



# CALIFORNIA STATE SCIENCE FAIR 2008 PROJECT SUMMARY

<b>Name(s)</b> <b>David C. Liu</b>	<b>Project Number</b> <b>S1314</b>
<b>Project Title</b> <b>Content-Based Image Retrieval</b>	
<b>Abstract</b> <b>Objectives/Goals</b> With the astounding number of digital images available, there is an increasing need to search image collections. Many systems such as Google and Flickr use text-based search, but the vast majority of photos (particularly family albums) have no text description available. Content-based image retrieval searches images using visual similarity, rather than text. This research investigates methods to improve the performance of image retrieval systems. <b>Methods/Materials</b> A novel technique is presented for automatic combination of features using machine learning to improve retrieval accuracy. Perceptual characteristics (color and texture signatures) are extracted as a mathematical representation of images. Color signatures are extracted based on k-means clustering of Lab color space coordinates, and texture signatures are extracted using k-means clustering of Gabor filter dictionaries of 4 scales and 6 orientations. Signature dissimilarities are measured using the Earth Mover's Distance, and integrated through normalized linear weighting. k-nearest neighbor supervised learning is used to predict weights based on statistical characteristics of color and texture signatures: "color spread" and "texture busyness". Unlike other research in which entire images are analyzed, this research indexes images by using specifically tagged regions. This eliminates irrelevant content and backgrounds and allows users to specify what they are looking for and make more precise queries. <b>Results</b> It was found that the learning model significantly improves retrieval accuracy: by 9.32% (6.56 percentage points) over using color signatures alone, and 37.06% (20.81 percentage points) over texture signatures alone. This is a statistically significant improvement ( $p$ -value $< 0.001$ ). An extensible framework is also presented, which visualizes color and texture signatures. This helps researchers understand the relationship between the optimal weights and signature characteristics. It also includes an intuitive user interface for tagging regions and querying. <b>Conclusions/Discussion</b> Content-based image retrieval is a very active research topic that has the potential to significantly change digital image management as well as image search engines. This research contributes a new technique to improve retrieval accuracy.	
<b>Summary Statement</b> This project explores image searching by visual similarity. A novel algorithm is presented to improve retrieval accuracy through machine learning of color and texture characteristics.	
<b>Help Received</b>	