|  |  |
| --- | --- |
| **PROJECT** | A GESTURE BASED TOOL FOR STERLILE BROWSING OFRADIOLOGYIMAGE |
| **TEAMID** | PNT2022TMID50430 |

DataGatheringforGestureRecognitionSystemsBasedonSingleColor-,StereoColor-andThermalCameras

***Abstract***

In this paper, we present our results of automatic gesture recognitionsystemsusingdifferent types of cameras in order to compare them in reference to their performances insegmentation. The acquired image segments provide the data for further analysis. Theimages of a single camera system are mostly used as input data in the research area ofgesture recognition.In comparison to that, the analysisresultsof a stereocolorcameraanda thermal camera system are used to determine the advantages and disadvantages ofthese camera systems. On this basis, a real-time gesture recognition system is proposed toclassify alphabets (A-Z) and numbers (0-9) with an average recognition rate of 98% usingHiddenMarkovModels(HMM).

***Keywords:****GestureRecognition,StereoCameraSystem,ThermalCamera,ComputerVision&ImageProcessing,PatternRecognition.*

# Introduction

Gesturerecognitionisanimportantareafornovelhumancomputerinteraction(HCI)systems and a lot of research has been focused on it. These systems differ in basic approachesdepending on the area in which it is used. Basically, the field of gestures can be separated intodynamic gestures (e.g. writing letters or numbers ) and static postures (e.g. sign language ). Thegoalofgestureanalysisandinterpretationistopushtheadvancedhuman-machinecommunication inorderto bring the performanceofhuman-machine interaction closertohuman-human interaction. Themost important component of gesture recognitionsystemsisthe exact segmentation and recognition of the hands and the face which depends on the datagathering. Therefore, different types of cameras are established inthis researcharea(e.g.singlecolor-, stereocolor-, thermalcameras).

Mostresearchersusesinglecolorcamerasfordataacquisition.Abigadvantageofthese cameras is that they are fast and simple to control, so it is possible to realize a suitablegesture recognition system also in real-time applications. Additionally, the color map of theimage can be used for skin color recognition in order to improve the segmentation results.However the robustness of such a system can suffer from a complicated real background. Andthe separation of region of interest will still be a challengingproblem if onlyone camerasystemisused.

Binh etal.used asingle colorcamerafordataacquisition and afterwardspseudo twodimensional HiddenMarkovModels (P2-DHMMs)to recognizeposture with good results.TheyusedKalmanfilterandblobsanalysisofhandsforhandtracking.Undercontrolled

conditionstheyachievedarecognitionrateof98%fortheclassificationof36posturesofASL(AmericanSignLanguage)inreal-time.However,aslowmovementofgestureisnecessaryandocclusionsbetweenhandandfacehavetobeavoided.Also,wearinglongsleeves and thepresence ofahomogenous background are preconditions in theirsystemwhichcannotalwaysbeassuredinareal-timeapplication.Liuetal.developedagesturerecognitionsystemtorecognize26alphabetsfromadatabasebyusingdifferentHMMtopologies.Theyachievethebestrecognitionrateof89.6%byusingtheLeft-Righttopology.Stereo cameras are rarely usedinthefieldof gesturerecognition.Theiradvantagesareoften ignoredorrathernotutilized.Malassiotisetal.had demonstrated acomplete stereobasedsystemfortherecognitionof20statichandposturesfromtheGermansignlanguagealphabet basedona3Dsensor.Theyobtainedthebest resultsusingprinciplecomponentanalysis (PCA)approach based on 3Dposecompensation.Also Elmezain et al.hasbeenworkingwithdifferentHMMtopologiesforgesturerecognition byusingastereocolorcamerasystem.Theyanalyzed36isolatedgesturesandachievedthebestresultsusingLRB

topology.

