The University Of New South Wales

School of Electrical Engineering and Telecommunications

Automatic Video Surveillance Systems

Summary for: “XXVII Convocatoria de premios de Ingenieros de Telecomunicación”

Gabriel Moreno Rodriguez

A Thesis submitted in fulfilment of the requirements
of a Bachelor of Engineering

THE UNIVERSITY OF
NEW SOUTH WALES

SCHOOL OF ELECTRICAL
ENGINEERING &
TELECOMMUNICATIONS

Submission Date: November 2005

Supervisor: Dr. David Taubman.

COIT membership nº: C 13761 / A 14449

Contents

I.- Abstract

I.1.- Motivation

As engineers, our first aim when choosing a project should be to find out the real problems which lie underneath it. Although we are developing an engineering project, we should never forget its goal and we should always look for an appropriate background were the subsequent project should be developed.

This is the main reason why our first step when facing a project should be to compile as much relevant information as we can. Why are Video Surveillance Systems needed? Which is their contribution to the society? Which sectors in our society need them? Which are the main objectives that these Surveillance Systems should cope with?

Probably, in order to solve some of these questions, we will have to read not only engineering books, but also sociological, philosophical or economical ones. Once, that we have discovered the main problems ("why do human beings need security?") we should solve them as engineers.

In 1943, Abraham Maslow's article called "A Theory of Human Motivation" was published in Psychological Review. Based upon his clinical experiences with human beings, Maslow attempted to formulate a needs-based framework for human motivation. The basis of Maslow's theory is that human beings are motivated by unsatisfied needs, and that certain lower needs must be satisfied before the higher ones. This way, Maslow created his own pyramid of needs.

In this regard, Abraham Maslow's model indicates that basic, low-level needs (such as physiological requirements and "safety") must be satisfied before higher-level needs (such as self-actualization) are pursued.

Hence, it can be understood why a project about security will be important enough. Surveillance systems perform a good solution to avoid risk situations and their use has become popular during the last decades, overall in those sectors where a guaranteed safety is really needed.

Throughout this report we will study the main features of Surveillance Systems. We will discover economical and social problems that have driven to the installation of security systems. Not only we will face with Signal Processing problems, but also, we will deal with networking problems and with real implementations of some surveillance products. All this information will be used in order to construct the background where our final system will be developed.

Surveillance systems are not a new fancy trend of the society. Indeed, there has been a continuous evolution from the traditional technologies towards the use of the newest ones.

Notwithstanding, there are still a big amount of museums, corporations, banks, schools and other security technologies users who keep on investing in old analogical systems, where a man has to be aware of some screens. Even so, video surveillance systems in large museums are still cost-prohibitive and ineffective. With hundreds of galleries to watch over, a big museum would require a big quantity of trained technicians simply to monitor the feeds. This is the reason why most large museums use “Containment security”, as it is pointed out in some best-sellers novels.

It is clearly shown that it is a very tiresome task for a technician as he will become easily tired and absent-minded after being aware of a huge amount of monitors for a long period of time in order to decide whether an event is important or not. So, in the case of large corporations, managers should employ a huge amount of staff who should be “always-on” trying to identify the threatening elements in their screens. And even more important, in those old technological systems, everything has to be recorded onto tapes. This means that corporations need big rooms to store everything and each time that someone needs to re-view a video, someone has to go to that room and pick up the selected tape (which can be damaged or even ruined).

I.2.- Objective

With the use of newer technologies some of these problems can be overcome. For instance, automatic systems will automatically notify when some important events occur. Digital video compression based on MPEG4 or H.263 can be used to store everything on our databases. These databases can be managed from different locations throughout Secure Network Protocols. And even, we can have some back-ups for the information in diverse locations to avoid losing risks.

Digital Video Applications have powerful impact on Video Surveillance Systems. As previously explained, the tedious tasks performed formerly by human staff, will be replaced by an automatic detection, localization and tracking of moving objects. Highly inefficiency was introduced in those former systems due to the fact that the store video data was normally redundant, thus wasting operator time, money and storage space.

