Appendix B

Report on Video Analytics in the Greater Toronto Area

An appendix to the 'Smart' Private Eyes in Public Places? report

Prepared for	Office of the Privacy Commissioner of Canada
Prepared by	Joseph Ferenbok & Silvia Valdman
	Information Policy Research Program, Faculty of Information, University of Toronto
Corresponding Author	joseph.ferenbok@utoronto.ca
Date	23 July 2011

Table of Contents

1. Purpose

The purpose of this report is to outline the literature, technologies and practise of the growing area of what is labelled here as video analytics—or the automated processing of video to derive 'meaningful' information. The literature gives a broad perspective on video analytics (VA) while the survey of technologies and practises was conducted mainly throughout the Greater Toronoto Area (GTA). We also specifically evaluate the Xiris video analytic system for de-identifying faces.

2. Introduction & overview

Increasingly, video surveillance systems integrate one or more layers of algorithmic analytics that mediate the flow of information between the operator(s) and the camera(s). This intermediate video processing, which may start at the camera but continue long after information has been recorded and stored, is referred to here as Video Analytics (VA). Although vendors also use terms such as smart video surveillance, Intelligent Video Analytics (IVA), Intelligent Analytics (IA), Video Content Analysis (VCA) to describe a range of video image processing techniques the term video Analytics will be used here to incorporate a range of functionality “from systems that classify and store simple data, through more complex systems that compare the captured data to other data and provide matches, to systems that attempt to predict events based on the captured data” (Norris and McCahill, 2006). Video Analytics as a form of digital signal processing allows for algorithmic surveillance (Introna and Wood, 2004) that is hidden from even the most observant surveillance subject.

What is video analytics?

Conventional, analogue CCTV cameras were limited by the required balance between the volume of information multiple networked cameras generated and the amount of resources necessary to screen, store and analyze the visual data. The deployment of digital video surveillance through Internet Protocol (IP) networks have and continue to change the balance between captured information and our ability to analyze and interpret data. Video Analytics is the umbrella term used to include a range or assemblage of technologies that convert visual data into information for purposes of automated surveillance (Gorodnichy, 2010).

Video Analytics rely on a foundation of hardware infrastructures. Analogue surveillance systems may still be found but increasingly suppliers are selling

digital equipment with increasingly more sensitive sensors, increasingly higher storage capacities and cheaper operating costs. It is not surprising that the retail sector is using digital cameras and storage for store level surveillance. Once infrastructures are in place however, it becomes difficult to assess how that information is processed. Video Analytics is conceptualized here as the software layer that turns the collection of digital video into meaningful information for automated decision making.

Understanding video analytics as software however is still itself a major task because programming computers to detect, classify, code and semantically label temporal and spatial patterns of pixels in a video stream is a complex problem. Video Analytic software involves a series of tasks that help computers to visually 'understand' contextual events like objects moving to a pedestrian falling. Video Analytic software or smart computer vision has a range of applications from ensuring proper welds in factories to detecting forest fires, to tracking customer patterns and identities.

Perhaps because of the complexity and scope of 'teaching' machines to see, video analytic architectures tend to be modular -- that is a series of software modules strung together -- where outputs of one module are used as inputs for the next. Generally an early processing stage includes a module that attempts to distinguish and track objects across successive video images. Differences between frames can help demarcate moving objects in the foreground from the background. Once detected, the background is subtracted, or discarded, so subsequent object and event detection modules can process the pixels that remain, for instance by comparing them with patterns in earlier frames or pre-defined reference models.

The image information can then pass on to object and event classification modules where patterns are classified and alerts triggered automatically. The criteria for alerts are based on predefined or user-specified models for various types of objects and events, such as the presence of a gun, a person running or an unexpected package. In the classification stages video analytic engines produce meta-data that both guides the visual processing and provides semantic descriptions of the video inputs. This classification of objects and events can also produce an index of objects and recognized events, which in turn can be used as meta-data for future search and retrieval and as data for further processing.

Based on the semantic coding of object recognized—e.g. a vehicle, body, or face—other algorithms may be employed to identify objects individually and track them between frames and across different video streams. If an object is classified as a license plate, for example, automatic number plate recognition (ANPR) (also known as automatic license plate recognition (ALPR)) software can then be used to correct any distortion and apply optical character recognition (OCR) to identify the registration.

An object determined to be a body can be analyzed for individual identity or behavioural pattern (e.g. 'loitering'). Though perhaps further off than ALPR in terms of routine adoption, but at least as challenging with respect to policy and individual rights, are the analytic techniques for behavioural analysis and biometrics such as gait or face recognition. Tracking and identifying people in

video is an active area of research, development and commercialization (Cai et al., 2009, Bojkovic and Samcovic, 2006, Gorodnichy, 2006, Shaokang et al., 2008, Suman, 2008, Goffredo et al., 2008).

II. Technologies and practices in GTA

The surveillance industry is a complex network where multiple actors shape the landscape the overall industry. Among these stakeholders are those who provide video surveillance services and equipment across the Greater Toronto Area (GTA). These include: retailers, commercial services (suppliers and consumers), wholesalers, manufacterers, and software developers.

Four site visits to venders of video surveillance equipment in the GTA suggests that currently few are carrying products augmented by 'video analytics'. When the concept of automated surveillance was discussed, retailers inevitably turned to products that included software for managing one or more surveillance cameras in a network. Two software suites were reviewed by the researcher that are packaged with wireless IP cameras (one Cisco, one Sony) and in both packages the software included options like remote IP video viewing, automatic 'patrol' and motion detection, but more suffisticated algorithmic analytics were included. Part of the issue seems to be that many of the cameras on sale are still measured in TV lines (400 — 540 TV lines is currently still a typcial range). This resolution provides only very coarse grain information and makes object recognition, event detection, human detection and identification, more difficult—garbage in, garbage out. Although HD cameras are available (1.3 megapixles or more) they are still relatively expensive and feature-poor.

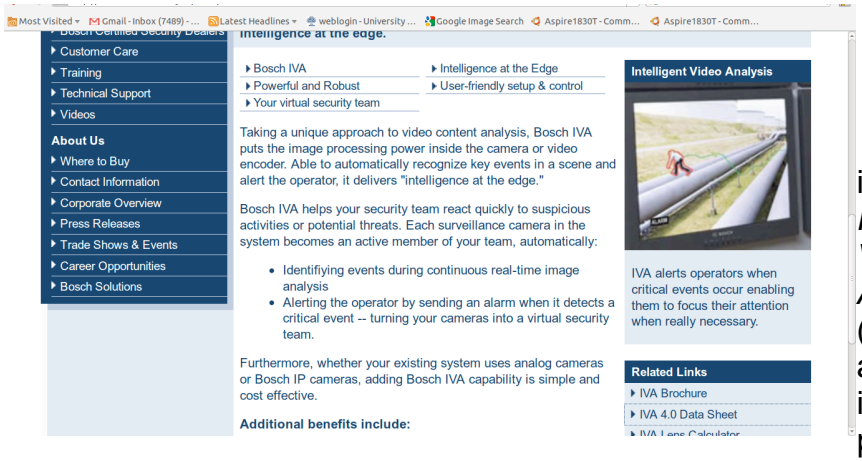
In our field work of retailers who employ video surveillance, only one retailer openly divulged that they employ automated people counters as part of their video surveillance—the Apple Store. But the employee also hat their video surveillance is networked and has remote access capabilities. So the infrastructure for harvesting video analytic information is already in place. It is only an incremental step to include software on the backend of an existing system—one that can remain entirely hidden.

A review of suppliers of commercially available video surveillance services suggests again that video analytics is still only at the very cusp of deployment. Security Companies in the GTA were unresponsive to requests to discuss video analytic technologies. During one site visit, the Director of Loss Prevention said “I not currently intersted in video analytics. I looked at it a while back but I can't see any advantages at the moment. But I think that [competitor X] is looking into it for their stores”. A site visit to one of this competitor's store suggested that none of the front-line security personnel were aware of any automated system used by the retailer. A subsequent information request initated by the researcher and the research assistant did not suggest that the retailer has any analytic capabilities.

Video Analytics vendors

Bosch

Looking at the manufacturers of surveillance cameras tells a different story. Bosch, a leading supplier of video surveillance equipment in North America,



includes *Intellent Video Analysis* (IVA) along side its core product offerings.

Bosch is directly claming that Intelligent Video Analysis can help augment human performance—addressing an established problem with human monitoring: attention span. The company claims that their system can recognize perimeter breaches and atypical behavours to bring to the foreground cameras where incidents may be occuring.

Bosch states that “no matter how few or how many cameras your system uses” their Intelligent Video Analysis can help you use your resources more efficiently. In Bosch's system the IVA is performed on a chip in the camera itself rather than

at a centralized monitoring station. The embedded digital signal-processing allows for objects to be recognized and coded in a meta-data stream (see figure 2) that may then be used to trigger events and process or aggregate data. The 'meta-data' stream can be decoupled from the original video for further processing, aggregation, modelling and other forms of 'intelligent' viewing.