In the last few years, thermal cameras are often used inthefieldof face recognition, E.g.Socolinsky et al. proposed the combination of thermal and visual cameras for face recognition.Their analysis shows that thermal cameras achieved the similar performance as normal visualcameras. And thermal cameras havetheirown advantagesunderuncontrolledilluminationconditions (indoorandoutdoorscene).Still,theuse of thermal cameras for the segmentationin thefield of gesture recognitionis rare. Based on that, this paper will give a fundamentalanalysisof usingdifferentcamerasystemsfor thedatagathering.

This paper introduces a novel system to recognize continuous hand gestures from alphabetsA-Z and numbers from 0-9in real-time by using themotion trajectoryof a single hand withthe aid of HMM. Forsegmentation,threedifferent typesof cameras (single color-, stereocolor-andthermalcamera)areanalyzedfortheiradvantagesanddisadvantages.Thesesegmentation results buildthe basis for the feature extraction and the tracking of hands andface.The orientation between two following pointswasextracted and used asa basicfeaturefor HMM. These HMM are trained by Baum-Welch (BW) algorithm and the sequences areanalyzed by Viterbi path. We have built adatabase with at least 30video sequencespergesture(A-Z&0-9)andused20sequencesfortrainingand10sequencestotestouralgorithm.Additionally,wehaveaccomplishedalotofonlinetests.

This paper is organized as follows: advantages and disadvantages of the three differentcamera types for segmentation are analyzed and evaluated in section 2. In section 3, the real-time gesture recognition system and experimental results are presented. Finally, the summaryand conclusionarepresentedinsection4.

# Gesturerecognitionsystem

Theproposedsysteminthispaperhasthreeprocessingsteps.Thefirststepisthesegmentation which depends on the evaluation of data gathered by different camera systems.Then,the image segmentsof the first stepprovide the datainput intothe feature extractionand tracking process. Then the third step of this system is the classification process usingHiddenMarkovModels.

### Datagatheringandevaluation

Inthefieldofthesegmentation,differenttypesofcamerasareusedfor datagathering (Fig. 1)tobuildupagesturerecognitionsystem.Thesinglecolorcameraisthemostcommontypeofdataacquisition toolbecauseof itseaseofuse andfastdataevaluationcapabilityeven withhighresolutionimages.Thestereoimageevaluationformsanotherapproach.Inadditiontothecolorinformation,thedepthinformation,whichisdeterminedthroughadisparitycalculation,isused.Nevertheless,thedisadvantageofstereocalculationistheincreasedcomputationalcost. The thermalforms the third type of data production.Inthiscase,thetemperatureofanobjectiscapturedwiththehelpofinfraredradiationsandafterwardsshownintheimage.



**Segmentation**

Bodytemperature

Skincolordetection

ReceiverOperating

Characteristic(ROC)

**Evaluation**

**Data gathering**Single color cameraStereocolorcamera

Thermalcamera

Figure1.Datagatheringandevaluationforgestureandposturerecognition

### Segmentation

**Single/Stereo color camera.** For skin color segmentation in colored images, at first twodecisions must be made. The choice of the color space is the most important feature. The nextallocation of a pixel as skin or non skin is only a problem ofclassification.Thereforeasuitable skin color model is generated to check the affiliation to a skin color class. The mostsuitable color spaces are the ones which are oriented towards perception, becausetheyarecloseto the humanperception system in the way thatthecolor-and brightnessinformationareseparatefromeachother.

In our approach we use YCbCr colorspace where Yrepresents brightness andCbCrrefersto chrominance. We ignore Y channel in order to reduce the effect of brightness variation anduse only the chrominance channels which fully represent the color. The human skin color isfound in a smallareaof thechrominanceplane; so apixel can be classified as skin ornonskin by using a Gaussian model. A large database of skin pixels is used to train the Gaussianmodel, which is characterized by the mean vector μ (eq. 1) and covariance matrix σ (eq. 2).Since are skin color model is based on the chrominance plane Cband Cr, a two dimensionalvector with xi=[CbCr]Twas used as an input. The Mahalanobis-Distance (eq. 3) describes thedistributionofanellipticalareawiththehighestprobabilitywithameanvalueμ.Theprobabilityp(eq.3)iscontinuouslydecreasingwhereasorientationandslopewillbedescribedbythevaluesofthecovariancematrix σ.

