

Visual and semantic context modeling for scene-centric image annotation

ABSTRACT

Automatic image annotation enables efficient indexing and retrieval of the images in the large-scale image collections, where manual image labeling is an expensive and labor intensive task. This paper proposes a novel approach to automatically annotate images by coherent semantic concepts learned from image contents. It exploits sub-visual distributions from each visually complex semantic class, disambiguates visual descriptors in a visual context space, and assigns image annotations by modeling image semantic context. The sub-visual distributions are discovered through a clustering algorithm, and probabilistically associated with semantic classes using mixture models. The clustering algorithm can handle the inner-category visual diversity of the semantic concepts with the curse of dimensionality of the image descriptors. Hence, mixture models that formulate the sub-visual distributions assign relevant semantic classes to local descriptors. To capture non-ambiguous and visual-consistent local descriptors, the visual context is learned by a probabilistic Latent Semantic Analysis (pLSA) model that ties up images and their visual contents. In order to maximize the annotation consistency for each image, another context model characterizes the contextual relationships between semantic concepts using a concept graph. Therefore, image labels are finally specialized for each image in a scene-centric view, where images are considered as unified entities. In this way, highly consistent annotations are probabilistically assigned to images, which are closely correlated with the visual contents and true semantics of the images. Experimental validation on several datasets shows that this method outperforms state-of-the-art annotation algorithms, while effectively captures consistent labels for each image.

Keyword: Automatic image annotation; Visual diversity; Mixture model; Visual context; Semantic context