ISSN:0975-9646
K. Bapayya et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 6 (5) , 2015, 4563-4567

RTSP Based Video Surveillance System Using IP Camera for Human Detection in OpenCV

K. Bapayya[1] ,K. Sujitha[2] , Mr. SD. Akthar Basha[3]

[1] *Asst. Professor, Department of ECE, CVR College of Engineering, Hyderabad-501510*

[2] *M-Tech, Department of CSE, QUBA College of Engineering and Technology, Venkatachalam-524320*

[3] *Associate Professor, Department of CSE, QUBA College of Engineering and Technology, Venkatachalam-524320*

***Abstract*—The Real Time Streaming Protocol or RTSP, is an application-level protocol for control over the delivery of data with real-time properties. RTSP provides an extensible framework to enable controlled, on-demand delivery of real-time data, such as audio and video. The protocol can be intelligently chosen for multi thread environment where each parameter of information such as audio, video and text can be separated and processed concurrently and executed in parallel. The proposed system is able to perform on MPEG-4, H.264 and several other popular formats where video is undergoing encoding and decoding to save the bandwidth up to 80% over channel. HOG is a feature descriptor used in computer vision and image processing for the purpose of object detection like human. The technique detects occurrences of gradient orientation in localized portions of an image. In H.264 RTSP video, frames are identified as I, B, P and HOG is performed using GPU hardware, OpenCV and CUDA programming with HOG algorithm in Linux platform helps to automate and find the presence of human and alarm is generated.**

Keywords— ***RTSP, IP Camera, GPU, HOG, H.264***

I. INTRODUCTION

RTSP is Real Time Streaming Protocol which provides advanced control features in its streaming protocols like play, pause, rewind, forward kind of facilities for video streams. On –Time delivery is the specialty in this protocol where data is delivered in real-time. Commonly data is in the form of audio and video. Sources of data can include both live camera data in network or a CCTV feeds and footage video clips. The protocol supports the multiple data delivery sessions connected in the network, it provide a means for choosing delivery schemes such as multi cast UDP , UDP and TCP/IP, and provide a means for choosing delivery mechanisms based upon RTP. The Real-Time Streaming Protocol (RTSP) establishes and controls both a single or several time-synchronized streams of continuous data flow media such as audio and video. It does not typically deliver the continuous streams itself, although interleaving of the continuous media stream with the control stream is possible RTSP acts as a "network remote control" for multimedia servers in the network. HOG algorithm is applied on the video frames to detect human in the frame and processes to alarm the object detection and generation of signal.

II. VIDEO COMPRESSION IN H.264

H.264 is I,B,P H.264 coding technique is the modern and most effective video compression standard. High quality video streaming can be achieved with low bit rates. The H.264 compressed video stream needs just 10 percent of an MJPEG compressed in network bandwidth. It is seen that H.264 video quality was approximately 95 percent superior when compared to MJPEG video compression method.

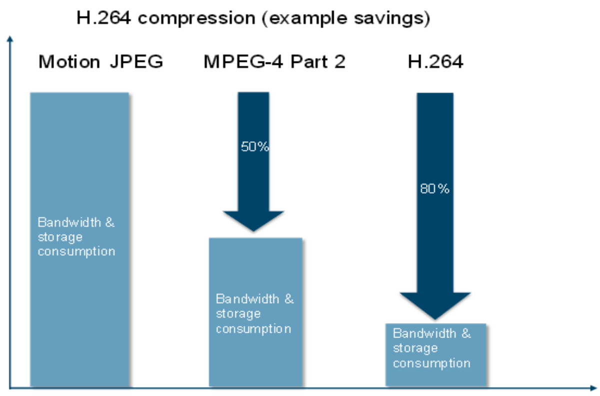


Figure1. H.264 compression importance

Figure -1 shows the H.264 compression scheme which is superior to all other latest coding techniques. A high bandwidth saving is also observed. Video streams encoded with H.264 compression methods resulted in significantly reduced network storage requirements over MJPEG compressed streams. Even though these tests did not measure the amount of storage required to save these images, there is a direct correlation between the amount of network bandwidth required to transmit the compressed data across the network and the amount of space required to store the data. For IT and security system managers, The

Lab recommends using H.264 compression technique, where two continuous frame difference is non-zero that is motion frame is only transmitted in the channel is shown in Figure-2.

