Abstract—In this paper, we survey the state-of-the-art in the development of automated visual surveillance systems and discuss on how this can be applied for the national security. Video surveillance system is the very complex task that needs to combine multiple technologies together. We summarize the progress of this technology in term of sensor and devices, infrastructure, and intelligent technologies. The intelligent technology for video analysis is emphasized more deeply on the ability of object and human recognition techniques, to understand their actions and interactions. The increasing need for intelligent surveillance in the national security, especially in the south of Thailand, for law enforcement and military applications makes automated video surveillance systems become very important. We discuss on why the CCTV based system installed in the area is only a basic necessary step for the better system with more powerful features needed for nation security protection.

Keywords— CCTV, surveillance system, video analysis

I. INTRODUCTION

The past decade has been the rise of video surveillance and its deployment spread around the world. The transition from analog to digital systems is marked by decreasing camera and recording equipment’s costs. The advances in digital sensor, compression technology, and IP network infrastructure reduce the cost of transporting video data over large distances.

In Thailand, the system of video surveillance has been deployed massively during the last few years, especially in the southern part, which mainly serves for law enforcement and the internal security protection. This article reviews the technical terms of future video surveillance systems, focused on the video intelligence, that are taking place to meet the theses needs. From our point of view, we group the research technologies for video surveillance system into three groups: the video analysis framework as an online duty heavy processing system, video processing and filtering as the intelligent detection methods, and video retrieval as video indexing and searching part. These three grounds of technology are discussed in the sections two, three and four respectively. In the section five, we conclude: How far the existing system from new technologies? Is it possible to be deployed? Will it improve or help for the better protection of national security?

II. VIDEO ANALYSIS FRAMEWORK

In the third-generation video surveillance systems, the video information, as a stream or sequence of digital images, provides an intelligent and efficient approach for automatic system analysis. Motion detection, object tracking, or abnormal event detection are some of the video analysis techniques needed as basic functions for modern surveillance systems.
Content-based video analysis for surveillance systems requires basically realtime processing with respect to some quality of services, such as accuracy, spatial/temporal resolution, and latency. Effectively, the challenging task needs a particular architecture to be supported [1]. In real situation, multiple camera sensors can connect to system for several analysis tasks: detection, tracking, recognition, etc. System’s resources, for example CPU memory and network bandwidth, are necessary to be allocated dynamically. Certainly, the scalability of architecture is required in order to guarantee a certain quality of services. Besides, the capability of distribution of analysis tasks to match with the allocating additional resources in an optimum way will increase systematically the efficiency of the system. A distributed scalable architecture is basically adopted as our conceptual application design, implemented by parallelizing and distributing different components to different cores and processors. The re-configurability of the system is another important requirement in order to reflect the necessity of usages in the real world. Content-based video analysis for surveillance system needs to be adaptable for the specific tasks, for example, local analysis with signal sensor, cooperative analysis with multiple sensors, and global analysis for overall decision system [2].

Many researches [3][4] have proposed new surveillance system architectures for supporting these technologies. For example, they [5][6] present cooperative target tracking methods using multiple cameras. The surveillance systems offered by private companies and government organizations range greatly in price and quality. However, most of them support only their own products and camera infrastructure.

The specification of a good system [7] could be defined based on the surveillance application, which concerns initially the video sensors, and should satisfy the criteria as follows:

- Support the multiple cameras platform, such as Axis, D-Link, Panasonic, Sony, Stardot, Pixord, etc.
- Support the multi-cores and multiprocessors architecture.
- Support the standard functions of the Network Video Recorder (NVR), such as viewing, recording (full, motion based), and playback.
- Support the additional plug and play modules of image analysis filters and encoders.

We found that, in the actual situation, the majority of systems installed in the southern of Thailand are close systems. A camera from a place cannot link to the other locations having different platform.

