Human Recognition in a Video Network

Bir Bhanu
Center for Research in Intelligent Systems
University of California, Riverside, CA 92521, USA

ABSTRACT

Video networks is an emerging interdisciplinary field with significant and exciting scientific and technological challenges. It has great promise in solving many real-world problems and enabling a broad range of applications, including smart homes, video surveillance, environment and traffic monitoring, elderly care, intelligent environments, and entertainment in public and private spaces. This paper provides an overview of the design of a wireless video network as an experimental environment, camera selection, hand-off and control, anomaly detection. It addresses challenging questions for individual identification using gait and face at a distance and present new techniques and their comparison for robust identification.

Keywords: Video networks, Camera Selection, Hand-off and Control, Anomaly Detection, Face Recognition, Gait Recognition, Multi-modal Fusion.

1. INTRODUCTION

Sensor networks has been a very active area of research in recent years. However, most of the sensors used in the development of these networks have been local and nonimaging sensors such as acoustics, seismic, vibration, temperature, humidity, etc. The development of emerging video sensor networks poses its own set of unique challenges, including high bandwidth and low latency requirements for real-time processing and control. For example, the cameras in a network can cooperate with each other and perform various tasks in a collaborative manner. Multiple cameras enable us to have different views of the same object at the same time, such that we can choose one or some of them to monitor a given environment. This helps to solve the occlusion problem to some extent, as long as the field-of-views (FOVs) of cameras have some overlap.

In Section 2 we present a systematic approach for the design, implementation, and evaluation of a large-scale, software reconfigurable wireless camera network, suitable for a variety of practical real-time applications. We take into consideration issues related to the hardware, software, control, architecture, network connectivity, performance evaluation, and data processing strategies for the network. We perform multi-objective optimization on settings such as video resolution and compression quality to provide insight into the performance trade-offs when configuring such a network.

One of the most basic tasks in a video network is the tracking of objects, which requires mechanisms to select a camera for a certain object and hand-off this object from one camera to another so as to accomplish seamless tracking. In Section 3, we provide a comprehensive comparison of current and emerging camera selection and hand-off techniques. We consider geometry, statistics, and game theory-based approaches and provide both theoretical and experimental comparison using centralized and distributed computational models.

Most existing methods that used the human actions or trajectories to analyze the human activity assume overlapping field-of-views, In Section 4 we use the appearance and travel time-based human activity classification in the camera network of non-overlapping field-of-views. The mixture of Gaussian-based appearance similarity model incorporates the appearance variance between different cameras to address changes in varying lighting conditions. To address the problem of limited labeled training data, we propose the use of semi-supervised Expectation-Maximization algorithm for activity classification.

It has been found to be difficult to recognize a person from arbitrary views in changing environmental conditions when a non-cooperative subject is walking at a distance. Some of the challenges include low resolution of the video from single/multiple cameras, changing pose of the subject and uncontrolled illumination conditions. In section 5 we address the problems associated with recognizing people at a distance using side face and gait. Finally in Section 6 we provide the conclusions of the paper.

2. VIDEOWEB – DESIGN OF A WIRELESS NETWORK OF VIDEO CAMERAS

We describe the development of a new laboratory called VideoWeb to facilitate research in processing and understanding videos in a wireless environment. While research into large scale sensor networks has been carried out for various applications, the idea of massive video sensor networks consisting of cameras connected over a wireless network is largely new and relatively unexplored [1]. The VideoWeb laboratory [2] entails constructing a robust network architecture for a large number of components, including wireless routers and bridges, video processing servers, database servers, and the video cameras themselves. Hardware and equipment selection needs to take into account durability, performance, and cost. In addition, VideoWeb requires a number of software applications including those for data recording, video analysis, camera control, event recognition, anomaly detection, and an integrated user interface. Challenges for the design of VideoWeb include creating a wireless network robust enough to simultaneously support hundreds of high-bandwidth video cameras at their peak performance, providing power and connectivity to cameras, building a server farm capable of processing all the streaming data in real-time, implementing a low-latency control structure for camera and server control, and designing algorithms capable of real-time processing of video data.

