International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Impact Factor (2012): 3.358

A Review on Conversion of Image to Text As Well As Speech Using Edge Detection and Image Segmentation

Mrunmayee Patil[1], Ramesh Kagalkar[2]

[1]Department of Computer Engineering, Dr.D.Y.Patil School of Engineering & Technology, University of Pune, India

[2]Professor, Department of Computer Engineering, Dr.D.Y.Patil School of Engineering & Technology, University of Pune, India

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 blind as well as physically challenging people to understand the scenario from the images. Core idea for image to text and speech conversion is to overcome the challenges faced by a blind person in real life. The techniques of image segmentation and edge detection play an important role in implementing this system. We formulate the interaction between image segmentation and object recognition in the framework of Canny algorithm. The system goes through various phases such as preprocessing, feature extraction, object recognition, edge detection, image segmentation and text-to-speech (TTS) conversion. The database of this system consists of huge set of sample images which help to identify similar kind of objects in every different image. 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 converts natural language into speech synthesis.***

Keywords: Image Processing, Feature Extraction, Edge Detection, Image Segmentation, Object Recognition, Text-to-Speech (TTS)

1. Introduction

Image processing is one of the most growing field in research and technology in today's world. Image processing uses hardware and software as computing resources to provide an efficient interface to process an image. Image processing uses various techniques such as image filtering, image pre-processing, image segmentation, image compression, image editing and manipulation, feature extraction, object recognition. An image can be defined in a function of two real variables f(r, w) where f as the amplitude (e.g. brightness) and of the image at the real co-ordinate position (r, w). The image can be of any file formats. These file format helps us to discriminate different types of images. In today's world there are around 285 million people who are visually impaired; out of which 39 million are blind and 246 have low vision. Such people have very low scope to understand what exactly is going on in their current environment. There is no such interface which is easily available for such disabled people to interact with the world. Providing efficient interface for such people is of great need.

In our proposed work of image to text and speech conversion system we develop a cost efficient and user friendly interface for blind people. The primary motivation is to provide a blind person with a friendly speech interface with computer and to allow such people which are physically and visually challenged to use the system for understanding any type of scenario. One important approach to develop this system is to make any visually blind person to analyze what is going on around him/her. Blind people usually rely on their partners or sense the scenario by their senses. In order to make a blind people more and more independent we developed this system. Many challenges are faced by a blind person in his/her day-to-days life while interacting with the world every day. Our proposed framework goes through various phases 1) Pre-Processing. 2) Feature Extraction. 3) Image Segmentation. 4) Text Conversion. 5) Text-to-Speech synthesis.

1.1 Edge Detection

A set of connected pixels that forms a boundary between two disjoint regions is known as an edge. The task of segmenting an image into regions of discontinuity is done using edge detection. Edges usually occur on the boundary of two different boundaries in an image. Edge detection helps to clearly identify the changes in region of an image where gray scale and texture change in the regions of an image. There are many available edge detection techniques for extracting edges from images such as Robert, Prewitt and Sobel which were not much efficient. Then in 1986 John. F. Canny developed an algorithm which provided high probability of edge detection and error rate.

1.2 Canny Algorithm

This algorithm focuses mainly on three main aims of low error rate, minimize distance between real edge and detected edge and minimum response i.e. one detector response per edge to detect the edges in an image.

1.3 Image Segmentation

Image segmentation is another important aspect necessarily required to divide an image into regions or categories which then helps to identify correctly the object in an image. Segmentation functions on the properties shown by the pixels

in an image, every pixel which belongs to same category has similar gray scale value whereas pixels of different categories have dissimilar values. Segmentation is often one of the critical steps in analyzing the images because additional overhead of moving to each new pixel of an image while working with object in an image. Once image segmentation is done successfully, the other stages in image analysis are much easier. While considering a fully automatic conversion algorithm, the success of image segmentation is partial and sometimes requires manual intervention. Segmentation mainly has two main objectives: 1) divide or decompose the image into parts for further processing, 2) perform change in organizing the pixels of image into higher-level units so that the objects become more meaningful.