This thesis aims to perform an innovative video security system using digital video processing tools and so, reducing the human component inside of the system as automatic computer-vision based surveillance systems normally perform the following functions:

- Automatic acquisition of images from a camera.
- Auto-calibration of the camera.
- Differentiation between foreground and background pixels.
- Tracking of foreground objects.
- Object recognition.
- Storage and retrieval of important information.

The latest advances in DSP hardware make it even possible to embed many of the low-level functions listed above into a camera. Cameras with such embedded functionality might have the option to provide not only raw pixels, but also metadata that describe the locations and motion of objects viewed by the camera. As cameras with embedded metadata production can transmit data over a variety of media, use of computer networks is convenient because it uses established infrastructure as well as exploits the abundance of network devices that are already available.

Something also to take into account is that a video system is desired to get the minimum false alarms as possible while keeping the sensitivity of the system as high as possible. This will require a minimum of knowledge about the scene.

Therefore, the objectives that are going to be pursued within this thesis can be summarised along the following lines:

- The first step will be the creation of an interface between the camera and the user in order to be able to manipulate the images and to extract all the available information that they have.
- Then, we will differentiate between foreground and background pixels within a scene. We will detect the possible objects in the area under surveillance, also known as blobs. And we will get rid of those objects that are not important for the operator, that is, we will perform noise removal operations. Afterwards, we will make motion estimation and we will track the blobs throughout the scene. This way, we will count them; having an exact record of how many people have passed through the scene and taking into account their direction.
- Simultaneously, an automatic saving module will be created. This module will only save those images in which some important events happen.

- There are some facts that will impose constraints to our system. People can cross with each other in the area under surveillance. The way of walking varies from an individual to another, that is, some people walk faster than others. Moreover, several people can be simultaneously inside the same scene.
- Finally, as we previously mentioned, we would like to perform all these operations in real-time while reducing as far as possible the number of false alarms and the probability of error. But by keeping the system sensitivity as high as possible.

These considerations will drive us into the report's aim: "The implementation of an easy-reconfigurable and automatic video-surveillance system for counting people in a real-time environment".

I.3.- System Overview

The Automatic Video Security System developed within this thesis can be split into the following components as shown in Figure 1:

1. Image Acquisition: This component is the one that implements the interface between the camera and the computer. We will be capturing images in a RGB base so that we can manipulate them afterwards. We will construct a GUI to easy interact with the program.
2. Background Subtraction: The second component aims to label pixels as either foreground or background, thus creating a bi-level image.
3. Morphological Filtering: Within this section we will use morphological filtering to perceive pixels as regions rather than isolated entities.
4. Alerting Phase: In this phase, the aim is to classify the objects and decide whether they are important to be tracked or not.
5. Tracking Phase: Next section is referred to the tracking along the scene of those objects that were considered as relevant. We will be able to track movement along the scene.
6. Interpretation Phase: Finally, we will use a module to decide when to count people and the interpretation of what has happened inside of the area under surveillance.
7. In addition, two different databases will be used; one to automatically store images in it and another to retrieve previously stored information to use it as a controlled test environment.

Note: If you require any further and or technical detail about any of the structural items and how they were configured in the final system, please refer to the full thesis version.

