Automated Video Surveillance System: Integration

The design and implementation aspects of individual prototype systems for real-time motion detection, focused region extraction, and object tracking have been discussed with details in previous chapters. These prototype systems can either detect motion or extract focused region or track a moving object, but are not capable of doing a combination of multiple tasks such as motion detection in focused region, automatic region of interest/target detection and tracking, etc. For this purpose, in this chapter, we have integrated the three individual prototype systems developed earlier in chapters 3-5 into a complete FPGA-based prototype automated video surveillance system. Before presenting integration of the three prototype systems, we also present the implementation of a prototype system for generating the history of a live incoming video stream using motion detection system developed in chapter 3 and some extra hardware modules.

6.1. Video History Generation System

In some cases of video applications, a single frame where motion has occurred does not deliver the requisite information. In case of motion of a large object, motion may appear localized to the parts moving only in a single frame such as some human body parts like hands, head, etc. For such cases, it is important to gather information from the previous or the next frames to gain a better understanding of the situation. This process of merging information from multiple video frames is called video history generation in the context of automated video surveillance systems and is explained in Figure 6.1.

Video history/summary generation has two major advantages. Firstly, it provides more meaningful and relevant information about areas of motion and the target moving object in a video scene. For example, when a new object/person enters in a video scene, the output of motion detection does not indicate trajectory of the object in motion. It primarily identifies the motion blocks (Figure 6.1b). By merging motion blocks from multiple output frames, the characteristic motion pattern showing the trajectory of the object of interest can be obtained (Figure 6.1c). Secondly, as the summary of a video is generated by merging
multiple video frames showing significant motion into one frame, therefore, the output frame rate is reduced, thereby reducing further communication and processing overheads. In remote video surveillance scenarios with limited communication bandwidth more meaningful information can be transferred using history generation.

![Original Frames of Input Video Sequence](image1)

![Motion Detection Outputs](image2)

![History Generation Outputs](image3)

Figure 6.1: (a) Original Frames of Input Video Sequence; (b) Motion Detection Outputs; (c) History Generation Outputs.

Video history generation is in respect of video frames showing significant motion in a video sequence. Therefore, the memory efficient motion detection architecture, designed and presented in chapter 3 of the thesis, has been used for generating the history of live incoming video streams. Our proposed and implemented video history generation architecture is shown in Figure 6.2. In addition to all the modules of motion detection system, three additional blocks (shown with green color background) are required to record the history of motion blocks in a video sequence. These are Motion History Memory, Multiplexer (MUX), and OR gate. Motion History Memory module stores the merged motion information of incoming video frames. The OR operation is performed on 1-bit Motion Detection output (MDF) of current block and 1-bit Motion History Memory
output (HMOP) of current block. Its 1-bit result (HMIN) is stored back in *Motion History Memory*. This process merges the current frame motion information with the previous frames’ motion information, stored in *Motion History Memory*.

Video history can be generated for a given number of frames (N). When the frame number reaches N, the output of *Motion History Memory* (HMOP) is stored in *Output Memory* of motion detection architecture. The 1-bit video history information from *Output Memory* is read. If this output is ‘1’ then the original video pixel data is sent for display on monitor through *DVI Display Controller* and if this output is ‘0’ then 0 (black color pixel data) value is sent for display. This is achieved by using a 2:1 multiplexer (MUX) for which control signal is the 1-bit output of *Output Memory*. At the end of Nth frame, ‘0’ value is written to all the memory locations of *Motion History Memory* in order to reset existing history information in the memory. The video history generation process has used one additional two-input OR gate, one 2:1 multiplexer (MUX), and one *Motion History Memory* in addition to FPGA resources utilized by the memory efficient motion detection system. The *Motion History Memory* size is 25920 bits \((720 \times 576)/16\) for PAL (720x576) size color video. Thus, it uses one extra 36Kb- Block RAM on the FPGA. This memory size is independent of the number of frames for which the history is generated. This implementation is capable of generating video history for a live incoming video stream in real-time.

