A Strategy for Image Description and Speech Synthesis Generation in Hindi

Vasundhara kadam, Arti Waghmare

[1] *ME Second Year Student,Dr D Y Patil School of Engineering and technology, Lohegaon, Pune*

[2] *Research Scholar and Asst. Professor, Computer Engg. Department,Dr. D Y Patil School of Engg and Technology, Lohegaon, Pune,Maharashtra,India*

---***---

Abstract - *Picture preparing is a quickly developing field of exploration. Pictures are of various record positions and of various things, places, people, logical, visionary and some such. A picture is a gathering of a few pixels organized in lines and segments. These pictures are caught, handled and put away for different employments. For basic individuals it is anything but difficult to recognize and investigate general pictures yet for the visually impaired and physically incapacitated individuals it is troublesome. Lamentably, there is no earlier medium or interface for such penniless individuals to speak with the world. Blind or outwardly hindered individuals are generally those individuals who are disregarded by the general public, so there is dependably a need to help such individuals. Henceforth, we propose another procedure of changing over pictures into content and in addition discourse utilizing methods gave by picture preparing like pre-handling, picture division, edge identification, object discovery and discourse amalgamation. In this paper we first acquaint picture with content change requirement for visually impaired individuals and framework diagram of picture to content and discourse transformation framework. Edge recognition assumes a vital part in this framework where canny edge discovery calculation is utilized to distinguish objects from pictures. Object acknowledgment is done on the premise of shading, size, surface and state of the article.*

***Key Words*: Image Processing, Image Segmentation, Speech Synthesis, Text to Speech Conversion, Edge Detection.**

1.INTRODUCTION

Image processing is one of the majority increasing field in research and technology in today's world. Image processing use hardware and software as computing resources to provide an interface for image processing. It uses a variety of techniques such as image filtering, image pre-processing, image segmentation, image compression, image editing and manipulation, 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 image at the real co-ordinate position (r, w). Images of dissimilar file formats. These file format helps us to distinguish dissimilar types of images. In today's world in attendance around 285 million people who are visually impaired; out of which 39 million are visually impaired and 246 have low vision. Such people have extremely low scope to recognize what precisely happening in their surroundings. There is no such crossing point which is easily obtainable for such disabled people to relate with the world. Only if efficient interface for such people is of great need.

The proposed work of image to text and speech translation system is used to develop a user friendly interface as well as cost efficient for visually impaired people. The main inspiration is to give blind person having a friendly speech interface with computer and to permit those people who are physically and visually challenged to employ the system for under-system.

Image processing is a fast growing field in all aspect and all fields. Images are a huge concern to be discussed worldwide. Images can be of any neighboring, scene, things or humans. These images can assist to simply recognize things, objects, place, human beings etc. for ordinary people. But the visually impaired people and disabled people cannot do these things. Now days there above 285 million of people who are visually impaired, out of which 39 million are visually impaired and 246 have low vision. Due to low vision this people have less scope to understand what exactly is going on in their current environment. So there was a huge requirement to implement an interface for such people. This system will eventually help visually impaired people to identify objects and things around them. They often depend on common people or their relatives for leading day to day life. That is they are dependent. They cannot simply go out of their places and lead life in an independent manner as other common people do. So in order to make the life of a visually impaired person additional simple we propose this system. Images are being rehabilitated into related text and speech which will be heard by the visually impaired person and they will act accordingly.

2. Literature Review

Literature reviews are minor sources and as such, do not report any new or original trial work. The ground of image processing and computer vision is on great demand in today's world.

Girish Kulkarni et al. provided a system to robotically produce text descriptions from images [1]. This method consists of two components, (1) The content coming up with, smoothes the output of laptop that based on vision detection and recognition algorithms with statistics strip-mined from massive pools of visually descriptive text to work out the simplest content words to employ to explain a image, (2) Surface realization, choose words to build text sentences supported the predictable content and common statistics from text. Also multiple approaches for the surface realization step and assess every exploitation automatic procedures of parallel to human generated reference descriptions.

Early work on relating words and pictures focused on associating individual words with image regions for tasks such as clustering, auto-annotation or auto-illustration [2]. Other work has finished make use of text as a source of noisy labels for predicting the content of an image. This works particularly well in inhibited recognition scenario for recognizing fussy classes of objects such as for tagging faces in news photographs with associated captions [3],[4] or characters in television or movie videos with related scripts [5],[6]. Other object classes that have been considered comprise animal images from the web [7-9] where text from the containing web page can be utilized for better image classification.

In [11] projected an image parsing to text description that generates text for images and video content. Two major task of this framework are Image parsing and text description. It computes a graph of most possible interpretations of an input image. Three formatted breakdown content from various scenes, picture or any parts those will cover all pixels of image. Natural language creation constitutes one of the basic investigation problems in Natural Language Processing (NLP) and is heart to a wide range of such applications such as dialogue systems, summarization, machine translation, and machine-assisted revision. Regardless of considerable progression within the last decade, natural language creation still residue an open research problem. Most preceding work in this filed on automatically producing captions or descriptions for images is based on recovery and summarization.