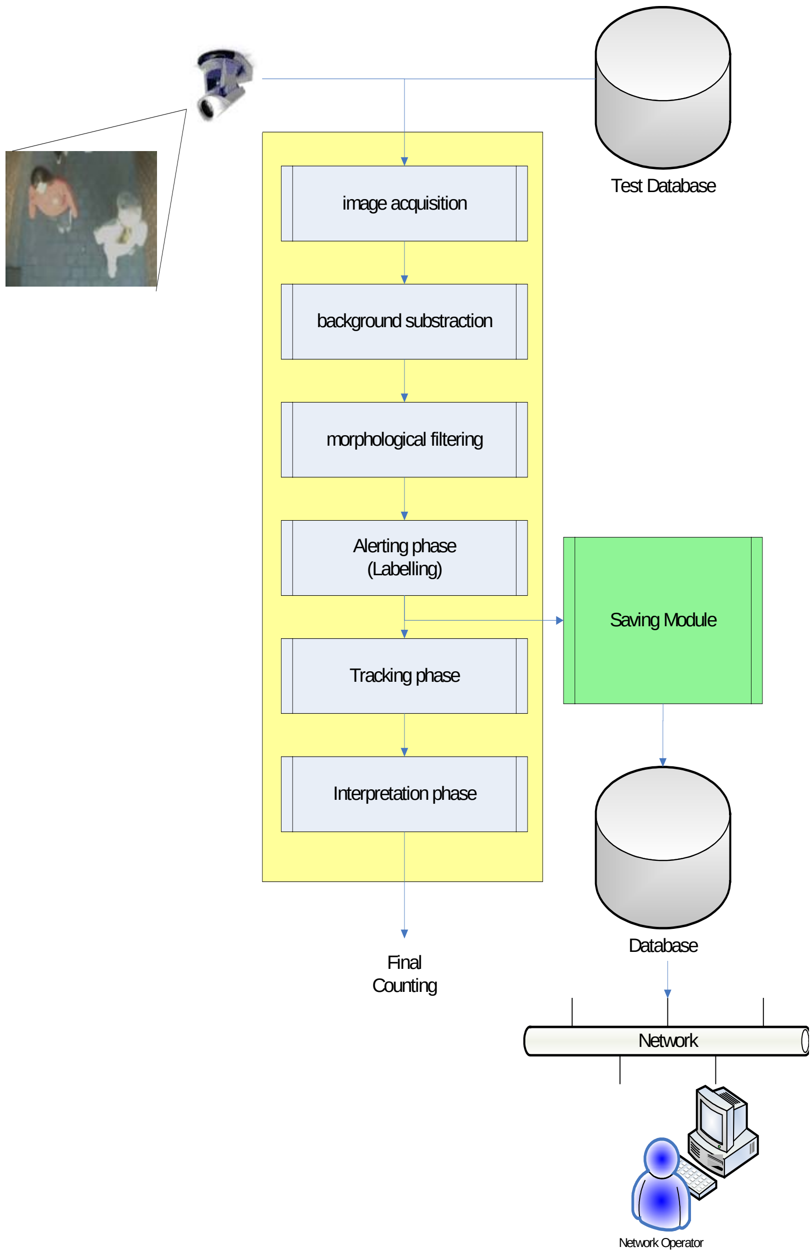

Figure 1: System Overview

I.4.- Conclusions and Further Work

Along this thesis we presented the state of the art for surveillance systems, the importance of security systems in our society, the security industry growth and its future revenues.

It was within this environment where appeared the idea of applying video processing techniques to the video stream provided by a camera. This way, an automatic system was built and so, a reduction in time and costs was achieved.

Firstly, the most notable feature in our system is its real-time capability. This is its most valuable characteristic. We should never forget that our system was developed to be run in a common laptop with a common USB webcam. So, this is an open and reconfigurable system and it can be easily run in both public and private sectors in a cost-effective and intelligent way. On the one hand, big industries, museums or shops could decide to use an automatic surveillance system as it helps to reduce their expenses and to build a secure environment. On the other hand, this system could be also installed in a private house, as it only needs a webcam and a common computer to be run. There is no need to use a high-complexity or expensive system as our solution provides a good way to tackle the surveillance problem.

Moreover, system benefits do not stop at this real-time capability. Further benefits can be found if we think about the storage reduction. In the former ages, “everything” had to be recorded into tapes but now our system records automatically “the important data” in a hard disk. Another benefit is the reduction of the human effort in the system.

The system satisfies a good level on a PEST (Political, Economical, Social and Technological) analysis so it could be easily commercialized. As a summary of this analysis we can point out:

- Political framework: It helps to provide enhanced security standards to the population.
- Economical framework: Its use will dramatically reduce the company expenditures.
- Social framework: It will contribute to guarantee people’s welfare state.
- Technological framework: It comprises different video and signal processing techniques, so it can be considered as a first step in order to achieve an enhanced video security system.