![Figure 6.2: System Level Architecture for Video History Generation.](image)
6.2. Integration of Developed Hardware Architectures and Interfaces

In this section we describe the integration of all the hardware architectures and interfaces, designed and developed earlier in the thesis, into a single prototype automated video surveillance system implemented on Xilinx ML510 (Virtex-5 FX130T) FPGA development platform. In chapter 2, we have designed and developed four input/output interfaces, namely – Camera Interface, DDR2 External Memory Interface, Camera Movement Controller, and DVI Display Controller. Three respective hardware architectures have been presented in chapters 3-5 for real-time motion detection, focused region extraction, and object tracking, in live PAL (720x576) resolution color video streams. We have also presented a real-time video history generation system using motion detection architecture in the previous section of this chapter. All the four input/output interfaces and all the four developed architectures (motion detection, focused region extraction, object tracking, and history generation) have been integrated to design a complete automated video surveillance system, addressed earlier in chapter 2 of the thesis. Block level dataflow diagram of the complete implementation of automated video surveillance system is shown in Figure 6.3.

Figure 6.3: Dataflow Diagram of Complete Automated Video Surveillance System Designed, Implemented, and Presented in the Thesis.
Detailed descriptions and functionalities of each module of the implemented prototype automated video surveillance system (shown in Figure 6.3) except for Motion Detection in Focused Regions module have been presented earlier in the thesis. The Motion Detection in Focused Regions module takes inputs from Motion Detection module and Focused Region Extraction module to detect motion in focused regions only. The output of Motion Detection module is 1-bit motion information for 4x4 pixel block and the output of Focused Region Extraction module is 1-bit focus information for 1-pixel. Therefore, before processing the input received from Focused Region Extraction module, it buffers the information of three rows and starts processing when the fourth row is received. To achieve this buffering, three 36Kb Block RAMs and 16 1-bit registers have been used. After buffering, the Focused Region Extraction module performs the OR logic for Motion Detection output and Focused Region Extraction output for each 4x4 pixel block and sends the processed output for display. This block also allows the filtering of frames of interest based on motion detection only in the focused regions in a live video stream. If the motion threshold for focused regions in a video frame is crossed then the frame is filtered for further processing or storage. The 1-bit output signal of Motion Detection in Focused Regions module is also made available on FPGA GPIO (General Purpose Input Output) pin. This signal is high (‘1’) if the motion threshold is crossed in the extracted focused regions in a video frame. It can be used to generate an alarm signal for security personnel to pay attention to the occurrence of movement in restricted areas in a scene, monitored using video surveillance cameras. The output of Motion Detection in Focused Regions module is also used for automatic detection of a moving target in the incoming live video stream, and, the detected moving target is then tracked by Object Tracking module in subsequent frames. The moving target object for Object Tracking module can be selected either manually by moving the camera on the target object or automatically based on the output from Motion Detection in Focused Regions module.

While in automatic moving target detection and tracking mode, the implemented system starts detecting motion in focused regions at power-on. The size of relevant motion detection region varies, depending on the size of moving object present in the scene. However, the implemented Object Tracking module is designed to track a target of maximum 100x100 pixel size. For this region, a control logic has been implemented to select the upper most part of size 100x100 pixels of moving object. After selecting the upper most part of size 100x100 pixels of the moving object, it draws the rectangle on selected
region and passes the co-ordinate information to the *Object Tracking* module. In subsequent frames, the *Object Tracking* module tracks the detected 100x100 pixel size moving target with purposive camera movement to follow the tracked target object.

The implemented prototype automated video surveillance system is capable of automatically performing real-time motion detection, real-time video history generation, real-time focused region extraction, real-time motion detection only in focused regions, real-time object tracking of a manually selected target with automatic purposive camera movement, and real-time automatic moving target detection and tracking with purposive camera movement. Output of any of the five processing modules can be displayed by selecting the output of the associated module for display and sending it to *DVI Display Controller*. This output selection is done by a multiplexer (*MUX*). The output of a particular block can be selected for display based on the value of the 3-bit control signal to the multiplexer. The control signal is provided using GPIO (General Purpose Input Output) switches available on the Xilinx ML510 (Virtex-5 FX130T) FPGA development platform.

