

Learning Attributes from Human Gaze (Supplemental Material)

Nils Murrugarra-Llerena

Department of Computer Science
University of Pittsburgh

{nineil, kovashka}@cs.pitt.edu

Adriana Kovashka

As supplemental material, we provide qualitative results comparing our single template (ST) and multiple template (MT) approaches. Also, we provide detailed results for an application of our approach to discovering the “schools of thought” among viewers in terms of their attribute perception.

1. Single vs multiple templates

We provide qualitative results comparing our single-template and multiple-templates approaches in Figures 1 (for shoes) and 2 (for faces). For our multiple-template approach, we select one meaningful template per image. Each subfigure contains two images: the left one shows the single template, and the right one shows a selected template from the multiple-templates method.

In Figure 1, we see that MT captured high-heel as a cue for the attribute “feminine,” while ST focused on the entire shoe. For the “formal” attribute, MT concentrated on the shoe center, while ST focused on the entire shoe. For “pointy,” MT focused on the front of the shoe, and for “open,” it concentrated on the center of the shoe, where the open attribute resides. Finally, for “sporty,” MT highlighted shoelaces, which are a relevant part of any sporty shoe. In contrast, for these three attributes, ST could not determine a specific relevant part for the attribute.

On our face data (Figure 2), MT focused on people’s eyes for the “Asian” attribute. Similarly, for “Indian”, it concentrated on the eyes and nose, while ST covered a wider area and picked the mouth also. For “chubby” and “big-nosed,” MT found a smaller relevant area concentrated on the cheeks and nose, respectively. For “baby-faced,” MT determined that the eyes, cheeks and nose are relevant; the template is better localized than the one found by ST. Finally, for the “attractive,” “masculine” and “youthful” attributes, MT found the same face components as ST, however MT templates were a bit better localized and covered a smaller area.

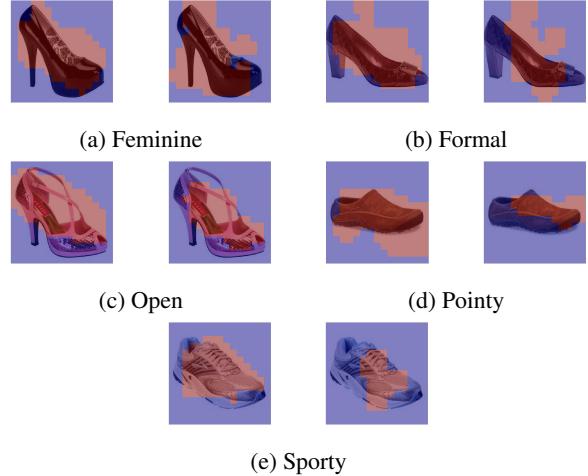


Figure 1: A comparison of the single and multiple template methods, for shoe attributes. Left = ST, right = MT.

2. Using gaze to find schools of thought

A complete table comparing the original schools of thought approach of Kovashka and Grauman with our gaze-based approach is shown in Table 1.

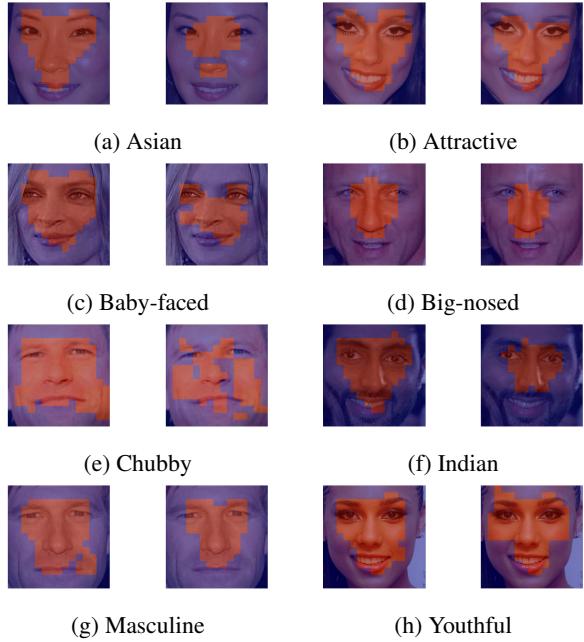


Figure 2: A comparison of the single and multiple template methods, for face attributes. Left = ST, right = MT.

	original approach	gaze-based approach
feminine	0.36	0.43
formal	0.40	0.44
open	0.52	0.58
pointy	0.36	0.43
sporty	0.41	0.43
asian	0.43	0.34
attractive	0.13	0.19
baby	0.49	0.52
big-nosed	0.29	0.41
chubby	0.38	0.35
indian	0.46	0.43
masculine	0.35	0.40
youthful	0.29	0.26
avg	0.37	0.40

Table 1: Quantitative comparison of the original schools of thought approach and our gaze-based approach.