Security Verification of Video Telephony System Implemented on the DM6446 DaVinci Processor

Deepak Ghimire
Computer Engineering, Chonbuk National University
Jeonju, Jeollabuk-do 561-756, Rep. of Korea

Jooncheol Kim
Electronics Engineering, Seonam University
Namwon, Jeollabuk-do 590-711, Rep. of Korea

Joonwhoan Lee
Computer Engineering, Chonbuk National University
Jeonju, Jeallabuk-do 561-756, Rep. of Korea

ABSTRACT

In this paper we propose a method for verifying video in a video telephony system implemented in DM6446 DaVinci Processor. Each frame is categorized either error free frame or error frame depending on the predefined criteria. Human face is chosen as a basic means for authenticating the video frame. Skin color based algorithm is implemented for detecting the face in the video frame. The video frame is classified as error free frame if there is single face object with clear view of facial features (eyes, nose, mouth etc.) and the background of the image frame is not different then the predefined background, otherwise it will be classified as error frame. We also implemented the image histogram based NCC (Normalized Cross Correlation) comparison for video verification to speed up the system. The experimental result shows that the system is able to classify frames with 90.83% of accuracy.

Keywords: DaVinci Processor, Code Composer Studio, Face Detection, Facial Features, Normalized Cross Correlation, Image Authentication.

1. INTRODUCTION

Video telephony comprises the technology for the reception and transmission of audio-video signals by user at different locations, for communication between people in real time. Sometimes it is necessary to monitor the video communication between users especially in security sensitive places. In this paper we develop a system for verifying the video frames during video telephony communication based on some predefined criteria. The system is implemented using DM6446 DaVinci Processor. Here, we will focus on how to verify the image frames that are transmitted during communication. If the image frame is classified as error frame it will be blocked otherwise it will be transmitted to the destination.

In today's advancing market, the growing performance and decreasing price of embedded processors are opening many doors for developers to design highly sophisticated solutions for different end applications. The complexities of these systems can create bottlenecks in the form of longer development times, more complicated development environments and issues with application stability and quality [1]. So, in order to develop the embedded system a special care about algorithm complexity, memory management, etc. need to be taken. An overview of the DaVinci technology for digital video is described by G. Frantz et al. [2]. Y-S Jeon et al. [3] proposed a smart camera hardware design based on the TI DaVinci processor. This platform combines the DM6446 DaVinci DSP from Texas Instruments with the low-cost Spartan-3A FPGA from Xilinx. It provides the sufficient hardware resource for implementing security functions in smart camera like object masking and encryption. Other vision based embedded systems are also developed by using DaVinci processor. F. Zhao et al. [4] has implemented Adaboost algorithm for face detection on a TI's TMS320DM642 DSP processor. P. Ayyalasomayajula et al. [5] also uses DM6446 DaVinci processor for implementing an image recognition algorithm. They called their embedded device 'PictoBar', which recognizes barcodes or images, such as pictograms or pictures. It also plays a sound message associated with the

recognized barcode or image. A near infrared imaging detection system based on DaVinci DSP platform is implemented by Lu Hua et al. [6]. They described the schematic, hardware structure and image data processing algorithm of the near-infrared imaging detection system in detail. Recently the design and implementation of network video surveillance system based on DM6446 and OMAP3530 is proposed by P. Han et al. [7]. DM6446 works as the IP camera and OMAP3530 as the client-end. Both DM6446 and AMAP3530 are DaVinci Media System-on-Chips which contains an ARM and a DSP core.

In our developed system the human face is the basic means for deciding the video frames as legal or illegal. Therefore, first it is necessary to find the human face in the image. Mainly, the face detection techniques in the image are classified into four categories [6]: knowledge based methods; feature invariant approach; template matching methods and appearance based methods. In this paper we are using skin color based face detection technique, which falls into feature invariant approach. Using skin color as a feature to detect a face has several advantages. Color processing is much faster than processing other facial features. A survey on pixel based skin color detection technique can be found in [9]. Here we are using skin color and edge based face detection technique to improve the detection accuracy. Face detection in every frame is not possible because of time complexity. So we will find the face in one frame and look the consecutive similar frames by comparing with the last face detected legal frame. Normalized cross correlation (NCC) based image histogram comparison is used for this propose. As soon as the NCC threshold not satisfied, the face detection module is applied again and the process will be repeated.