There are numerous “optimal” ways to configure a network. For instance, maximizing video resolution and quality may be paramount for biometrics, particularly in face recognition where a large number of pixels on the face is beneficial to identifying features. Surveillance and alarm systems, on the other hand, may find reliability more important. For instance, it may be more important that every moment is recorded with minimal skipping (not only for evidence in the event of an incident, but also because security applications often employ vision-based motion detection). Object tracking in turn, may benefit most by sacrificing resolution in exchange for a high sustained frame rate. Configuring the network may consist of changing camera parameters (e.g., resolution, compression) as well as physical network parameters (e.g., number of cameras per bridge, number of bridges per router, number of routers per square foot).The later is helpful in introducing a metric for minimizing labor and monetary cost. We define 5 metrics (resolution, compression, frame rate, standard deviation of frame rate and longest skip time between two complete frames) for measuring camera network performance.

We use the concept of Pareto efficiency to define which configuration of parameters is “better” than another. While this does not always tell a user which configuration should be used for a particular application, it serves to reduce the large number of possible configurations by showing which of those are usually “inferior”; a user only has to consider a configuration from the (potentially) much smaller Pareto set rather than every possible combination. We have designed a software-reconfigurable architecture for a wireless network of a large number of video cameras and implemented a working system by building the servers, installing the cameras, writing the software, and configuring the network to support it. Further, we gained insight into configuring the network’s cameras by defining a set of metrics and discovering Pareto-efficient camera configurations by performing multi-objective optimization on a large volume of real data recorded by the system. The idea persists that if one has a camera network with 30FPS cameras, one will be able to obtain the said 30 frames per second regardless of network configuration or parameters. Though this may be true in a controlled test environment, the performance expectation should not be so optimistic for real-world implementations. Even using the most preferred Pareto-efficient configurations on a non-congested network, it is shown that frame rates will most certainly suffer and that tradeoffs must be made. In line with this confirmation is the need to emphasize that *partial frames are important*. Rather than having algorithms which assume that the data consists entirely of complete video frames (and are only capable of processing such frames), real time computer vision algorithms should take advantage of as much information as is available to them; a stream of partial frames which may only be missing the last few rows of data can still be tremendously useful for a number of applications.

3. CAMERA SELECTION, HANDOFF AND CONTROL

An object may be seen in several cameras and multiple cameras may be involved over long physical distances. Also we may need control of cameras to get a better view. Thus, we have to deal with the problems of camera selection, handoff and control. *Camera handoff* is the process of finding the next best camera to see the target object when it is leaving the FOV of the current camera which is being used to track it. This has been an active area of research and many approaches have been proposed. Some camera networks require switches (video matrix) to help monitor the scenes in different cameras. The control can be designed to switch among cameras intelligently. Both distributed and centralized systems are proposed. Some researchers provide hardware architecture design, some of which involve embedded smart cameras, while others focus on the software design for camera assignment and algorithm development.

The research work in camera selection and handoff for a video network consisting of multiple cameras can be classified according to many different aspects, such as whether it is embedded /PC-based; distributed/centralized; calibration-needed/calibration-free; topology-based or topology-free; statistics-based/statistics-free, etc.

Four key approaches for camera selection, handoff and control are shown in Table 1. They are chosen as typical approaches because these approaches cover both distributed systems and centralized systems. Although none of these approaches needs camera calibration, some of them do a geometry correspondence [6] while some do not [5, 7]. Approaches such as [4, 5] provide a more systematic approach to camera selection and handoff.

Various experiments were performed – Case *1:* 2 cameras 3 persons, indoor. *Case 2*: 3 cameras 5 persons, indoor. *Case 3:* 4 cameras 6 persons, outdoor. Table II and Figure 1 show that the game theoretic approach is more flexible to perform camera handoffs based on different criteria.

