IJITCE

ISSN 2347- 3657

International Journal of
Information Technology & Computer Engineering

Email : ijitce.editor@gmail.com or editor@ijitce.com

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 12 , Issue 2, April 2024

OCR BASED IMAGE TEXTTO SPEECH CONVERSION

Ms. Anusha[1] Nalla Niranjan Reddy[2] Manala Bharadwaja[3] P.Indraja[4] T.Raju[5]
Assistant Professor/CSE(DS) TKR College of Engineering and Technology Telangana, India Anusha18@gmail.com
IV Final Year of CSE (DS) TKR College of Engineering and Technology Telangana, India Niranjanreddy2206@gmail.com
IV Final Year of CSE (DS) TKR College of Engineering and Technology Telangana, India manlabharadwaja@gmail.com
IV Final Year of CSE (DS) TKR College of Engineering and Technology Telangana, India parikpalliindraja@gmail.com
IV Final Year of CSE (DS) TKR College of Engineering and Technology Telangana, India rajureddythada789@gmail.com

ABSTRACT

This paper emphasizes the crucial roleofcustomerretentioninthecontemporarybankingsector,recognizingthesubstantialexpensesassociatedwithacquiringnewcustomers.Imagesmay containimportanttextual data that the user may need to edit orstoredigitally.Manualentryofdataistimetaking andmaycontainerrors.Therearemillions of blind people in the world who arevisually impaired. Disability to read has a largeimpact on the life of visually impaired people.TheProposedsystemhelpsthevisuallyimpairedperson to hear the text.The main idea of thisproject is optical Character recognition which isusedtoconverttextcharacterintotheaudiosignal. Tkinter for GUI, pytesseract for imageprocessingandcharacterrecognitionandpyttsx3foraudioconversionareused.

Keywords:

OCR, Tesseract, video text,Image Capturing,ImageToTextConverter, PythonTSX3TextToSpeechSynthesizer,TranslationandSpeechOutput.

INTRODUCTION

In the recent era, visual text in natural ormanmade scenes might carry very important anduseful information. Therefore, the scientists havestartedtodigitizetheseimages,extractandinterpretthedata byusingspecific techniques.

Opticalcharacterrecognitionisemployedtorecognizeandextractthewordsandfinallytheextractedtext is convertedtoappropriate speechusing text-to-speech synthesizer. -to-speech (TTS)conversionistheprocessofconvertingwrittentext into spoken words using computer software. Theresultingaudiooutputcanbecustomizedbyadjustingthespeed,pitch,andotherparametersTextto matchthe user'spreferences.

LITERATURESURVEY

SurveyofmajorareasofprojectDeepLearningDeeplearningisasubsetofmachinelearning that involves the use of artificial neuralnetworks to learn from large volumes of data. Thistechnologyhasbeenaroundfordecades,butonlyin recentyearshasitseensignificantgrowthanddevelopment.Withtheriseofbigdataandadvancements in computing power, deep learninghas become increasingly popular in solving complexproblems across a wide range of industries. Deeplearning is modeled after the structure and functionof the human brain, using artificial neural networksto process and analyze data. These neural networksconsist of layers of interconnected nodes, each ofwhich performs a specific function. The first layerofnodesreceivestheinputdata,whichisthenprocessed through subsequent layers to produce afinal output. The more layers a network has, thedeeperitis,hencethename"deeplearning."Oneofthemainadvantagesofdeeplearningisitsabilitytoautomatically learn and improve from experience.Thisisaccomplishedthroughaprocesscalledbackpropagation, which allows the network to adjustitsparametersbasedontheerrorbetweenitspredictions and the actual outputs. By iterativelyupdating the weights and biases of the network, itcan gradually improve its accuracy over time.