Bosch lists among the advantages of this system, cameras with on-board digital signal-processing, that you don't need analysis servers or operating systems; and no need for centralized hardware.

Cognovision

Cognovision, now a part of Intel, is another piece of the future of Video Analytics. Instead of augmenting human performance, they measure it. CognoVision's core competencies centred around audience 'measurement' and & retail intelligence. According to their website, “CognoVision helps retailers & digital signage networks measure the effectiveness of in-store marketing and

understand shopper behaviour" (retrieved July 20, 2011) .

Amongst CognoVision's focus is retail intelligence by monitoring and classifying different levels of customer engagement. The company measures dwell time and traffic flow all with the promise to increase business—ultimately sales. Dixon (2010) has discussed how some large American retailers already employs such customer surveillance techniques in their stores. This trajectory seems to beg the question as to what degree is a corporation's ability to profit from an individuals personal information habits, before that corporation is breaching the right to information self determination.

Summary

The retail and security video surveillance market appears to be lagging behind the consumer imaging technologies like laptops and cellphones. Video surveillance vendors sell High Definition cameras of 1.3 megapixels or more, but these are relatively new and generally cost prohibative for small business or mass deployment. Surveillance camera retailers still discuss resolution interms of TV lines (400 — 540 is typical), while 1.3 mpx webcams (more than three times the resolution of 540TV lines[1]) are typical on most laptops and netbook computers.

Companies are currently offering systems that are supplementing traditional CCTV practices with video analytic features. Whether called Intelligent Video Analysis or Video Analytics, these products are already being offered. As the cost of sensors comes down, and higher resolution networked cameras become typical, adding video-analytics software becomes an incremental cost

1 Assuming a 3:4 aspect ration.

that may offer more granular and meaningful information. As video information captured in publically accessible spaces becomes more common and valuable, these issues of an individual's right to information self-determination will undoubtedly become more prevelant.

Video Analytics (VA), is an active research area and a developing technological edge in surveillance and consumer metrics. It is also clear that large stakeholders like Sony, Intel and Bosch are already trying to deploy 'intelligent' video surveillance systems particularly for large scale zone surveillance, traffic monitoring and for consumer metrics (see Bosch and CognoVision discussions). The features of these advertised systems go well beyond people counting to a range of analytic measures of an individual, event, behaviour etc.,. Though they do not seem to currently promise ubiquitous tracking across multiple zones and cameras or one to many identification at a distance it is clear that the sophistication of the techniques and technologies is improving. It is also clear that these algorithms are only the first barrage of video analytic strategies, applications and granularity. What remains unclear is the success of deployment or the actual effectiveness of these systems to perform as promised. As of yet, there is little evidence that these systems are being actively deployed in the GTA by retailers in publically accessible spaces.

Based on observations and field work it appears that the retail surveillance in publically accessible spaces has hardware that lags behind current generations of consumer electronics. In the GTA video equipment suppliers are still selling equipment that measures image quality in TV lines rather than Megapixels. Camera side analytics will likely remain bound to event detection and recognition applications like traffic monitoring and perimeter security, until equipment is upgraded to include higher resolution sensors. As cameras are updated video analytics and digital signal processing incorporated into cameras may become the norm.

III. Framework to Assess Object-level Coding

In trying to assess a Privacy Enhancing Technology (PET), it seems appropriate to turn to established Privacy Principles. These Privacy principles provide at least one lens to inform whether a new technology 'enhances' privacy. Particularly appropriate to any PET are the following four privacy principles first entrenched in the US Privacy Act (1974):

1. Accuracy:
Does the system help reduce errors or improve accuracy?
Does the system introduce new sources of potential error?

2. Access:
Does the system provide appropriate access?
Does the system allow for access control / managment?

3. Security:
Is the system as secure as, or more secure, than existing solutions?

4. Accountability:
Does the system provide mechanisms for improved oversight and compliance?

These principles serve as the framework for assessing face de-identification through Object-Level-Encoding as a Privacy Enhancing Technology.

IV. Evaluation of Object Level Encoding (Xiris Solution)

Background

Secure Object Level Coding is a technique developed at the University of Toronto that is currently undergoing commercialization (Martin and Plataniotis, 2008) by Engineered Privacy Inc. (EPI). EPI is a joint venture between Xiris Automation Inc. and the University of Toronto and Martin and Plataniotis continue as contributors. Their system employs secure visual object coding to encrypt objects identified in an image on a separate information layer. The encrypted object is then obscured on the original image layer. Combined with face detection this approach allows the privacy of individuals to be protected while retaining all data for retrieval with the proper access controls.

This software allows objects detected in video signals, like bodies and faces, to be obscured. What an “operator” sees is simply a blob (or any other graphic, logo, or image) instead of a face (see figure 1). This software may be performed on a chip built into a surveillance camera or performed before storage or viewing (not unlike the Bosch camera discussed above). Performing the encryption before transmission or storage would provide for optimal privacy protection. For example, Figure 1 is part of footage provided as part of a PIR. The video includes images of other people who have not been de-identified. Figure 2: shows an image of a woman with a stroller whose data was provided as part of the PIR. The footage provided has no faces obscured and hence a privacy violation. Using the system by EPI even faces at the back of the frame are obscured.

Figure 1: Obscured faces

Figure 2: Face in the Crowd

Object-level Encoding detects faces in video and replaces them with other information. In Figure 1, the system has detected the faces of several of the subway passengers and has replaced them with a white mask. The face data within the white masks is secured through an encryption algorithm and then either added on a separate layer of video for transmission or stored separately from what the "operator" sees. In the case of police warrant for the data, a system administrator may be asked to provide the "key" to the system and the face data can be decoded and reintegrated to create the original data image. The entire process only adds approximately 5% overhead, or additional bandwidth to the original data stream from a surveillance camera (depending on the number of faces and their relative size and duration in the video).

Figure 3: Still image from original PIR video of the Principal Investigator

When processed using the systems being developed by EPI the same video masks out facial information. As the PI comes into the frame a white mask instantly blocks out his facial features. Anyone viewing the video without proper permission would not be able to see more than the image below (Figure 4). However, when proper passwords are entered the image can be restored to its original form.

Figure 4: White mask obscures face in processed video

Secure object level encoding allows for privacy by default and access to the visual data only using proper permissions or access control. However, its effectiveness depends in part on the function of the overall system. For example if the face detection modules have difficulty detecting faces at angles between a face profile and an direct (forward) orientation, then portions of the final 'encoded' data my not fully obscure faces in the video at these angles. This then may lead to only partial information encoding and there may be frames where faces are partially visible (see Figure 4).

Figure 5: Limits to Masking

Although this limitation is relatively minor compared to entirely unprotected images--especially since subject's face is already partially obscured by the jacket—this may raise the concern that super-sampling the image may one day allow possible de-identification without proper permission. However it is likely that this limitation will be improved as the system is further developed. Furthermore an easy way around this is to apply secure coding to the entire body rather than just a face. Entire bodies, moving objects or events may be obscured and replaced by background data from other frames. This approach could be used to obscure all visual data in a frame that is not part of the background.

In Situ Testing

The above stills from the PIR video were processed in a controlled setting on at the company labs. We invited the company to a research day. The images in Figure 6 and 7 are the unprocess and processed stills from video captured that day.

Figure 6: Research Day audience

Figure 7: Audience after Encryption

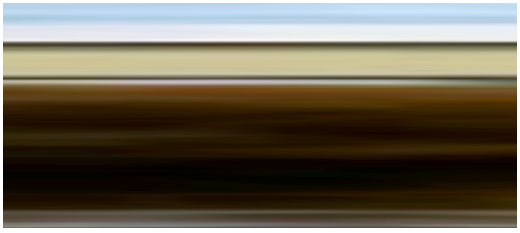


Figure 9: Background obscured

The above still demonstrate the potential of this system for encrypting privacy on a secure protected video information layer. The algorithm also demonstrates that it is not without some kinks. Figure 8 demonstrates a frame where the system mistakenly obscured an area above the individual's cup. However since the video information is still preserved on a separate layer and not changed in any

way, the accuracy of the original information is uncompromised. So although the system may itself have limitations currently (also see figure 5), and some infromation is not properly processed, it has no fundamental effect on the accuracy of the video information origially recorded when the process is reversed and the information is decrypted.

Access:
Since the system is designed to place personal infromation on a data layer separate from the original video layer it introduces a level of access control that the unprocessed video does not contain. The encryption and obscuring of the visual information prevents unautherized (re)viewing and can introduce a granular level of access control. The system can be set to obscure all but one specific face or a series of faces throughout a video sequence. If faces that are obscured are accessed without proper permissions, the image will only show as noise.