The remainder of this paper is organized as follows. Section 2 briefly describes about DaVinci system. Section 3 describes about the image processing algorithm used in this system for image verification. Some experimental results and discussion are given in section 4. Finally the conclusion is given in section 5.

2. OVERVIEW OF DaVinci PROCESSOR

The DaVinci technology-based processor consists of scalable, programmable system-on-chip (SoCs) tailored from DSP and RAM cores, accelerators and peripherals that are optimized to match performance and feature requirements for a broad spectrum of digital video and equipments. In the TMS320DM6446 DaVinci processor the ARM is responsible for handling many system functions such as system-level initialization, configuration, user interface, user command execution, interface and control of the DSP subsystem, and overall system control. The TM320C64x+ DSPs are the highest performance fixed point DSP generation in the TMS320C6000 DSP platform. The TMS320C64x+ DSP is an enhancement of the C64x+ DSP with added functionality and an expanded instruction set [10].

The DM6446 device includes a Video Processing Subsystem (VPSS) with two configurable video/imaging peripherals: 1 Video Processing Front-End (VPFE) input used for video

capture, 1 Video Processing Back-End (VPBE) output with imaging co-processor (VICP) used for display.

In our telephony system the system initialization, user interface, video input and output are handled through ARM processor and the algorithm for verifying the video frame is implemented in DSP processor. In this paper we mainly focus on how to verify video frame as legal or illegal frame according to our requirement. So the DSP processor reads the image frame from the memory, process it, and write back the decision about the video frame to the shared memory between ARM and DSP processor. Fig. 1 shows the DaVinci processor board along with the Color TFT LCD Monitor.

3. IMAGE PROCESSING AND VERIFICATION

This section is divided into three subsections. First we will describe about the skin color based face detection method implemented in our system. After that the method for detecting facial features is described. Finally, the procedure for video frame verification is explained.

Fig. 1. DaVinci processor board with color LCD monitor

3.1 Face Detection

In our system we are using skin color and edge information for face detection. Fig. 2 depicts the complete system of face detection. It starts with skin segmentation and edge detection. Numerous researches have been done to find the optimal color space for skin color distribution. The explicit separations of luminance and chrominance components make YCbCr color space very popular. Again, here input image format is YCbCr with 4:2:2 chroma sub sampling, i.e. Cb0 Y1 Cr0 Y0. Therefore here we have also chosen YCbCr space for skin segmentation. The Cb and Cr channel are used for skin segmentation and the Y channel is used for edge detection using Sobel edge detector.
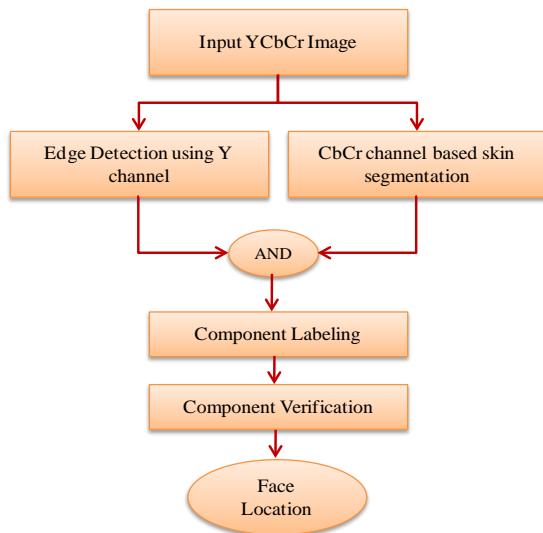
Fig. 2. Flow chart of face detection system

Fig. 3. An input color image and corresponding skin segmented image

Fig. 4. Result of face detection with different facial poses

According to D. Chai et al. [11], the ranges of Cb and Cr for all types of input color images to find skin pixels are [77, 127] and [133, 173] respectively. Actually the input image size is 720x480 pixels and because of chroma sub sampling if we segment the Cb Cr channel for skin region detection, the output binary image will be of size 360x480 pixels. Processing this binary image for searching face region will take long time, therefore we down sample this image in vertical direction twice and in horizontal direction once, and made the binary skin segmented image of size 180x120 pixels. Again the edge image calculated from the Y channel is also down sampled to the size of 180x120 pixels. The result images from two ends are combined using logical AND operation as shown in Fig. 2, to get the final segmented image. Here the edge image is used to disconnect the face skin region occluded with hands or other skin like regions, which improves the performance of face detection. Fig. 3 shows the input color image and corresponding skin segmented binary image.