Besides, the surveillance system developed within the project has provided a good performance during the field trial tests. As far as possible, we got a low probability of false alarms and the system accuracy was high enough.

To conclude, just to point out that this thesis does not attempt to be a final solution, but the first step in the development of an automatic, cost-effective, real-time and reliable surveillance system. Further work to achieve an improved system is needed. However, more than 80% system accuracy was achieved during the trials.

Therefore, although the main objective of this thesis has been accurately covered, it can only be seen as the first step in the difficult process of developing a trustworthy system which will improve one of the major needs of human beings: Safety.

II.- Novelties – Originality

The originality of this thesis was found in both the methods that we applied to solve the problem and the compilation of information about the surveillance system state of the art. Here, we will roughly point out how we manage to solve the different modules in order to build an automatic video surveillance system.

Our first step on the thesis elaboration is to deal with the acquisition of the image through the camera. To cope with this task, we have developed our own application interface capable of handling images from the camera, showing them in the computer and recording them in our hard disk. This camera-computer-user interface was developed in Microsoft .NET (Visual C++).

Once that the image has been captured by the camera, we will be able to apply our algorithms to it. Nevertheless, the captured image was given in a RGB format and we worked with the luminance pattern (YIQ colour model) to perform the subsequent detection.

At this point we began the pixel-level processing. This processing will be used to determine whether a pixel belongs either to the foreground or to the background, thus creating a bi-level image. Many authors, who have already stated some relevant problems that should be eradicated when facing the background subtraction, have already tackled this problem. Problems such as the following ones has already been tackle thanks to our algorithms:

- Changes in scene illumination.
- Foreground aperture.
- Moving objects in the background:
- The “occlusion” problem.
- Absence of training:

We decided to build an “Adaptive Background Model” to exploit the good properties of the already known “derivative algorithms” and “static background algorithms.” Our “3-Weights adaptive background model algorithm” was designed to address the foreground acquisition problems in a computationally efficient manner. This algorithm is based on a feedback loop to upgrade the actual background model according to the image changes.

In the following figure, a diagram of the algorithm can be observed:

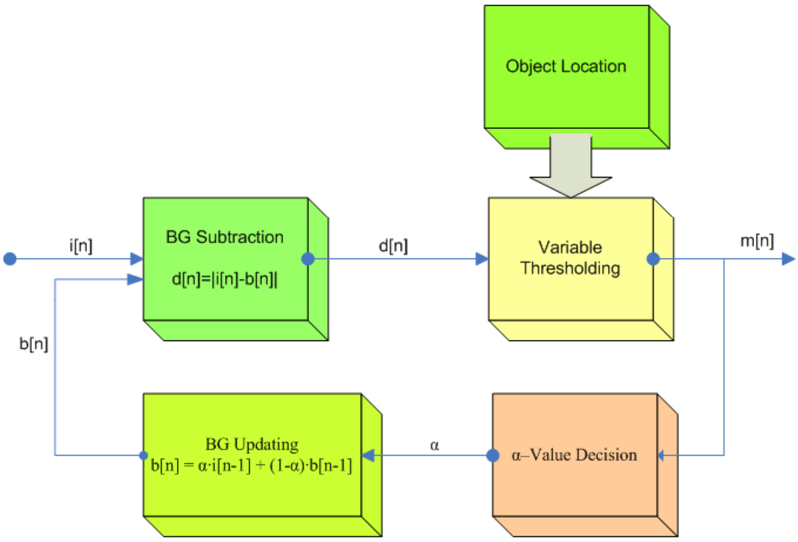

Figure 2: 3-Weights Adaptive Background Model Algorithm

The threshold value will be variable and depending on:

- Sensitivity to object presence in the area under surveillance:
- Sensitivity to the likely position of a given object:

This way, we will construct "threshold valleys" in the most likely positions. Furthermore, in the 3-Weights algorithm, instead of using a single value of α, we will use the previous information from the motion image (m[n]) to switch among 3 different α values (α_t , α_s , α_f) according to the image properties.