### 6.3. Synthesis Results

A top level design module was created which invoked all the five hardware architectures and all the four input/output interfaces. A User Constraint File (UCF) was created to map the input/output ports of the design on the actual pins of the FPGA. All the above mentioned modules of the implemented system were coded in VHDL and simulated using ModelSim. This top level design was synthesized using Xilinx ISE (Version 12.1) tool chain. The resulting configuration (.bit) file was stored in the Flash Memory to enable automatic configuration of the FPGA at power-on. A standalone complete prototype of automated real-time video surveillance system was built and is shown in Figure 6.4. The components of this system are Xilinx ML510 (Virtex-5 FX130T) FPGA platform, Sony EVI D-70P Camera, and display monitor.

FPGA resources utilized by individual hardware architectures and input/output video interfaces, maximum operating frequency for each design, and achieved frames rates by designed architectures were presented earlier in chapters 2-5 of the thesis. FPGA resources utilized by the complete automated video surveillance prototype system are given in Table
6.1. Maximum operating frequency of the complete integrated system is 125.8 MHz, and, maximum possible frame rate for PAL (720x576) size color video is 244 frames per second.

Figure 6.4: Complete Hardware Setup of Implemented Automated Video Surveillance System.

Table 6.1: FPGA Resource Utilization by Complete Implemented Prototype Automated Video Surveillance System.

<table>
<thead>
<tr>
<th>Resources</th>
<th>Complete Automated Video Surveillance System</th>
<th>Total Available Resources</th>
<th>Percentage of Utilization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Slice Registers</td>
<td>28838</td>
<td>81920</td>
<td>35.20%</td>
</tr>
<tr>
<td>Slice LUTs</td>
<td>39916</td>
<td>81920</td>
<td>48.72%</td>
</tr>
<tr>
<td>Route-thrus</td>
<td>9729</td>
<td>163840</td>
<td>5.93%</td>
</tr>
<tr>
<td>Occupied Slices</td>
<td>13588</td>
<td>20840</td>
<td>65.20%</td>
</tr>
<tr>
<td>BRAMS 36K</td>
<td>216</td>
<td>298</td>
<td>72.48%</td>
</tr>
<tr>
<td>Memory (Kb)</td>
<td>7776</td>
<td>10728</td>
<td>72.48%</td>
</tr>
<tr>
<td>DSP Slices</td>
<td>3</td>
<td>320</td>
<td>0.94%</td>
</tr>
<tr>
<td>IOs</td>
<td>292</td>
<td>840</td>
<td>34.76%</td>
</tr>
<tr>
<td>DCMs</td>
<td>2</td>
<td>12</td>
<td>16.67%</td>
</tr>
</tbody>
</table>

Synthesis results reveal that the complete implemented prototype automated video surveillance system utilized approximately 66% FPGA slices and 73% Block RAMs (on-chip memory) on Xilinx ML510 (Virtex-5 FX130T) FPGA development platform. Thus,
there is further scope of implementing a few addition features of an automated video surveillance system on the same FPGA development platform.

The implemented system works in real-time for a standard PAL (720x576) resolution live color video stream. All architectures of the implemented automated video surveillance system are adaptable and scalable for other video resolutions also and the results for the same have been presented in earlier chapters of the thesis. The seven configuration (.bit) files – one each corresponding to motion detection, history generation, focused region extraction, motion detection in focused regions, object tracking with purposive camera movement for manually selected target, automatic moving target detection and tracking with purposive camera movement, and complete integrated system are stored in the Compact Flash memory. With the help of three configuration switches available on the Xilinx ML510 (Virtex-5 FX130T) FPGA development platform, the desired configuration (.bit) file can be downloaded onto the FPGA at power-on and the system can be used for that purpose.

6.4. **Visual Results**

The implemented prototype automated video surveillance system can cater to various real world video surveillance applications such as motion detection, focused region extraction, video history generation, motion detection in focused regions, filtering of frames of interest, object tracking with purposive camera movement, and automatic moving target detection and tracking. Its real-time test results for motion detection, focused region extraction, and object tracking with purposive camera movement for a manually selected target have already been presented in earlier chapters of the thesis. The results for video history generation, filtering of frames of interest based on motion detection in focused regions only, and automatic moving target detection and tracking are presented in this section.

