

The author(s) shown below used Federal funds provided by the U.S. Department of Justice and prepared the following final report:

**Document Title: Learning Models for Predictive Behavioral Intent
 and Activity Analysis in Wide Area Video
 Surveillance**

Author(s): Shishir K. Shah

Document No.: 250273

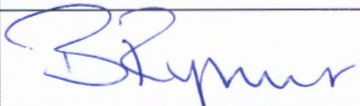
Date Received: October 2016

Award Number: 2009-MU-MU-K004

This report has not been published by the U.S. Department of Justice. To provide better customer service, NCJRS has made this federally funded grant report available electronically.

<p>Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.</p>

COVER PAGE

Federal Agency and Organization Element to Which Report is Submitted	National Institute of Justice Office of Science and Technology
Federal Grant or Other Identifying Number Assigned by Agency	2009-MU-MU-K004
Project Title	Learning Models for Predictive Behavioral Intent and Activity Analysis in Wide Area Video Surveillance
PD/PI Name, Title and Contact Information (e-mail address and phone number)	Shishir K. Shah Associate Professor University of Houston 4800 Calhoun, 564 P. G. Hoffman Houston, TX 77204-3010 sshah@central.uh.edu 1-713-743-3360
Name of Submitting Official Title, and Contact Information (e-mail address and phone number), if other than PD/PI	Ms. Beverly Rymer Exec. Director, Office of Contracts and Grants 316 E. Cullen Bldg. University of Houston Houston, TX 77204-2015 brymer@uh.edu 1-713-743-5773
Submission Date	September 30, 2015
DUNS and EIN Numbers	DUNS: 03-683-7920 EIN: 74-6001399
Recipient Organization (Name and Address)	University of Houston 316 E. Cullen Bldg. 4800 Calhoun Houston, TX 77204
Recipient Identifying Number or Account Number, if any	-
Project/Grant Period (Start Date, End Date)	October 1, 2013 - September 30, 2015
Reporting Period End Date	September 30, 2015
Report Term or Frequency (annual, semi-annual, quarterly, other)	Final
Signature of Submitting Official (signature shall be submitted in accordance with Agency-specific instructions)	 9/30/2015

MANDATORY REPORTING CATEGORIES

What are the major goals and objectives of the project?

The goal of this research is to develop an intelligent, non-obtrusive, real-time, continuous monitoring system for assessing activity and predicting emergent suspicious and criminal behavior across a network of distributed cameras. It is envisioned that such a system would consist of two main modules, namely: 1) A non-obtrusive tracking system that can continuously: i) track all objects across a network of distributed cameras, ii) analyze the spatio-temporal movement pattern of each object, and iii) detect and measure descriptive information continuously of each tracked object. 2) A decision system that can: i) correlate each object's spatio-temporal patterns with others and generate models of suspicious/criminal activity, and ii) generate activity alerts for security personnel who monitor and make critical decisions. Figure 1 depicts a schematic of the envisioned system. To realize our envisioned goal, the specific tasks undertaken in our effort include:

1. Functionalize a state-of-the-art distributed surveillance camera network and collect data for algorithm development and testing.
2. Evaluate and improve tracking algorithm capable of robust object label management and trajectory generation.
3. Evaluate and improve algorithm for object reacquisition across non-overlapping cameras.
4. Evaluate and improve model and algorithm for activity analysis.
5. Develop algorithm for recovering motion trajectories of tracked objects across non-overlapping cameras.

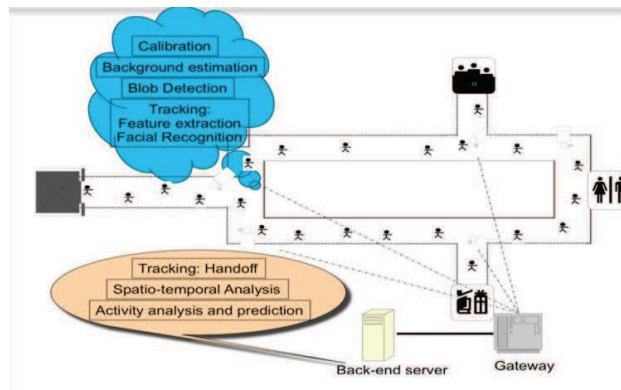


Fig. 1: A systems view of wide area distributed intelligent surveillance system with smart camera nodes performing local processing and the back-end node performing intelligence gathering and reporting.

What was accomplished under these goals?

1 Functionalize a state-of-the-art distributed surveillance camera network and collect data for algorithm development and testing

We collected significant data for evaluation of tracking, re-acquisition, and activity recognition algorithms. We augmented this dataset with additional collections for evaluation and validation of developed algorithms.

