Aggregation of Social Photographs Using Visual and Multimodal Features







Unaiza Ahsan and Irfan Essa Georgia Institute of Technology

7. Experimental Results Compute visual & textual features 7.1 Which metadata is most useful? Transform the two feature sets via Feature combinations NMI 0.7137 dates + tags kernel canonical correlation analysis 0.8439 descriptions + tags 0.8509 tags + usernames tags + titles 0.8411 0.8266 SIFT + tags 'August 2007 Electrelane Field Day 'Summer Sonic SummerSonic2007 Tokyo' 'John Strand capture the flag sans 7.2 Visual Feature Analysis Compute text features (e.g. Analysis of color descriptors on clustering event TFIDFs) $Y = \{y_1, y_2, ..., y_n\}$ Transform into ____ 🦊 $Y \mapsto \phi(Y)$ kernel space [S+RGB-H+RGB-S] S+RGB-H+RG-SI Kernel IRGB-H+RGB-S $K_{y} = YY$ space IRGB-S+RG 0.74 kernel CCA 7.3 Tags or Usernames? $\omega_{v} = Y'\beta$ space with fewer dimension Which non-visual feature contributes most to No initial clusters using any metadata clustering score? Assume minimal metadata availability Vocabulary size 1024 Choosing between tags and usernames analysis on 300k images Spatial pyramids of 2 levels. Descriptors RGB-S+ RGB-S + [U+T cluster ids with each image





8. Conclusion and Future Work

Visual features still lag behind text and metadata combination for producing unique social event clusters The difference between the two scores (0.12) is promising Plan to refine the dataset and simplify this problem