Most image processing and computer vision tasks are very computer intensive and require parallel processing. A digitized image usually

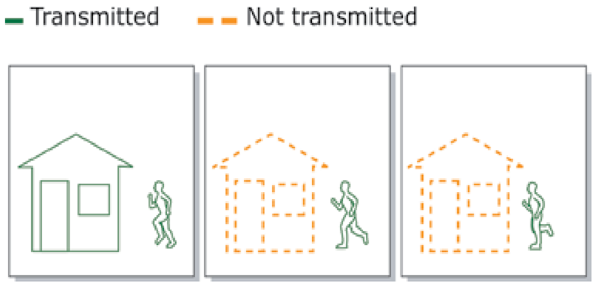


Figure 2: Video with H.264 codec in RTSP

contains 256 k pixels. For a typical computer vision problem, there are lots of images to be created and processed within a limited time. A traditional computer is no longer suitable for such a real-time vision system. Designing parallel algorithms for solving computer vision problems is therefore both of theoretical interest and of practical importance. Parallel algorithms for computer vision problems such as moment computation, histogramming, histogram modification.

III. HOG ALGORITHM

Histograms of Oriented Gradients where feature vector is calculated for object image. The feature vector of the object in various environmental conditions is recorded and preserved under conditions like in tilt position and orientation formats. The purpose is to give positive signal for existence of the object in the frame. This makes the classification easy. SVM is the classifier which separates the feature vector from the reference points used while training. Training is performed on huge collection of data set and positive and negative samples in vector are generated and noted. The same vector is again calculated for test image as shown in Figure-3 and its subsequent results and various poses is shown in Figure-4. Further the comparison is made with feature vector or trained image and test image and SVM classifies with a valid margin. Here the INRIA standard data is reference vector and test vector is the set of images taken from video frames continuously on which HOG is performed.

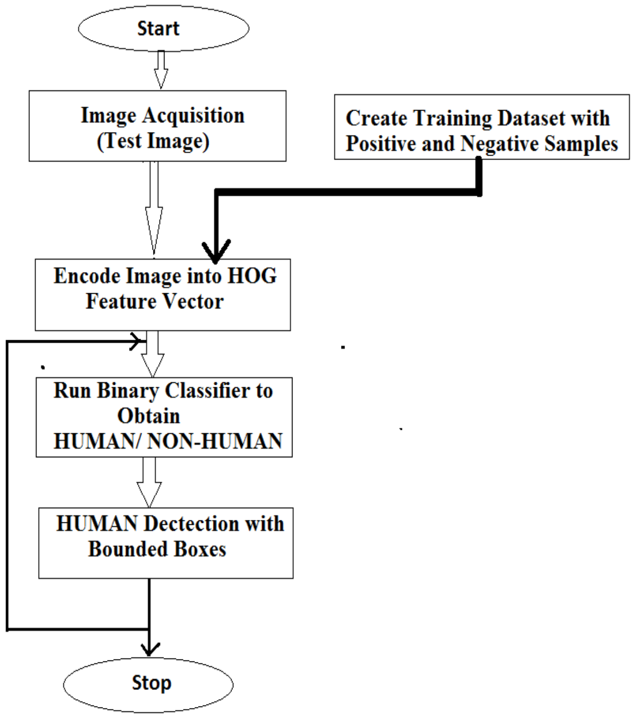
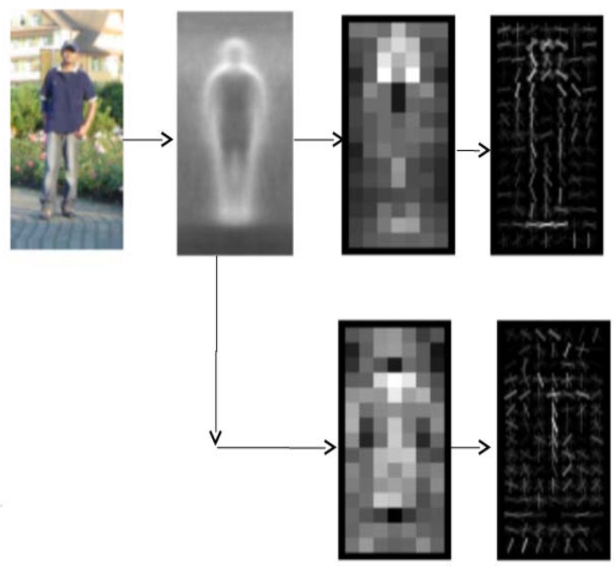


Figure 3: Flow chart for the HOG algorithm



.Figure 4: Results after HOG algorithm for human in standing position.

The HOG person detector is very convenient method to understand. One of the main reasons for using HOG than SIFT is that it uses a “global” feature for a reference dataset like for human rather than a collection of “local” features.