We have been able to install a camera network system with multiple cameras providing indoor and outdoor views. This network has been instrumental in our ability to collect data and evaluate the developed algorithms. Figure 2 shows the layout of the established network.

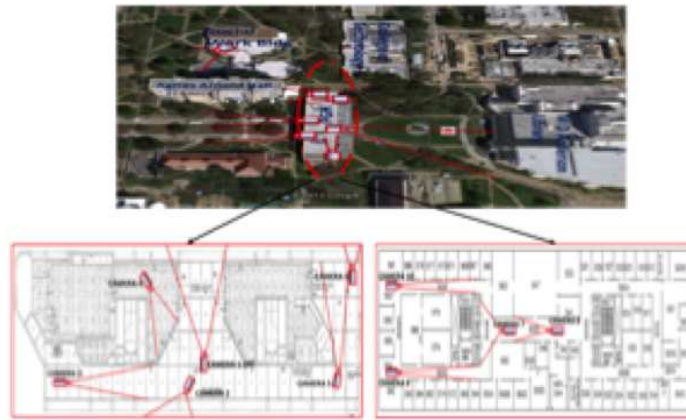


Fig. 2: Coverage provided by the camera network established at University of Houston.

2 Develop tracking algorithm capable of robust object label management and trajectory generation

The tracking methods we have developed have shown good success in tracking multiple targets in a wide range of scenes. Visual tracking of multiple targets in complex scenes captured by a monocular, potentially moving, and uncalibrated camera is a very challenging problem due to measurement noise, cluttered-background, uncertainty of the target motion, occlusions, and illumination changes. While traditional methods for tracking have focused on improving the robustness of motion models and predictive filters, recent advances in methods for object detection have led to the development of a number of tracking-by-detection approaches. However, varying visual properties of the object of interest often results in false positives and missed detections. Hence, the resulting prediction and association problem has to be resolved by inferring between-object interactions using incomplete data sets. Our work has focused on developing novel ensemble framework that leverages redundancy and diversity between tracking and detection for robust multi-target tracking. Our approaches have exploited the discriminative power of the tracker and detector. Moreover, we have developed human motion models that apply to individuals or groups of people and used these models to further the state-of-the-art in visual trackers.

Our work in tracking has resulted in the following peer-reviewed publications, which provide specific details of the developed methods as well as provide a thorough analysis of the performance gains compared to existing methods.

- X. Yan, I. A. Kakadiaris, and S. K. Shah, Modeling Local Behavior for Predicting Social Interactions towards Human Tracking, *Pattern Recognition*, vol. 47-4, pp. 1626–1641, 2014.
- X. Yan, A. Cheriyyadat, and S. K. Shah, Hierarchical Group Structures in Multi-Person Tracking, in *Proc. of 22nd International Conference on Pattern Recognition*, pp. 2221–2226, 2014.
- X. Yan, I. A. Kakadiaris, and S. K. Shah, What Do I See? Modeling Human Visual Perception for Multi-person Tracking, in *Proc. of European Conference on Computer Vision*, pp. 314–329, 2014.

- X. Yan, X. Wu, I. A. Kakadiaris and S. K. Shah, To Track or To Detect? An Ensemble Framework for Optimal Selection, in *Proceedings of European Conference on Computer Vision*, pp. 594–607, 2012.
- X. Yan, I. A. Kakadiaris and S. K. Shah, Predicting Social Interactions for Visual Tracking, in *Proceedings of British Machine Vision Conference*, pp. 102.1–102.11, 2011.

3 Develop algorithm for object reacquisition across non-overlapping cameras

Person re-identification is defined as a process of establishing correspondence between images of a person taken from different cameras. It is used to determine that instances captured by different cameras belong to the same person, in other words, assign a stable ID to different instances of the person. This is critical in deploying useful camera network that can provide intelligent information. Nonetheless, this is a non-trivial problem. Figure 3 shows the function of the re-identification system as a stand alone component of automated video analytics.

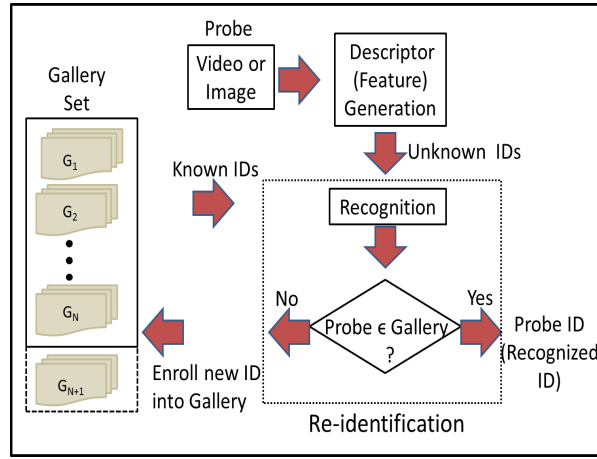


Fig. 3: Re-identification System diagram.