Security:
First it should be said that the video encoding portion of the overall security infrastructure is only as secure as other parts of the system. The strategy Xiris is proposing is most effective when incorporated at the video capture stage. If the de-identification is done on a chip built into the camera then the transmission of the video through IP and wireless networks becomes considerably more secure. If however the software is applied at the server or client side of the monitoring, then the storage and subsequent access may be considered more secure, but the transmission of the personal video information remains unaffected.

Accountability:
Accountability is arguably the principle most changed by the implementation of a deidentification system like Object-level Coding. With the ability to establish multiple information hierarchies and multiple levels of access to the encoded information if deployed as part of an infrastructure that promotes privacy and accountability the Xiris system imposes accountability on video surveillance systems—no individual can view the encoded data without proper permissions as well as recorded logs of the activity. However, with multiple layered accountibility and information coding it may become more difficult for individuals to gain access to their specific records as the information custodians intentionally or unintentionally make information requests harder to execute.

Overall, when deployed in a privacy sensitive framework face de-identification as a privacy enhancing strategy will likely prevent unauthorized access without compromising security. Once the viability of this technique is established, then by the data minimization principle those that do not incorporate this feature will be non-compliant, and potentially forced to adopt de-identification as the standard.

Works Cited

Bojkovic, Z. and A. Samcovic (2006).Face detection approach in neural network based method for video surveillance.

Cai, Y., D. Kaufer, et al. (2009). Semantic Visual Abstraction for Face Recognition Computational Science - Iccs 2009, Part I. G. Allen, J. Nabrzyski, E. Seidelet al. **5544:** 419-428.

Dixon, P., (2010). The *One-Way-Mirror Society:* Privacy Implications of the new Digital Signage Networks. World Privacy Forum. www.ftc.gov/os/comments/privacyroundtable/544506-00112.pdf *Retrieved: February 10, 2011.*

Gorodnichy, D. O. and E. Dubrofsky (2010). "VAP/VAT: Video Analytics Platform and Testbed for testing and deploying video analytics." Proceedings of SPIE-The International Society for Optical Engineering**7709**(Journal Article): 77090T-77090T.

Introna, L. D. and D. Wood (2004)."Picturing Algorithmic Surveillance: the politics of facial recognition systems."Surveillance & Society**2**(2/3): 177-198.

Introna, L. D. and H. Nissenbaum (2009).**Facial Recognition Technology:** A Survey of Policy and Implementation Issues, Lancaster University, UK; Centre for the Study of Technology and Organization.

Martin, K., Plataniotis, K., “Privacy Protected Surveillance Using Secure Visual Object Coding”, IEEE Transactions on Circuits and Systems for Video Technology, Vol. 18, pp. 1152-1162, 2008.

Norris, C. and M. McCahill (2006). "CCTV: Beyond penal modernism?" British Journal of Criminology**46**(1): 97-118.

Appendix A. Survey & Taxonomy of Video Analytics

Where is the state of the art?

Understating Video Analytics as a series of tasks allows for a selective review of research in areas including signal process, object recognition and classification, etc., to gain a sense of the state of the art in the field. Video analytic software may be divided into more developed activities (modules in the process chain) and higher level functions. The more established activities include signal processing of image quality, and automated object detection, tracking, recognition and classification.

Signal processing

Signal processing is itself a robust engineering area of research that includes a range of strategies to analyze and improve the quality of an image or video. Signal processing also referred to as Digital Signal Processing, is used to refer to a range of software and hardware systems that reduce signal noise, improve contrast and picture quality or improve data compression or transmission.

Ablavsky, V., M. Snorrason, et al. (2002).Real-time autonomous video enhancement system (RAVE).Image Processing. 2002. Proceedings. 2002 International Conference on.

The ability to autonomously enhance low-quality or corrupted streaming video data is essential in a number of important civilian and defense scenarios. Applications include visual surveillance, motion picture restoration, and remote control of unmanned aerial vehicles. We have developed a prototype of RAVE: real-time autonomous video enhancement system. It consists of a suite of video artifact detection algorithms and corresponding correction algorithms. The system is autonomously controlled by an intelligent software agent. Our prototype has been successfully validated on several video sequences from different application domains and is being matured into a fully-functional, real-time embedded system.

Deligiannidis, L., A. P. Sheth, et al. (2006). Semantic analytics visualization Intelligence and Security Informatics, Proceedings. 3975: 48-59.

Dockstader, S. L. and M. Tekalp (2001). "On the tracking of articulated and occluded video object motion." Real-Time Imaging 7(5): 415-432.

Kage, H., M. Seki, et al. (2007). Pattern recognition for video surveillance and physical security.

Kim, J. O., J. S. Kim, et al. (2005). "On a video surveillance system with a DSP by the LDA algorithm." LECTURE NOTES IN COMPUTER SCIENCE 3597(Journal Article): 200-207.

Lin, L., M. L. Shyu, et al. (2009). Mining High-Level Features from Video using Associations and Correlations.

Manap, N. A., G. Di Caterina, et al. (2010). Smart surveillance system based on stereo matching algorithms with IP and PTZ cameras. 3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON), 2010.

In this paper, we describe a system for smart surveillance using stereo images with applications to advanced video surveillance systems. The system utilizes two smart IP cameras to obtain the position and location of objects. In this case, the object target is human face. The position and location of the object are automatically extracted from two IP cameras and subsequently transmitted to an ACTi Pan-Tilt-Zoom (PTZ) camera, which then points and zooms to the exact position in space. This work involves video analytics for estimating the location of the object in a 3D environment and transmitting its positional coordinates to the PTZ camera. The research consists of algorithms development in surveillance system including face detection, block matching, location estimation and implementation with ACTi SDK tool. The final system allows the PTZ camera to track the objects and acquires images in high-resolution quality.

Object Tracking

Another significant area of research particular to video analytics is object tracking between frames. Even when a system detects an 'object' in a frame, in real-time video signals, systems face the challenge of then tracking a particular object or multiple objects between video frames.

Amer, A. (2005). "Voting-based simultaneous tracking of multiple video objects." Circuits and Systems for Video Technology, IEEE Transactions on 15(11): 1448-1462.

This paper proposes an automatic object tracking method based on both object segmentation and motion estimation for real-time content-oriented video applications. The method focuses on the issues of speed of execution and reliability in the presence of noise, coding artifacts, shadows, occlusion, and object split. Objects are tracked based on the similarity of their features in successive frames. This is done in three steps: feature extraction, object matching, and feature monitoring. In the first step, objects are segmented and their spatial and temporal features are computed. In the second step, using a nonlinear two-stage voting strategy, each object of the previous frame is matched with an object of the current frame creating a unique correspondence. In the third step, object changes, such objects occlusion or split, are monitored and object features are corrected. These new features are then used to update results of previous steps creating module interaction. The contributions in this paper are the real-time two-stage voting strategy, the monitoring of object changes to handle occlusion and object split, and the spatiotemporal adaptation of the tracking parameters. Experiments on indoor and outdoor video shots containing over 6000 frames, including deformable objects, multi-object occlusion, noise, and coding and object segmentation artifacts have demonstrated the reliability and real-time response of the proposed method.

Chun-Ming, L., L. Yu-Shan, et al. (2005). Moving object segmentation and tracking in video. Machine Learning and Cybernetics, 2005.Proceedings of

2005 International Conference on.

Cossalter, M., M. Tagliasacchi, et al. (2009). Privacy-Enabled Object Tracking in Video Sequences Using Compressive Sensing. Advanced Video and Signal Based Surveillance, 2009. AVSS '09.Sixth IEEE International Conference on.

In this paper we propose a new coding scheme suitable for video surveillance applications that allows tracking of video objects without the need to reconstruct the sequence,thus enabling privacy protection. By taking advantage of recent findings in the compressive sensing literature, we encode a video sequence with a limited number of pseudo-random projections of each frame. At the decoder, we exploit the sparsity that characterizes background subtracted images in order to recover the location of the foreground object. We also leverage the prior knowledge about the estimated location of the object, which is predicted by means of a particle filter, to improve the recovery of the foreground object location. The proposed framework enables privacy, in the sense it is impossible to reconstruct the original video content from the encoded random projections alone, as well as secrecy, since decoding is prevented if the seed used to generate the random projections is not available.

Guo, L. and Y. Zhang (2006).Video Object Tracking Method Based on Snake Model Using Object's Histogram Information.Communications, Circuits and Systems Proceedings, 2006 International Conference on.

Jian, W., Y. Heng-jun, et al. (2010). Video Object Tracking Method Based on Normalized Cross-correlation Matching. Distributed Computing and Applications to Business Engineering and Science (DCABES), 2010 Ninth International Symposium on.