1 *n*

 *xi*

*n*1*i*1

(1)

1 *n*

*n*1*i*1

(*xi*

)(*xi*

)*T*

(2)

*p*(*y*|*j*) 1 exp1 (*y*)*T*1(*y*)

 *j j j*

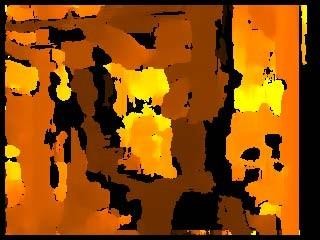
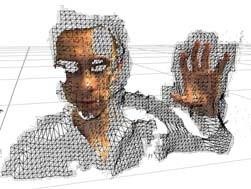
2  2

*j*

 (3)

Now,allskin-coloredareas can be segmentedusing theskin colormodelin theimage.This already showsthe first deficitin single cameras. With aninhomogeneous background,like infigure 2(a), it is notpossible to segmentthe hands and the face perfectly.Further, it isnot possible to separate ambiguously overlapping face and hands. These disadvantages can beovercome with a stereo camera system byusingdepthinformation. The camera calibrationdata and the image features that are matched based on the cross correlation of the left and theright images can be used to estimate the depth information of a 3D point P(X, Y, Z). Also anunequivocalseparationoftheuserfromaninhomogeneousbackgroundispossiblebyutilizing thedepthmap (Fig. 2b). Furthermore,the hands can be held in front of the face andall areas are assigned ambiguously to each other. However, this approach introduces someproblems. Large areas like a closed hand or a head etc. can be easily segmented, but it can beproblematictogetdepthinformationofsmallerareaslikesinglespreadfingers,whichdependsonthecalculation ofdisparity.



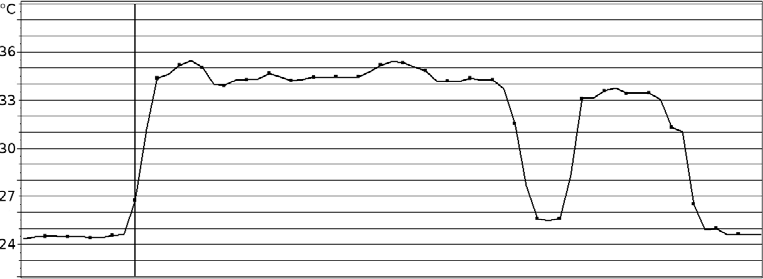
(a) (b) (c)

Figure2.(a)Originalsingle/stereocolorcamera(b)depthinformation

(c)3Dstructurefromastereocamera

**Thermalcamera.**Aninfraredcameraisadevicethatdetectsinfraredradiation(temperature) from the target object and converts it into an electronic signal to generate athermal picture on a monitor or tomake temperature calculationsonit.The temperaturewhich iscaptured by an infraredcameracan bemeasuredorquantifiedexactly,so thatnotonly the thermal behavior can be observed but also the relative magnitude of temperature-relatedproblemscan berecognizedand noted.

In figure 3a, a normal scene of human interaction captured by a thermal camera is shown.Normally, the background can be neglected because the human temperature can be found in asmall thermal area. As shown in figure 3a, the areas of the head and the articulating hands canbe well separated from the background. Besides, the objects have very sharp contours and iteven allows a good and clear segmentation of very smallareas. The graph in figure 3b showsthe temperature course of the straight line in figure 3a. Clearly, the sharp edges and the exactarea of the hand are recognizable. However, like in the single camera case, the information ofoverlapping of handsorthe face is stillnotpossible to extractdue to themissing depthinformation.