It allows for person tracking across multiple cameras, whether or not the cameras have overlapping field-of-views. This enables the analysis of long term activities and behaviors of people in the scene, which is required for high level surveillance tasks like activity detection, event detection and analyzing crowd movement. Figure 4 shows a typical surveillance area monitored by multiple cameras with non-overlapping fields-of-views and the role of re-identification. Automated surveillance systems with these capabilities can have several safety and security applications: suspicious behavior prediction, criminal tracking, customer tracking in stores, searching for lost children, monitoring the elderly, and so on. Re-identification can be used for surveillance applications within a single camera as well, to detect repetitive behavior, for example, to determine if a person visits a shop-window multiple times or if the same person or another person picks up a bag left by someone. The ability to assign a consistent label to multiple observations of a person improves the semantic coherence of analysis. Re-identification is pervasive in other applications like robotics and multimedia, for instance, photo browsing and photo tagging.

Re-id at the operational level can be thought of as a two step process, capturing a unique person description or model and then comparing two models to infer either a match or a non-match. Our work has resulted in development of novel descriptors and models that can be computed from detections of individuals in images or videos. Further, matching functions have also been developed. Our work on re-id has resulted in the following peer-reviewed publications, which provide specific details of the developed methods as well as provide a thorough analysis of the performance gains compared to existing methods.

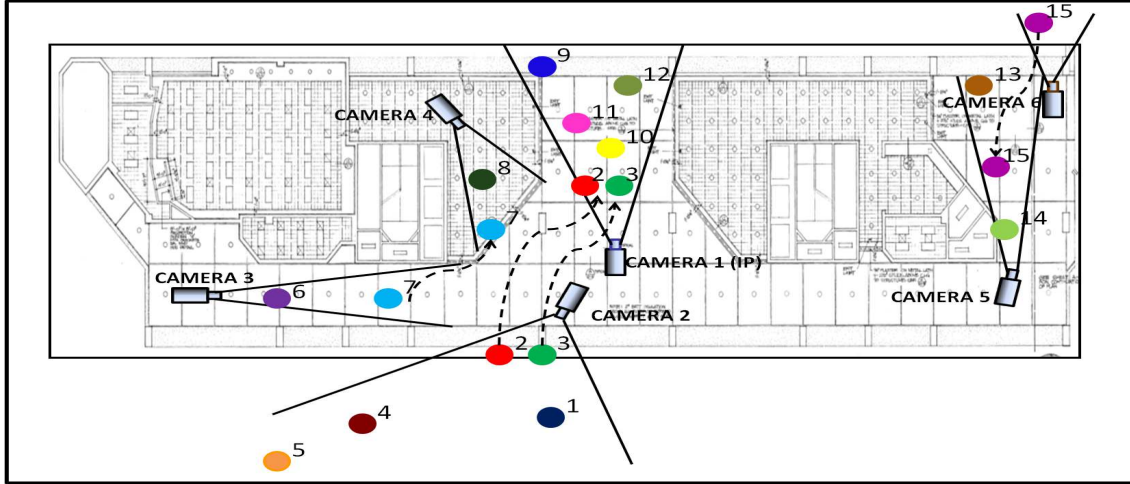


Fig. 4: Multi-camera surveillance network illustration of re-identification.

- L. Wei and S. K. Shah, Subject Centric Group Feature for Person Re-identification, in *Proc. of Workshop on Group and Crowd Behavior Analysis and Understanding (Computer Vision and Pattern Recognition)*, 2015.
- P. Mantini and S. K. Shah, Person Re-Identification using Geometry Constrained Human Trajectory Modeling, in *Proc. of IEEE Symposium on Technologies for Homeland Security*, pp. 1–6, 2015.
- S. K. Shah and A. Gala, **Person Re-identification in Wide Area Camera Networks**, *Wide Area Surveillance*, Eds. V. Asari, Springer, 2014.
- A. Bedagkar-Gala and S. K. Shah, A Survey of Approaches and Trends in Person Re-identification, *Image and Vision Computing*, vol. 32-4, pp. 270–286, 2014.
- A. Gala and S. K. Shah, Gait-assisted Person Re-identification in Wide Area Distributed Surveillance, in *Proc. of Workshop on Human Identification for Surveillance (Asian Conference on Computer Vision)*, pp. 633–649, 2014.
- A. Bedagkar-Gala and S. K. Shah, Part-based Spatio-temporal Model for Multi-Person Re-identification, *Pattern Recognition Letters*, vol. 33-14, pp. 1908–1915, 2012.