Combing with specific temporal information of video, this paper proposes a kind of video object tracking method based on normalized cross-correlation matching by using the high precision characteristics of normalized cross-correlation image matching. Firstly, extract video background from the temporal information of video. Then, acquire the region of moving object using background subtraction. Lastly, carry out related matching and updating towards the extracted moving object by means of normalized cross-correlation. Experimental result shows that the adaptability of our method is strong, which can well solve the tracking problems when tracking objects have scale transform. It also has good anti-interference ability and robustness, and can track moving objects accurately under the condition of noise interference, lens dithering and background mutation.

Khan, Z. H., I. Y. H. Gu, et al. (2009). Joint anisotropic mean shift and consensus point feature correspondences for object tracking in video. Multimedia and Expo, 2009.ICME 2009.IEEE International Conference on.

We propose a novel tracking scheme that jointly employs point feature correspondences and object appearance similarity. For selecting point correspondences, we use a subset of scale-invariant point features from

SIFT that agree with a pre-defined affine transformation. The selected consensus points are then used for pre-selecting candidate regions. For appearance similarity based tracking, we employ an existing anisotropic mean shift, from which the formula for estimating bounding box parameters (width, height, orientation and center) are derived. A switching criterion is utilized to handle the situation where only a small number of point correspondences is found. Experiments and evaluation are performed on tracking moving objects on videos where objects may contain partial occlusions, intersection, deformation and pose changes among other transforms. Our comparisons with two existing methods have shown that the proposed scheme has yielded marked improvement, especially in terms of reducing tracking drifts, of robustness to occlusions, and of tightness and accuracy of tracked bounding box.

Kim, T., S. Lee, et al. (2011). "Combined shape and feature-based video analysis and its application to non-rigid object tracking." Image Processing, IET 5(1): 87-100.

Many video object tracking systems use block matching algorithm (BMA) because of its simple computational structure and robust performance. The BMA, however, exhibits fundamental limitations resulting from non-rigid shapes and similar patterns to the background. The authors propose a combined shape and feature-based non-rigid object tracking algorithm, which is tightly coupled with an adaptive background generation to overcome the limit of block matching. The proposed algorithm is robust to the object's sudden movement or the change of features. This becomes possible by tracking both feature points and their neighbouring regions. Combination of background and shape boundary information significantly improves the tracking performance because the target object and the corresponding feature points on the boundary can be easily found. The shape control points (SCPs) are regularly distributed on the contour of the object, and the authors compare and update the centroid during the tracking process, where straying SCPs are removed, and the tracking continues with only qualified SCPs. As a result, the proposed method becomes free from potential failing factors such as spatio-temporal similarity between object and background, object deformation and occlusion, to name a few. Experiments have been performed using several in-house video sequences including various objects such as a moving robot, swimming fish and walking people. In order to demonstrate the performance of the proposed tracking algorithm, a number of experiments have been performed under noisy and low-contrast environment. For more objective comparison, performance evaluation of tracking surveillance 2002 data sets were also used.

Ritch, M. and N. Canagarajah (2007).Motion-Based Video Object Tracking in the Compressed Domain.Image Processing, 2007.ICIP 2007.IEEE International Conference on.

In this paper an algorithm for real-time unsupervised segmentation and tracking of a moving object is proposed. This is performed within the

compressed domain using motion information only. Initial object segmentation is done using iterative rejection, taking advantage of its computational efficiency. The system seeks to overcome its disadvantages, namely a delay in object macroblocks appearing after consistency checking and non-identification of macroblocks containing object boundaries, by taking a model based approach to object tracking. The output of iterative rejection is used to update the model after tracking has taken place in each frame. Experimental results on a number of MPEG-2 encoded sequences demonstrate its effectiveness in identifying and tracking an object of interest from a compressed video stream and that the system is better than purely using iterative rejection as a segmentation method.

Stamm, M. and K. J. R. Liu (2008). Live video object tracking and segmentation using graph cuts. Image Processing, 2008.ICIP 2008.15th IEEE International Conference on.
Graph cuts have proven to be powerful tools in image segmentation. Previous graph cut research has proposed methods for cutting across large graphs constructed from multiple layered video frames, resulting in an object being tracked across multiple frames. However, this research focuses on cutting graphs constructed from a prerecorded video sequence. In live video scenarios, frames cannot be layered to construct 3D volumes, since the contents of the subsequent frames are unknown. Instead, new graphs must be created and cut for each frame on demand. Resource limitations make this unfeasible on high-resolution videos. In addition, object tracking requires a method for incorporating the previous frame's object position and shape into the current graph. We propose a method for tracking and segmenting objects in live video that utilizes regional graph cuts and object pixel probability maps. The regionalization of the cuts around the tracked object will increase the speed of the tracker, and the object pixel probability maps will enable more flexible tracking.

Szczodrak, M., P. Dalka, et al. (2010).Performance evaluation of video object tracking algorithm in autonomous surveillance system.Information Technology (ICIT), 2010 2nd International Conference on.
Results of a performance evaluation of a video object tracking algorithm are presented. The method of moving object detection and tracking is based on background modelling with mixtures of Gaussian and Kalman filters. An emphasis is put on algorithm's efficiency with regards to its settings. Utilized methods of a performance evaluation based on a comparison of the algorithm output to manually prepared reference data are introduced. The experiments aimed at examining the performance achieved with various object detection algorithm parameter settings are presented and discussed.

Thirde, D. and G. Jones (2004). Hierarchical probabilistic models for video object segmentation and tracking. Pattern Recognition, 2004.ICPR 2004.Proceedings of the 17th International Conference on.

Wei, Y. and W. Badawy (2003). A novel zoom invariant video object tracking

algorithm (ZIVOTA). Electrical and Computer Engineering, 2003. IEEE CCECE 2003. Canadian Conference on.

Yi, L. and Y. F. Zheng (2005). "Video object segmentation and tracking using ψ-learning classification." Circuits and Systems for Video Technology, IEEE Transactions on 15(7): 885-899.

As a requisite of the emerging content-based multimedia technologies, video object (VO) extraction is of great importance. This paper presents a novel semiautomatic segmentation and tracking method for single VO extraction. Unlike traditional approaches, the proposed method formulates the separation of the VO from the background as a classification problem. Each frame is divided into small blocks of uniform size, which are called object blocks if the centering pixels belong to the object, or background blocks otherwise. After a manual segmentation of the first frame, the blocks of this frame are used as the training samples for the object-background classifier. A newly developed learning tool called ψ-learning is employed to train the classifier which outperforms the conventional Support Vector Machines in linearly nonseparable cases. To deal with large and complex objects, a multilayer approach constructing a so-called hyperplane tree is proposed. Each node of the tree represents a hyperplane, responsible for classifying only a subset of the training samples. Multiple hyperplanes are thus needed to classify the entire set. Through the combination of the multilayer scheme and ψ-learning, one can avoid the complexity of nonlinear mapping as well as achieve high classification accuracy. During the tracking phase, the pixel in the center of every block in a successive frame is classified by a sequence of hyperplanes from the root to a leaf node of the hyperplane tree, and the class of the block is identified accordingly. All the object blocks thus form the object of interest, whose boundary unfortunately is stair-like due to the block effect. In order to obtain the pixel-wise boundary in a cost efficient way, a pyramid boundary refining algorithm is designed, which iteratively selects a few informative pixels for class label checking, and reduces uncertainty about the actual boundary of the object. The proposed method has been applied on video sequences with various spatial and temporal characteristics, and experimental results demonstrate it to be effective, efficient, and robust.

Ying-Tung, H., C. Cheng-Long, et al. (2005). Robust Multiple Targets Tracking Using Object Segmentation and Trajectory Estimation in Video. Systems, Man and Cybernetics, 2005 IEEE International Conference on.

In this paper, a novel robust unsupervised video object tracking algorithm is proposed. The proposed algorithm combines several techniques: mathematical morphology, region growing, region merging, and trajectory estimation, for tracking several predetermined video objects, simultaneously. A modified mathematical morphological edge detector was employed to sketch the contour of the video frame; and an edge-based object segmentation algorithm was applied to the contour for partitioning the predetermined objects; moreover, according to the motion of the

objects, the proposed algorithm can estimate and partition the objects in following video frames, automatically. The proposed algorithm is also robustness against mobile cameras. The experimental results show that the proposed algorithm can precisely partition and track multiple video objects

Zhi, L., S. Liquan, et al. (2007).A Novel Video Object Tracking Approach Based on Kernel Density Estimation and Markov Random Field.Image Processing, 2007.ICIP 2007.IEEE International Conference on.
In this paper, we propose a novel video object tracking approach based on kernel density estimation and Markov random field (MRF). The interested video objects are first segmented by the user, and a nonparametric model based on kernel density estimation is initialized for each video object and the remaining background, respectively. A temporal saliency map is also initialized for each object to memorize the temporal trajectory. Based on the probabilities evaluated on the non-parametric models, each pixel in the current frame is first classified into the corresponding video object or background using the maximum likelihood criterion. Starting from the initial classification result, a MRF model that combines spatial smoothness and temporal coherency is selectively exploited to generate more reliable video objects. The nonparametric model and the temporal saliency map for each video object are updated and propagated for the future tracking. Experimental results on several MPEG-4 test sequences demonstrate the good segmentation performance of our approach.