IV. THE PROPOSED SYSTEM

The proposed is represented in Figure-5 where the input is generated as part of image/video acquiring from an IP camera. IP camera is the network camera which is connected in the network

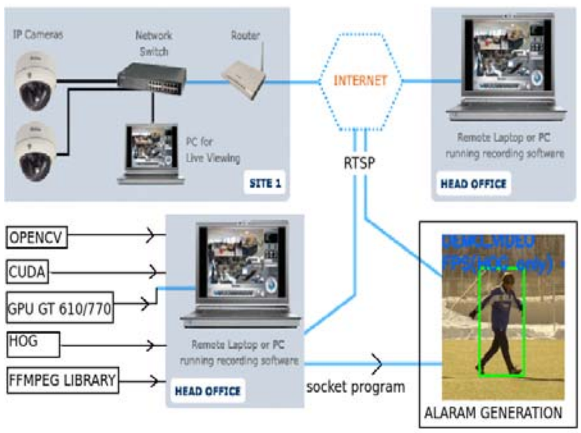


Figure 4: Block Diagram of the proposed system

and operated online, it is also having facility of PTZ (Pan, Tilt and Zoom).Camera generates H.264 video file which gives up to 80% compression ratio by transmitting the only frames which has motion with B-frame, P and I frames occasionally. Figure-2 shown the logic used in H.264 codec where the two continuous frame difference is non-zero is transmitted.

IP camera of Axis is having facility of PoE (Power over Ethernet) for which camera is powered on and starts working as soon as it is connected to network of the switch and the compressed video can be accessed from any part of the world through its IP address and Internet connectivity. IP camera based RTSP, real time streaming protocol video is further accessed by keywords typing in the browser as below.

RTSP://182.16.20.167/username/password

IP camera acts as Server for the whole system and delivers the footage of the events occurring in office site where the IP camera is mounted for area of interest. At night times the human pedestrian is limited and only officials are allowed. IP camera upon gives the footage of the site to the user program where HOG like algorithm is applied and the human existence with his behavior of intention can be noticed. An alarm is generated in the form of blinking of lights or buzzer to alert the officials to aware. The whole system is automated and no manual operator is required for this system. Figure-5 shows the Server end of the system.

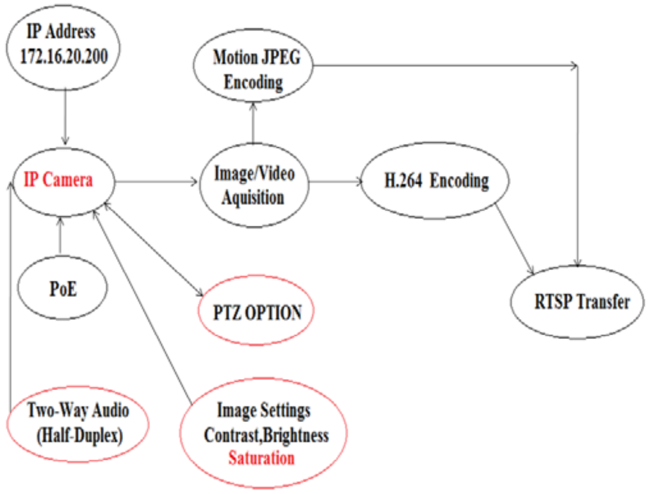


Figure 5: IP Camera at Server End

Client system is the other end where the processing of the video is performed. Client system is a high-end CPU-i-7 (Central Processing Unit) with GPU (Graphical Processing Unit) as part of it, here GT610 and GTX 770 GPU was used which is having 1536 cores inside it resulting fast processing of video input. CUDA 6.0(Compute Unified Device Architecture) tool is used to work with GPU cores. Further OpenCV (Open Source Computer Vision) has number of functions in its library mainly aimed at real-time computer vision applications. OpenCV was installed in open source Operating system that is Ubuntu Linux 12.04.Client system is responsible for following main events steps as shown in Figure-7:

a. Decoding H.264 video frames,
b. Detect the presence of Human
c. Generate the alarm

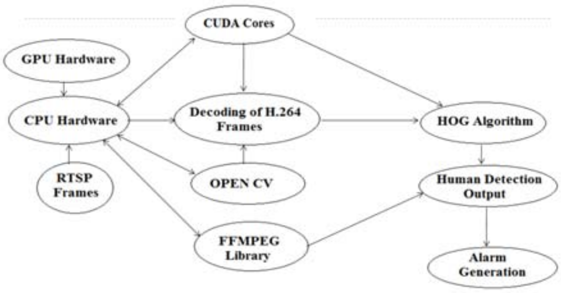


Figure 7: CPU and GPU at Client End

HOG algorithm functions are written and used in the program to find the existence of human in the footage. More libraries can be added in Ubuntu Linux by which media players like VLC and FFMPEG players was installed which will run the program video output. On human detection alarm is as rectangle surrounding human, generated by blinking LED of Axis IP Camera.