III. VIDEO PROCESSING AND FILTERING

The video processing techniques for visual monitoring, such as human activity understanding, is the key for the intelligent surveillance system in the future. A lot of work was done in the domain. From the basic techniques of processing video, such as object detection, tracking and recognition, trajectory analysis and object classification, to the very complex tasks are explored. Some specific problems are frequently invoked, for example, face analysis, license plate recognition, under vehicle scanner, helmet detector, gesture analysis and recognition, abandoned and removed objects, intrusion, pedestrian detection, etc. More advanced problems using multiple cameras or moving cameras are also discussed, such as, the 3D localization, people’s interaction, etc. Privacy issues in surveillance system are also being largely examined, such as how the automatic processing can protect privacy?
**A. Human Activity Analysis**

Recognizing human activities from video sequence is one of the most challenging problems of surveillance application. In the paper [8] is proposed a method for recognizing the basic actions of human activities, which is necessary for the event of interest detection, for example abnormal or rare events. Here, “action” is referred to simple motion patterns normally executed by a single person, such as walking, standing, laying, bending and sitting, and “activity” refers to the complex sequence of actions performed by several humans.

A large number of publications work for action recognition [8]. Efros et al. [9] attempt to recognize a set of simple actions (walking, running with direction and location), using a set of features that are based on blurred optic flow. Robertson and Reid [10] propose an approach where complex actions can be dynamically composed out of a set of simple actions by building a hierarchical system that is based on reasoning with belief networks and HMMs. Yilmaz and Shah [11] extract information such as speed, direction and shape by analyzing the differential geometric properties of space-time relation. [13] analyze the space-time volume. Zelnik-Manor et al. [15] define dynamic actions as long-term temporal objects at multiple temporal scales features. Bradski et al. [14] develop MHI for motion segmentation that allow determination of the optical flow.

**B. Abnormal Event Detection**

The abnormal event detection is another domain of research that many laboratories try to solve the specific problems. Y. BenezethI and all [16] explore a location based approach for behavior modeling and abnormality detection. They precede event characterization and behavior modeling at the pixel(s) level, based on motion labels obtained from background subtraction. Since events are temporally and spatially dependent, they learn the co-occurrence statistics for normal events across spacetime. The Markov Random Field (MRF) model is used to describe the probability of observations within the same spatio-temporal volume. The MRF distribution implicitly accounts for speed, direction, as well as the average size of the objects passing in front of each key pixel. The learned normal cooccurrence distribution can be used for abnormal detection.

Claudio Piciarelli and all [17] propose the techniques that detect the anomalous events differing from typical patterns. The proposed work addresses anomaly detection by means of trajectory analysis, which is based on single-class support vector machine (SVM) clustering using the identification of anomalous trajectories. Particular attention is given to trajectory classification in absence of a priori information on the distribution of outliers.

Weiyao Lin and all [18] present a method for automatic recognition of group activities. They propose to use a group representative to handle the recognition with flexible or varying number of group members, and use an Asynchronous Hidden Markov Model (AHMM) to model the relationship between two people. A group activity detection algorithm is introduced, which can handle symmetric and asymmetric group activities.

Ying-li Tian and all [19] present a framework to robustly and efficiently detect abandoned and removed objects in complex environments for real-time video surveillance. The background is modeled by a mixture of Gaussians. A model is employed to detect the static foreground regions (i.e., static blobs potentially corresponding to abandoned or removed objects). Additional improvement is implemented for shadow removal, quick lighting change adaptation, reduction of fragmented foreground regions. A matching method is proposed to trigger alerts indicating abandoned and removed objects.
C. Privacy

The privacy issue is a very important topic that is dominated by the question: how to protect in private appearance of people being monitored? Winkler, T. and all [20] propose a security enhanced smart camera based on trusted computing, which support the integrity and authenticity by cryptographically signing frames with a key before they are streamed. It supports the authenticity and confidentiality of image data with multiple levels of privacy protection, together with access control by encrypting sensitive image regions to maintain confidentiality using public binding keys.

Hosik Sohn and all [21] propose a privacy-preserving watch list screening system applying the concept of the anonymous biometric access control that guarantees both security of public places and privacy protection in a video surveillance. The face images detected at camera side are preprocessed, encrypted, and transmitted to the server side, which has a watch list containing face features. The similarity is measured in the encryption domain, and encrypted decision is send back to the camera for alert. All encryptions are performed using a public key, while decryption can be done only at the camera side using a private key.