1. Originalandthermalimage
2. Thermalgradientfromthemarkedlinein(a)

Figure3.Solutionfromthermalcamera

### Evaluationofsegmentationresults

Fortheproposedgesturerecognitionsystem,threedifferenttypesofcamerasarereviewed.In this manner,ourexperimentaldataiscapturedsynchronouslybyasinglecolor, a stereo color and athermalcamerasystem incomplexrealsituationswherethesystemhadtorecognizedifferenttypesofgestures.Theevaluationoccursbyusingthe*Receiver OperatingCharacteristic(*ROC)curves. Therefore, therequired groundtruthsofrealskinpixelsweremarkedbyhand.

Figure 4ashowsafrequencydistributionwhichisseparatedintotwoclasses(inthiscase the areasof'skin'and'nonskin')byusingathreshold.TPindicatesthesumof*True Positive*pixels,which arecomingfromthegroundtruthandidentifiedasaskinpixel by the system. Also FN is definedasthesumof*FalseNegative*,FPas*FalsePositive*,andTNas*TrueNegative*pixels.Asinmostcasesnounequivocaldifferentiation of the classes is possible by using a normalthreshold.Ifthethresholdisgetting smallerthe numberof falseaspositive(FP)valuesdecreases,meanwhilethenumberofFNincreases. The effectofshifting thethreshold can berepresentedbytheReceiverOperatingCharacteristiccurvesbyusingthe*TruePositiveRate*(TPR),the*FalsePositiveRate*(FPR)andthe*Accuracy*(ACC),whichcanbecalculatedasfollows:

*TPR*

*FPR*

*TPTP**FN*

*FPFP**TN*

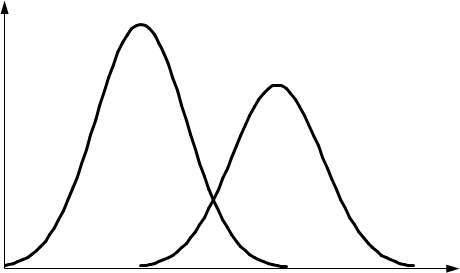
(4)

(5)

*ACC* *TP**TN TP**FN**FP**TN*

(6)

Infigure4banexampleofdifferenttypesofclassifiersispresented,whereasthecurves differsignificantlyinthecurvature.Furthermore,differentworkingpointsA1,*Error Equal Rate* (ERR) and A2 aremarked.A1 isaworkingpointwhereahighTPRwith a low FPRexists.IncontrasttoA1 thepointEERdescribestheareawherenovalue (TPRorFPR)ispreferred,i.e.atheoreticaloptimum.Thisoptimum isnearbythe45° line. In general, themoretheworkingpointgetsto100%TPRand0%FPRthebettertherecognitionis.



Threshold

A2

EER

A1

TP

45°

TN

FPFN

Decisionvalue

0%

(a)

FalsePositiveRate

(b)

100%

n

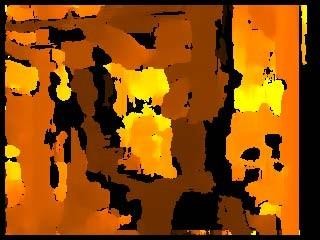
100%

TruePositiveRate

Figure4.(a)showsafrequencydistributionwhichisseperatedintotwoclassesand(b)showsdifferentReceiverOperationCharacteristic(ROC)curves

Figure5aillustrates an imageof ourdatabasecapturedbyasinglecamera.Thereby,parts of the background are also segmented by the skin colormodel (Fig. 5b). Hence, incomparisontotheothercameratypesunderconsideration,thesinglecamerasystemhasa relatively low mean TPR of 79.86%, ACC rateof92.56%andahighmeanFPRof7.07%. Without anypriorior additionalinformation thisisnotenoughforarealtimehumancomputerinteraction(HCI)systemwhenusinginhomogeneousbackgrounds.