We analyzed existing and emerging techniques for the camera selection and handoff problem. Pros and cons of distributed and centralized systems are discussed. Four selected approaches are discussed in detail. Both theoretical and experimental comparisons are provided for these approaches [4]. It is shown that the utility-based game theoretic approach is more flexible and has low computational cost. However, it is a centralized algorithm unlike the CSP approach and the fuzzy-based approach. The COR approach is not applicable when the scenario is complicated.

TABLE I. COMPARISON OF KEY APPROACHES FOR CAMERA SELECTION AND HANDOFF APPROACHES.

Approaches	Pros	Cons
Utility-based Game Theoretic Approach [3, 4]	Provides a mathematical framework; Can deal with the cooperation and competition among cameras; Can perform camera selection based on user-supplied criteria; No need for overlapping FOVs.	Communication among cameras is not involved, can be extended for distributed computation; The local utility has to be designed that will align with the global utility in a potential game.
Co-occurrence-to-Occurrence Ratio Approach [6]	Intuitive efficient approach; Acceptable results when there are few occlusions and few cameras and objects.	Time consuming correspondence of point pairs; When correspondence fails or occlusion happens, there is handoff ambiguity and the error rate increases; Computing structure becomes complicated with the increase of # of camera nodes/objects; FOVs have to be overlapped.
Constraint Satisfaction Problem Approach [5]	Provides a distributed system design; Camera nodes can cooperate by forming coalition groups; Conflicts among cameras are solved by the CSP; No requirement for overlapping FOVs.	The backtracking approach is time consuming for solving the constraint satisfaction problem; Only simple constraints are provided; Only simulation (no real video) results are provided.
Fuzzy-based Approach [7]	Distributed approach; Camera state transition and handoff rules are both intuitive; No requirement for overlapping FOVs.	Only simulation results are provided; Tracking has to be accurate; Not robust when occlusion happens; No guarantee for convergence in a large-scale network.

TABLE II: ERROR RATES OF THE SLECTED APPROACHES.

	Utility-based	COR	CSP	Fuzzy-based
Case 1	3.86%	4.23%	3.92%	4.64%
Case 2	4.98%	10.01%	6.33%	7.11%
Case 3	7.89%	45.67%	12.96%	21.33%

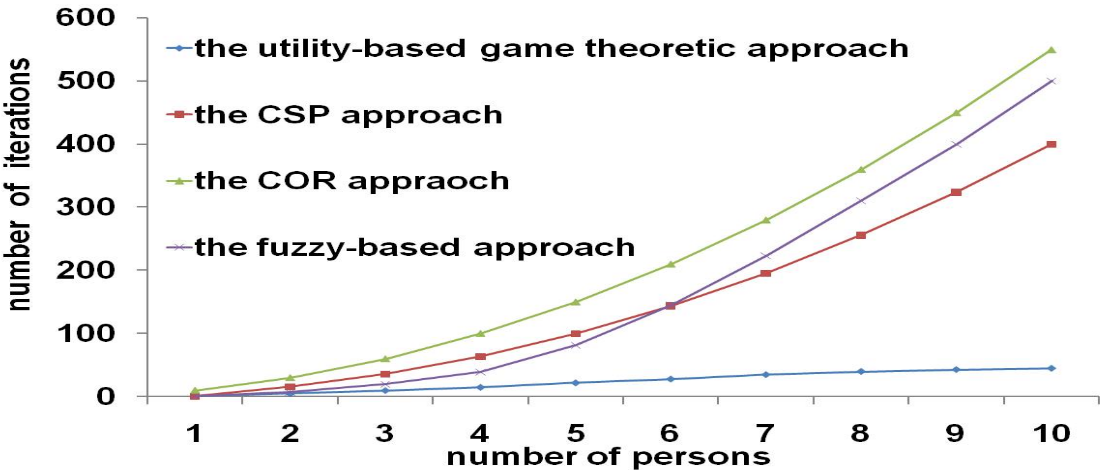

Figure 1. Comparison for the number of iterations with a fixed number of cameras (10) and various numbers of the objects.