We could summarise the benefits of this algorithm:

- Computationally simple.
- Minimal memory consumption.
- Combines the benefits of the derivative and the static background model algorithms.
- It is able to handle with transient lighting changes.
- The possibility of choosing α is also a major advantage of this algorithm.
- It has a variable Thresholding feature.

Although "Background Subtraction" will provide us a good approximation of those pixels belonging to the foreground, we will not be able to distinguish among whole objects yet. The fact is that we do not perceive pixels as isolated entities; rather we perceive regions. Therefore, we should manipulate the foreground pixels to obtain possible foreground regions by eliminating noise and by agglutinating spread pixels into a unique region. We will achieve it by means of morphological operations. We should apply an Opening operation to eliminate possible noise and transients from the scene. And immediately afterwards, a Closing operation should be also applied to merge possible disjoint parts of the same object and to smooth the final image. However, this solution brought some problems in the timing for our real time system. The solution to this problem can be achieved by using the properties of the morphological operators.

Once that the objects are defined, in the alerting phase, we will classify those objects as entities, calculate their centroids and decide whether they are important for the tracking phase or not.

In the tracking phase, an algorithm to follow the objects that are going to change their position from one frame to the next one was developed. Object tracking is a difficult task. As in previous chapters, we could use different image features in order to track the objects (or even a mixed of them). We will follow this schema:

- Prediction of the next position of the blob in the following frame.
- Use of distance criteria to determine the best match nearby the predicted position.
- If a match is found, refine the position and upgrade the templates for the next searching phase.

Finally, during the interpretation phase, the system is going to decide what is happening in the scene. Basically, a set of three possible events were defined:

- A new object can appear in the scene.
- An object can change its position within the scene.
- An object can disappear from the scene.

All this algorithms were fully understood and implemented during the thesis work and they constitute the basis for the final results.

III.- Field Trial – Main Results

Along this chapter, we will introduce some of the experiments that we performed with our system. We will point out how our solution overcomes some of the problems and the complexity of the proposed algorithms. In the project CD, the reader can find all the information needed in order to run again these tests.

One of the test aims was to find out the program complexity in terms of number of operations. Taking a first glance to the complexity problem, the morphological filtering was the most significant phase. A bad performance in the morphological filtering means that we will lose the real-time capabilities. This was the main reason to split the opening and closing operations into simpler operations such as erosion and dilation with a simple structuring element instead of a huge element. Thanks to both decisions, the number of operations for the morphological filtering phase will improve around 85.4%. This was clearly reflected during our tests by checking the real-time performance.

This drives us into a complexity comparison of each of the program modules. This comparison has been made in an intuitive way, so the percentage assigned to each section might not be accurate enough. The following figure tries to represent the complexity of each program part in relation with the total program complexity:

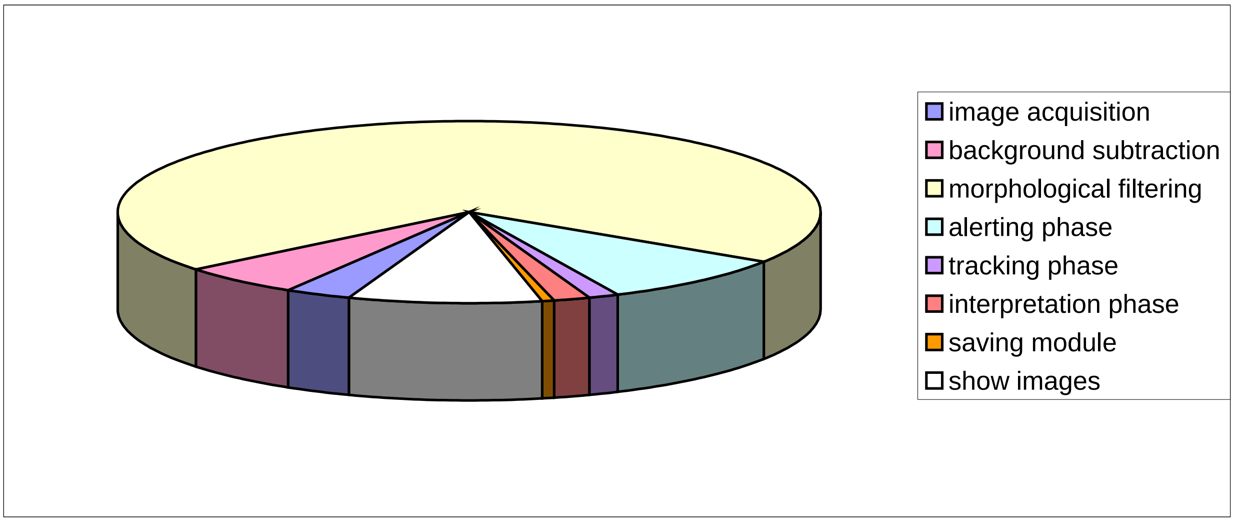

Figure 3: Program Complexity

Two different tests were performed with the final system. The first one was a simulated environment that used the recorded images of the testing database. The second one was a test in a real environment (an apartment corridor).

Controlled Environment (Test Database)

Our system was aimed to count those people that fully crossed the area under surveillance. Moreover, it will distinguish between two different directions: people heading from the upper part towards the bottom part and people heading the other way round.

We can say that the whole system performance was good, as it was able to solve most of the cases that we dealt with. Furthermore, we identified why the system could not cope with one test, and we suggested one feasible solution to it.

These examples state the importance of the splitting and merging problem that we have further developed during the full version of the thesis and they can be repeated if you install the program version that is available in the CD.

The next table is a summary of the tests performance using the database:

Test Name	Description	Performance	Comments
Unoup	One person coming from the upper part of the scene and going down	Good	
Dosupuno	Two people coming separately from the upper part of the scene towards the lower part	Good	
Dosdownuno	Two people coming separately from the lower part of the scene towards the upper part	Good	Some noise (shadows) detected, but it did not affect to the system, as it was easily erased
Dosupdos	Two people coming from the upper part of the scene towards the lower part and simultaneously in the area under surveillance	Good	
Dosdowndos	Two people coming separately from the lower part of the scene towards the upper part and simultaneously in the area under surveillance	Good	Some noise detected, but it doesn't affect to the system
Unomedio	One person that goes to half of the area under surveillance and then he decides to come back towards the same entrance (so he is not counted up)	Good	
Dosmedio	Two people that appear inside of the area under surveillance, they begin talking and after a while they come back to the same entrance (therefore, they are not added to the counters)	Good	
Doscruce	Two people crossing each other in the area under surveillance (one of them comes from the upper part and the other comes from the lower part)	Bad	Example where the use of a splitting/merging module will give an enhanced system. Besides, we could apply newer constraints to the matching criteria of the centroids
Trescruce	In this example three people were simultaneously in the area under surveillance (two coming from the upper part and one coming from the lower one)	Good	
Morethanone	More than one people walking along as a group	Good	Different environmental conditions, the camera was located higher, so the Np value was smaller

Table 1.: Test Performance

Real Environment Test

The final test was performed in a real environment. That is, the camera was located in a corridor during two different days and people were not warned about its location. A total of five hours were processed by the system. When an event occurred, the system processed it and the network operator noted down if the system was able to solve it properly. A total of 82 people were processed by our system and only 16 had a bad performance. Therefore, we can say that the overall result was good enough as we obtained more than 80% in the system accuracy.

Indeed, through to the interpretation of the results, we can extract some information to improve the system.

A total of 57 people walked alone into the area under surveillance. But, only 8 of these people were not counted. Thanks to the annotations that we made during the test, all these people were wearing dark clothes. These dark clothes were easily melted with the background (formed mainly by grey tiles). In order to solve this problem, the programmer could use different techniques for the background differentiation. For instance, techniques based on the image texture could improve the final performance.

The system was able to count 10 people when there were at least two people in the area under surveillance at the same time. But, it did not work properly for another 7 people. Normally, the problem was also the dark colour of the clothes.