The resulting image, which is combination of skin tone image and edge image, is now searched for connected components according to the adjacent 8-neighbor pixels. Each of the connected components is now analyzed to judge whether it is face or non-face. Suitable geometric discriminative criteria are defined, which are extracted from prior knowledge about geometric structure and shape features. Here the preference is given to the largest connected area. In our system if there is non-face skin blob with large area then the image frame is classified as error frame because this type of skin blob can be any unknown objects. So to locate a skin blob as face, it should be largest blob from the available skin regions and should satisfy the facial geometric criteria like bounding box proportions, centroid, extent etc. In our system the distance of face from the camera should be in some predefined range, therefore the face is not detected if it is very far or very near from the camera. The threshold values for the range distance are set according to the requirements. After analyzing the skin blobs we can locate the face region by the bounding box around the face skin blob. Fig. 4 shows some results of face detection with different pose.

3.2 Facial Feature Detection

The facial features are detected to verify the detected candidate face region as face or non face. The system is very sensitive to security; therefore the image is verified only when both eyes, and mouth are detected. Here we try to make algorithm for facial feature detection as simple as possible. First the face region is segmented using intensity thresholding for separating eyes region and other face region, because the region around eyes is darker than the other face region. After this the resulted binary image is labeled for searching the eyes region. Two eyes are detected using the postulate that the eyes are located upper part of the face region and two eyes are symmetric to each other. In addition we also used the standard deviation value of the candidate eyes region in the Y channel image. The candidate region with large standard deviation is selected from both left and right side of the face image. Again the mouth region is located using the geometry of two eyes and mouth. The distance between the two eyes and the distance from the middle of two eyes to the mouth in the vertical direction is approximately equal. First the mouth region is located and verified if this region is red dominated, i.e Cr value is larger than the Cb value around the mouth, because the lips are red color dominated. Fig. 5 shows the process of facial feature detection used in our system. Fig. 6 shows some results of eyes and mouth detection.
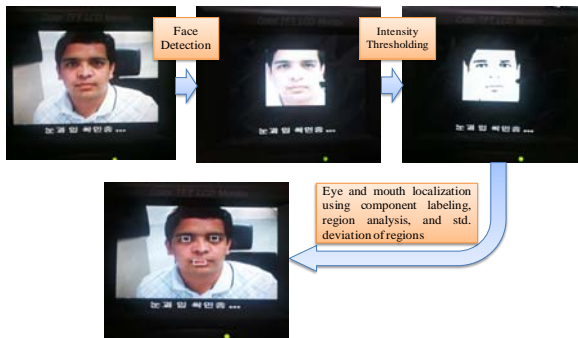
Fig. 5. Facial feature (eyes and mouth) detection procedure
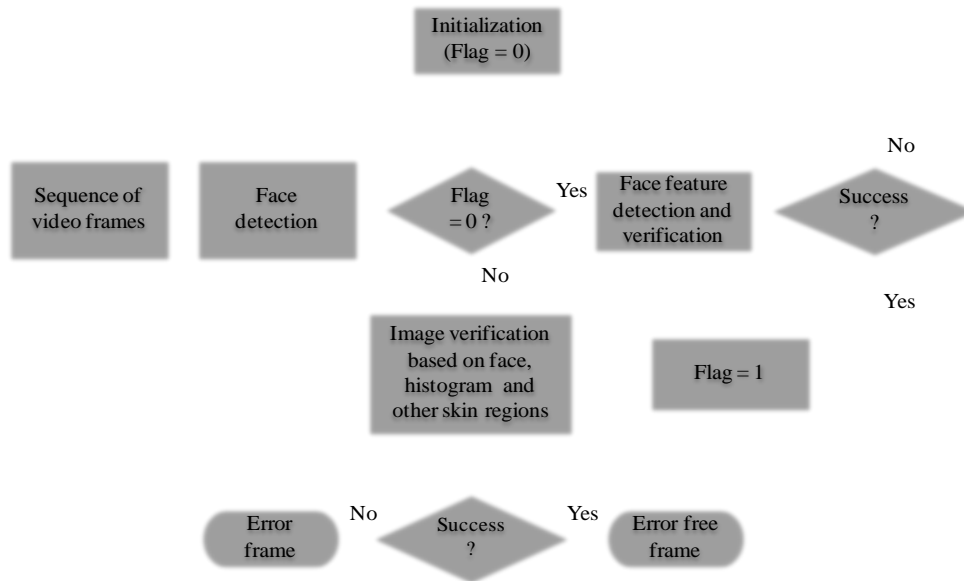