4. ANOMALOUS ACTIVITY CLASSIFICATION

As the number of cameras grows, it is becoming humanly impossible to analyze a large number of video feeds effectively. Therefore, we need methods that can automatically analyze the human activities in the video sequences collected by a network of cameras. Suppose there are humans walking in the scene consisting of the conference room, the hallway, the patio and the doors to the stairs. Since the space is divided by the walls and rooms, the paths people can take are relatively constrained. Also, the travel times between the entry and exits of the key areas are relatively fixed depending on the characteristics of the pedestrians. The violation of the common paths and travel times constitutes the anomalous activities. For instance, there is someone taking the emergency exit of the conference room instead of the main door for convenience. It results in the unusual travel time between the conference room and the stairs much shorter than it is supposed to be. Another example is that someone climbs over the wall to circumvent the access control installed at the main entrance. In this case the object suddenly appears at the door of the conference room without previously being detected at the main entrance. Other examples includes: suspicious long stay in the conference room and the sudden disappearance of the subject after showing at the entrance which means that the subject might hide somewhere. All these human activities mentioned above can be categorized to four main types: *break-in*, *stay*, *sudden appearance/disappearance* and *normal*. Among them, the first three anomalous activities require further attention or human involvement.

One possible way is to track the objects (humans) across the overlapping field-of-views (FOVs) of different cameras and determine the types of human activities based on the observed tracks and travel times. However, the assumption of overlapping FOVs requires a huge number of cameras to cover a large area. The data volume increases exponentially along with the equipment cost making such an idea impractical. On the contrary, non-overlapping cameras overseeing the entry/exits in the environment greatly reduce the complexity of the surveillance system. However, the data correspondence problem also arises since there are multiple objects moving in the space and there exist “blind” areas or “gaps” between the FOVs.

We build upon these ideas to develop a framework for analyzing the activity patterns of a group of pedestrians given the inferred network topology and appearance similarity distribution [8, 9]. We use the appearance similarity and travel times observed from much fewer cameras with non-overlapping FOVs to classify the human activities into four different classes: *normal*, *break-in*, *stay*, and *sudden appearance/disappearance*. We employ color histogram-based appearance similarity to establish the correspondence between departure and arrivals at different nodes, and use the statistical model of appearance similarity to incorporate the uncertainty and variance of appearances between different FOVs under varied lighting. Moreover, for a traditional learning-based classification scheme, sufficient labeled training data is the prerequisite for satisfactory classification performance. However, it is really expensive to manually label a large volume of video sequences. Thus, we have used a semi-supervised Expectation-Maximization (SS-EM) algorithm to classify the human activities the limited labeled data. We use a mixture of Gaussian-based statistical model of appearance similarity for correspondence. The GMM parameters are learned on the labeled and unlabeled data by using the SS-EM algorithm. Then, in the testing phase, the estimated Gaussian mixture model is used by the naïve Bayes classifier for anomalous activity classification.

5. INDIVIDUAL RECOGNITION AT A DISTANCE

5.1. Face Recognition in Video Acquired at a Distance

There is a growing interest in face recognition and identification for surveillance systems, information security, and access control applications. In many of the above scenarios, the distance between the objects and the cameras is quite large, which makes the quality of video usually low and face images quite small. Low resolution is one of the challenges in video-based face recognition. Enhancing low resolution (LR) images from the video sequence has been studied by many researchers in the past. Traditional approaches in this area first perform tracking in each frame and then use a super-resolution (SR) method for obtaining increased resolution of the imagery. This process does not pass on the benefits of the SR result to the tracking module and inhibits the entire system from reaching its maximum performance potential. However, in real applications, small size images not only make the recognition task more difficult, but also affect the accuracy of face tracking**.**

