Face Discovery with Social Context

Yong Jae Lee
https://webspace.utexas.edu/yl3663/~ylee/

Kristen Grauman
http://www.cs.utexas.edu/~grauman/

University of Texas at Austin
Austin, TX, USA

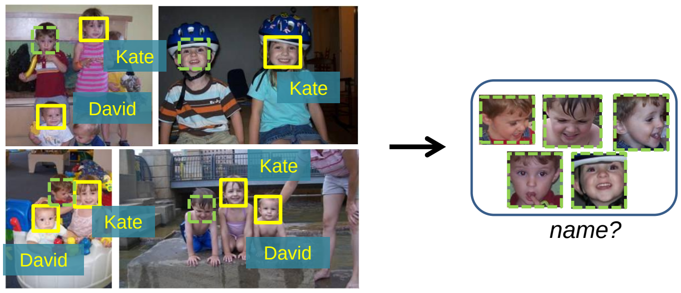

Figure 1: Main idea. For any unfamiliar face not recognized by the system (in dotted green), we use the co-occurrence cues from familiar faces nearby (in solid yellow) to produce more reliable groups. In this example, an appearance-based method that clusters the unfamiliar faces would likely fail to recognize the many instances of the boy, given their variability. In contrast, by also representing the *social context* of familiar people, our approach computes more reliable clusters.

Datasets	# Unknowns	Ours	No-Context	App-Context
Mixture	15	**0.30**	0.26	0.28
Wang1	16	**0.25**	0.20	0.21
Wang2	104	**0.24**	0.23	0.21

Figure 2: Face discovery as judged by the F-measure. Higher values are better. Our method outperforms both baselines in all cases, showing the impact of modeling the co-occurrence information of surrounding familiar people for discovery.

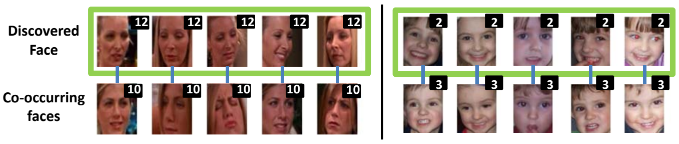

Figure 3: Face discovery examples. The first row shows representative faces of the dominant person for a discovered face, with their respective co-occurring faces below. The second row faces belong to a known person—their social context helps to group the diverse faces of the same person in the first row.

We present an approach to discover novel faces in untagged photo collections by leveraging the social context of co-occurring people. The goal is to perform unsupervised clustering to determine batches of photos likely of the same individual, so a user can efficiently tag or prune them with minimal effort. In contrast to previous face clustering algorithms (e.g., [1, 2, 3]), we propose to expand the representation of the detected faces to include not just their appearance, but also their *social context*. Co-occurrence cues from people in the same image allow the system to produce more reliable groups (see Figure 1).

Why do co-occurrence cues help? New faces in a collection appear with some strong social context, as users' photos tend to dwell within different cliques of people: families, friends, co-workers, etc. This means the context of "familiar people" can both help disambiguate people with similar appearance, and help the system realize that instances of faces in different poses or expression are actually of the same person.

Approach We first train SVM classifiers for N initial people, $\{c_1,\dots,c_N\}$, for whom we have tagged face images. These classifiers will allow us to identify instances of each familiar person in novel images. We use those predictions to describe the social context for each *unfamiliar* face.

For any unlabeled photo, we detect the people in it, and then determine whether any of them resembles a *familiar person*. To compute the known/unknown decision for a face region r in an unlabeled image, we apply the N trained classifiers to the face to obtain its class membership posteriors $P(c_i|r)$, for $i = 1,\dots,N$, where c_i denotes the i-th person class. To distinguish which faces should be considered to be unknown, we compute the entropy: $E(r) = -\sum_{i=1}^{N} P(c_i|r)\log P(c_i|r)$. Faces with low entropy values will likely belong to familiar people, while those with high values will likely be unfamiliar.

For each unfamiliar face, we want to build a description that reflects that person's co-occurring familiar people, at least among those that we can already identify. Having such a description allows us to group faces that look similar and often appear among the same familiar people.

Suppose an image has T total faces: $r_1,\dots,r_T$. We define the social context descriptor $S(r)$ as an N-dimensional vector that captures the distribution of familiar people that appear in the same image:

$$S(r) = \left[\sum_{j=1}^{T} P(c_1|r_j),\dots,\sum_{j=1}^{T} P(c_N|r_j)\right]. \qquad (1)$$

If our class predictions were perfect, with posteriors equal to 1 or 0, this descriptor would be an indicator vector telling which other people appear in the image. When surrounding faces do belong to previously learned people, we will get a "peakier" vector with reliable context cues, whereas when they do not appear to be a previously learned person the classifier outputs will simply summarize the surrounding appearance.

Finally, we cluster all faces that were deemed to be unknown, using spectral or agglomerative clustering. We want the discovered groups to be influenced both by the appearance of the face regions themselves, as well as their surrounding context. Therefore, given two face regions r_m and r_n, we evaluate a kernel function K that combines their appearance similarity and context similarity:

$$K(r_m,r_n) = \alpha\cdot K_{\chi^2}\left(S(r_m),S(r_n)\right) + (1-\alpha)\cdot K_{\chi^2}\left(A(r_m),A(r_n)\right), \qquad (2)$$

where $A(r)$ is the appearance descriptor, α weights the contribution of social context versus appearance, and each K_{χ^2} is a χ^2 kernel function for histogram inputs x and y.

Results We compare our method to a **no-context** baseline that simply clusters the face regions' low-level texture features, and an **appearance-context** discovery method that uses the appearance of surrounding faces as context. These are important baselines to show that we would not be as well off simply looking at a model of appearance using image features, and to show the impact of social context analysis versus a low-level appearance context description for discovery.

We validate on three datasets of consumer photo collections composed of 1,000 to 12,000 images and 23 to 152 people. We partition each dataset into two random subsets. The first is used to train N classifiers for the initial "knowns". On the second subset, we perform discovery using the N categories as context to obtain our set of discovered categories. This reflects the real scenario where a user has tagged only some of his/her family members and friends.

Figure 2 shows discovery results. Our method significantly outperforms the baselines, validating our claim that social context leads to better face discovery. Our substantial improvement over the appearance-context baseline shows the importance of representing context with models of familiar people. Figure 3 shows qualitative discovery examples.

In the paper, we study several other aspects of interest including (1) how accurately we predict novel instances to be familiar or unfamiliar, and (2) how our discovered faces can be used to predict tags in novel photos. Our results show that the models learned from faces discovered using social context generalize better on novel face instances than those learned from faces discovered using appearance alone. This is evidence that our approach can indeed serve to save human tagging effort.

[1] T. Berg, A. Berg, J. Edwards, M. Maire, R. White, Y. Teh, E. Learned-Miller, and D. Forsyth. Names and Faces in the News. In *CVPR*, 2004.

[2] Y. Song and T. Leung. Context-Aided Human Recognition Clustering. In *ECCV*, 2006.

[3] Y. Tian, W. Liu, R. Xiao, F. Wen, and X. Tang. A Face Annotation Framework with Partial Clustering and Interactive Labeling. In *CVPR*, 2007.