S.-C.S. Cheung and all [22] describe a prototype system for privacy-protected video surveillance. An inpainting algorithm to erase privacy information from video is introduced. An optimized data-hiding scheme integrated with the H.263 compression algorithm is used to embed the identified private information into the modified video. Data is hidden by modifying the luminance DCT coefficients, which typically occupy the largest portion of the bit stream. With proper authentication, the embedded data can be retrieved along with the inpainted video for recovering the original data. A secure infrastructure similar to a Digital Right Management system is provided for complete control of privacy information with grant access level using a combination of an asymmetric public-key cipher (1024-bit RSA) and a symmetric cipher (128-bit AES).

Jie Shen[23] presents scrambling video information for privacy protection in realtime video, based on MPEG-4 standard using the symmetric key encryption and public-key encryption techniques to ensure privacy information security. The scrambling is applied on the intra macro block and outside of the motion estimation and motion compensation loop. The sign of transform DCT coefficients of Intra macro-block is pseudorandomly flipped based on a pseudorandom number generator (PRNG) with a seed state, encrypted by symmetric key, and so only the authorized persons are allowed to correctly decode the code-stream.

Frédéric Dufaux and all [24] address more specifically the cases of MPEG-4 as the standard in video surveillance equipment. They introduce two efficient approaches to conceal regions of interest (ROIs) based on transform-domain and codestream-domain scrambling. In the first technique, the scrambling is applied on the quantized DCT coefficients during the encoding process. In the second method, in the scenario where the video stream is already compressed by an IP camera, it is favorable to apply the scrambling directly in the codestream domain in order to save computational complexity, to avoid fully decoding and reencoding the video. In MPEG-4, DCT coefficients are coded using run-length as codewords of variable length code (VLC). The scrambling is done by randomly inverting some bits of the codestream. Both techniques successfully hide private data of ROIs while the scene remains comprehensible. Additionally, the amount of noise introduced by the scrambling process can be adjusted. It is shown to be secured against brute-force or error concealment attacks. The coding efficiency performance
Integrated Surveillance System for National Security - A Review

is small, and the required computational complexity is negligible.

IV. VIDEO RETRIEVAL

One of the important goals of surveillance system is to be able to retrieve any interesting information during a short period of time. The video retrieval systems are then designed and developed as an integrated multimedia information retrieval and database management infrastructure. It mostly emphasizes on the following two directions: 1) retrieval technologies as multimedia content representation by image processing and computer vision techniques, in order to extract the salient features from visual media, and 2) retrieval framework as video database management system to support efficient storage of multimedia objects and retrieval based on the degree of match.

A. Retrieval Framework

Many works have been done successfully. Roberto Vezzani and all [25] introduce the ViSOR system, which is a dynamic repository of annotated video sequences related to surveillance applications. The hierarchical taxonomy of the video concept is introduced: a “concept” describes either the context (e.g., indoor, traffic surveillance, sunny day), or the content including physical objects (e.g., building, person, animal) or actions/events (e.g., falls, explosion, interaction between people). With this taxonomy, the video surveillance concepts can belong to three semantically different categories: Physical Object, Action/Event, and Context. A video annotation can be considered as a set of instances of these classes. They are connected to the entity with a “IS-A” relation (a sort of specialization of the “object” class e.g., man, woman) or “HAS-A” relation (descriptive features e.g., contour, color, position).

Thi-Lan Le and all [26] focus on the method for relating recognized video contents (objects, events, and frames) with visual words. Object features based on color histogram and trajectory are used. An event is described with confidence value. Queries are formed, SQL-like grammar such a SELECT <objects/events/frames > FROM < Database time/location> WHERE <Conditions>. Different levels of queries can be made, such as finding objects with time and location, searching an object related to events etc.

Arun Hampapur and all [27] present a combining framework for video searching, which includes video parsing, indexing and query mechanisms. The Smart Surveillance Engine (SSE), hosting a wide range of video analytics like behaviour analysis, face recognition, license plate recognition etc, is used for parsing surveillance video into time intervals corresponding to “events of interest”. The Middleware for Large Scale Surveillance (MILS) extract meta-data descriptors for these intervals and indexing it into database tables. Solutions provide the query interface and result reporting. The SSE and MILS systems are developed by IBM. The “presence of people index” is established as the combination of face and people detection to ensure a very low rate of false negatives mechanisms for surveillance events. This search can be qualified by one or more of the following: object color, object class, object size, object shape, object location, object movement, time of event of occurrence and event duration.