4 Develop model and algorithm for activity analysis

Human action recognition is a challenging problem that has received considerable attention from the computer vision community in recent years. Its applications are diverse, spanning from its use in activity understanding for intelligent surveillance systems to improving human-computer interaction. The challenges in solving this problem are multi-fold due to the complexity of human motions, the spatial and temporal variations exhibited due to differences in duration of different actions performed, and the changing spatial characteristics of the human form in performing each action. We have developed generative representations of the motion of human body-parts to learn and classify human actions. Our contributions combine the advantages of both local and global representations, encoding the relevant motion information as well as being robust to local appearance changes. In addition, we have extended our work to address the challenges in understanding activities exhibited by groups or crowds. We have developed novel methods and models to detect groups in videos as well as capture human behavior of groups of people.

Our work on human action/activity recognition has resulted in the following peer-reviewed publications, which provide specific details of the developed methods as well as provide a thorough analysis of the performance gains compared to existing methods.

- K. N. Tran, A. Gala, I. A. Kakadiaris, and S. K. Shah, Activity Analysis in Crowded Environments using Social Cues for Group Discovery and Human Interaction Modeling, *Pattern Recognition Letters*, vol. 44, pp. 49–57, 2014.
- X. Wu, and S. K. Shah, Regularized Multi-View Multi-Metric Learning for Action Recognition, in *Proc. of 22nd International Conference on Pattern Recognition*, pp. 471–476, 2014.
- K. N. Tran, A. Bedagkar-Gala, I. A. Kakadiaris, and S. K. Shah, Social Cues in Group Formation and Local Interactions for Collective Activity Analysis, in *Proceedings of 9th International Conference on Computer Vision Theory and Applications*, 2013. **(Best Paper Award)**.
- K. Tran, I. A. Kakadiaris, and S. K. Shah, Part-based Motion Descriptor Image for Human Action Recognition, *Pattern Recognition*, vol. 45-7, pp. 2562–2572, 2012.
- K. N. Tran, I. A. Kakadiaris and S. K. Shah, Modeling Motion of Body Parts for Action Recognition, in *Proceedings of British Machine Vision Conference*, pp. 64.1–64.12, 2011.

We are presently developing methods for analyzing group actions and activities in unconstrained environments.

5 Develop algorithm for recovering motion trajectories of tracked objects across non-overlapping cameras

A human trajectory is the likely path a human subject would take to get to a destination. Recovering this information in unobserved spaces require the development of trajectory forecasting algorithms that try to estimate or predict this path. While our focus has been on developing such methods for recovering motion trajectories of tracked objects across non-overlapping cameras, their applicability spans applications in robotics, computer vision and video surveillance. Understanding the human behavior can provide useful information towards the design of these algorithms. Human trajectory forecasting algorithm is an interesting problem because the outcome is influenced by many factors, of which we believe that the geometry of the environment plays a significant role. In addressing this problem, we have developed models to estimate the occupancy behavior of humans based on the geometry and behavioral norms. We have also developed a trajectory forecasting algorithm that understands this occupancy and leverages it for trajectory forecasting in previously unseen geometries. Results obtained suggests a significant enhancement in the accuracy of trajectory forecasting by incorporating the occupancy behavior model. We have even tested our model for its applicability in improving person re-identification.

Our work in this area has resulted in the following peer-reviewed publications, which provide specific details of the developed methods as well as provide a thorough analysis of the performance gains compared to existing methods.

- P. Mantini and S. K. Shah, Person Re-Identification using Geometry Constrained Human Trajectory Modeling, in *Proc. of IEEE Symposium on Technologies for Homeland Security*, pp. 1–6, 2015.
- P. Mantini and S. K. Shah, Human Trajectory Forecasting in Indoor Environments using Geometric Context, in *Proc. of Indian Conference on Computer Vision, Graphics and Image Processing*, doi: 10.1145/2683483.2683547, 2014.

What opportunities for training and professional development has the project provided?

Over the duration of this project, 7 students have worked with the PIs in addressing various aspects of this research. Each student has benefited through deeper understanding of the problem and its challenges. They have received training in algorithm development and have been able to develop advanced skills in video analytics. Research accomplished has resulted in various publications and the students have been able to attend relevant workshops, conferences, and seminars. 3 students have graduated with a Ph.D. and 4 have made significant progress towards their Ph.D. degrees.

How have the results been disseminated to communities of interest?

