Automated Keywording/Tagging of Images Using Machine Learning

ID# 2006-3242





grass, people, animal, horse, rural, dog, landscape, tribal, plant

Categories Suggested for a Photo

Technology Summary

Categorized images are used to train a dictionary of hundreds of concepts automatically based on statistical modeling. Images of any given concept-category are regarded as instances of a stochastic process that characterizes the category. The likelihood of occurrence of the image based on the stochastic process derived from the category is computed. A high likelihood indicates a strong association, and thus associates the concept-category textual descriptors with the image. A portfolio of US patents have been granted around this technology, including: 7,394,947; 9,646,226; 7,941,009.

Application & Market Utility

Current content-based image retrieval systems aim at general-purpose image indexing and retrieval focused on searching images visually similar to the query image. However, because of the difficulty recognizing large numbers of objects, these systems lack the capability to assign comprehensive textual descriptions automatically to images. The disclosed software/algorithms transcend these limitations. Applications: digital libraries, medical imaging, video surveillance, web searching. Fields: biomedicine, military & law enforcement, education, electronic commerce.

Next Steps

Seeking research collaboration and licensing opportunities.

TECHNOLOGY READINESS LEVEL

4-7

Seeking

Licensing |

Keywords

- image indexing
- image recognition
- web searching
- image tagging
- machine learning

Researchers

Jia Li

Professor of Statistics

Online Bio Website

James Wang

Professor of Information Sciences and Technology Website

Originating College

College of Information Sciences and Technology

Office of Technology Management Contact

Swope, Brad bas101@psu.edu 814-865-6277