We have developed an incremental super-resolution (ISR) technique where SR and tracking are linked together in a closed-loop system. We assume that a 3D generic model is available. The super-resolved texture that is fed back improves the accuracy of pose and illumination estimation, which, in turn, improves the SR result in subsequent frames. Unlike a traditional approach which treats registration and SR steps separately, our approach feeds the super-resolved 3D facial texture back to the tracking algorithm, thus increasing the overall quality of tracking and super-resolving the texture over time. The more accurate tracking, in turn, improves the output of the SR algorithm to acquire better SR texture. Our approach generates SR video by updating the super-resolved texture with the incoming frames.

Our experimental results demonstrate that our closed-loop approach can significantly improve the accuracy of motion estimation and the quality of SR results compared with traditional open-loop approaches. In various experiments [10, 11] we find that in spite of large changes of pose and lighting, the final super-resolved texture can reach a PSNR in the range of 26-29 dB, the tracking can achieve sub-pixel accuracy with a mean of 0.5 pixel and face recognition can improve over 10-20%. We can treat the entire face as a single unit or treat it in terms of its parts (eyes, lips, eyebrows, and rest of the face). Since we use 3D face model, our approach can integrate the information over multiple frames from a video sequence as parts of the face become visible from being invisible at the beginning.

5.2 Gait-based human Recognition at a Distance in Video

We have developed a representation, called Gait Energy Images (GEI) [12] to recognize individuals by their gait as observed in video. GEI is a spatio-temporal compact representation of gait in video. In this representation the entire gait sequence is divided into cycles according to gait frequency and phase information. GEI captures the major shapes of silhouettes and their changes over the gait cycle. Silhouettes in each frame can be obtained using a physically based approach for moving object detection [13, 14]. GEI accounts for human walking at different speeds. It has several advantages over the gait representation of binary silhouette sequence. It is not sensitive to incidental silhouette errors in the individual frames. Moreover, with such a 2D template, we do not need to consider the time moment of each frame, and , therefore, the incurred errors can be avoided.

Given the preprocessed binary gait silhouette sequence in the complete cycle(s), the grey-level gait energy image (GEI) is obtained by averaging the normalized and aligned silhouette images in the gait cycle(s). Various dimensionality reduction techniques such as the Principal Component Analysis (PCA) and subspace methods can be used to develop a compact set of features from GEI for gait-based individual human recognition. For example, we have used PCA and multi-discriminant analysis (MDA) and their various combinations for feature level fusion. A complete set of results on HumanID database, together with their comparison with the state-of-the art algorithms, is given in [12]. In the next subsection we combine gait with the side face for human recognition.

5.3. Side Face and Gait Recognition at a Distance in Video

A fusion system, which combines face and gait cues from a video sequence, is a promising approach to accomplish the task of human recognition at a distance. The general solution to analyze face and gait video data from arbitrary views is to estimate 3-D models. However, the problem of building reliable 3-D models for non-rigid face, with flexible neck and the articulated human body from low resolution video data remains a hard one. In recent years, integrated face and gait recognition approaches without resorting to 3-D models have achieved some success.

The fusion of face and gait is promising in real world applications because of their individual characteristics. Compared with gait, face images are readily interpretable by humans, which allows people to confirm whether a biometrics system is functioning correctly, but the appearance of a face depends on many factors: incident illumination, head pose, facial expressions, moustache/beard, eyeglasses, cosmetics, hair style, weight gain/loss, aging, and so forth. Although gait images can be easily acquired from a distance, the gait recognition is affected by clothes, shoes, carrying status and specific physical condition of an individual. The fusion system is relatively more robust compared with the system that uses only one biometrics. For example, face recognition is more sensitive to low lighting conditions, whereas gait is more reliable under these conditions. Similarly, when the walker is carrying a heavy baggage or he/she is injured, the captured face information may contribute more than gait.