Results of performed research have largely been disseminated through publications and presentations. The specific publications are:

- L. Wei and S. K. Shah, Subject Centric Group Feature for Person Re-identification, in *Proc. of Workshop on Group and Crowd Behavior Analysis and Understanding (Computer Vision and Pattern Recognition)*, 2015.
- P. Mantini and S. K. Shah, Person Re-Identification using Geometry Constrained Human Trajectory Modeling, in *Proc. of IEEE Symposium on Technologies for Homeland Security*, pp. 1–6, 2015.
- P. Mantini and S. K. Shah, Human Trajectory Forecasting in Indoor Environments using Geometric Context, in *Proc. of Indian Conference on Computer Vision, Graphics and Image Processing*, doi: 10.1145/2683483.2683547, 2014.
- X. Yan, I. A. Kakadiaris, and S. K. Shah, Modeling Local Behavior for Predicting Social Interactions towards Human Tracking, *Pattern Recognition*, vol. 47-4, pp. 1626–1641, 2014.
- X. Yan, A. Cheriyyadat, and S. K. Shah, Hierarchical Group Structures in Multi-Person Tracking, in *Proc. of 22nd International Conference on Pattern Recognition*, pp. 2221–2226, 2014.
- X. Yan, I. A. Kakadiaris, and S. K. Shah, What Do I See? Modeling Human Visual Perception for Multi-person Tracking, in *Proc. of European Conference on Computer Vision*, pp. 314–329, 2014.
- S. K. Shah and A. Gala, **Person Re-identification in Wide Area Camera Networks**, *Wide Area Surveillance*, Eds. V. Asari, Springer, 2014.
- A. Bedagkar-Gala and S. K. Shah, A Survey of Approaches and Trends in Person Re-identification, *Image and Vision Computing*, vol. 32-4, pp. 270–286, 2014.
- A. Gala and S. K. Shah, Gait-assisted Person Re-identification in Wide Area Distributed Surveillance, in *Proc. of Workshop on Human Identification for Surveillance (Asian Conference on Computer Vision)*, pp. 633–649, 2014.
- K. N. Tran, A. Gala, I. A. Kakadiaris, and S. K. Shah, Activity Analysis in Crowded Environments using Social Cues for Group Discovery and Human Interaction Modeling, *Pattern Recognition Letters*, vol. 44, pp. 49–57, 2014.
- X. Wu, and S. K. Shah, Regularized Multi-View Multi-Metric Learning for Action Recognition, in *Proc. of 22nd International Conference on Pattern Recognition*, pp. 471–476, 2014.
- K. N. Tran, A. Bedagkar-Gala, I. A. Kakadiaris, and S. K. Shah, Social Cues in Group Formation and Local Interactions for Collective Activity Analysis, in *Proceedings of 9th International Conference on Computer Vision Theory and Applications*, 2013. **(Best Paper Award)**.

- X. Yan, X. Wu, I. A. Kakadiaris and S. K. Shah, To Track or To Detect? An Ensemble Framework for Optimal Selection, in *Proceedings of European Conference on Computer Vision*, pp. 594–607, 2012.
- A. Bedagkar-Gala and S. K. Shah, Part-based Spatio-temporal Model for Multi-Person Re-identification, *Pattern Recognition Letters*, vol. 33-14, pp. 1908–1915, 2012.
- K. Tran, I. A. Kakadiaris, and S. K. Shah, Part-based Motion Descriptor Image for Human Action Recognition, *Pattern Recognition*, vol. 45-7, pp. 2562–2572, 2012.
- X. Yan, I. A. Kakadiaris and S. K. Shah, Predicting Social Interactions for Visual Tracking, in *Proceedings of British Machine Vision Conference*, pp. 102.1–102.11, 2011.
- K. N. Tran, I. A. Kakadiaris and S. K. Shah, Modeling Motion of Body Parts for Action Recognition, in *Proceedings of British Machine Vision Conference*, pp. 64.1–64.12, 2011.

OPTIONAL CATEGORIES

PRODUCTS: What has the project produced? This project has focused on development of new methods and algorithms for wide area surveillance and understanding of human activities in large distributed camera networks. The main product of this work is research publications that span algorithms for tracking individual and groups, activity recognition of individuals and groups, and re-identification of individuals and groups. Each of the developed algorithms have been developed and tested on public and small benchmark datasets. These are standalone evaluations and have not yet been integrated into a system solution.