Inthenextpartofourexperiments,weareusingaBumblebee2stereocamerasystembyPointGreyResearch[14]whichcancalculatethestereodata,i.e.disparitycalculation, withinhardwareandprovidesimageswitharesolutionof320×240withupto30fps.However,itisnotpossibletoworkwithhigherresolutionsvaluese.g.1024×768 in real-time. In comparison totheothercameratypesweachievedthebestresultswithameanACCof99.14%withameanTPRof78.24%and FRPof0.267%

here.Theimprovementofthestereocamerasisthe depthinformation(Fig. 5d).Therebythehighrecognitionrateofskincoloredpixelsresultsfromfadingoutthebackgroundinformation(Fig.5c),whichisonlyoneoftheadvantagesof stereocamerasystems, described in section2.1. Thethirdkindofcamerawasanuncooledmobilethermal camerasystem(FLIRSC600)withapixelresolutionof640×480and30fps.Thecameraspecificationsare:spectralsensibilityof 7.5-13.5 μm and thermalsensibilityof<45mK.Infigure6theROCcurveisgraphicallypresentedfordifferentthresholdsfrom33.6°Cto31.4°Cbystepsof0.2°C.Forsegmentation,themaximaltemperaturewaschosenasathresholdvalue.



(a)Original

(b)Singlecamera

(c)Stereocamera

(d)Depthinformation

(e)Thermalcamera

Figure5.Acomparisonofsegmentationresultsofdifferentcamerasystems.

* + 1. theoriginalcapturedimage,(b)theanalyzedimagewithoutdepthinformation(d),(c)theanalyzedimagebyusingdepthinformationand(e)

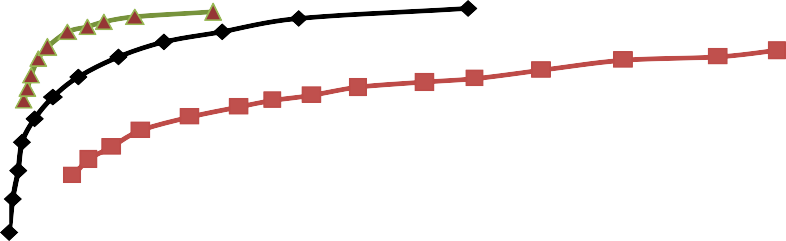
theimagefromathermalcamera

Inourexperimentswe achieved an average ACC rateof ≈96%.These arenot optimalresults,becauseclothescanreceivethebodytemperatureandarepartiallywarmerthanextremitieswherethebloodpressurevalueislower(e.g.hands)asshowninfigure5e.If the face shouldbesegmentedonly,thermalcamerasachievegoodresults.Becauseunder normalconditionsthefacehasahighbloodpressurevalueandownswithinahighbodytemperature.However,normallythebackgroundcanalsobeignoredbythermal cameras. Anadvantageofthermalcamerasisthattheycanbeusedindarknessorunderlowilluminationconditionswhereskincolormodelsarenotsuitable.

The acquired image segments will nowprovidethedatabaseforfurtheranalysistorealizeanaturalhumancomputerinteractionsystembyusingHiddenMarkovModel(HMM).



Figure6.ROCcurvesforasinglecolor-,stereocolor-,andthermalcamerasystembyusingdifferentthresholds.



**ReceiverOperatingCharacteristic**

100%

90%

80%

70%

60%

50%

40%

30%

0,0% 1,0% 2,0% 3,0% 4,0% 5,0% 6,0% 7,0% 8,0%

**FalsePositiveRate(FPR)**

Singlecolor- Stereocolor- Thermalcamera system

**TruePositiveRate(TPR)**

### Featureextraction

There isno doubt about the factthat significantfeaturesplay amajorrole in the detectionofhandgesticulation.Andthesefeaturesmustbeindependentofpersonandlocationinorderto realize the natural HCI system. Inthis mannerlocation, orientation and velocityare themain basicfeatures whichcan be extractedfrom hand trajectory.Ourprevious workshowedthat using the orientation information as the input data to our system gives the best results intermsof accuracy.Hence,in thiswork, theangleθtwas selected between two consequentpoints of the gesture path which consistsof the center of gravityfrom the respective hand(xc(t),yc(t))orthelocationoffingertips.