Besides, the system managed to solve two objects formed by two people. So, the system counts 4 people after these objects crossed the scene. However, there was one case where the system counts only one person but 2 people formed the blob. To solve this problem, we could reduce the value of Np in our algorithm.

Finally, it is also remarkable the fact that the system managed with a 3-people crossing (two going up-down and one going down-up).

Summing up, the total system performance was quite accurate and we filtered those cases in which we obtained a lower performance level as we detected the black-points to be improved in the system.

IV.- Real Implementations

IV.1- The Graphical User Interface (GUI)

The GUI provides the way of interacting with the main program and it shows the system information in real-time to the user. This GUI was fully implemented during the thesis work by using Microsoft Visual Studio .NET 2003. More specifically, we developed the application with Visual C++. Its appearance can be seen in the following figure:

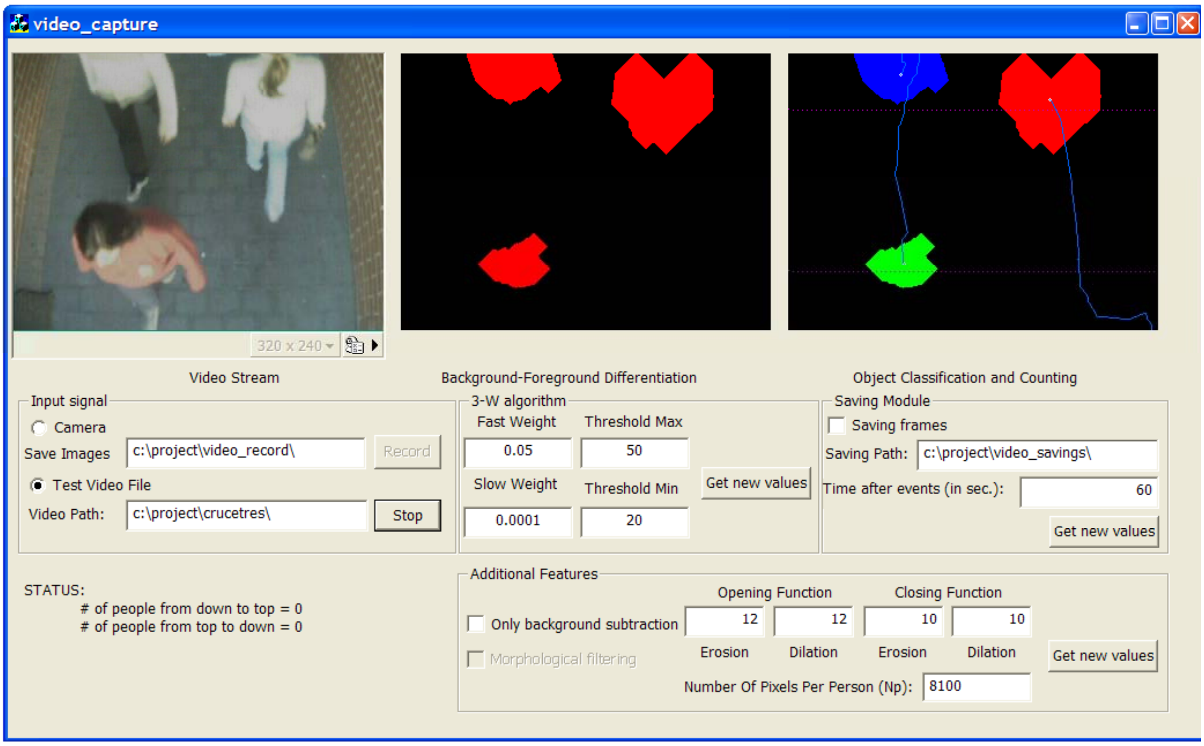

Figure 4: Graphical User Interface (GUI)

The GUI is comprised of three different screens:

- In the left screen, the user can check in real time what is happening in the area under surveillance.
- In the centre screen, the program shows the background and foreground pixels (marked as black and red respectively), once that the system has performed the background subtraction and the morphological filtering.
- Finally, in the right screen, the program shows the object classifications, the object centroids and the track-lines.