Considering an integrated solution, we have developed a prototype code for person re-identification that facilitates generation of a gallery/probe from input videos and performs matching of an input observation to the IDs in the gallery. While this is not an integrated system for re-id, the developed prototype has two modular components. The first module takes in a video and performs human detection and tracking. The result of tracking allows for the generation of either a gallery or probe dataset. The second module includes human parts-based re-id, wherein the images in the specified gallery and probe datasets are individually segmented to identify body parts. An appearance descriptor is generated as a model for each body part and integrated over multiple images of the person, if available. This model is used to establish a match between IDs in the gallery and probe folders. The resultant of the program is an error matrix, which in turn can be used to generate a CMC curve if the ground-truth data is available. If not, the error matrix is used to generate a rank ordered set of matched IDs.

PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS: Who has been involved?

Name: Shishir Shah

Project Role: PI

Nearest person month worked: 1 (over past 6 months)

Contribution to Project: The PI performed algorithm development and system evaluation.

Collaborated with individual in foreign country: No

Name: Ioannis Kakadiaris

Project Role: Co-PI

Nearest person month worked: 0 (over past 6 months)

Contribution to Project: The Co-PI assisted in algorithm development and system evaluation.

Collaborated with individual in foreign country: No

Name: Can Cao

Project Role: Graduate Student

Nearest person month worked: 1 (over past 6 months)

Contribution to Project: Mr. Yan has performed work on human tracking and social models for motion prediction.

Collaborated with individual in foreign country: No

Name: Seyyedeh Mirsharif

Project Role: Graduate Student

Nearest person month worked: 1 (over past 6 months)

Contribution to Project: Ms. Mirsharif has performed work on human motion models.

Collaborated with individual in foreign country: No

Name: Pranav Mantini

Project Role: Graduate Student

Nearest person month worked: 3 (over past 6 months)

Contribution to Project: Mr. Mantini has performed work on scene modeling and integration of human occupancy models based on geometric constraints as applied to human tracking.

Collaborated with individual in foreign country: No

Name: Li Wei

Project Role: Graduate Student

Nearest person month worked: 1 (over past 6 months)

Contribution to Project: Mr. Wei has performed work on reidentification models for managing subject IDs across disparate observations within a camera network.

Collaborated with individual in foreign country: No

Name: Apurva Gala

Project Role: Graduate Student

Nearest person month worked: 0 (over past 6 months)

Contribution to Project: Dr. Gala contributed to development of reidentification models for managing subject IDs across disparate observations within a camera network.

Collaborated with individual in foreign country: No

Name: Xu Yan

Project Role: Graduate Student

Nearest person month worked: 0 (over past 6 months)

Contribution to Project: Dr. Yan contributed to development of visual tracking methods.

Collaborated with individual in foreign country: No

Name: Khai Tran

Project Role: Graduate Student

Nearest person month worked: 0 (over past 6 months)

Contribution to Project: Dr. Tran contributed to development of human action/activity recognition methods.

Collaborated with individual in foreign country: No

IMPACT: What is the impact of the project? How has it contributed?

Video is by far the most ubiquitous sensing modality available in large scale monitoring of scenes and activities and can be configured to passively record video and playback as needed, or be actively monitored by security personnel. Several studies have cited the role of such systems in deterring criminal activities, with active monitoring being the most efficient. However, recent studies indicate that traditional passive video surveillance acts mostly in triggering a perceptual mechanism in a potential offender and has little effect on crime. Video data from existing surveillance systems are mostly used after an event as a forensic tool, thereby reducing its primary value as an active, real-time medium. The need for intelligent visual surveillance is paramount; providing automatic monitoring of surveillance video to help security officers detect and predict suspicious activities so that they have enough time to take specific actions to prevent a crime or mitigate a security threat. Intelligent visual surveillance in complex and dynamic environments is aimed to detect, recognize, and track subjects in video sequences in an attempt to understand and describe subject

behavior and predict activities of the subject. The ability to model simple activity patterns from a single camera's video data has been demonstrated for a variety of applications. However, to address the security of large-scale critical infrastructures (e.g., roadway systems, borders, port authorities) surveillance systems are distributed over large geographical spans using numerous cameras with non-overlapping field of views. Intelligence gathering and activity understanding in these environments requires us to scale our ability to understand activities and complex behaviors beyond local camera views to the global spatial extent of the system. Considering the sheer amount of video data generated from such networks, intelligence gathering capability needs to be performed both at each local camera node as well as across the network of cameras. Each camera in turn is required to be a "smart" camera exhibiting intelligent behavior. Hence, there is an urgent need for a new systems paradigm for smart cameras and computational tools to process video imagery from distributed video surveillance camera network for the purpose of real-time activity reporting, and detection and prediction of emergent behaviors. Existing approaches and methods do not adequately address these challenges. The reason is two fold: 1) there is lack of robust automatic techniques to track subjects across cameras with non-overlapping field of view, and 2) there is a lack of automatic techniques to extract and model spatio-temporal patterns of subject activity for the purpose of understanding threats and behaviors. Research supported under this project has enabled us to develop new approaches and models to understand the motion of humans dependent on their interactions with both animate and inanimate objects in the scene along with local scene geometry. We have successfully developed tracking algorithms that integrate such observations and have designed approaches for recognition of human activities. We have been able to create an infrastructure for collecting data for a distributed camera network and test our developed algorithms in realistic scenarios, both for tracking and recognition of individual and group human activities.