Deeplearninghasseentremendoussuccessinawiderange of applications, from image recognition tonaturallanguageprocessingandspeechrecognition. Forexample,imagerecognitionmodelscanbetrainedtoaccuratelyidentifyobjectswithinphotos,

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 12 , Issue 2, April 2024

allowing them to be used for tasks such as self-drivingcars,securitysurveillance,andmedicalimaging . Natural language processing models can betrainedtounderstandandinterprethumanlanguage,enabling applications such as chatbots and voiceassistants.Oneofthekeyadvantagesofdeeplearning is its ability to handle unstructured data,suchasimages,text,andaudio.Traditionalmachinelearningalgorithmsrelyonstructureddata,whichistypicallyrepresentedinatabularformat.Deeplearning, on the other hand, can learn directly fromraw data, allowing it to extract patterns and insightsthat may not be immediately apparent in structureddata. 4 Despite these challenges, deep learning hasseen significant growth and development in recentyears.Advancesinhardware,suchasthedevelopment of specialized deep learning chips andcloud-basedcomputingresources,havemadeitmoreaccessibletoawiderrangeofusers.Additionally, the development of new algorithmsandtechniques,suchastransferlearningand reinforcementlearning,haveopenedupnewpossibilities for deep learning in a wide range ofapplications.Inconclusion,deeplearningisapowerful subset of machine learning that has theability to learn and improve from experience. Itsabilitytohandleunstructureddatahasmadeitparticularlyvaluableinfieldssuchasimagerecognition and natural language processing. Whilethere are challenges associated with deep learning,suchastheneedforlargevolumesofdataandcomputing resources, advances in technology andalgorithmshavemadeitmoreaccessibleandapplicable to a wider range of users and applications.PytesseractPytesseractusesacombinationofseveralalgorithmstoperformopticalcharacterrecognition (OCR) on images. The main algorithmused by pytesseract is the Tesseract OCR engine,which was originally developed at Hewlett-PackardLaboratoriesinthe1980sandhassincebeendevelopedandmaintainedbyGoogle.TheTesseractOCR engineusesacombinationofimageprocessingtechniques,statisticalmodels,andmachinelearningalgorithms to recognize characters in an image. Theengine works by first preprocessing the image toenhanceitscontrastandremovenoise.Itthensegments the image into individual characters orwords and applies a set of rules to recognize eachcharacter or word. To improve the accuracy of OCR,pytesseract also uses several other algorithms andtechniques. For example, pytesseract can performadaptive thresholdingto improve the contrastofthe image and reduce noise. It can also apply language-specificdictionariesandmodelstoimprovetherecognitionofwordsandphrasesInadditiontothesealgorithms, pytesseract also provides several optionsand parameters that can be used to fine-tune theOCR process. For example,you canspecify thelanguage of the text to be recognized, adjust thecontrastandbrightnessoftheimage,andcontroltheamountofnoisereductionappliedtotheimageOverall,pytesseractusesacombinationofalgorithmsandtechniquestoperformOCRonimages with high accuracy and reliability. 5 TheTesseractOCRengineisthebackboneofthepytesseractlibrary,butotheralgorithmsandtechniquesareused toenhancetheOCRprocessandimprovetheaccuracyof theresults.OCROCRstands for Optical Character Recognition, which isthe technology used to extract text from scanneddocuments,images,orotherdigitalfiles.Itinvolvesanalyzingthetextcharactersinanimageandconvertingthemintomachine-readabletextthatcanbe edited, searched, and processed by computers.OCR can handle various types of fonts, includingstandard fonts like Times New Roman and Arial, aswellaslesscommonfontslikehandwritingorcursive. OCR can be used with various file formats,suchasPDFs,JPEGs,TIFFs,andPNGs.OCRtechnologycanbetrainedtorecognizespecificlanguagesorscripts,suchasLatin,Chinese,orArabic.OCRaccuracycanvarydependingonvariousfactors,suchasimagequality,fonttype,andcharactercomplexity.However, modernOCRsystemscanachievehighlevelsofaccuracy,rangingfrom95%to99%.OCRtechnologycanbeenhancedwith additional features, such as natural languageprocessing,documentclassification,anddata extraction.Thesefeaturescanhelpimprovetheaccuracy and efficiency of OCR applications. OCRtechnologyhasadvancedsignificantlyinrecentyears,withthedevelopmentofdeeplearningalgorithmsandneuralnetworks.Thesetechnologiescanimprovetheaccuracyandspeedof OCRapplications, especially for complex documents orhandwriting.OCRisnotfoolproofandcanstillmake errors, especially with poorly scanned or low-quality images. Therefore, it is always a good ideato review and verify the OCR output to ensure itsaccuracy. Overall, OCR is a versatile and powerfultechnologythatcanhavesignificantbenefitsinvariousindustries,includingdocumentmanagement, dataentry,andaccessibility.Withcontinuedadvancementsin OCRtechnology,we

