Abstract

An image can be defined as a matrix of square pixels arranged in rows and columns. Image processing is a leading technology, which enhances raw images received from gadgets such as camera or a mobile phone in normal day to day life for various applications. An image to text and speech conversion system can be useful for improving accessibility of images for visually impaired as well as physically challenging people understand the scenario from the images and also train the system as that of human brain. The techniques of image segmentation and edge detection play an important role in implementing proposed system. The system generates text descriptions for an input image given by the user. Object wise generation of sentences, preposition and conjunction mapping is a challenging task. We formulate the interaction
between image segmentation and object recognition in the framework of Canny algorithm. The system goes through various phases such as pre-processing, feature extraction, object recognition, edge detection, image segmentation and Text To Speech (TTS) conversion. The proposed system database consists of huge set of sample images, which help to perform training of database. The accuracy of proposed system is achieved due to the proper recognition of objects and sentences are formed making use of the recognized objects. These sample images consists of several categories of images. The system mainly consists of two main modules such as image to text and text to speech. An image to text module generates text descriptions in natural language based on understanding of image. A text to speech module generates speech synthesis in English from description of natural language.

References

- Benjamin Z. Yao, Xiong Yang, Liang Lin, Mun Wai Lee and Song-Chun Zhu, "I2T: Image Parsing to Text Description", IEEE transactions on image processing, 2008.
- Mina Makar, Member, IEEE, Vijay Chandrasekhar, Member, IEEE, Sam S. Tsai, Member, IEEE, David Chen, Member, IEEE, and Bernd Girod, Fellow, IEEE, "Interframe Coding of Feature Descriptors for Mobile Augmented Reality", IEEE Transactions On Image Processing, Vol. 23, No. 8, August 2014.
- A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain,
A New Approach for Generating Text Description of Images and Speech Synthesis

  - S. Feng, D. Xu, X. Yang, "Attention-driven salient edge(s) and region(s) extraction with application to CBIR", Signal Processing 90, pp. 1–15, 2010.

**Index Terms**

Computer Science  
Image Processing

**Keywords**

Image Processing; Image Segmentation; Speech Synthesis; Text To Speech Conversion; Edge Detection.