Future Work: There are several challenges that have been realized from the performed research. The two areas that have emerged are related to person re-identification and the understanding of human behavior that can be modeled to impact all aspects of wide area video surveillance. The latter is especially true since there is an overall lack of understanding of human behavior, whether it be that of an individual person or a group of individuals.

Person Re-Identification

Re-identification is a relatively new problem, far from being solved definitively, and strongly dependent on the quality of the used sensory data. While much of the existing work on this problem has relied to appearance features of a person, recent trends are beginning to address this limitations. For example, the usage of soft biometric cues as features, alternative or complementary attributes to classical appearance-based re-id, is envisaged as one of the current trends in this field. Actually, soft biometrics systems mostly deal with subjects that do not have a strong collaborative behavior, which would be applicable in certain scenarios. The key underlying challenge for re-id is the extraction of reliable features from data with partial occlusions and large variations of appearance. For constrained scenarios, discriminant cues could be extracted from range data acquired by RGB-D cameras, such as the MS Kinect or Asus Xtion PRO, which are able to acquire depth information in a fast and affordable way. The idea here is to consider more implicit human body characteristics, in particular, to extract a set of features computed directly on the range measurements given by the sensor related to specific anthropometric measurements.

In the following, a subdivision of the re-id issues among those about to be solved in the short term (about 2-3 years), medium term (about 5 years), and longer term is presented. The research challenges are very dependent upon the quality of the input data.

Short-term Possibilities

Input:

In this case, input data consists of high quality images without significant occlusion and clutter, and the scenarios considered are rather constrained and cooperative. Constrained could mean that the people are limited to certain regions (e.g., walkways) and their features can be identified and extracted. Cooperative means that they are not trying to evade identification; in the extreme, this could mean that they are willing to provide their images in a fixed setup (e.g., a security checkpoint).

Research tasks:

In such conditions, the pre-processing part of the re-id process is relatively reliable, that is, *detection and tracking* of moving objects, is possible. Accurate *foreground and body part extraction* can be considered as effective; hence *feature extraction*, the building of the *signature descriptor*, and the consequent actual *recognition*, are problems which are going to be solved in the short-term.

Datasets:

Better (i.e., more realistic) experimental datasets are actually needed to test and validate the current algorithms in real-world applications. One aspect that should be considered is data collected over multiple cameras and longer time periods. That will help analyze how consistent the results are even in these relatively constrained environments.

Medium-term Challenges

Input:

If we consider input images of slightly lower quality, with some partial occlusions, still in a cooperative but relatively unconstrained scenarios, other problems arise which require a more thorough and longer-term investigation. Examples could be people walking on a not-too-busy street and not objecting to having their images taken.

Research tasks:

The *extraction of reliable features* is a major challenge when dealing with data with partial occlusions and large variations of appearance. In such scenarios, moderate body pose variations and the associated appearance and resolution changes may require the use of more robust cues, like 3D features and soft biometric data, including distinctive signs which may characterize univocally an individual (e.g., a tattoo). The *integration of attributes* (e.g., a carrying bag), and in general, of *contextual information* will actually become necessary for disambiguating individuals in a more robust and effective way. Given the more realistic conditions, detection and tracking cannot be considered anymore to be reliable procedures, so *the propagation of errors* from such modules should be taken into account.

The necessity of *better benchmark, more practical, datasets and related performance measures*, still remains a requirement for the deployment of the re-id technology in the real world. In these cases, one may consider the *novelty detection* problem, i.e., to decide when an individual (probe) is not in the gallery and the automatic enrollment of a probe image in the gallery set. Besides, one may consider inclusion of *human in the loop* and this will entail additional issues to be studied like, for instance, the human feedback for the refinement of the query. *Active image acquisition*, possibly through the human, is another aspect that can be studied.

Datasets:

Currently, there are no good datasets to rigorously study these problems. A task will be to collect such datasets, annotate them, and make them available widely. The design of such datasets needs to be carefully thought, if they are acquired in staged scenarios.

Long-term Challenges

Input:

Full real-world scenarios, low-quality images, unconstrained and uncooperative conditions are challenges that will need a longer time horizon. This will involve dealing with natural videos with high clutter in the data, and severe variations in the environmental conditions. An example could be a busy city scene, where a person needs to be recognized as he walks through several blocks with large blind areas in between.