Fig. 6. Result of facial feature detection

Fig. 7. Overall flowchart of image frame verification system

3.3 Image Verification

Fig. 7 shows the overall flowchart of the system describing the image verification procedure. Once the system initializes, first for some number of frames the system will check the facial feature. Now, if face and facial feature are detected correctly on some number of frames the image verification starts. During the facial feature verification the image will not be transmitted. But is depends on the system setting, we can also send those frames which are verified with face, facial feature and some other criteria during facial feature verification process.

Once facial feature verification is completed, the system enters into second stage: image verification. Now each incoming frame is classified as legal or illegal frame using the verification procedure. If the frame is classified as legal frame it will be transmitted to the destination otherwise it will not be transmitted. Here the problem is how to define the rules to classify the frame as legal or illegal frame. It is subjective matter and depends upon the system requirement. In our system the current image frame is classified as legal frame if there is only one face in the image, face is not very close to the camera, not very far from the camera, the user is sitting normally watching towards the camera and so on. The main criterion is, if there is no face in the image it is strongly classified as illegal frame. Also the background is predefined and the plane background is preferred. The objects in the background are not allowed. If the user is covering his face by hands or other objects the face will not be detected correctly, and the frame is classified as illegal frame. Again if there is skin color like background the image frame is illegal, because it can be detected as face image, but more than one face is not allowed. These entire requirements can be fulfilled by first detecting the face in the image and checking for the background against the predefined background.

This algorithm is implemented in the DM6446 DSP processor, the ARM is used to read the video frame and to write in the memory, after finishing the image processing to decide either it is legal or illegal frame, the resultant image with the decision is write back to the memory from DSP, so that the ARM can display the image on the display monitor. Face detection contains many operations including skin segmentation, edge detection, noise removal, component labeling, component analysis etc. Doing all those operations in every frame makes the system slower. So we also developed some other scheme to speed up the system. This operation is described in the following subsection.

3.3.1 Normalized Cross Correlation (NCC) based Image Comparison

The NCC can be used to compare two signals to find their closeness. The more the NCC value of two signals is close to 1, the more similar the two signals are. So NCC can be used to compare two images to find how much those images are similar. But direct NCC calculation between two images also takes long time. Here first we define the Region of Interest (ROI) in the image and find the histogram of this ROI. Now the 1-D 32-bin histograms of ROI of two images are compared using the NCC method. To use this concept in our system for image verification propose, we first store the histogram of the region around detected face position of the legal frame, which is defined as ROI, in the buffer and for the incoming frame the histogram of the same ROI is calculated. Now the NCC between the buffer histogram from the previous frame and the histogram of the current frame is calculated using Eq. (1).

$$NCC = \frac{\sum_{i=0}^{n-1}\left(h_r(i)-\overline{h}_r\right)\left(h_c(i)-\overline{h}_c\right)}{\sqrt{\sum_{i=0}^{n-1}\left(h_r(i)-\overline{h}_r\right)^2}\sqrt{\sum_{i=0}^{n-1}\left(h_c(i)-\overline{h}_c\right)^2}} \quad (1)$$

Where h_r is the buffer histogram, h_c is the histogram from the current frame, and $\overline{h}$ is the mean histogram value.

