From Bikers to Surfers: Visual Recognition of Urban Tribes

Iljung S. Kwak[1]
iskwak@cs.ucsd.edu
Ana C. Murillo[2]
acm@unizar.es
Peter N. Belhumeur[3]
belhumeur@cs.columbia.edu
David Kriegman[1]
kriegman@cs.ucsd.edu
Serge Belongie[1]
sjb@cs.ucsd.edu

[1] Dept. of Computer Science and Engineering
University of California, San Diego, USA.
[2] Dpt. Informática e Ing. Sistemas - Inst. Investigación en Ingeniería de Aragón.
University of Zaragoza, Spain.
[3] Department of Computer Science
Columbia University, USA.

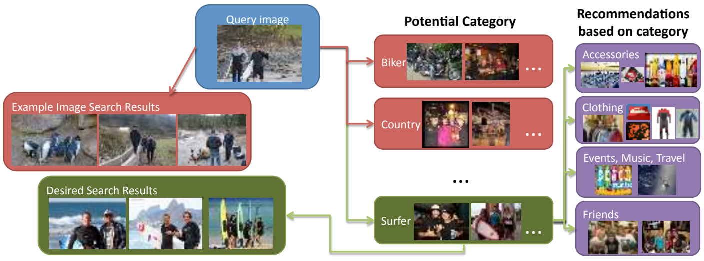


Figure 1: The social groups influence the appearance of their members. This work leverages this intuition to classify images of groups of people into social categories. Automatic urban tribe classification can improve recommendation systems and user experience with social media, and image search engines can take advantage of this classification and provide more meaningful search results.

Motivation. In 1985 Maffesoli described *urban tribes* [3] as a group of people who have similar visual appearances, personal style, and ideals. Visual cues shared by tribe members provide the basis for our work; members from the same urban tribe are expected to look more similar than members of different tribes. As shown in Fig 1, with automatic classification of urban tribes more inspiring image searches can be conducted; more relevant advertisements can enhance the web experience of both businesses and consumers; social networks can provide better recommendations [6]. Unfortunately, this categorization problem is difficult because of the ambiguous nature of social categories and the high intraclass variance. Social categories can evolve and fracture into separate groups; individuals within groups may exhibit features of multiple urban tribes or certain individuals may not present a visually salient style at all.

Contributions. This work attempts to recognize the content of an image from a social perspective, in line with the growing area of social signal processing [4]. A key element common to our work and related works is the need for both global image statistics as well as more semantic individual level attributes. Many recent works recognize the presence or absence of facial attributes, for instance of facial attributes to perform higher level tasks, such as face verification [2]. Rather than approaching this problem of social group image categorization by classifying isolated individuals (as most recent fashion/style analysis works do [5]), we focus on calculating meaningful group features and models. For example, the importance of group structure has been previously studied and shown useful to detect gender, age and even family relationship [1]. Following these ideas, we present a novel recognition pipeline and evaluate different modeling approaches, following common frameworks used for recognition tasks. We present the Urban Tribes dataset, with around 100 labeled images per class from 11 different classes, to facilitate further research on social categorization of group pictures[1]. Fig. 2 and Table 1 show examples from our dataset.

Results. We set up classification experiments where training and testing images were randomly selected from each of the categories for 50 different iterations. For each experiment, a fixed number of the images from each class are used for learning the models and the rest of images are used for testing. An image is correctly classified if the most likely group label matches the ground truth labels. In Fig. 3, we provide confusion matrices for classification results using two different modeling approaches.

[1] `http://vision.ucsd.edu/content/urban-tribes`

Figure 2: Examples of social groups in our Urban Tribes dataset. The images show a wide range of inter-class and intra-class variability. More details can be seen at the dataset website.

Label	# im	# people	Label	# im	# people	Label	# im	# people
biker	114	443	hip-hop	90	253	club	100	365
country	107	347	hipster	102	288	formal	103	414
goth	99	226	raver	116	305	casual/pub	125	459
heavy-metal	102	266	surfer	100	333			

Table 1: Summary of Urban Tribes dataset.

Both cases show that our techniques have promising recognition results. These experiments show that it is possible to extract semantic meaning from social media group photos, opening opportunities for the previously mentioned applications. Future work can include incorporating semantic attributes to improve the classification performance and analyzing the contribution of different types of semantic attributes.

[1] A. Gallagher and T. Chen. Understanding images of groups of people. In *CVPR*, 2009.

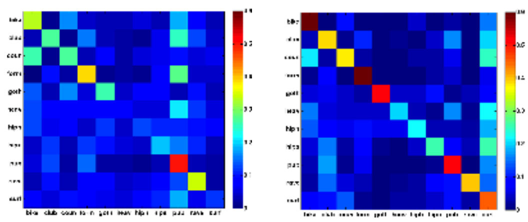
[2] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar. Attribute and Simile Classifiers for Face Verification. In *ICCV*, 2009.

[3] M. Maffesoli. *The Time of the Tribes: The Decline of Individualism in Mass Society*. Sage Publications, 1996.

[4] A. Vinciarelli, M. Pantic, and H. Bourlard. Social signal processing: Survey of an emerging domain. *Image and Vision Computing*, 2009.

[5] K. Yamaguchi, M. H. Kiapour, L. E. Ortiz, and T. L. Berg. Parsing clothing in fashion photographs. In *CVPR*, 2012.

[6] J. Yu, X. Jin, J. Han, and J. Luo. Collection-based sparse label propagation and its application on social group suggestion from photos. *ACM Trans. Intell. Syst. Technol.*, 2011.



(a) (b)

Figure 3: Confusion matrix for classification results obtained with (a) Bag of Parts modeling (BoP_{30k}) and (b) SVM classification framework (SVM_8), using 80% of the data for training. The rows show results in alphabetical order of the labels.