Patrizia Asirelli and all [28] present the MultiMedia Metadata Management (4M) developed at CNR ISTI, which is an infrastructure to support the integration of media from different sources. This infrastructure enables the collection, analysis and integration of media for semantic annotation, search and retrieval consisting of five units: a feature extraction and processing unit, an XML database management unit, an algorithms ontology unit, a multimedia semantic annotation unit, and an integration unit. They found that MPEG-7 is the most mature and widely
recognized standard for multimedia description. The eXist database provided the most stable implementation and the critical features. They also proposed the utility of algorithm ontology to record and describe available algorithms for application to image analysis.

B. Retrieval Techniques

Many researches focus on how to extract the description of video used for searching, browsing or retrieval. Sangoh Jeong [29] investigated six histogram-based search methods using three types of distance, such as Euclidean, intersection, and quadratic, for two different color spaces: RGB and HSV. He showed experimentally that images retrieved by using the global color histogram may not be semantically related even though they share similar color distribution. Searching by color distribution will discard about object location, shape, and texture. In general, histogram-based retrievals in HSV color space gives better performance than in RGB color space. The intersection-based with HSV color space was found to be most desirable.

Li Zhao and all [30] propose an approach for video retrieval by querying with key frame or video example. They consider both, feature extraction and distance computation. A video shot is represented by key-frames determined by breakpoints of feature trajectory of video shot, and use the lines passing through these points to represent the shot, defined as feature lines. The distance between a query image and a shot is measured based on the so called Nearest Feature Line (NFL).

Jing Huang and all [31] introduce a new image feature called the color correlogram, which is both effective and inexpensive for content-based image retrieval. A color histogram captures only the color distribution in an image and does not include any spatial correlation information. This feature extracts the spatial correlation of colors. The correlogram gives distances in probability between any pixels of color spatially. The correlogram robustly tolerates large changes in appearance and shape caused by changes in viewing positions and camera zooms.

V. DISCUSSION AND CONCLUSIONS

We review in this paper the research on actual trends of video surveillance system. We can classify the corresponding research into three major domains: video analysis framework, video processing and filtering, and video retrieval. The overall system of video surveillance actually tends to concentrate mostly on the video processing method in order to detect a specific event, but the object retrieval from stored video are also very much emphasized.

This trend is confirmed by considering the system of video surveillance actually deployed in the observed area, in Hat Yai and Songkhla district. These systems are currently robust and stable at the level of video analysis framework. Streamed video from IP cameras are sent to the server, in which the video analysis framework analyses the very simple video processing tasks, such as motion detection and its related variations (i.e., direction control and automatic observation of a specific area). However, the framework is far from a standard. In some cases the camera connection is definitively locked with the platform that leads to the difficulty of relating video data from the different areas.

If we consider the intelligent features, the actual system is very distant from research domain. Most human activity analysis is not implemented yet. Only basic object considerations, such as position, size, or histogram, has been explored. Accordingly, the object retrieval features of systems are very poor. In the most cases, only searching and browsing of videos by using the general information (i.e., date, time, and location) is allowed. The limited features of surveillance system actual installed is then becomes sometimes a
problem by itself. When the videos from thousands of cameras are recorded on different systems 24 hours a day during a few weeks, how can we find efficiently or easily the interested events as a small shot of video from a huge video database?

There is only one way to solve the problem: to deploy little by little the intelligent video processing techniques for either detecting event or indexing video. However, this objective is not easy to attain. Video processing is a very time consuming process, especially for the complex event detection or abnormal detection or human activity recognition. Deploying such system means more advanced hardware needed to be installed that cost certainly much more money. Moreover, some limitations of video processing technique in the real environment are not well declared in most of research. Generally, the environmental factors, such as light condition, raining, shadow, etc., will make it fail. We can conclude at this stage that the visual surveillance system is waiting for the big step in the near future when the robust intelligent methods are implemented on higher performance hardware and at a small price.

REFERENCES


[16] Claudio Piciarelli, Christian Micheloni, and Gian Luca Foresti, Trajectory-Based Anomalous Event