If the detected NCC value is greater than some threshold, in our system we use 0.94, the previous and current frame are similar and classified as error free frame. For the next frame the process is repeated, but if the NCC value is less than the threshold, we go back to the face detection module and the image frame is classified as legal or illegal frame. As soon as the legal frame is detected, again the histogram of the ROI is saved in the buffer and NCC based comparison is processed. If legal frame is not detected the frames are checked using face detection procedure until legal frame is detected. This NCC based comparison helps to speed up the system if there are large number of similar consecutive frames, this can happen only when the user is not moving in the video frames. In normal situation there can have large number of similar consecutive frames.

4. EXPERIMETN AND DISCUSSION

In our current system we have divided the error situation into six basic categories. Actually there are no strict common assumptions that which situation will become error and which will become non error condition, it depends upon the system requirements. Therefore according to system requirement we can assume the error situations and can adjust the program to fulfill those requirements. Table 1 shows the six main categories that we have implemented in our system as error situations.

Table 1. Error Types

Error Type	Description
Absence of face	If there is no user in front of camera, this situation is clearly classified as error situation. At this situation there is no necessary to transmit video frame.

Face far from camera	There is some predefined range between camera and user position. User distance from the camera is approximately identified by finding the face area in the current frame. If it is very small it is assumed that user is sitting far from the camera position and image is classified as error frame.
Face very close to camera	If user is very close to camera, the face area in the current frame is larger than predefined threshold, in our system this is classified as error situation.
Presence of multiple face objects	Multiple users in front of camera are not allowed. If number of face objects is greater than 1, this is error situation. As well as if the background is different than predefined one, this is also classified as illegal condition.
Unclear face object	Sometimes non face object in a scene can be detected as face object, but if the facial features are not detected accurately, we classify this situation as error situation. Also if the user is not looking towards the camera, like looking up, left/right, down etc, the face is not detected accurately and classified as error situation. This means to be perfect condition user should look towards camera.
Occlusion of face	Sometimes user can hide his face by his hands, some other objects, documents etc, and the face will be occluded by those objects. We classified this as error situation because these kinds of objects can be any kind of legal or illegal objects.

The software platform used to implement the system is Code Composer Studio (CCStudio) v3.3 with emulator for testing proposes. In order to test our system, we recorded some video during the testing phase. While using the system, user makes most of the possible actions in front of the camera. The system analyzes each video frames and classifies them into legal or illegal frames according to the defined rules. Later we analyzed all the recorded video frames and classified them into legal or illegal images by human evaluation. Now we can compare the frames that are classified by our system and classified by the expert according to our system requirement. The goal is nearly achieved. Table 2 shows the statistics of the comparison between human evaluation and our system classification. On average 92.22% of legal frames are classified as legal frames, 88.96% of illegal frames are classified as illegal frames. In overall the system gives the 90.83% of accuracy. While we use the system in real time, as soon as some error frames are detected the system is aborted and the communication is disconnected. The important thing is, sometimes the legal frame can be classified as illegal frame but, it is strictly restricted to classify the illegal frame as legal frame.

Table 2. Classification Performance

	Legal Frames	Illegal Frames	Correctly Classified Legal Frames	Correctly Classified Illegal Frames	Average System Accuracy
Video1	390	201	362 (92.82%)	184 (91.45%)	92.38%
Video2	229	122	210 (91.70%)	109 (89.34%)	90.88%

Video3	280	259	258 (92.14%)	223 (86.10%)	89.23%
Average			**92.22%**	**88.96%**	**90.83%**

Fig. 8 shows some images that are captured while we experiment the proposed system. Input video sequence is processed and the result of system verification is displayed on the LCD monitor. At the bottom of the image on LCD monitor the status of the current image frame is indicated. If image frame is error frame, corresponding error message is displayed. Here we have includes some of the image frames that are categorized as error frames according our system settings.

Fig. 8. Some captured images while testing the system for image verification, message at the bottom indicates error frame

In order to make the system to operate in real time, the program is optimized and written on linux system, we do not use emulator anymore. The program is made as simple as possible. The system can operate in 30 frames per second and video after processing is smooth and operating in real time.