Object recognition

Tracking an object between frames is compounded by the challenge of trying to 'recognize' dimensionally shifting shapes. Object recognition within images is a key part of real-time event detection and response. Research in this area is moving to object recognition on mobile devices to help with augmented reality applications.

Amer, A., E. Dubois, et al. (2002). Context-independent real-time event recognition: application to key-image extraction. Pattern Recognition, 2002.Proceedings.16th International Conference on.

Fuerstenberg, K. and V. Willhoeft (2001). Object tracking and classification using laserscanners-pedestrian recognition in urban environment. Intelligent Transportation Systems, 2001.Proceedings.2001 IEEE.
Current car safety systems are passive systems. Modern car assistance systems are based only on vehicle data. Future safety systems will also include object recognition in the near frontal area of the vehicle to detect dangerous situations. Therefore, special sensors and algorithms are needed. The paper discusses a system using a laserscanner and a video camera

Gal, L., M. Rudzsky, et al. (2010). "Video Event Modeling and Recognition in Generalized Stochastic Petri Nets." Circuits and Systems for Video Technology, IEEE Transactions on 20(1): 102-118.
In this paper, we propose the surveillance event recognition framework using Petri Nets (SERF-PN) for recognition of event occurrences in video.

The Petri Net (PN) formalism allows a robust way to express semantic knowledge about the event domain as well as efficient algorithms for recognizing events as they occur in a particular video sequence. The major novelties of this paper are extensions to both the modeling and the recognition capacities of the Object PN paradigm. The first contribution of this paper is the extension of the PN representational capacities by introducing stochastic timed transitions to allow modeling of events which have some variance in duration. These stochastic timed transitions sample the duration of the condition from a parametrized distribution. The parameters of this distribution can be specified manually or learned from available video data. A second representational novelty is the use of a single PN to represent the entire event domain, as opposed to previous approaches which have utilized several networks, one for each event of interest. A third contribution of this paper is the capacity to probabilistically predict future events by constructing a discrete time Markov chain model of transitions between states. The experiments section of the paper thoroughly evaluates the application of the SERF-PN framework in the event domains of surveillance and traffic monitoring and provides comparison to other approaches using the CAVIAR dataset , a standard dataset for video analysis applications.

Hoogs, A., J. Rittscher, et al. (2003). Video content annotation using visual analysis and a large semantic knowledgebase.Computer Vision and Pattern Recognition, 2003.Proceedings.2003 IEEE Computer Society Conference on.

We present a novel approach to automatically annotating broadcast video. To manage the enormous variety of objects, events and scenes in video problem domains such as news video, we couple generic image analysis with a semantic database, WordNet, containing huge amounts of real-world information. Object and event recognition are performed by searching WordNet for concepts jointly supported by image evidence and topic context derived from the video transcript. No object- specific or event-specific training is required, and only a few object models and detection algorithms are required to label much of the significant content of news video. The hierarchical structure of WordNet yields hierarchical recognition, dynamically tailored to the level of supporting image evidence. The potential of the approach is demonstrated by analyzing a wide variety of scenes in news video.

Nguyen Dang, B. (2009).Autonomous Learning for Tracking and Recognition.Computing and Communication Technologies, 2009.RIVF '09.International Conference on.

We present an efficient approach for autonomous learning an object model from video or image sequences. The idea is to employ online boosting technique to adaptively learn an object representation from only as few as one labeled training sample. Our main contributions are: (1) A robust updating strategy of a discriminative classifier, which allows effective learning of an object model for tracking and recognition; (2) Learning and

tracking are performed in a single procedure with possibility of reducing drifting and ability to recover tracking failure; and (3) a simple yet reliable framework for object recognition. Our main concern is to use the approach for the problem of hand and face tracking and gesture recognition. However, the proposed framework can be applied to other objects. Experiments on different data sets (publicity available) show the efficiency of our approach over very recent published approaches on different objects.

Shuji, Z., F. Precioso, et al. (2010). STTK-based video object recognition. Image Processing (ICIP), 2010 17th IEEE International Conference on.
In this paper, we extend our video object recognition system to multiclass object recognition context, dealing with unbalanced data sets and comparing our resuls to state-of-the-art methods. Our approach is based on a Spatio-Temporal data representation, a dedicated kernel design and statistical learning techniques for object recognition. From video tracks made of segmented object regions in the successive frames, we extract sets of spatio-temporally coherent SIFT-based features, called Spatio-Temporal Tubes. To compare these complex tube objects, we integrate a Spatio-Temporal Tube Kernel (STTK) function into a multi-class classification framework with balancing process for unequal classes. Our approach is successfully evaluated on episodes from “Buffy, the Vampire Slayer” TV series which have been used in other works targeting same objectives. Our method proved to be more robust than dictionary based, facial feature based and key-frame based approaches. Our method is also tested on a small car database and preliminary results for car identification task illustrate its generalization potential.

Taehee, L. and S. Soatto (2010). Feature tracking and object recognition on a hand-held. Mixed and Augmented Reality (ISMAR), 2010 9th IEEE International Symposium on.
We demonstrate a visual recognition system operating on a hand-held device, with the help of an efficient and robust feature tracking and an object recognition mechanism that can be used for interactive mobile applications. In our recognition system, corner features are detected from captured video frames in a multi-scale image pyramid, and are tracked between consecutive frames efficiently. In order to perform object recognition, local descriptors are calculated on the tracked features, and quantized using a vocabulary tree. For each object, a bag-of-words model is learned from multiple views. The learned objects are recognized by computing the ranking score for the set of features in a single video frame. Our feature tracking algorithm and local descriptors are different than the Lucas-Kanade algorithm in image pyramid or the SIFT descriptor, however improving the efficiency and accuracy. For our implementation on a mobile phone, we used an iPhone 3GS with a 600MHz ARM chip CPU. The video frame is captured from a camera preview screen at a rate of 15 frames per second using the public API. The task of object recognition on a mobile phone runs at around 7 frames per second, including the feature tracking

and descriptor calculation.
Tan, F., Q. Guan, et al. (2009).A method for robust recognition and tracking of multiple objects.Communications, Circuits and Systems, 2009.ICCCAS 2009.International Conference on.
This paper presents an accurate and flexible method for robust recognition and tracking of multiple objects in video sequence. We calculate color moments and wavelet moments for each detected object. Based on the extracted moment features, the SVM achieves optimal object recognition performance. The object recognition rate is above 98.53%. Since the tracking accuracy of feature matching method could be degraded by occlusion, we add a Kalman filter tracking framework based on object recognition to improve multiple objects tracking. The previous object recognition module improves the performance and the accuracy of the Kalman filter tracking framework. Results obtained suggest that our tracking algorithm is very effective and robust even in challenging tracking conditions like occlusion and background clutter.
Tie, L., Y. Zejian, et al. (2011)."Learning to Detect a Salient Object." Pattern Analysis and Machine Intelligence, IEEE Transactions on 33(2): 353-367.
In this paper, we study the salient object detection problem for images. We formulate this problem as a binary labeling task where we separate the salient object from the background. We propose a set of novel features, including multiscale contrast, center-surround histogram, and color spatial distribution, to describe a salient object locally, regionally, and globally. A conditional random field is learned to effectively combine these features for salient object detection. Further, we extend the proposed approach to detect a salient object from sequential images by introducing the dynamic salient features. We collected a large image database containing tens of thousands of carefully labeled images by multiple users and a video segment database, and conducted a set of experiments over them to demonstrate the effectiveness of the proposed approach.
Walls, B. (2010)."Cascaded Automatic Target Recognition (Cascaded ATR)." Proceedings of SPIE-The International Society for Optical Engineering 7696(Journal Article): 76960W-76960W.
Wechsler, H. (2007). Robust Recognition-by-Parts Using Transduction and Boosting with Applications to Biometrics. Systems, Signals and Image Processing, 2007 and 6th EURASIP Conference focused on Speech and Image Processing, Multimedia Communications and Services. 14th International Workshop on.

Event recognition

Another area of research included in Video Analytics is event recognition. Even recognition allows for the automatic (usually contextual) interpretation of recognizable objects over time. This area of research spans a range of application domains from fire-alarms to theft prevention.