2. Literature Survey

Benjamin Z. Yao, Xiong Yang, Liang Lin, Mun Wai Lee and Song-Chun Zhu [1] proposed an image parsing to text description that generates text for images and video content. Image parsing and text description are the two major tasks of his framework. It computes a graph of most probable interpretations of an input image. This parse graph includes a tree structured decomposition contents of scene, pictures or parts that cover all pixels of image.

Over past decade many researchers form computer vision and Content Based Image Retrieval (CBIR) domain have been actively investigating possible ways of retrieving images and videos based on features such as color, shape and objects[2][3][4][5][6].

Paper [7] introduced by Yi-Ren Yeh, Chun-Hao Huang, and Yu-Chiang Frank Wang presents a novel domain adaptation approach for solving cross domain pattern recognition problem where data and features to be processed and recognized are collected for different domains

S. Shahnawaz Ahmed, Shah Muhammed Abid Hussain and Md. Sayeed Salam [8] introduced a model of image to text conversion for electricity meter reading of units in kilo-watts by capturing its image and sending that image in the form of Multimedia Message Service (MMS) to the server. The server will process the received image using sequential steps: 1) read the image and convert it into a three dimensional array of pixels, 2) convert the image from color to black and white, 3) removal of shades caused due to nonuniform light, 4) turning black pixels into white ones and vice versa, 5) threshold the image to eliminate pixels which are neither black nor white, 6) removal of small components, 7) conversion to text.

In [10] Fan-Chieh Cheng, Shih-Chia Huang, and Shanq-Jang Ruan gave the technique of eliminating background model form video sequence to detect foreground and objects from any applications such as traffic security, human machine interaction, object recognition and so on. Accordingly, motion detection approaches can be broadly classified in three categories: temporal flow, optical flow and background subtraction.

Iasonas Kokkinos and Petros Maragos [11] formulate the interaction between image segmentation and object recognition using Expectation-Maximization (EM) algorithm. These two tasks are performed iteratively, simultaneously segmenting an image and reconstructing it in terms of objects. Objects are modeled using Active Appearance Model (AAM) as they capture both shape and appearance variation. During the E-step, the fidelity of the AAM predictions to the image is used to decide about assigning observations to the object. Firstly start with oversegmentation of image and then softly assign segments to objects. Secondly uses curve evolution to minimize criterion derived from variational interpretation of EM and introduces AAMs as shape priories.

Mina Makar, Vijay Chandrasekhar, Sam S. Tsai, David Chen and Bernd Girod [13] proposed that streaming mobile augmented reality applications requires both real-time recognition and tracking of objects of interest in a video sequence. A temporally coherent keypoint detector and design efficient interframe predictive coding techniques for canonical patches, feature descriptors and keypoint locations. Mobile Augmented Reality (MAR) Systems are more important with growing interest in applications that use image based retrieval on mobile devices. Streaming MAR applications require real-time recognition and tracking of objects of interest.

3. Basic System Architecture

The system architecture of image to text as well as speech conversion system is illustrated in Fig. 1. There are various phases in which our system will work. They are:

1) Input image: This image is the image entered as input image to the system.
2) Pre-processing: In this phase pre-requesit processing on the input images such as removing noise and making it more usable to be recognizable for the system are carried out.
3) Feature extraction: This phase is one of the important one. Extracting priliminary features and dividing them into geometric elements like arc, line and circle and comparing these elements with known set of characters which are store in the database.
4) Matching with database: After feature extraction, system requires assistance of database in order to recognize the objects in the image, so matching is done.
5) Generate text: After successful recognition of objects, it is now important task of the system to generate appropriate text for every input image.
6) Speech output: The appropriate speech output for the generated text is given in the final phase.

