Automatic Video Descriptor for Human Action Recognition

Minoli Perera
Department of Computing
Informatics Institute of Technology
Sri Lanka
minoliperera91@gmail.com

Cassim Farook
Department of Computing
Informatics Institute of Technology
Sri Lanka
cassim.f@iit.ac.lk

A.P. Madurapperuma
Department of Electrical and
Computer Engineering
Open University of Sri Lanka
Sri Lanka
apmad@ou.ac.lk

Abstract— **Assistive software such as screen readers are unable to describe images or videos for visually impaired people. Although recent research have found ways to describe an image automatically, describing the content of a video is still an ongoing issue. Visually impaired people find it difficult to understand video content without an indication of sound. The current solution of video description is only provided through digital television and for selected programs and movies. As an initiative to describe video content for visually impaired people, the solution acts as a video player which automatically understands the ongoing human action on screen, associates textual descriptions and narrates it to the blind user.**

The human actions in the video should be recognized in real time, hence fast, reliable feature extraction and classification methods must be adopted. A feature set is extracted for each frame and is obtained from the projection histograms of the foreground mask. The number of moving pixels for each row and column of the frame is used to identify the instant position of a person. Support Vector Machine (SVM) is used to classify extracted features of each frame. The final classification is given by analyzing frames in segments. The classified actions will be converted from text to speech.

Keywords— *Human Action Recognition, Classification, Support Vector Machine, Machine Learning, Feature Extraction, Action Detection*

I. INTRODUCTION

According to World Health Organization [1] 2014 statistics, 285 million people are estimated to be visually impaired worldwide: 39 million are blind and 246 have low vision. Blindness can be of two types: total blindness and partial blindness. Totally blind people completely lack light perception and form perception, while partially blind are able to see to some extent with corrective aid. Both these types can be addressed as visual impairment. Blindness can occur from birth, due to diseases or accidents. The National Federation of the Blind encourages blind persons to use alternative methods to engage in activities that normal people with vision would do using their eyes. Although it is possible for a blind person to ask for help from others, most blind people prefer navigating the world as independently as possible.

"Accessibility to digital information for all, including people with disabilities, is one of the major social challenges of our society" [2]. There are many gadgets and assistive software invented to support blind people with their day today activities. Screen reader is such a software, which helps blind people to interact with computers. It enables blind users to know what is on the computer screen through speech. A screen reader makes it easier for the blind to accomplish tasks such as writing a document, checking email, or using social networks by reading texts on screen, but screen readers are unable to describe images or videos.

According to the recent research in Newsroom.fb [3] and Next at Microsoft [4], Facebook and Microsoft have created an application to describe images, but describing a video is still an ongoing issue. Amount of videos available on the internet is continually growing and its consumption is continuously increasing. Video content appears as one of the first-choice mediums to share information. Therefore, video accessibility for the blind is a current need in the world.

As an initiative to enhance video content accessibility to visually impaired persons, the proposed solution will allow computers to describe human actions happening on screen.
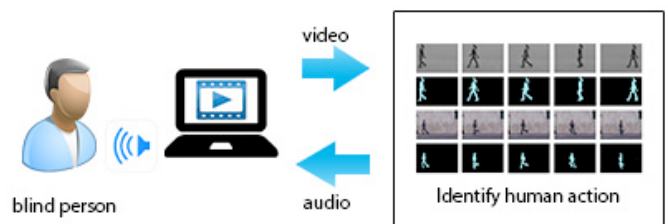

Fig. 1.   Rich picture

II. BACKGROUND AND RELATED WORK

A. Existing Solutions

Table I lists the existing solutions which blind people use to overcome the problem of video accessibility along with the advantages and disadvantages.