"Smart CCTV raises bush-fire alarm." New Scientist 205(2749): 19-19.
Abrams, D., S. McDowall, et al. (2007).Video content analysis with effective

response.
Loney, G. (2007). Border Intrusion Detection: Thinking outside the perimeter. Security Technology, 2007 41st Annual IEEE International Carnahan Conference on.
Maciejewski, R., S. Kim, et al. (2008). Situational awareness and visual analytics for emergency response and training.
Pratl, G., L. Frangu, et al. (2007). Smart nodes for semantic analysis of visual and aural data. 2007 5th Ieee International Conference on Industrial Informatics, Vols 1-3: 1027-1032.
Venkoparao, V. G., R. N. Hota, et al. (2009). Flare Monitoring for Petroleum Refineries.
Wang, Y. and G. Mori (2010). "Hidden Part Models for Human Action Recognition: Probabilistic vs. Max-Margin." Pattern Analysis and Machine Intelligence, IEEE Transactions on PP(99): 1-1.
We present a discriminative part-based approach for human action recognition from video sequences using motion features. Our model is based on the recently proposed hidden conditional random field~(HCRF) for object recognition. Similar to HCRF for object recognition, we model a human action by a flexible constellation of parts conditioned on image observations. Different from object recognition, our model combines both large-scale global features and local patch features to distinguish various actions. Our experimental results show that our model is comparable to other state-of-the-art approaches in action recognition. In particular, our experimental results demonstrate that combining large-scale global features and local patch features performs significantly better than directly applying HCRF on local patches alone. We also propose an alternative for learning the parameters of an HCRF model in a max-margin framework. We call this method the max-margin hidden conditional random field~(MMHCRF). We demonstrate that MMHCRF outperforms HCRF in human action recognition. In addition, MMHCRF can handle a much broader range of complex hidden structures arising in various problems in computer vision.
Asif, M. and J. Soraghan (2008)."Video Analytics for Panning Camera in Dynamic Surveillance Environment." MONOGRAPH OF THE COTSEN INSTITUTE OF ARCHAELOGY, UCLA(Journal Article): 79-82.
Janoos, F., S. Singh, et al. (2007). Activity Analysis Using Spatio-Temporal Trajectory Volumes in Surveillance Applications.Visual Analytics Science and Technology, 2007.VAST 2007.IEEE Symposium on.
In this paper, we present a system to analyze activities and detect anomalies in a surveillance application, which exploits the intuition and experience of security and surveillance experts through an easy- to-use visual feedback loop. The multi-scale and location specific nature of behavior patterns in space and time is captured using a wavelet-based feature descriptor. The system learns the fundamental descriptions of the behavior patterns in a semi-supervised fashion by the higher order singular value decomposition of the space described by the training data.

This training process is guided and refined by the users in an intuitive fashion. Anomalies are detected by projecting the test data into this multi-linear space and are visualized by the system to direct the attention of the user to potential problem spots. We tested our system on real-world surveillance data, and it satisfied the security concerns of the environment.

Marraud, D., B. Cepas, et al. (2009). Semantic Browsing of Video Surveillance Databases through Online Generic Indexing.

Venkoparao, V. G., R. N. Hota, et al. (2009). Flare Monitoring for Petroleum Refineries.

Human tracking

Object tracking and recognition are only a few of the modules that contribute to Human tracking, a basic requirement for analysing human behaviours in retail settings for example. When the object tracking and classification modules code an object or objects in the foreground as human or parts of a human, then other modules take over to perform such tasks as tracking the individual across frames, between cameras and record and analyze information like gender, age and even ethnicity in addition to other biometrics such as gait and physiological or behavioural based identification.

Bocchetti, G., F. Flammini, et al. (2009). Dependable integrated surveillance systems for the physical security of metro railways.

Dawson, D., P. Derby, et al. (2009). A Report on Camera Surveillance in Canada: Part Two. Kingston, Queen's University

De Angelis, D., R. Sala, et al. (2009). "A new computer-assisted technique to aid personal identification." International journal of legal medicine 123(4): 351-356.

Deisman, W., P. Derby, et al. (2009). A Report on Camera Surveillance in Canada: Part One. Kingston, Queen's University.

Dixon, P. (2010). The One-Way-Mirror Society: Privacy Implications of the new Digital Signage Networks, World Privacy Forum.

Elder, J. H., S. J. D. Prince, et al. (2007)."Pre-attentive and attentive detection of humans in wide-field scenes." International Journal of Computer Vision 72(1): 47-66.

Everingham, M. and A. Zisserman (2005).Identifying individuals in video by combining 'generative' and discriminative head models.Computer Vision, 2005.ICCV 2005.Tenth IEEE International Conference on.

The objective of this work is automatic detection and identification of individuals in unconstrained consumer video, given a minimal number of labelled faces as training data. Whilst much work has been done on (mainly frontal) face detection and recognition, current methods are not sufficiently robust to deal with the wide variations in pose and appearance found in such video. These include variations in scale, illumination, expression, partial occlusion, motion blur, etc. We describe two areas of innovation: the first is to capture the 3-D appearance of the entire head, rather than just the face region, so that visual features such as the hairline

can be exploited. The second is to combine discriminative and 'generative' approaches for detection and recognition. Images rendered using the head model are used to train a discriminative tree-structured classifier giving efficient detection and pose estimates over a very wide pose range with three degrees of freedom. Subsequent verification of the identity is obtained using the head model in a 'generative' framework. We demonstrate excellent performance in detecting and identifying three characters and their poses in a TV situation comedy

Gagnon, L., F. Laliberte, et al. (2006). "A system for tracking and recognizing pedestrian faces using a network of loosely coupled cameras - art. no. 62460N." PROCEEDINGS OF THE SOCIETY OF PHOTO-OPTICAL INSTRUMENTATION ENGINEERS (SPIE) 6246(Journal Article): N2460-N2460.

Greenberg, J. and S. P. Hier (2009)."CCTV Surveillance and the Poverty of Media Discourse." Canadian Journal of Communication 34(3): 461-486.

Saptharishi, M. and D. Marman (2009). An Information Value Driven Architecture for Urban Video Surveillance in Data and Attention Bandwidth Constrained Environments.

Candamo, J., M. Shreve, et al. (2010). "Understanding Transit Scenes: A Survey on Human Behavior-Recognition Algorithms." Ieee Transactions on Intelligent Transportation Systems 11(1): 206-224.

Chen, Y., Z. Yiwen, et al. (2008). Visual mining of multimedia data for social and behavioral studies.Visual Analytics Science and Technology, 2008.VAST '08.IEEE Symposium on.

With advances in computing techniques, a large amount of high-resolution high-quality multimedia data (video and audio, etc.) has been collected in research laboratories in various scientific disciplines, particularly in social and behavioral studies. How to automatically and effectively discover new knowledge from rich multimedia data poses a compelling challenge since state-of-the-art data mining techniques can most often only search and extract pre-defined patterns or knowledge from complex heterogeneous data. In light of this, our approach is to take advantages of both the power of human perception system and the power of computational algorithms. More specifically, we propose an approach that allows scientists to use data mining as a first pass, and then forms a closed loop of visual analysis of current results followed by more data mining work inspired by visualization, the results of which can be in turn visualized and lead to the next round of visual exploration and analysis. In this way, new insights and hypotheses gleaned from the raw data and the current level of analysis can contribute to further analysis. As a first step toward this goal, we implement a visualization system with three critical components: (1) A smooth interface between visualization and data mining. The new analysis results can be automatically loaded into our visualization tool. (2) A flexible tool to explore and query temporal data derived from raw multimedia data. We represent temporal data into two forms - continuous variables and event variables. We have developed various ways to visualize both

temporal correlations and statistics of multiple variables with the same type, and conditional and high-order statistics between continuous and event variables. (3) A seamless interface between raw multimedia data and derived data. Our visualization tool allows users to explore, compare, and analyze multi-stream derived variables and simultaneously switch to access raw multimedia data. We de- - monstrate various functions in our visualization program using a set of multimedia data including video, audio and motion tracking data.

Hampapur, A., R. Bobbitt, et al. (2009). Video Analytics in Urban Environments.

Jones, C., M. Ogawa, et al. (2009). VIDI surveillance - embassy monitoring and oversight system.Visual Analytics Science and Technology, 2009.VAST 2009.IEEE Symposium on.

Human Identification

Though not necessary for all application domains involving human tracking and behaviour analysis, video analytics also includes strategies for human identification using video data. A particular focus of human identification has been HumanID at a distance. The face features prominently among other identification or biometric strategies like gait recognition, with a large number of contributing research strands. Face Recognition a significant focus of computer vision over the last decade is itself a composite of modular research areas the stretch back into the 1970's. Face Recognition Technologies (FRT) include face detection, tracking and recognition solutions.

Face detection

Solutions for detecting faces in visual date may be trased back to the 1970's (Gates, 2004) and existing algorithms found on commercial cameras and phones are robust and feature rich. However research continues in this area to improve the scope and accuracy of identifying faces from various angles, under differing conditions and of multiple subjects.