*yt*1*yt*

*t*arctan

## 

*xt*1

;t=1,2,…,T -1 (7)

* *xt*

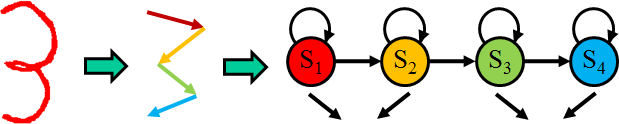
where T represents the length of the gesture path. Foranoptimalcharacterizationofmotiontrajectory,theangleθtwasquantizedinstepsof20°togenerateacodewordfrom1to

18. The codeword also included zerocode which is specified in. In contrast to otherfeatures,anevaluationofthequantizedvectorispossibleateachpointoftime.Thisisa majoradvantage,sinceitisnotnecessarytowaittilltheendofthemotion.Thisdiscretevectoristhen usedasaninputintotheclassificationprocess.

### Classification

The classification is based on a Hidden Markov Model which is a mathematical model of astochastic process and it includes three parameters Λ = (Π, A, B). Thereby Π is defined as theinitial vector; A refers to the translation matrix and B represents the emission matrix. ThesevaluesdependontheconstructionofHMMwhichcanvaryaccordingtotheapplication.For

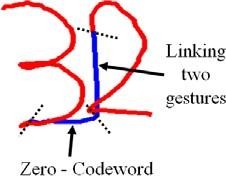
complex applications, a Fully Connected (Ergodic) model can be used, where any state can bereached from other states. In speech recognition, primarily the Left-Right (LR) model is usedwhere any state can change to itself or following states. We have used the Left-Right-Banded(LRB) model in our approach where a state can reach itself or just the next state as shown infigure 7.In order to use HMMthreemain problemsmust be solved: Evaluation, DecodingandTraining.Theseproblemscan beovercomebyusingForward-Backwardalgorithm,ViterbialgorithmandBaum-Welch algorithmrespectively.

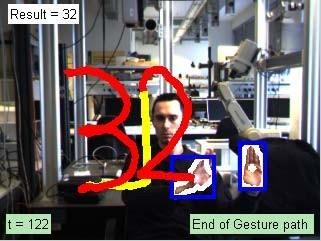


Codeword: 16 11 16 11

Figure7.ExampleofgeneratingtheCodewordbyLRBmodelwith4states.

The isolated and continuous gesture paths are recognized by their discrete vectors using theForward algorithm to calculatethe probabilityof the best Viterbi path. Moreover, Baum-WelchalgorithmisusedtodoafulltrainingfortheinitializedHMMparameterstoconstructa gesture database. The number of states, an example can be seen in figure 7, is based on thecomplexity of each gesture and is determined by mapping each straight-line segment into asingleHMMstate. Moredetails aboutHMMcan befoundin .





(a) (b)

Figure8.GesturepathofacontinuousgesturebyusingtheZero-Codeword.

In our system, isolated gestures can be recognized from continuous gestures by using theZero-codeword.Eachgestureendswithalinesegment,whichisassignedtoaZero-codeword. There are many gestures (e.g. E, Z and 2) which contain Zero-codewords in somesegmented parts and produce separation problems. To overcome these problems, we assigncontinuous velocity as a threshold. After detecting the Zero-codeword, a small part of linkagebetweentwo gesturesmustbeignoredasshowninfigure8a.