Research tasks:

The main task identified above, that of robust feature extraction, remains the key and will need to be achieved in far more challenging conditions. This may call for the development of novel features. Tools that could be useful for this purpose include *image restoration* (e.g., super resolution) techniques to improve the quality of the acquired images/videos. It is to be expected that a *larger use of context*, like the joint re-id of groups and individuals, can be helpful. *Semantically meaningful attributes* could play a role in providing the required robustness. The development of online learning can be a major step in the feature extraction problem. Methods (like deep learning and sparse coding) which can *automatically learn the best features*, rather than using hand-engineered ones, hold promise in this respect. The use of *multi-modal multi-sensory* input (beyond optical, e.g., multi-spectral, infrared) should also be considered to cope with real scenarios, as well as the potential exploitation of *active acquisition systems and mobile platforms* for collecting more information-rich data. This could allow capturing salient parts of the human body as a way to acquire features that would enable recognizing people in such harsh scenarios. All these techniques should *scale gracefully* with large numbers of cameras, and cope with wider space-time horizons.

Datasets:

The construction of large datasets (*in the wild*) will be a necessity to validate the developed re-id technologies. However, collecting such datasets that will be meaningful and annotating them reliably will be a challenge.

Human Behavior in Groups

After years of research on the analysis of individuals by automatic methods, there is an acknowledged need to increase our understanding on the new issue of analyzing/modeling gatherings of people, commonly referred as groups or crowds, depending on the number of people involved. The research done on these two topics has brought about many diverse ad-hoc methodologies and algorithms, yielding to an increasing trend in the literature, witnessed by the growing number of papers on this subject in the last ten CVPR, ICCV, ECCV venues. This is due to many reasons. From one side, the advancement of the detection and filtering strategies, running on powerful hardware encouraged the development of algorithms able to deal with hundreds of different individuals, providing promising results. Another reason is the increase in the number of cameras that provide coverage of public spaces. Such sensory devices make it possible to observe people from radically different points of view, in a genuine ecologic, noninvasive manner, and for long durations: from low-angle direct view settings to bird-eye views of people. Moreover, the advancement of

social signal processing has brought in the computer vision and pattern recognition community new models imported from the social sciences, able to read between-the-lines of simple locations and velocities assumed by the individuals, using advanced notions of proxemics and kinesics. Hence, there is a need for innovative ideas and solutions for exploiting the potential synergies emerging from the integration of the two domains.

Emerging Challenges:

While the community and our work has started addressing some of these challenges, much work is needed in these specific topics:

- Group/crowd detection
- Crowd counting
- Information fusion for crowd modeling
- F-formation/free conversational groups recognition
- Group detection in a crowd

Methods for analyzing the behavior of groups and crowd, that is, once they have been detected, how to extract semantic information from them is a nascent problem. Predicting/tracking the movement of a group, the formation or disaggregation of a group/crowd, together with the identification of different kinds of groups/crowd depending on their behavior has to be taken into account. Specific challenges are, but not limited to:

- Group/crowd tracking
- Tracking in the crowd
- Group/crowd behavior understanding and activity recognition
- Group profiling
- (Collective) Head orientation, gesture recognition in groups and crowd
- Jointly focused/commonly focused gathering recognition
- Causal, spectator, protest crowd recognition and modeling
- Abnormality detection in a group/crowd
- Crowd forecasting

Finally, identifying and promoting datasets for group/crowd analysis and modeling, developing metrics for evaluating the pros and cons of the various models and methods will be needed to support research and understanding in this domain. Advanced models for group and crowd simulation should also be considered. The particular issues are, but not limited to:

- Metrics for group and crowd modeling
- Video surveillance and sensor networks applications on group/crowd modeling
- Group and crowd datasets
- Group and crowd simulation

References

- [1] Multi-camera object tracking challenge, <http://mct.idealtest.org/datasets.html> (August 2014).
URL <http://mct.idealtest.org/Datasets.html>
- [2] M. Farenzena, L. Bazzani, A. Perina, V. Murino, M. Cristani, Person re-identification by symmetry-driven accumulation of local features, in: Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on, IEEE, 2010, pp. 2360–2367.
- [3] O. Barnich, M. Van Droogenbroeck, Vibe: A universal background subtraction algorithm for video sequences, Image Processing, IEEE Transactions on 20 (6) (2011) 1709–1724.
- [4] P. Grother, P. J. Phillips, Models of large population recognition performance, in: Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on, Vol. 2, IEEE, 2004, pp. II–68.