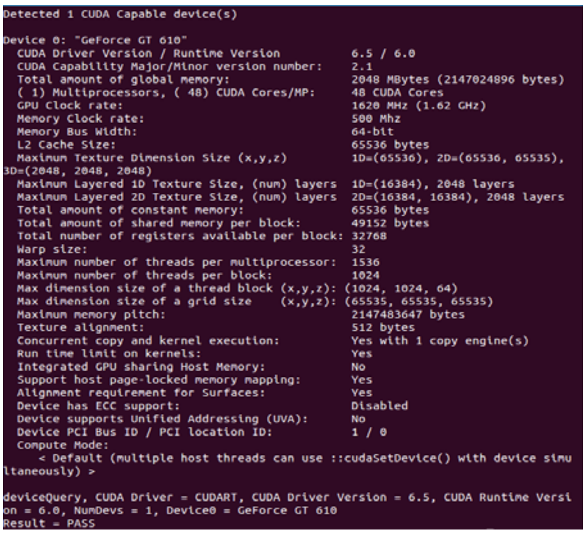
Hundreds of cameras can be connected in the network as client server model, at this stage bandwidth of the network comes to the picture, because H.264 format is being used it saves a lot of usage of channel and improves the video transfer rates and the respective videos are processed. The events can be monitored from any place of the world by username and password which is the service available in Axis IP camera provider by enabling the option and presence of human is detected without any intervention of operator.

V. MPLEMENTATION DETAILS

RTSP based video is generated by AXIS M3004-V IP (Figure-8) camera which generated the desired H.264 format with username and password. The code uses the FFMPEG library to decode and run the video.

Figure 8: Axis M-3004V IP camera

GTX770 GPU was installed inside Intel Core-2 duo CPU. CUDA 6.0 tool kit was installed under Ubuntu 12.04 Linux for the use of GPU which has 1536 cores, a pass report is shown in Figure-9. FFMPEG libraries for streaming, OpenCV 2.4.9 for application program development with g++ 2.4.6 support.



```
Detected 1 CUDA Capable device(s)

Device 0: "GeForce GT 610"
  CUDA Driver Version / Runtime Version          6.5 / 6.0
  CUDA Capability Major/Minor version number:    2.1
  Total amount of global memory:                 2048 MBytes (2147024896 bytes)
  ( 1) Multiprocessors, ( 48) CUDA Cores/MP:     48 CUDA Cores
  GPU Clock rate:                                1620 MHz (1.62 GHz)
  Memory Clock rate:                             500 Mhz
  Memory Bus Width:                              64-bit
  L2 Cache Size:                                 65536 bytes
  Maximum Texture Dimension Size (x,y,z)         1D=(65536), 2D=(65536, 65535),
3D=(2048, 2048, 2048)
  Maximum Layered 1D Texture Size, (num) layers  1D=(16384), 2048 layers
  Maximum Layered 2D Texture Size, (num) layers  2D=(16384, 16384), 2048 layers
  Total amount of constant memory:               65536 bytes
  Total amount of shared memory per block:       49152 bytes
  Total number of registers available per block: 32768
  Warp size:                                     32
  Maximum number of threads per multiprocessor:  1536
  Maximum number of threads per block:           1024
  Max dimension size of a thread block (x,y,z): (1024, 1024, 64)
  Max dimension size of a grid size    (x,y,z): (65535, 65535, 65535)
  Maximum memory pitch:                          2147483647 bytes
  Texture alignment:                             512 bytes
  Concurrent copy and kernel execution:          Yes with 1 copy engine(s)
  Run time limit on kernels:                     Yes
  Integrated GPU sharing Host Memory:            No
  Support host page-locked memory mapping:       Yes
  Alignment requirement for Surfaces:            Yes
  Device has ECC support:                        Disabled
  Device supports Unified Addressing (UVA):      No
  Device PCI Bus ID / PCI location ID:           1 / 0
  Compute Mode:
     < Default (multiple host threads can use ::cudaSetDevice() with device simu
ltaneously) >

deviceQuery, CUDA Driver = CUDART, CUDA Driver Version = 6.5, CUDA Runtime Versi
on = 6.0, NumDevs = 1, Device0 = GeForce GT 610
Result = PASS
```

Figure 9: GPU pass report for CUDA 6.0/6.5

VI. RESULTS

HOG algorithm is able to detect the human presence for various poses, an event snapshot is shown in below figures, Figure -10, 11, 12 respectively. A compatible and efficient setup model was proposed and designed so as to meet the streaming, speed of decoding and algorithm process to detect human for H.264 video. The system is able to generate the alarm in the form of rectangular boxes and tracked in the video, further blinking of LED of camera. The system generates alarm in real-time for CIF, QCIF, VGA, SVGA video resolutions.