Underneath these screens, the system shows the system parameters to interact with the program and the actual status of the system.

The video input to our program was fed by using a Logitech QuickCam. This camera provides a 24-bit RGB image at a resolution of 320·240 pixels. It was physically connected to the computer through the USB port and it was interfaced with the software thanks to the Logitech SDK. The camera frame rate was variable, over 25 to 30 frames per second (fps). So, in order to operate at a fixed rate, a timer was used to subsample the images at 20 fps.

Besides, some of the program functions were initially developed in MATLAB. These functions were tested using single images instead of a video stream. This way, it was easier to find bugs in the function. Finally, we only had to translate the code from MATLAB to C++ .NET.

IV.2- Real market implementations

In the previous section we have dealt with the GUI developed for this thesis. The PEST discussion has also been taken in previous sections and it showed that a real implementation of this proposal was easily commercialised in the public transports, airports, shopping centres, or theme park industries. Such systems may be used to perform the following activities:

- To monitor after-hours activities.
- To perform access decisions.
- To improve safety in a parking garage.
- To implement publicity control (finding the most effective advertising times and locations).
- To estimate queuing time.
- To compare between number of customers and sales per customer.
- To calculate the lease value (according to the number of people that decide to have a look to a given property).
- To redirect visitors in crowded areas.
- To accurately allocate the staff (according to people flows).

Biodata Ltd, a UK company, offers a system that copes with a wide range of difficult situations, such as busy exhibitions, changes of light, people turning sharply, groups of people or the problem of having people of different heights.

Acorel, a French company, is also implementing automatic surveillance systems similar to the ones that we are facing in this project. They have already installed their equipment in train and bus stations, theme parks, emblematic spots and shops.

Dilax is another clear example of an enterprise that is focused on Automatic Surveillance Systems for counting people. They have installed a stationary people counting system in the casino of Ibiza (Spain). Five sensors installed in a sensor bar above the door detect all entering and leaving visitors. Real-time visualisation and monitoring of counting data are also done. The same system (only with 10 sensors for 2 doors) will be installed at the clothes shop "Polaris World" in Murcia (Spain). Furthermore, Transports Metropolitans of Barcelona (TMB) in Spain has placed an order regarding a people counting system for the main entrance of an underground station in Barcelona. And there is more specific cases, for instance, two more systems will be installed at the direct access to underground lines number 2 and 5 to control the fraud and distribution of passengers.

We have discussed some real implementations of this kind of systems as well as some probably uses. For instance the Australian industry called Beonic suggests that this kind of systems could be used to revive "dead" retail areas by measuring how changes to store layout and promotion can improve customer traffic and product sales in those areas. Also, this kind of systems helps to optimise visitors traffic flows by eliminating bottlenecks and reviving "dead" areas. Moreover, they could be used to reduce queues by allowing retailers to set alerts that trigger requests for extra cash register staff when queues reach a pre-determined length. By analysing patterns in pre-sale visitor behaviour, one can predict how changes to retail area layout and centre promotions can influence that behaviour favourably. The people counting solutions provide vital management information to assist decision-making for a wide range of public facilities, including government-run locations, tourist attractions, entertainment venues, libraries and museums. For instance, let us suppose a big event in a given stadium. In case of possible congestion in a given corridor, an alarm could be raised. This alarm could be immediately sent to the staff. So, they could be able to apply measures to control that possible congestion.

All the recorded information by our system could be further processed using other programs. For instance, we could add compression features to our final system. Furthermore, we could build a management system. This system could interact with the camera from a remote location through secure and reliable protocols like IPsec. Different network technologies could be used: wireless, optical fibre or coaxial, among others. Even, we could think about an alerting system. So, when an event occurs, an alerting message would be sent to the network operator. This alerting message could even be an sms to the operator's mobile or a mail to his PDA. All these features will contribute to build a reliable always-on automatic surveillance system.