5. CONCLUSION

This paper described a video telephony verification system depending upon the human face and facial features. The system is implemented in DaVinci processor, specially the video verification algorithm is developed in DSP processor. NCC based scheme is also proposed to speed up the system. The video telephony system is operating in real time with 30 frames per second. The future work is focused on reducing the system error in image verification stage as well as study and implementation of more error situations to make system more stable and accurate.

6. REFERENCES

[1]   J. Coombs, R. Prabhu, "OpenCV on TI's DSP+ARM® Platforms: Mitigating the Challenges of Porting OpenCV to Embedded Platforms", *White Paper*, http://www.ti.com/lit/wp/spry175/spry175.pdf.

[2]   G. Frantz, L. Adams, "DaVinci™, Technology for Digital Video", *White Paper*, http://www.ti.com/lit/wp/spry067/spry067.pdf.

[3]   Y-S. Jeon, M-S. Kim, Y-S. Kim, J-W. Han, "Smart Camera Hardware Design based on the TI DaVinci Processor", *Int. Technical Conf. on Circuits Systems, Computers and Communications*, pp. 622-624, 2009.

[4]   F. Zhao, Li Yang, Y. Zhu, P. Liao, "Enhancing the Implementation of Adaboost Algorithm on a DSP-based Platform", *Int. Conf. on Scalable Computing and Communications; The Eighth Int. Conf. on Embedded Computing*, pp. 393-395, 2009.

[5]   P. Ayyalasomayajula, S. Grassi, N. Deurin, P-A. Farine, T. Gueguen, "Implementation of an Image Recognition Algorithm on the DM6446 DaVinci Processor", *Proc. of the 4th European DSP in Education and Research Conference (EDERC 2010),* pp. 175-179, Dec. 1-2, 2010.

[6]   ] Li Hua, Z. Shi-Chao, H. Chao, Z. Ming, M. Xiao-Feng, "A Near Infrared Imaging Detection System Based on DaVinci Plateform", *The 9th Int. Con. on Electronics Measurement and Instruments (ICEMI 09),* pp. 154-159 Aug. 16-19, 2009.

[7]   P. Han, Z. Ye, S. Yang, "The Design and Implementation of Network Video Surveillance System Based on DaVinci Chips", *Advances in Information Technology and Education: Communication in Computer and Information Science*, vol. 201, pp. 296-302, 2011.

[8]   M.-H. Yang, D. J. Kruegman, and N. Ahuja, "Detecting Faces in Images: A Survey", *IEEE Trans. on PAMI*, vol. 24, pp. 34-58, Jan 2002.

[9]   V. Vezhnevets, V. Sazonav, A. Andreva, "A Survey on Pixel-Based Skin Color Detection Technique", *Proc. of GraphiCon*, pp. 85-92, 2003.

[10] TMS320DM6446 Digital Media System-on-Chip, *TI's Data Sheet*, http://www.ti.com/lit/ds/sprs283h/sprs283h.pdf

[11] D. Choi, and K. N. Ngan, "Face Segmentation Using Skin-Color Map in Videophone Applications", *IEEE Trans. on Circuits System and Video Telephony*, Vol. 9, pp. 71-77, June 1999.

Deepak Ghimire
He received the B.E. degree in computer engineering from Pokhara University, Nepal in 2007 and M.S. degree in computer science and engineering from Chonbuk National University, Rep. of Korea in 2011. Currently he is pursuing his Ph.D. degree in computer science and engineering at Chonbuk National University, Rep. of Korea from 2011. His main research interests include image processing, computer vision, and pattern recognition.

JoonCheol Kim
He received the B.S., M.S., and Ph.D. degrees in electronic engineering from Chonbuk National University, Jeonju, Korea, in 1986, 1988, and 1995, respectively. From 1988 to 1991, he was researcher at LG Industrial System R&D Laboratory. Since 1993, he is now with the School of the Engineering in Seonam University. His research interests include image processing, computer vision.

Joonwhoan Lee
He received his BS degrees in Electronic Engineering from the University of Hanyang University in 1980. He received the MS degree in Electrical and Electronic Engineering from KAIST in 1982, and the Ph.D. degree in Electrical and Computer Engineering from the University of Missouri in Columbia in 1990. He is currently a Professor in the Department of Computer Engineering of the Chonbuk National University. His research interests include image processing, computer vision, emotion engineering.