Figure 10: Detecting Human in Outdoor

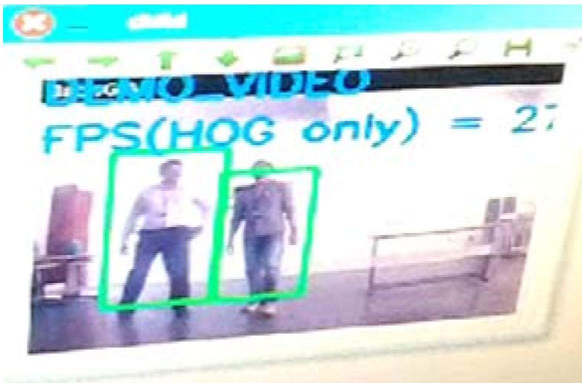


Figure 11: Detecting Human in Indoor

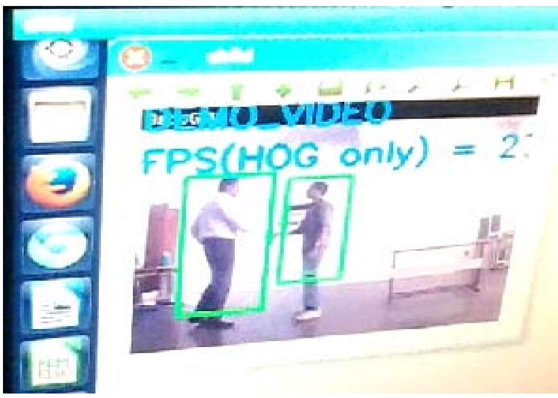


Figure 12: Detecting Humans in Indoor with different

VII. Conclusion And Future Scope

The proposed system is able to process the H.264 data format, The HOG algorithm is able to detect the presence of human in the live video with very less chance for false alarm up to 8% is seen. The system is fully automated where no human intervention is required and an operator working for 24x7 for tens of screen is eliminated. Results are produced(Alarms) in the form of LED, rectangular box of OpenCV, Buzzer can be easily connected to hardware of camera. By this surveillance system is made easy. High frame rate up to 170 fps is achieved for recorded videos.

OpenCV opens a chance to detect the face of human, his distance and direction can be calculated and multiple cameras connecting and processing of data will be upcoming challenges for the system.

References

[1] N. H. Mackworth, The breakdown of vigilance durning pro-longed visual search,Q. J. Exp. Psychol.,vol. 1, no. 1, pp. 621, Apr. 1948.

[2] Implementation of HTTP Live Streaming for an IP Camera using an Open Source Multimedia ConverterInternational Journal of Software Engineering and Its Applications Vol.8, No.6(2014), 39-50 pp.

[3] Navneet Dalal and Bill TriggsHistograms of oriented gradients for human detection.Computer Vision and Pattern Recogni-tion, IEEE Computer Society Conference on, 1:886893, 2005.

[4] David G. Lowe. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision, 60:91110, 2004

[5] Christian Wojek, Gyuri Dorko, Andre Schulz, and Bernt Schiele. Sliding-windows for rapid object class localization: A parallel technique. In Proceedings of the 30th DAGM sympo-sium on Pattern Recognition, pages 7181, Berlin, Heidelberg, 2008. Springer-Verlag.

[6] Mao, Y.F., et al. (2011) Research and Design of a Web DVR Based on RTSP. Computer Engineering and Design, 32, 2523-2526.

[7] Li, X.L., et al. (2011) Design and Implementation of Wireless Video Surveillance System Based on RTP/RTCP, RTSP. Video Engineering, 35, 89-92

[8] Design and implementation of a measurement system for power station video and environment surveillance system Zhengwei Changa , Dongsheng Caib *, Wei Zhena , Qi Huangb

[9] http://en.wikipedia.org/wiki/Pedestrian_detection

[10] http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.122.1149 &rep=rep1&type=pdf