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 12 , Issue 2, April 2024

canexpecttoseeevenmoreinnovativeapplicationsand usecasesinthefuture. OCRtechnologies.

Simultaneously,industriesareincreasinglyadoptingdata-driven decision-making, driven by the growthof big data and analytics capabilities. In the airlineindustry,bigdataanalyticsholdspromiseforoptimizing aviation operations, enhancing customerservice,andrevenuemanagement.Airfarepriceprediction,acomplextask,hasadvancedwithmachinelearningalgorithmssuchasLinearRegression,SupportVectorMachines,andRandomForests. Recent studies explore more sophisticatedmodels like Deep Regressor Stacking to improveairfare price prediction accuracy, showcasing thecontinuous evolution of predictive modeling in theairlineindustry.

However,challengespersistindataavailabilityandorganization,particularlyconcerning airline ticket data. While datasets likeT100 and DB1A/1B exist, the limited associationbetweenpricesandspecificflightinformationposesobstacles to comprehensiveanalysis.Researchersoften rely on web crawling or private data sources,makingreplicationandcomparisonchallenging.

1.PROBLEMDEFINATION

Implementthesolutionusingappropriateprogramming languages, libraries, and frameworksfor image processing, machine learning, and speechsynthesis.Considerdeployingthesystemonplatforms suitable for the intended use case, such asdesktopapplications,mobiledevices,orcloudservices.

By addressing the outlined components, challenges,andsolutionapproach,youcandevelop aneffectiveOCR-basedimagetext-to-speechconversionsystem.

Limitationsofexistingsystem

Accuracy:Measuretheaccuracy oftextextraction andspeechsynthesisagainstgroundtruthdata.Efficiency:Evaluatethesystem'sspeedandresourceusage,especiallyforreal-timeapplications.User Experience: Gather feedback onthenaturalnessandunderstandabilityofthesynthesizedspeech.Robustness:Testthesystem's performanceundervariousconditions,includingdifferenttypesofimagesandtext.

Proposedsystem

ProposedsystemusespytesseractforOptical CharacterRecognitionwhichinternallyusesLongShort-TermMemorywhichisapartofRecurrent Neural Network which is more suitableforhandwrittentextsandtorecognizelargeportionof textual data. Not only from the image but it alsoextracts the text from the videos by splitting thevideointoframesusingOpenCV.Furthertheextracted text from images and videos is convertedinto speechbyusingpyttsx3.