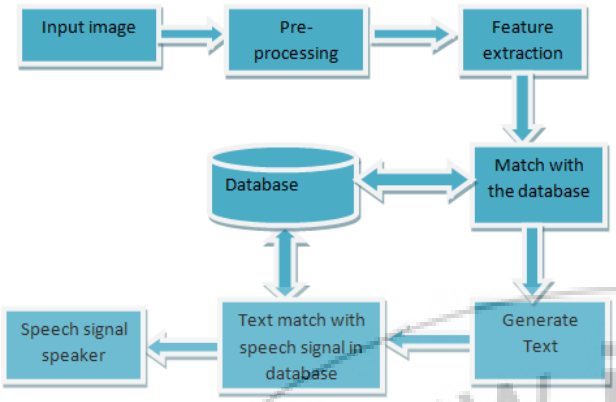

Figure 1: System architecture of image to text as well as speech conversion

4. Conclusion

In this way we surveyed many techniques which are necessary to implement image to text as well as speech system. Our contribution towards this work will surely be helpful for blind as well as physically disabled people of our society. This is a small help form our side for such people to make them more interact able with real world. More focus is on recognition of object in an image. This will ultimately result in identifying important objects from an image. This paper contains an abstract view of various technique proposed in recent past year for image to text conversion and text to speech conversion.

5. Acknowledgment

The authors would like to thank Chairman Shri Ajeenky D. Patil of D Y Patil Knowledge city and also all the management committee. Thankful to the Principal Dr. Uttam Kalawane, Guide, Head, Coordinators, Colleagues of the Department of Computer Engineering, Dr. D. Y. Patil School of Engineering and Technology, Charoli (B.K.) via Lohegaon, Pune, Maharashtra, India, for their support, encouragement and suggestions.

References

[1] Benjamin Z. Yao, Xiong Yang, Liang Lin, Mun Wai Lee and Song-Chun Zhu, "I2T: Image Parsing to Text Description" IEEE Conference on Image Processing, 2008 .

[2] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta and R. Jain, "Content-based image retrieval at the end of the early years," IEEE Transactions PAMI, vol 22, no. 12, 2000.

[3] Y. Rui, T. S. Huang, and S. F. Chang, "Image retrieval: Current techniques, promising directions, and open issues," Journal of Visual Communication and Image Representation, vol. 10,1999.

[4] M. S. Lew, N. Sebe, C. Djeraba, and R. Jain, "Content-based multimedia information retrieval: State of the art and challenges," ACM Transactions on Multimedia Computing, Communications, and Applications, vol. 2, no. 1, pp. 1-19, Feb. 2006.

[5] C. Snoek and M. Worring, "Multimedia video indexing: A review of the state-of-the-art," Multimedia Tools Appl, vol. 25, no. 1, 2005.

[6] R. Datta, D. Joshi, J. Li, and J. Z. Wang, "Image retrieval: Ideas, influences, and trends of the new age," ACM Computing Surveys, vol. 40, no. 2, pp. 1-60, Apr. 2008.

[7] Yi-Ren Yeh, Chun-Hao Huang, and Yu-Chiang Frank Wang, "Heterogeneous Domain Adaptation and Classification by Exploiting the Correlation Subspace," IEEE Transactions on Image Processing, vol. 23, no. 5, May 2014.

[8] S. Shahnawaz Ahmed, Shah Muhammed Abid Hussain and Md. Sayeed Salam, "A Novel Substitute for the Meter Readersin a Resource Constrained Electricity Utility" IEEE Trans. On Smart Grid, vol. 4, no. 3, Sept. 2013.

[9] A. Abdollahi, M. Dehghani and N. Zamanzadeh, "SMS-based reconfigurable automatic meter reading system," in Proc. 16th IEEE Int. Conf. Control Appl. Part IEEE Multi-Conf. Sust. Control Singapore, Oct. 1-3, 2007, pp. 1103-1107.