For instance, Aker and Gaizauskas [12] rely on Global Positioning System (GPS) metadata to entry applicable text documents and Feng and Lapata assume appropriate documents are provided.

The process of generation then becomes one of combining or cutting relevant documents, in some cases determined by keywords expected from the image content [13].

From the computer vision society, work has considered matching a whole input image to a database of images with captions [14], [15]. The caption of the best matching image up fashion, initial from what computer vision systems distinguish in an image and then constructing a original caption around those predictions, with text figures to smooth these (sometimes) earsplitting vision predictions.

3. Proposed Architecture

In the proposed system our major focus is on identifying input image and converting it into appropriate text description in Hindi language. So initially we generate a database of images with description and it contains object and its actions performed in image. The input to the system can be images form database or captured image. These images when in use as input from the system, system checks for kind of images and in database in organize to identify the objects from the image. Once objects are detected after that it identifies the image and system gives text description as output. This kind of approach may help to improve ease of access of images for the visually impaired and creating text-based indexes of visual data for civilizing image recovery algorithms.

3.1 Architecture Overview

The proposed system consists of two major module Training and Testing is shown in figure 1.

3.1.1 Training Module:

The training section is employ to instruct images and stored on database with objects and proceedings which describing object. This section is handled by admin who is responsible for data training. This stage comprised of two stages. In the first, content development, the occasionally deafening output of computer vision recognition algorithms is smoothed with statistics collected from visually descriptive natural language.

Once the content to be used in formation is chosen, the next stage is exterior insight, finding words to explain the chosen content. Once again text statistics are used to choose shell realization that is more similar to constructions in commonly used language. After that each image is processed out and converts it hooked on matched description and stored into database.

3.1.2 Testing Module

This module trial the images and gets result only if at least one image is trained. In this stage all methods on images are performed same as in training phase such as object discovery, Stuff discovery, and feature extraction. The final stage of the system is feature matching of images with all that are stored into database.

3.2 Proposed Architecture Diagram

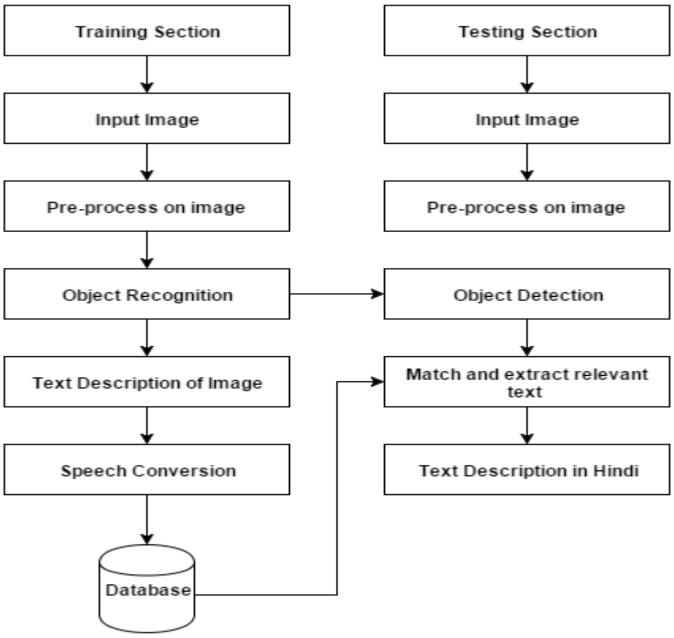

Fig. 1. Proposed Architecture

4. Expected Results

For evaluation at least 100 images are stored into database after processing. Objects from images are detected and suitable text is associate to image. The data dictionary involves of tables generated in the database by methods formed command files, modified for every maintained back-end DBMS.

TABLE I. EXPECTED RESULTS

Sr.	Expected Results		
No.	*Image*	***Actual Result***	***Predicted Result***
1		यह एक आकाश एक सड़क और एक बस की तस्वीर है	आकाश बस सड़क
2		यह दो कुत्तों की तस्वीर , दूसरी प्यारे कुत्ते के पास पहला कुत्ता है	कुत्ते के पास दूसरा कुत्ता है
3		हवाई जहाज के साथ आसमान के नीचे दो व्यक्ति	आसमान दो व्यक्ति
4		आकाश के नीचे एक गाय है	आकाश के नीचे एक गाय
5		घास के साथ बकरी	घास के साथ बकरी

5. Conclusion

In proposed system we have apply a easy and fast technique which works duly for identify image and change it into text as well as speech. It is low time-consumption approach, so that the real time recognition ratio is achieved easily. In the projected system, Canny edge detection algorithm is used which will distinguish the input image by detecting the edges of objects in the image. It is able to handling the dissimilar input images and translates them into text and speech. The projected system is intended to interpret the dataset contains the number of images that are taken from multiple user of diverse size which helps to identify the correct output to several user using the system. The projected system is trained on predefined dataset.