FIGURES

Figure1:SYSTEMARCHITECTUR

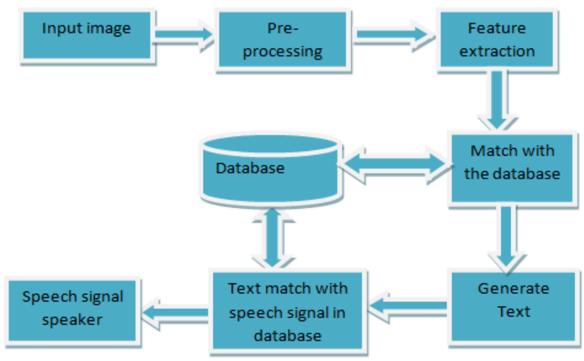


Figure2

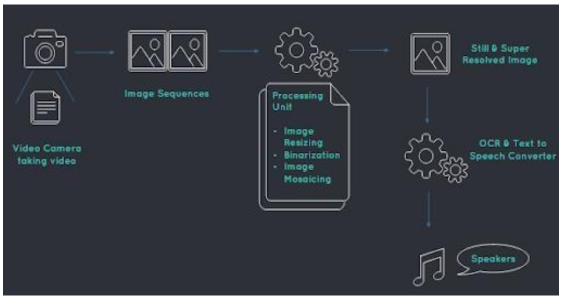


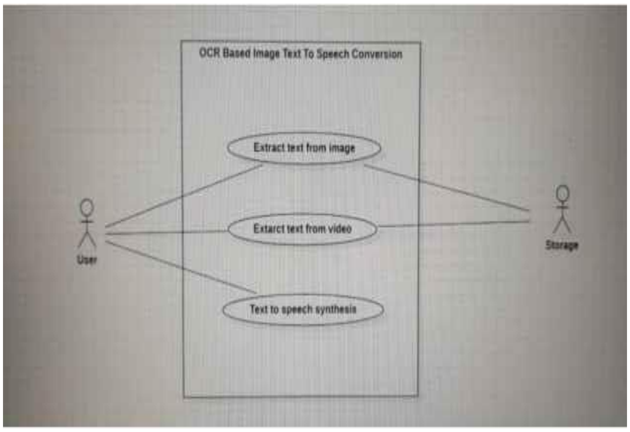
Image working modeloftheimagesandvideos

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 12 , Issue 2, April 2024

UseCaseDiagram



MODULES

1. ImageInputModule:

Responsibleforacceptinginputimagescontainingtext.Handles image loading, whether from files, cameras,orothersources.

2. ImagePreprocessingModule:

PerformspreprocessingtechniquestoenhanceOCRaccuracy. Includesoperationslikeresizing,noisereduction,contrastenhancement, and binarization.

3. OpticalCharacterRecognition(OCR)Module:

Text Detection: Identifies regions of text within the inputimage.Text

Recognition:Recognizesandextractstextfromthedetected regions.

Post-processing:RefinestheOCRoutputtoimproveaccuracyandformatting.

4. TextProcessingModule:

Text Cleaning: Removes noise, artifacts, and unwantedcharactersfromthe extracted text.

Language Identification: Determines the language of thetextforaccuratepronunciationduringspeechsynthesis.

Text Normalization: Standardizes the format and structureofthe textfor betterTTSperformance.

5. TextToSpeechConversion:

Text Analysis: Analyzes the processed text for prosody,intonation,andemphasis.

VoiceSelection:Optionallyallowsuserstochoosedifferentvoicesoraccents for speechsynthesis.

SpeechSynthesis:ConvertstheprocessedtextintoaudiblespeechusingTTSenginesormodels.

6. OutputModule:

Handlestheoutputofsynthesizedspeech.Playsthesynthesizedspeechthroughspeakersorheadphones.Provides a graphical user interface (GUI) foruser interactionandfeedback.

7. IntegrationandControlModule:

Orchestratestheflowofdataandcontrolbetweendifferentmodules.Manages error handling, logging, and feedbackmechanismsforsystemmonitoringandimprovement.Provides APIs or interfaces for integrationwithotherapplicationsor services.

8. PerformanceOptimizationModule:

Optimizesthesystemforefficiencyandspeed.Utilizestechniques like parallel processing, model optimization,andhardwareaccelerationforimproved performance.

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 12 , Issue 2, April 2024

Monitorsandoptimizesresourceusagetoensurescalabilityandreliability.

9. DeploymentModule:

Handlesdeploymentaspectssuchasplatformcompatibility,scalability,andsecurity,Supportsdeploymentonvariousplatforms,includingdesktop,mobile,andcloudenvironments.Implementssecuritymeasures to protect user data and ensure privacy duringimageprocessingand speechsynthesis.

DEPLOYMENT

ChooseDeploymentPlatform:

Determine the platform(s) on which the system will bedeployed, such as desktop, mobile, or cloud. Considerfactorsliketargetaudience,scalabilityrequirements,andresourceconstraints.