Alajel, K. M., W. Xiang, et al. (2010). "Face Detection Technique Based on Skin Color and Facial Features." Mathematics and Computers in Science and Engineering(Journal Article): 192-199.

Bao, P. T., J. Y. Kim, et al. (2005). "Fast multi-face detection in color images using fuzzy logic." International Symposium on Intelligent Signal Processing and Communication Systems-ISPACS(Journal Article): 777-780.

Bojkovic, Z. and A. Samcovic (2006). Face detection approach in neural network based method for video surveillance.

Chen, T.-W., W.-K.Chan, et al. (2007).Efficient face detection with segmentation and feature-based face scoring in surveillance systems.

Dockstader, S. L. and A. M. Tekalp (2000). "Real-time object tracking and human face detection in cluttered scenes." PROCEEDINGS OF THE SOCIETY OF PHOTO-OPTICAL INSTRUMENTATION ENGINEERS (SPIE) 3974(Journal Article): 957-968.

Elder, J. H., S. J. D. Prince, et al. (2007)."Pre-attentive and attentive detection of

humans in wide-field scenes." International Journal of Computer Vision 72(1): 47-66.
Feris, R. S., T. E. de Campos, et al. (2000). "Detection and tracking of facial features in video sequences." LECTURE NOTES IN ARTIFICIAL INTELLIGENCE 1793(Journal Article): 127-135.
Hjelmas, E. and B. K. Low (2001). "Face detection: A survey." Computer Vision and Image Understanding 83(3): 236-274.
Hota, R. N., V. Venkoparao, et al. (2006). Face detection by using skin color model based on one class classifier.
Hsu, R. L., M. Abdel-Mottaleb, et al. (2002). "Face detection in color images." IEEE Transactions on Pattern Analysis and Machine Intelligence 24(5): 696-706.
Karungaru, S., M. Fukumi, et al. (2009)."DETECTION AND RECOGNITION OF VEHICLE LICENSE PLATES USING TEMPLATE MATCHING, GENETIC ALGORITHMS AND NEURAL NETWORKS." International Journal of Innovative Computing Information and Control 5(7): 1975-1985.
Kim, J. B., Y. H. Sung, et al. (2004). A fast and robust face detection based on module switching network.
Kim, J. O. and J. S. Kim (2005)."Real-time implementation of face detection for a ubiquitous computing." LECTURE NOTES IN COMPUTER SCIENCE 3480(Journal Article): 1187-1195.
Kim, J. O., S. J. Seo, et al. (2004). "Face detection by facial features with color images and face recognition using PCA." LECTURE NOTES IN COMPUTER SCIENCE 3043(Journal Article): 1-8.
Kim, T. K., S. U. Lee, et al. (2002). "Integrated approach of multiple face detection for video surveillance." INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION(Journal Article): 394-397.
Lin, D.-T.and M.-J. Liu (2006). "Face occlusion detection for automated teller machine surveillance." LECTURE NOTES IN COMPUTER SCIENCE 4319(Journal Article): 641-651.
Liu, Z. F., Z. S. You, et al. (2003). Face detection and facial feature extraction in color image.
Loney, G. (2007). Border Intrusion Detection: Thinking outside the perimeter. Security Technology, 2007 41st Annual IEEE International Carnahan Conference on.
Detecting infiltration across national borders is not simply a matter of deploying commercial-off-the-shelf perimeter intrusion sensors. The sheer length of border systems has led some to propose using wide area surveillance systems to reduce cost. Unfortunately the most common of these technologies: thermal infrared and visible wavelength sensors integrated with video analytics and ground radar have line-of-sight limitations and less than optimum nuisance alarm characteristics for real world border applications. The missing link is a cost-effective terrain following trip-wire sensor to cue these wide-area systems and mitigate their performance limitations. Buried ported coax, sometimes called leaky coax or guided radar sensors have protected high value perimeters for

over two decades. In theory, their high probability of detection, resistance to defeat and vandalism, invisible terrain following volumetric field and good nuisance alarm characteristics make them well suited to secure borders; but their high cost per zone and inability to accommodate different soil conditions have argued against their use. OmniTrax¿ is a new ultra wideband spread spectrum ranging guided radar which changes this equation with a lower cost per zone, one-meter target resolution and the ability to accommodate different soil types. Ongoing research may eventually result in the ability to track targets along the cables, determine the direction of travel of targets crossing the cables, surface mount the sensor cables in rocky terrain and cost effectively install the sensor cables in soil or sand using cable plows; any of which would only improve the technology's utility in border intrusion detection.

Lu, Y. Z., J. L. Zhou, et al. (2003). "A survey of face detection, extraction and recognition." Computing and Informatics 22(2): 163-195.

Manap, N. A., G. Di Caterina, et al. (2010). Face detection and stereo matching algorithms for smart surveillance system with IP cameras. Visual Information Processing (EUVIP), 2010 2nd European Workshop on.
In this paper, we describe a smart surveillance system to detect human faces in stereo images with applications to advanced video surveillance systems. The system utilizes two smart IP cameras to obtain the position and location of the object that is a human face. The position and location of the object are extracted from two IP cameras and subsequently transmitted to a Pan-Tilt-Zoom (PTZ) camera, which can point to the exact position in space. This work involves video analytics for estimating the location of the object in a 3D environment and transmitting its positional coordinates to the PTZ camera. The research consists of algorithm development in surveillance system including face detection, stereo matching, location estimation and implementation with ACTi PTZ camera. The final system allows the PTZ camera to track the objects and acquires images in high-resolution.

Miao, L. G., F. W. Wang, et al. (2009). Automatic License Plate Detection Based on Edge density and Color Model.

Nallaperumal, K., R. Subban, et al. (2006). Human face detection in color images using skin color and template matching models for multimedia on the web.

Oancea, R., S. Kifor, et al. (2009). Considerations on Skin Colour Algorithms used for Candidate Faces Detection.

Pai, Y.-T., S.-J.Ruan, et al. (2006).A simple and accurate color face detection algorithm in complex background.

Park, S. H., E. Y. Kim, et al. (2001). Face detection for security system on the Internet.

Phimoltares, S., C. Lursinsap, et al. (2007). "Face detection and facial feature localization without considering the appearance of image context." Image and Vision Computing 25(5): 741-753.

Ravi Kumar, C. N. and A. Bindu (2006)."An efficient skin illumination compensation model for efficient face detection." IEEE Industrial

Electronics Society(Journal Article): 3298-3303.
Tie, Y. and L. Guan (2009). "Automatic face detection in video sequences using local normalization and optimal adaptive correlation techniques." Pattern Recognition 42(9): 1859-1868.
Zhang, Q., S.-i.Kamata, et al. (2009). Face Detection and Tracking in Color images Using Color Centroids Segmentation.
Zhang, Q. and Z.-J.Liu (2006). Face Detection based on complexional segmentation and feature extraction.
Zhang, Q., J. Zhang, et al. (2008). "Face detection method based on color barycenter hexagon model." Lecture Notes in Engineering and Computer Science(Journal Article): 655-658.
Zhao, L., X. Sun, et al. (2006). "Face detection based on facial features." International Conference on Signal Processing-ICSP(Journal Article): 1758-1761.
Zuo, F. and P. H. N. de With (2003). "Fast human face detection using successive face detectors with incremental detection capability." PROCEEDINGS OF THE SOCIETY OF PHOTO-OPTICAL INSTRUMENTATION ENGINEERS (SPIE) 5022(Journal Article): 831-841.

Face recognition

Detecting a face in an image or video clip is one challenge, but the ability to correctly associate face data with specific individuals is more challenge. Face Recognition is perhaps one of the most active frontiers of biometric research since 9/11.

Cai, Y., D. Kaufer, et al. (2009). Semantic Visual Abstraction for Face Recognition. Computational Science - Iccs 2009, Part I. G. Allen, J. Nabrzyski, E. Seidelet al. 5544: 419-428.
Cheung, K.-W., J. Chen, et al. (2008). Pose-tolerant Non-frontal Face Recognition using EBGM.
Choi, J. Y., Y. M. Ro, et al. (2008). Feature Subspace Determination in Video-based Mismatched Face Recognition.
Choi, J. Y., Y. M. Ro, et al. (2009). "Color Face Recognition for Degraded Face Images."Ieee Transactions on Systems Man and Cybernetics Part B-Cybernetics 39(5): 1217-1230.
Harguess, J., H. Changbo, et al. (2009). Fusing face recognition from multiple cameras.Applications of Computer Vision (WACV), 2009 Workshop on. Face recognition from video has recently received much interest. However, several challenges for such a system exist, such as resolution, occlusion (from objects or self-occlusion), motion blur, and illumination. The aim of this paper is to overcome the problem of self-occlusion by observing a person from multiple cameras with uniquely different views of the person's face and fusing the recognition results in a meaningful way. Each camera may only capture a part of the face, such as the right or left half of the face. We propose a methodology to use cylinder head models (CHMs) to track the face of a subject in multiple cameras. The problem of face recognition from video is then transformed to a still face recognition

problem which has been well studied. The recognition results are fused based on the extracted pose of the face. For instance, the recognition result from a frontal face should be weighted higher than the recognition result from a face with a yaw of 30Â°. Eigenfaces is used for still face recognition along with the average-half-face to reduce the effect of transformation errors. Results of tracking are further aggregated to produce 100% accuracy using video taken from two cameras in our lab.