Acknowledgment

The authors would like to thank the researchers as well as publishers for making their resources available and teachers for their guidance. We are thankful to the Arti Waghmare for her valuable guidance and constant guidelines also thank full the computer department staff of DYPSOET, Lohegoan, Pune and support. Finally, we would like to extend a heartfelt gratitude to friends and family members.

REFERENCES

[1] irish Kulkarni, Visruth Premraj, Vicente Ordonez, Sagnik Dhar, Siming Li, Yejin Choi, Alexander C. Berg and Tamara L. Berg, "Baby Talk: Understanding and Generating Simple Descriptions",IEEE Transactions On Pattern Analysis And Machine Intelligence, Vol. 35, No. 12, December 2013.

[2] K. Barnard, P. Duygulu, N. de Freitas, D. Forsyth, D. Blei, and M. Jordan, Matching Words and Pictures, J. Machine Learning Research, Vol. 3, pp. 1107-1135, 2003.

[3] K. Barnard, P. Duyguly, and D. Forsyth, "Clustering Art", Proc. IEEE Conference On Computer Vision and Pattern Recognition, June 2001.

[4] T.L. Berg, A.C. Berg, J. Edwards, and D.A. Forsyth, "Whos in the Picture?", Proc. Neural Information Processing Systems Conference, 2004.

[5] T.L. Berg, A.C. Berg, J. Edwards, M. Maire, R. White, E. Learned- Miller, Y.-W. Teh, and D.A. Forsyth, "Names and Faces", Proc. IEEE Conference On Computer Vision and Pattern Recognition, 2004.

[6] M. Everingham, J. Sivic, and A. Zisserman, "Hello! My Name Is. Buffy Automatic Naming of Characters in TV Video", Proc. British Machine Vision Conference, 2006.

[7] C. Lampert, H. Nickisch, and S. Harmeling, "Learning to Detect Unseen Object Classes by Between-Class Attribute Transfer", Proc. IEEE Conference Computer On Vision and Pattern Recognition, 2009.

[8] T.L. Berg and D.A. Forsyth, "Animals on the Web", Proc. IEEE Conference On Computer Vision and Pattern Recognition, 2006.

[9] L.-J. Li and L. Fei-Fei, "OPTIMOL: Automatic Online Picture Collection via Incremental Model Learning", International Journal of Computer Vision, Vol. 88, pp. 147-168, 2009.

[10] F. Schroff, A. Criminisi, and A. Zisserman, "Harvesting Image Databases from the Web", Proc. 11th IEEE International Conference On Computer Vision, 2007.

[11] 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, 2010.

[12] A. Aker and R. Gaizauskas, "Generating Image Descriptions Using Dependency Relational Patterns", Proc. 28th Ann. Meeting Assoc. for Computational Linguistics, pp. 1250-1258, 2010.

[13] Y. Feng and M. Lapata,"How Many Words Is a Picture Worth? Automatic Caption Generation for News Images", Proc. Assoc. For Computational Linguistics, pp. 1239-1249, 2010.

[14] A. Farhadi, M. Hejrati, A. Sadeghi, P. Young, C. Rashtchian, J. Hockenmaier, and D.A. Forsyth, "Every Picture Tells a Story: Generating Sentences for Images", Proc. European Conference On Computer Vision, 2010.

[15] V. Ordonez, G. Kulkarni, and T.L. Berg,"Im2text: Describing Images Using 1 Million Captioned Photographs", Proc. Neural Information Processing Systems, 2011.

[16] G. Kulkarni, V. Premraj, S. Dhar, S. Li, Y. Choi, A.C. Berg, and T.L. Berg, "Babytalk: Understanding and Generating Simple Image Descriptions", Proc. IEEE Conference On Computer Vision and Pattern Recognition, 2011.

[17] 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, August 2009.

[18] Fan-Chieh Cheng, Shih-Chia Huang and Shanq-Jang Ruan, "Illumination- Sensitive Background Modelling Approach for Accurate Moving Object Detection", IEEE Transactions On Broadcasting, Vol. 57, NO. 4, December 2011.

[19] M. Oral and U. Deniz, "Centre Of Mass ModelA Novel Approach To Background Modelling For Segmentation Of Moving Objects", Image Vis. Comput., Vol. 25, pp. 13651376, August 2007.

BIOGRAPHIES

Vasundhara Kadam is M.E 2nd year student of Computer Engg. Department,Dr. DY Patil School of Engg and Technology,Lohegaon, Pune.
email:vasundharakadam99@gmail.com

Arti Waghmare is research scholar and asst. professor of Computer Engg. Department, Dr. D Y Patil School of Engg and Technology, Lohegaon, Pune