We distinguish a side face from a face problem. A face refers to the outline of the shape of a face as seen from the side. A side face includes not only the outline of the side view of a face, but also the entire side view of eye, nose and mouth, possessing both shape and intensity information. Therefore, a side face has more discriminating power for recognition than a face profile. For side face, an Enhanced Side Face Image (ESFI), a higher resolution image compared with the image directly obtained from a single video frame, is constructed as the face template. For gait, the Gait Energy Image (GEI), which is used to characterize human walking properties, is generated as the gait template.

We have developed several approaches that integrate information from side face and gait at the feature level and match score level to recognize non-cooperating individuals at a distance. Compared with the abundance of research work related to fusion at the match score level, fusion at the feature level is a relatively understudied problem because of the difficulties in practice. Multiple modalities may have incompatible feature sets and the relationship between different feature spaces may not be known. Moreover, the concatenated feature vector may lead to the problem of curse of dimensionality and it may contain noisy or redundant data, thus leading to a decrease in the performance of the classifier. However, pattern recognition and computer vision systems that integrate information at an early stage of processing are believed to be more effective than those systems that perform integration at a later stage. Therefore, while it is relatively difficult to achieve in practice, fusion at the feature level has drawn more attention in recent years. Among the existing research work, feature concatenation is the most popular feature level fusion methodology. Some of the schemes perform feature concatenation after dimensionality reduction while others perform feature concatenation before feature selection.

In our new feature level fusion scheme [15, 16] we propose to fuse information from side face and gait for human recognition at a distance in a single camera scenario. Multiple Discriminant Analysis (MDA) is carried out after the concatenation of face and gait features using PCA based analysis. This allows the generation of better discriminating features and leads to the improved performance. Face features are extracted from Enhanced Side Face Image (ESFI), which integrates face information over multiple frames in video. Similarly, gait features are extracted from Gait Energy Images (GEI). The concatenation of face and gait features generates better discriminating features for improved recognition performance.

The problem of curse of dimensionality is reduced since the feature vectors are of lower dimension than those in [8]. The problem of the curse of dimensionality is reduced in two ways: (a) PCA is used to transform high dimensional face and gait templates to low dimensional feature space; (b) synthetic features are generated based on all possible combinations of face and gait features from the same video sequence.

The proposed scheme [15, 16] is tested using two comparative data sets to show the effect of changing clothes and face changing over time. Moreover, the proposed feature level fusion (PCA followed by concatenation of features and then MDA) is compared with the match score level fusion and another feature level fusion scheme (PCA, MDA and then the concatenation of features). The experimental results demonstrate that the synthetic features, encoding both side face and gait information, carry more discriminating power thn the individual biometrics features. The experimental results show that the proposed feature level fusion scheme [16] is effective for individual recognition in video. It outperforms the previously published fusion schemes at the match score level (Sum and Max rules) and the feature level [15] for face- and gait-based human recognition at a distance in video.

6. CONCLUSIONS

Video network is a significant emerging area of research for human and vehicle identification. There are many problems of interest in such a network: understanding human actions, 3D model building from videos of vehicles and faces, super-resolution techniques to recognize people from a distance, optimal sensing and processing, building topological map of the environment by analyzing traffic patterns from non-overlapping cameras, detecting anomalous activity patterns, integrated analysis/synthesis and learning short/long term behavior [17].

We presented an overview of the design of a wireless video network and the problems of camera selection and handoff, anomaly detection and human recognition at a distance. Video-based human recognition at a distance remains a challenging problem for individual and multi-modal biometrics systems based on face, gait, side face and ear [18-20]. These biometrics can be used in both the low and high security scenarios and are of interest in networked applications.

ACKNOWLEDGEMENTS

This work was supported in part by NSF grants 0622176, 0551741. The contents of the information do not reflect the position or policy of the US Government.