Huang, H. and H.He (2011)."Super-Resolution Method for Face Recognition Using Nonlinear Mappings on Coherent Features." IEEE Transactions on Neural Networks 22(1): 121-130.

Huang, P. and Y. Wang (2009).The Impact of Changing Resolutions on Face Recognition.

Huang, Z.-K., W.-Z.Zhang, et al. (2008). Using Gabor Filters Features for Multi-Pose Face Recognition in Color Images.

Hulbert, W., C. Podilchuk, et al. (2008)."Face Recognition using a Pictorial-Edit Distance." IEEE International Conference on Image Processing (ICIP)(Journal Article): 1908-1911.

Jillela, R. R. and A. Ross (2009)."Adaptive Frame Selection for Improved Face Recognition in Low-Resolution Videos." IEEE International Joint Conference on Neural Networks (IJCNN)(Journal Article): 2835-2841.

Klare, B. and M. Burge (2010)."Assessment of H.264 Video Compression on Automated Face Recognition Performance in Surveillance and Mobile Video Scenarios." Proceedings of SPIE-The International Society for Optical Engineering 7667(Journal Article): 76670X-76670X.

Privacy Sensitive VA research

Despite the range of research in the areas of video analytics, proportionally little research actually incorporates privacy into the design. Notable exceptions are found below.

Coudert, F. and J. Dumortier (2008). Intelligent video surveillance networks: data protection challenges.

Gross, R., E. Airoldi, et al. (2006). "Integrating utility into face de-identification." LECTURE NOTES IN COMPUTER SCIENCE 3856(Journal Article): 227-242.

Matusek, F. and R. Reda (2008). Efficient Secure Storage of Privacy Enhanced Video Surveillance Data in Intelligent Video Surveillance Systems.

Matusek, F., R. Reda, et al. (2008). Efficient Secure Storage of Privacy Enhanced Video Surveillance Data in Intelligent Video Surveillance Systems.

Newton, E. M., L. Sweeney, et al. (2005). "Preserving privacy by de-identifying face images." IEEE Transactions on Knowledge and Data Engineering 17(2): 232-243.

Senior, A. (2008). "Privacy Enablement in a Surveillance System." IEEE International Conference on Image Processing (ICIP)(Journal Article): 1680-1683.

Senior, A. and Ieee (2008). PRIVACY ENABLEMENT IN A SURVEILLANCE SYSTEM. 2008 15th Ieee International Conference on Image Processing,

Vols 1-5: 1680-1683.
Vagts, H. and A. Bauer (2010).Privacy-Aware Object Representation for Surveillance Systems. Advanced Video and Signal Based Surveillance (AVSS), 2010 Seventh IEEE International Conference on.
Real-time object tracking, feature assessment and classification based on video are an enabling technology for improving situation awareness of human operators as well as for automated recognition of critical situations. To bridge the gap between video signal-processing output and spatio-temporal analysis of object behavior at the semantic level, a generic and sensor-independent object representation is necessary. However, in the case of public and corporate video surveillance, centralized storage of aggregated data leads to privacy violations. This article explains how a centralized object representation, complying with the Fair Information Practice Principles (FIP) privacy constraints, can be implemented for a video surveillance system.

"Smart" Video Surveillance

Higher level semantic decision driven responses using software are still being developed for deployment. A sample of some literature can be found below.

"Smart CCTV raises bush-fire alarm." New Scientist 205(2749): 19-19.
Abrams, D., S. McDowall, et al. (2007).Video content analysis with effective response.
Deligiannidis, L., A. P. Sheth, et al. (2006). Semantic analytics visualization.Intelligence and Security Informatics, Proceedings. 3975: 48-59.
Maciejewski, R., S. Kim, et al. (2008). Situational awareness and visual analytics for emergency response and training.
Pratl, G., L. Frangu, et al. (2007). Smart nodes for semantic analysis of visual and aural data. 2007 5th Ieee International Conference on Industrial Informatics, Vols 1-3: 1027-1032.
Venkoparao, V. G., R. N. Hota, et al. (2009). Flare Monitoring for Petroleum Refineries.

Automated Traffic Monitoring

A particularly active area of research in the area of video analytics includes traffic and transportation monitoring and Automated License-Plate Recognition or (ALPR). This form of analysis can range from tracking general patterns to specific violations, but can also be purposed to track movements of specific vehicles and by extension individuals.

Chen, Z. X., C. Y. Liu, et al. (2009). "Automatic License-Plate Location and Recognition Based on Feature Salience." Ieee Transactions on Vehicular Technology 58(7): 3781-3785.
Chen, Z. X., C. Y. Liu, et al. (2007). "Automatic license plate location and recognition." Indian Journal of Engineering and Materials Sciences 14(5): 337-345.
Fan, X., G. L. Fan, et al. (2007).Joint segmentation and recognition of license

plate characters. 2007 Ieee International Conference on Image Processing, Vols 1-7: 2049-2052.
Gonzalez, J., F. X. Roca, et al. (2009). "Research Steps Towards Human Sequence Evaluation." COMPUTATIONAL METHODS IN APPLIED SCIENCES 13(Journal Article): 105-115.
Huang, F., Z. M. Li, et al. (2008).A novel algorithm of character segmentation in vehicle license plates.
Kang, D. J. (2009). "DYNAMIC PROGRAMMING-BASED METHOD FOR EXTRACTION OF LICENSE PLATE NUMBERS OF SPEEDING VEHICLES ON THE HIGHWAY." International Journal of Automotive Technology 10(2): 205-210.
Karungaru, S., M. Fukumi, et al. (2009)."DETECTION AND RECOGNITION OF VEHICLE LICENSE PLATES USING TEMPLATE MATCHING, GENETIC ALGORITHMS AND NEURAL NETWORKS." International Journal of Innovative Computing Information and Control 5(7): 1975-1985.
Khan, N. Y., A. S. Imran, et al. (2007). Distance and color invariant automatic license plate recognition system.
Kulkarni, P., A. Khatri, et al. (2009). Automatic Number Plate Recognition (ANPR) System for Indian conditions.
Leibe, B., K. Schindler, et al. (2008). "Coupled Object Detection and Tracking from Static Cameras and Moving Vehicles." Pattern Analysis and Machine Intelligence, IEEE Transactions on 30(10): 1683-1698.
We present a novel approach for multi-object tracking which considers object detection and spacetime trajectory estimation as a coupled optimization problem. Our approach is formulated in a minimum description length hypothesis selection framework, which allows our system to recover from mismatches and temporarily lost tracks. Building upon a state-of-the-art object detector, it performs multiview/multicategory object recognition to detect cars and pedestrians in the input images. The 2D object detections are checked for their consistency with (automatically estimated) scene geometry and are converted to 3D observations which are accumulated in a world coordinate frame. A subsequent trajectory estimation module analyzes the resulting 3D observations to find physically plausible spacetime trajectories. Tracking is achieved by performing model selection after every frame. At each time instant, our approach searches for the globally optimal set of spacetime trajectories which provides the best explanation for the current image and for all evidence collected so far while satisfying the constraints that no two objects may occupy the same physical space nor explain the same image pixels at any point in time. Successful trajectory hypotheses are then fed back to guide object detection in future frames. The optimization procedure is kept efficient through incremental computation and conservative hypothesis pruning. We evaluate our approach on several challenging video sequences and demonstrate its performance on both a surveillance-type scenario and a scenario where the input videos are taken from inside a moving vehicle passing through crowded city areas.

Zopez, J., J. Gonzalez, et al. (2007). A versatile low-cost car plate recognition system.
Miao, L. G., F. W. Wang, et al. (2009). Automatic License Plate Detection Based on Edge density and Color Model.
Molder, C., M. Boscoianu, et al. (2008). Improved automatic number plate recognition system.Proceedings of the 1st Wseas International Conference on Visualization, Imaging and Simulation. M. Iliescu, R. I. Munteanu, J. FraustoSoliset al: 49-54.
Thornton, J., J. Baran-Gale, et al. (2009). An Assessment of the Video Analytics Technology Gap for Transportation Facilities.
Zidouri, A., M. Deriche, et al. (2008). RECOGNITION OF ARABIC LICENSE PLATES USING NN.