems

Lots of photos to label/group

Pure human labors $\rightarrow$ Unattainable

Pure machine labeling $\rightarrow$ Poor performance

Current Face Recognition Algorithm is not stable

1. Framework Overview

Picture with faces $\rightarrow$ Face Detection & Alignment $\rightarrow$ Normalized faces $\rightarrow$ Feature Extraction & Combination $\rightarrow$ Similarity Matrix

Efficient Labeling $\rightarrow$ Evident clusters & Background cluster $\rightarrow$ Partial Clustering

User Interaction

2. Contributions

Partial Clustering
Only group eviden good clusters in which similarity is of good quality, and leave all the other faces in the clutterbin, in which similarity measure is contaminated by noise and not reliable.

Efficient Labeling
Pick out a list of unlabeled faces, once at a time for the user to label. We regard labeling process as information influx to resolve ambiguity. Hence the ratio of information gain to estimated number of user interactions, or information efficiency, is maximized.
Yuandong Tian (Shanghai Jiaotong University)
Wei Liu (The Chinese University of Hong Kong)
Rong Xiao, Fang Wen, Xiaou Tang (Microsoft Research Asia)

Partial Clustering

Face Images

Similarity Matrix

Spectral Embedding to dimension M

Background Density

The "natural" uniform distribution on Sphere^M.
prob = the reciprocal of the "area" of Sphere^M

Initial labeling to get easy-labeled clusters

Efficient Labeling

The set of faces to be chosen

Q = arg max \( Q \)

Information Gain
if Q is labeled

Sets of labeled faces

Operations(Q)

Subset Saliency

How cohesive Q is, with respect to other unlabeled faces
Yuandong Tian (Shanghai Jiaotong University)
Wei Liu (The Chinese University of Hong Kong)
Rong Xiao, Fang Wen, Xiaou Tang (Microsoft Research Asia)

3. Experiment Result

Photos

Feature Extraction

Similarity

Partial Clustering

Evident & Background Clusters

Initial Labeling

Efficient Labeling

Annotated Photos

Data Preparation

Partial Clustering vs Spectral Clustering-KMean

Evident & Background Clusters

Efficient Labeling vs simple cluster-based labeling

Overall Performance vs Riya

We compare our framework with Riya’s. Ours outperform by about 46%. 