# Experimentalresults

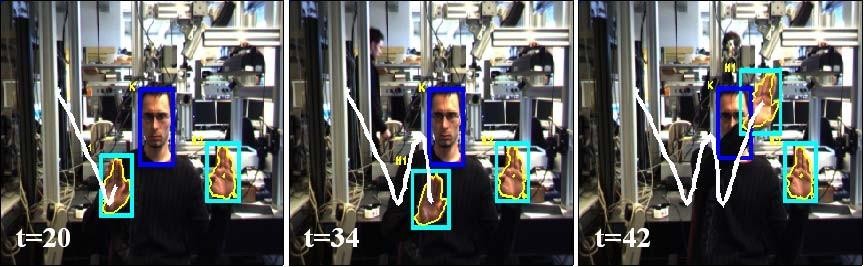
In the first part of ourproposed system, we describethe classificationof36isolatedgestures(A-Zand0-9)fromstereocolorimagesequencesoronlineexperiments.Afterdetectingthehandsand thefacebyusing askin colormodel,themotiontrajectoryisgenerated and afterwards analyzed by HMM.In ourprevious work,we designed differenttypes of HMM topologies for comparing and getting the best results byusing LRB topology[1]. In our experimental results, each isolated gesture is based onat least 30 video sequences;20 video samples for training and at least 10 video samples and additionally a lot of onlineexperimentsfortesting.Weachievedanaveragerecognitionrateof98%forisolatedgesturesbyusingtheLRBtopologywith 9states.Figure9showstheresultsoftheisolatedgesture'W'at different times. At t=20, the highest probability for getting the gesture 'V' and at t=34, thegesture'h'arecalculatedby thesystem.Finally att=42thegesture'W'wasrecognized.

Figure9.Resultofisolatedgesture‘W’atdifferenttimes.

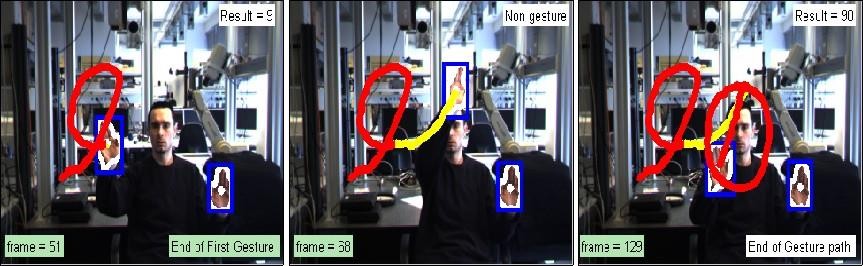
The second part is the classification of continuous gestures. For the separation into isolatedgestures a Zero-codeword detection is realized by using a constant velocity as threshold. Thesystem has been tested on 70 video sequences for continuous gestures with more than oneisolatedgestureandarecognitionrateof95.7%isachieved.Thesystemoutputforthecontinuous gesture of '90' is shown in figure 10. At t=51 the first gesture is ended with theoutput '9'. And the linkage between the two gestures is shown at t=68. The second gesture isended with the output '0' at t=129 and the final result is related to '90'. The input images arecapturedbyBumblebee2stereocamerasystemwith15fpsand320×240pixelimageresolution.Thisreal-timesystemwasimplementedin C++language.

Figure10.Resultofcontinuousgesture‘90’atdifferenttimeinstants.

# Summary andconclusion

In this paper, we propose an automatic system that recognizes continuous gestures (0-9 & A-Z). Thereby three different types of cameras (single color-, stereo color-andthermal camera)are compared in the area of segmentation. According to ouranalysesofthe advantages anddisadvantages and comparison of the receiver operating characteristic (ROC) curves, overall,stereo cameras give the best results for segmentation. Furthermore, it dependsonournovelideaof Zero-codeword detection.Asaresult,the developed system can be used in real-timeand allows the user to act in front of the camera online without the requirement of a secondperson.Ourdatabaseincludesmorethan30videosequencesforeachgesture.We haveachieved an average recognition rate of 98% for isolated gestures and a recognition rate of95.7%forcontinuousgesturesbyusingastereocamerasystemforsegmentationandinaddition to that, we have accomplished a lot of online tests. Our future research focuses on thearea of recognition from trajectoryof the fingertip instead of the center of the hand using amulticamerasystem.

# Acknowledgments

This work was supported by Transregional Collaborative Research Centre SFB/TRR 62(“Companion-TechnologyforCognitiveTechnicalSystems”)fundedbytheGermanResearch Foundation (DFG