Linguistic Indexing of Images with Database Mediation

Emmanuel Udoh
Indiana University – Purdue University, USA

INTRODUCTION

Computer vision or object recognition complements human or biological vision using techniques from machine learning, statistics, scene reconstruction, indexing and event analysis. Object recognition is an active research area that implements artificial vision in software and hardware. Some application examples are autonomous robots, surveillance, indexing databases of pictures and human computer interaction. This visual aid is beneficial to users, because humans remember information with greater accuracy when it is presented visually than when it originates in writing, speech or in kinesthetic form. Linguistic indexing adds another dimension to computer vision by automatically assigning words or textual descriptions to images. This augments content-based image retrieval (CBIR) that extracts or searches for digital images in large databases.

According to Li and Wang (2003), most of the existing CBIR projects are general-purpose image retrieval systems that search images visually similar to a query sketch. Current CBIR systems are incapable of assigning words automatically to images due to the inherent difficulty of recognizing numerous objects at once. This current situation is stimulating several research endeavors that seek to assign text to images, thereby improving image retrieval in large databases.

To enhance information processing using object recognition techniques, current research has focused on automatic linguistic indexing of digital images (ALIDI). ALIDI requires a combination of mathematical, statistical, computational, and graphical backgrounds. Many researchers have focused on various aspects of linguistic processing such as CBIR (Ghosal, Ircing, & Khudanpur, 2005; Iqbal & Aggarwal, 2002, Wang, 2001) machine learning techniques (Iqbal & Aggarwal, 2002), digital library (Witen & Bainbridge, 2003) and statistical modeling (Li, Gray, & Olsen, 20004, Li & Wang, 2003). A growing approach is the utilization of statistical models as demonstrated by Li and Wang (2003). It entails building databases of images to be used for supervised learning. A trained system is used to recognize and identify new images with statistical error margin. This statistical modeling approach uses a hidden Markov model to extract representative information about any category of images analyzed. However, in using computer to recognize images with textual description, some of the researchers employ solely text-based approaches. In this article, the focus is on the computational and graphical aspects of ALIDI in a system that uses Web-based access in order to enable wider usage (Ntoulas, Chao, & Cho, 2005). This system uses image composition (primary hue and saturation) in the linguistic indexing of digital images or pictures.

BACKGROUND

Current image indexing systems are text-based, relying on content-relevant text placed in proximity to images. There is need for Web-based automated linguistic indexing for digital images. This fact will likely accelerate the adoption of automated linguistic indexing for images in their native visual form, which basically assigns textual description automatically to images (Forsyth & Ponce, 2002; Li, Gray, & Olsen, 2000; Li & Wang, 2003). ALIDI is currently an active research area in data mining, and its application is growing in such fields as consumer photo managers, medical imaging databases and image search engines (Berman & Shapiro, 1997; Li & Wang, 2003; Tanev, Kouylekov, & Magnini, 2004; Zhang, Goldman, Yu, & Fritts, 2002).

In general, ALIDI systems aid computer object recognition and content-based image retrieval, despite inherent difficulties (Li & Wang, 2003). With statistical modeling and machine learning approach, especially supervised learning, much research progress has been achieved in this field, but it is obvious that no single method can be used to realize this endeavor. Due to varied uses, a complete linguistic indexing system is bound to implement different algorithms or methods for effectiveness. The focus of this project is on generic image composition, which can be applied for quality control in the food industry to determine intrinsic value of food products. The proposition involved parameter-based quality detection of consumable produce. Detecting such quality involved some analysis of primary hue and saturation of images. Color profiles would be constructed with acceptable levels of various quality-indicative color levels. Certain combinations of hue and saturation would yield unacceptable produce when compared with that which is acceptable. For instance, the apple fruit passes through different color stages in the ripening process. These stages can be captured and utilized in computer quality detection works. In this article,

the research is focused on assigning linguistic terms to any images based on composition, by using combinations of hue and saturation. It dwells on generic color identification and detection of images.

MAIN FOCUS

A current research theme on linguistic indexing is automating the process of assigning text to images using machine learning and statistical modeling techniques (Li & Wang, 2003). With some statistical considerations, this article focuses on the online assignment of color-based terms to images. The following describes the overall approach and Web access of the system developed by the author.

Overall Procedure

Linguistic indexing systems recognize images and textually describe them with human-readable terms like "beach" or "forest." In this article, the theoretical approach is to use image composition, and to describe the images using the color spectrum terms such as "blue" or "red." To index any image, the approach to decompose each image to pixels is adopted. After the decomposition, each pixel is assigned the values of triple composure using the red, green and blue (RGB) color model. The RGB value is then converted to the hue, saturation and brightness (HSB) color model before the color distribution in the image is determined. The quantification of the color value is carried out as opposed to the statistical similarity between the image and any concept. This approach has the potential to identify any digital image composition within a statistical error margin.

Elements of image composition require definition and human-encoded color names that are not necessarily native to computers, because computers process integers of bytes corresponding to varieties of hue, saturation, and brightness (HSB) or of the often overlapping values of red, green, and blue (RGB) from the additive color system. While RGB does reference colors directly in its indexing technique, HSB appeared to be most useful because it denotes definite color ranges with progressive hues. This informed our decision to convert RGB values to the HSB values with ease of manipulation. Due to the fact that RGB values do not translate exactly to HSB values, we statistically determined the errors associated with any conversion. It is worth mentioning that image composition offers two potential advantages over other architectural approaches: scalability and a simple programming model (Mello & Lins, 2002).

Figure 1. Implementation topology of a linguistic processing system