[10] Fan-Chieh Cheng, Shih-Chia Huang and Shanq-Jang, "Illumination-Sensitive Background Modeling Approach for Accurate Moving Object Detection," IEEE Trans. On Broadcasting, vol. 57, no. 4, Dec 2011.

[11] Iasonas Kokkinos and Petros Maragos, "Synergy between Object Recognition and Image Segmentation using the Expectation-Maximization Algorithm", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, no. 8, Aug. 2009.

[12] T. Cootes, G. J. Edwards and C. Taylor, "Active Appearance Models," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 23, no. 6, pp. 681-685, June 2001.

[13] Mina Makar, Vijay Chandrasekhar, Sam S. Tsai, David Chen and Bernd Girod, "Interframe Coding of Feature Description for Mobile Augmented Reality" IEEE Trans. Image Processing, vol. 23, no. 8, Aug 2014.

[14] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," International Journal of Computer Vision, vol. 60, no. 2, pp. 91-110, Nov. 2004.

[15] M. Nagamani, S. Manoj Kumar, S. Uday Bhaskar, "Image to Speech Conversion System for Telugu Language", International Journal of Engineering Science and Innovative Technology (IJESIT), vol. 2, Issue 6, November 2013.

[16] Yassin M. Y. Hasan and Lina J. Karam, "Morphological text extraction from images" IEEE Transactions on Image Processing, vol. 9, no. 11, Nov. 2000.

[17] Huiping Li, David Doermann and Omid Kia, "Automatic Text Detection and Tracking in Digital Video" IEEE Transactions on Image Processing, vol. 9, no. 1, Jan. 2000.

[18] Z. W. Tu, X. R. Chen, A. L. Yuille and S. –C. Zhu, "Image parsing: Unifying segmentation, detection and

Recognition," International Journal of Computer Vision, vol. 63, no. 2, pp. 184-203.

[19] J. Shi and J. Malik, "Normalized cuts and image segmentation," IEEE Transactions PAMI, vol. 22, no. 8, pp. 888-905, 2000.

[20] J. Canny, "A Computational Approach to Edge Detection," Reading in Computer Visions: Issues, Problems, Principles and Paradigms, pp. 184-203.

[21] S. J. Belongie, J. Malik and J. Puzicha, "Shape matching and object recognition using shape contexts," IEEE Trans. PAMI, vol. 24, no. 4, pp. 509-522, Apr. 2002.

[22] D. Xu and S. F. Chang, "Video event recognition using kernel methods with miltilevel temporal allignment," IEEE Transactions PAMI, vol. 30, no. 11, pp. 1985-1997, Nov. 2008.

[23] P. Felzenszwalb and D. Huttenlocher, "Pictoral Structures for Object Recognition," International Journal of Computer Vision, vol. 61, no. 1, pp. 55-79, 2005.

Author Profile

Mrunmayee G. Patil Research Scholar Dr. D. Y. Patil School of Engineering & Technology, Pune, University of Pune, Maharashtra, India. She has received her Bachelor's Degree in Information Technology from Padmashree. Dr. D. Y. Patil Institute of Engineering and technology, Pimpri, Pune, Maharashtra with first class distinction, University of Pune. Currently she is pursuing her Master's Degree in Computer Engineering from Dr. D. Y. Patil School of Engineering & Technology, Pune, University of Pune.

Prof. Ramesh. M. Kagalkar was born on Jun 1st, 1979 in Karnataka, India and presently working as an Assistant Professor, Department of Computer Engineering, Dr. D. Y. Patil School Of Engineering and Technology, Charoli, B.K.Via –Lohegaon, Pune, Maharashtra, India. He is a Research Scholar in Visveswaraiah Technological University, Belgaum. He had obtained M.Tech (CSE) Degree in 2005 from VTU Belgaum and he received BE (CSE) Degree in 2001 from Gulbarga University, Gulbarga. He is the author of text book Advanced Computer Architecture which cover the syllabus of Visveswaraiah Technological University, Belgaum. He has published many research paper in International and international conference.