Results of video history generation are shown in Figure 6.5. Figure 6.5a shows the original extracted video frames from three different live video sequences. Corresponding results produced by the implemented motion detection system for each frame are shown in Figure 6.5b. These results show only the motion blocks in each frame. The results produced by implemented video history generation system for the three video sequences are shown in Figure 6.5c. For each of the three input video sequences, video history frame is generated
for 60 frames. The output video history frame for each video sequence shows details about
the moving object and the trajectory of the moving object. As a single output video history
frame is generated for 60 input video frames, the output video frame rate is also reduced.
The results of filtering of the frames of interest based on motion detection only in focused regions are shown in Figure 6.6. Figure 6.6a shows four different images extracted from different live video streams. Corresponding results produced by focused region extraction system for each image are shown in Figure 6.6b. In Figure 6.6a, Image 1 and Image 3 are focused images and Image 2 and Image 4 are defocused/blurred images. Therefore, results produced by the implemented focused region extraction system for Image 2 and Image 4 are black images, as in both the images there are no focused regions. The results produced by the focused region extraction system for Image 1 and Image 3 show the extracted focused edge regions. Motion detection results for the four images are shown in Figure 6.6c. Moving objects are present in all the four images. Therefore, results produced by the implemented motion detection system show the motion blocks for all the four images, irrespective of the fact that they are focused images or defocused images. The module for motion detection in focused regions detects motion in focused regions only and filters the frames accordingly (i.e. filters the frames which show the motion in focused regions only). Frames filtered by the implemented motion detection in focused regions system are shown in Figure 6.6d. Output frames are filtered for Image 1 and Image 3 as these are focused images. For defocused Image 2 and Image 4, the output frames are not filtered and therefore, the results show black images.

Results of automatic moving target detection and tracking for live video sequences are shown in Figure 6.7. All images corresponding to three video sequences are extracted from live output results produced by the implemented system. In all the three video sequences there exists no moving object in the first image. In the second image of all the three video sequences, the moving object enters into the scene and motion detection in focused regions system detects the moving object (results of motion detection in focused regions are not shown as motion detection system is working internally and results are displayed for
tracking only). The moving object/person size in all the three video sequences is larger than 100x100 pixel size and the object tracking system implemented in the thesis is designed to track a target of maximum 100x100 pixel size. Therefore, the upper 100x100 pixel size part (i.e. person’s face in this case) of the moving object is selected and red color rectangle is drawn by the implemented control logic (as mentioned in section 6.2). This is shown in the third image of each video sequence. In the remaining three images of each video sequence, the selected 100x100 pixel size target (person’s face in this case) is tracked despite the presence of other moving objects/persons in the video scene. The background also changes due to purposive camera movement to follow the tracked target and to keep the tracked target in the middle of the frame.

![Figure 6.6: (a) Original Video Frames; (b) Focused Region Extraction Results; (c) Motion Detection Results; (d) Filtered Frames of Interest based on Motion Detection only in Focused Regions.](image-url)
Figure 6.7: Automatic Moving Target Detection and Tracking with Purposive Camera Movement.

These experimental results demonstrate successful operation of the implemented system.
6.5. Summary

In this chapter, we have described the system level architecture and results of complete automated video surveillance system developed on Xilinx ML510 (Virtex-5 FX130T) FPGA Board by us as a part of this thesis. The implemented system is capable of automatically performing real-time motion detection, real-time video history generation, real-time focused region extraction, real-time filtering of frames of interest based on motion detection only in focused regions, real-time object tracking of manually selected target with automatic purposive camera movement, and real-time automatic moving target detection and tracking with purposive camera movement. The system is designed to work in real-time for live color video streams of standard PAL (720x576) resolution, which is projected to be the most commonly used video resolution for next generation video surveillance cameras (for current generation surveillance systems most commonly used video resolution is QVGA (320x240) and CIF (352x288)). The developed standalone prototype system can be effectively used for surveillance applications.