

## SEMANTIC SENSITIVE REGION RETRIEVAL USING KEYWORD-INTEGRATED BAYESIAN REASONING

RONGRONG JI<sup>1</sup>, XIANBO LANG<sup>2</sup>, HONGXUN YAO<sup>1</sup> AND ZHEN ZHANG<sup>1</sup>

<sup>1</sup>School of Computer Science and Engineering  
Harbin Institute of Technology  
Harbin 150001, P. R. China  
{ rrji; yhx; z Zhang }@vilab.hit.edu.cn

<sup>2</sup>School of Computer Science and Engineering  
Harbin Engineering University  
Harbin 150001, P. R. China  
jameslang@gmail.com

Received November 2006; revised April 2007

**ABSTRACT.** *Inaccurate semantic representation is the main cause of semantic gaps in content-based image retrieval, owing to: 1. imprecise semantic description at image global scale; 2. inaccurate mapping from visual features to image perception. In this paper, we propose a region retrieval framework using keyword-integrated Bayesian reasoning. To address the first issue, a region-level visual dictionary is constructed using pre-labeled segmented images. To address the second issue, a keyword-integrated Bayesian reasoning is adopted for similarity ranking, together with the visual dictionary to precisely bridge the semantic gap. Furthermore, users' relevance feedback is utilized to adjust the semantic revealing ability of keywords. Based on this reasoning framework, both query-by-example and query-by-keyword scenarios are provided to facilitate user query. The promising experimental result indicates the effectiveness of our algorithm in enhancing the performance of region-based image retrieval and narrowing down the semantic gap.*

**Keywords:** Image retrieval, Image segmentation, Image annotation, Region matching, Bayesian reasoning

**1. Introduction.** Content-based image retrieval (CBIR) is a technique that effectively retrieves images based on their visual content. It is a long-standing research hot-spot over the past decade [1,2,16-18]. However, the performances of state-of-the-art CBIR systems are still unsatisfactory due to the semantic gap between image visual features and human image perception. Especially, the inaccurate semantic representation of image is the main cause of this gap, since: 1. Currently, most of the feature extraction and linguistic indexing methods are applied over image's global scale. Such strategies are imprecise when an image contains multiple objects and mixed semantic information. 2. The similarity calculation and ranking methods aiming at mapping visual features to human perception are inaccurate, due to the lack of high-level supervise information.

Much effort has been carried out to represent images more accurately at object level. Since human image discernment is strongly associated with interesting classes of objects in image [3], many CBIR systems have adopted a regional-based strategy to represent images and extract visual features more perceptually meaningful [4-6]. However, inaccurate