SoftwarePackaging:

Package the system components into deployable softwareartifacts suitable for the chosen platform(s).For desktopdeployment,createexecutablefilesorinstallerpackages.For mobile deployment, package the system as amobileapplicationforiOSorAndroid.Forclouddeployment, containerizethesystemusingDockerordeployitasserverlessfunctions.

Platform-specificConsiderations:

Desktop:Ensurecompatibilitywithvariousoperatingsystems (Windows,macOS,Linux)andprovideinstallationinstructionsfor end-users.

Mobile: Publish the application on app stores (e.g., AppleApp Store, Google Play Store) and comply with platform-specificguidelinesand requirements.

Cloud: Deploy the system on cloud platforms like AWS,Azure, or Google Cloud Platform, considering factors likescalability,availability,and cost.Scalabilityand

Performance:

Configure the deployment environment to handle varyingworkloadsandscaleresourcesdynamicallyasneeded. Optimizesystemperformanceforresponsivenessandreliability,especiallyduringpeakusageperiods.

TestingandQualityAssurance:

Conductthoroughtestingofthedeployedsystemtoensurefunctionality,performance,anduserexperiencemeet expectations.Performintegrationtesting,compatibilitytesting,anduseracceptancetestingacrossdifferentdeploymentplatforms.

DocumentationandSupport:

Providecomprehensivedocumentation,includinginstallation instructions, user guides, and troubleshootingtips.Offersupportchannels(e.g.,email,forums,helpdesk)foruserstoseekassistanceandreportissues.

MonitoringandMaintenance:

Implement monitoring tools and processes to track systemperformance, usage metrics, and user feedback. Regularlyupdatethesystemwithbugfixes,enhancements,and securitypatchestomaintainreliabilityand security.

2.ACKNOWLEDGMENTS

IwouldliketoexpressmysinceregratitudetoS.Rajarajacholanfortheirinvaluableguidance,mentorship, and support throughout this machine learningproject.Theirexpertise,insightfulfeedback,andencouragementhavebeeninstrumentalinshapingthedirectionand outcomesofthisendeavor.

IamdeeplythankfultoK.SaiKiranfortheircollaboration, dedication,andcontributionstovariousaspectsoftheproject.Theircollectiveeffortsandteamwork have enriched the project's development andfosteredaconduciveenvironmentforinnovationandlearning.

Iextendmyappreciationto[TKRCollegeofEngineering & Technology] for providing the necessaryresources, facilities, and opportunities that facilitated theexecutionofthisproject.Theircontinuedsupporthasbeenpivotal in our pursuit of excellence and achievement ofprojectmilestones

5.REFERENCES

[1] https://www.irjet.net/archives/V6/i2/IRJET-V6I299.pdf

[2] https://iaeme.com/MasterAdmin/Journal_uploads/IJARET/VOLUME_9_ISSUE_IJARET_09_05_015.pdf

[3] https://www.irjet.net/archives/V7/i7/IRJET-V7I7994.pdf

[4] https://www.iosrjournals.org/iosr-jce/papers/Vol20-issue2/Version1/E2002013643.pdf

[5] http://www.ijtrd.com/papers/IJTRD15941.pdf

[6] https://www.geeksforgeeks.org/python-gui-tkinter/

[7] https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=92https://www.engpaper.com/ijtsrd/text-extraction-from-image-using-python.html

[8] RaySmith,“AnoverviewofthetesseractOCRengine,”2005.

[9] Pratik Madhukar Manwatkar, Dr. Kavita R. Singh, “Atechnical review on text recognition from images,”IEEESponsored9thInternationalConferenceonIntelligentSystemsand Control(ISCO),2015.

[10] Akhilesh A.Panchal,Shrugal Varde, M.S. Panse,“Characterdetectionandrecognitionsystemforvisuallyimpairedpeople,”IEEEInternationalConferenceOnRecentTrendsInElectronicsInformation Communication Technology, May 20-21,2016.[11]NadaFarhani,NaimTerbeh,MounirZrigui, “Image to text conversion: state of the art andextendedwork,”IEEE/ACS14thInternationalConference on Computer Systems and Applications,2017