REFERENCES

1. H. Nguyen, B. Bhanu, A. Patel, and R. Diaz, "VideoWeb: Design of a wireless camera network for real-time monitoring of activities," *Third ACM/IEEE International Conference on Distributed Smart Cameras*, Como, Italy, 30 August - 2 September, 2009.
2. Videoweb http://www.vislab.ucr.edu/RESEARCH/Projects/VideoWeb/index.htm
3. Y. Li and B. Bhanu, "A comparison of techniques for camera selection and handoff in a video network," *Third ACM/IEEE International Conference on Distributed Smart Cameras*, Como, Italy, 30 August - 2 September, 2009.
4. Y. Li and Bhanu, "Utility-based dynamic camera assignment and hand-off in a video network," *International Conference on Distributed Smart Cameras*, Stanford University, September 7-11, 2008.
5. F. Bashir and F. Porikli, "Multi-camera control through constraint satisfaction for persistent surveillance," Proc. *IEEE International Conference on Advanced Video and Signal Based Surveillance*, pp.211-218, 2008.
6. Y. Jo and J. Han, "A new approach to camera hand-off without camera calibration for the general scene with non-planar ground," *Proc. of the 4th ACM international workshop on Video surveillance and sensor networks (VSSN06).* Santa Barbara, CA, October 2006, pp. 195–202.
7. K. Morioka, S. Kovacs, J. Lee, P. Korondi and H. Hashimoto, "Fuzzy-based camera selection for object tracking in a multi-camera system," Proc. *IEEE Conference on Human System Interactions, 2008, pp.767-772.*
8. *X. Zou, B. Bhanu, B. Song*, and A. K. Roy-Chowdhury. Determining topology in a distributed camera network. *Proc. IEEE Intl. Conf. on Image Processing*, pp. 133-136, Sept. 2007.
9. X. Zou and B. Bhanu. Anomalous activity classification in the distributed camera network. *Proc. IEEE International Conference on Image Processing*, pp. 781-784, 2008.

10. J. Yu, B. Bhanu, Y. Xu and A. Roy-Chowdhury. Super-resolved facial texture under changing pose and illumination, *Proc. IEEE International Conference on Image Processing*, pp. 553-556, 2007.
11. J. Yu and B. Bhanu, “Super-resolution of facial images in video with expression changes,” *IEEE International Conference on Advanced Video and Signal Based Surveillance*, Santa Fe, NM, September 1-3, 2008.
12. J. Han and B. Bhanu, “Individual recognition using gait energy image,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(2), 316-322, (February 2006).
13. J. Han and B. Bhanu, “Moving human detection by EO and IR sensor fusion,” *Pattern Recognition*, 40(6): 1771-1784 (2007)
14. S. Nadimi and B. Bhanu, “Physical models for moving shadow and object detection in video,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(8), 1079-1087, (August 2004).
15. X. Zhou and B. Bhanu. Integrating face and gait for human recognition at a distance in video, *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 37, No. 5, pp. 1119-1137, October 2007.
16. X. Zhou and B. Bhanu, “Feature fusion for video-based human identification,” *Pattern Recognition*, Special issue on Feature Generation and Machine Learning for Robust Multimodal Biometrics, Volume 41, Issue 3, Pages 778-795, March 2008.
17. N. Ghosh, B. Bhanu, and G. Denina, “Continuously evolvable Bayesian nets for human action analysis in videos,” *Third ACM/IEEE International Conference on Distributed Smart Cameras*, Como, Italy, 30 August - 2 September, 2009.
18. H. Chen and B. Bhanu,” Human ear recognition in 3D,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Special Issue on Biometrics, Vol. 29, No. 4, pp. 718-737, April 2007.
19. B. Bhanu and H. Chen, “Human Ear Recognition by Computer,” Springer, 2008.
20. B. Bhanu and J. Han, “Human Recognition at a Distance in Video,” Springer, 2009.