CHAPTER-2
REVIEW OF LITERATURE

2.1 Human Activity Recognition

During the past decade, there has been an exceptional development of microelectronics and computer systems, enabling sensors and mobile devices with unprecedented characteristics. Their high computational power, small size, and low cost allow people to interact with the devices as part of their daily living. That was the genesis of Ubiquitous Sensing, an active research area with the main purpose of extracting knowledge from the data acquired by pervasive sensors (A. J. Perez et al., 2010). Particularly, the recognition of human activities has become a task of high interest within the field, especially for medical, military, and security applications. For instance, patients with diabetes, obesity, or heart disease are often required to follow a well defined exercise routine as part of their treatment (Y. Jia et al.,2009) Therefore, recognizing activities such as walking, running, or cycling becomes quite useful to provide feedback to the caregiver about the patient's behavior. Likewise, patients with dementia and other mental pathologies could be monitored to detect abnormal activities and thereby prevent undesirable consequences (J. Yin et al., 2008).

In tactical scenarios, precise information on the soldiers' activities along with their locations and health conditions is highly beneficial for their performance and safety. Such information is also helpful to support decision making in both combat and training scenarios. The first works on human activity recognition (HAR) date back to the late '90s (O. X. Schlmilch et al.,1999 ;F. Foerster et al.1999). However, there are still many issues that motivate the development of new techniques to improve the accuracy under more realistic conditions. Some of these challenges are (1) the selection of the attributes to be measured, (2) the construction of a portable, unobtrusive, and inexpensive data acquisition system, (3) the design of feature extraction and inference methods, (4) the collection of data under realistic conditions, (5) the flexibility to support new users without the need of re-training the system, and (6) the implementation in mobile devices meeting energy and processing requirements (E. Kim et al 2010).

The recognition of human activities has been approached in two different ways, namely using external and wearable sensors. In the former, the devices are fixed in predetermined points of interest, so the inference of activities entirely depends on the voluntary interaction of the users with the sensors. In the latter, the devices are attached to the user. Intelligent homes (T. van Kasteren et al., 2010 ; A. Tolstikov et al.2011 ; J. Yang et al. 2011; J. Sarkar et al.2010) are a typical example of external sensing. These systems are able to recognize fairly complex activities (e.g., eating, taking a shower, washing dishes, etc.) because they rely on data from a number of sensors placed in target objects which people are supposed to interact with (e.g., stove, faucet, washing machine, etc.).

Nonetheless, nothing can be done if the user is out of the reach of the sensors or they perform activities that do not require interaction with them. Additionally, the installation and maintenance of the sensors usually entail high costs. Cameras have also been employed as external sensors for HAR. In fact, the recognition of activities and gestures from video sequences has been the focus of extensive research (P. Turaga et al, 2008 ; J. Candamo et al.2010 ; C. N. Joseph et al., 2010 ; M. Ahad et al. 2008). This is especially suitable for security (e.g, intrusion detection) and interactive applications. A remarkable example, and also commercially available, is the Kinect game console (J. Shotton et al, 2011) developed by Microsoft.

It allows the user to interact with the game by means of gestures, without any controller device. Nevertheless, video sequences certainly have some issues in HAR. The first one is privacy, as not everyone is willing to be permanently monitored and recorded by cameras. The second one is pervasiveness because video recording devices are difficult to attach to target individuals in order to obtain images of their entire body during daily living activities. The monitored individuals should then stay within a perimeter defined by the position and the capabilities of the camera(s). The last issue would be complexity, since video processing techniques are relatively expensive, computationally speaking, hindering a real time HAR system to be scalable. The aforementioned limitations motivate the use of wearable sensors in HAR. Most of the measured attributes are related to the user's movement (e.g.,

using accelerometers or GPS), environmental variables (e.g., temperature and humidity), or physiological signals (e.g., heart rate or electrocardiogram).

2.1.1 General Structure of HAR Systems

Similar to other machine learning applications, activity recognition requires two stages, i.e., training and testing (also called evaluation). Figure 2.1 illustrates the common phases involved in these two processes (Miguel A et al.,2014). The training stage initially requires a time series dataset of measured attributes from individuals performing each activity. The time series are split into time windows to apply feature extraction thereby filtering relevant information in the raw signals. Later, learning methods are used to generate an activity recognition model from the dataset of extracted features.

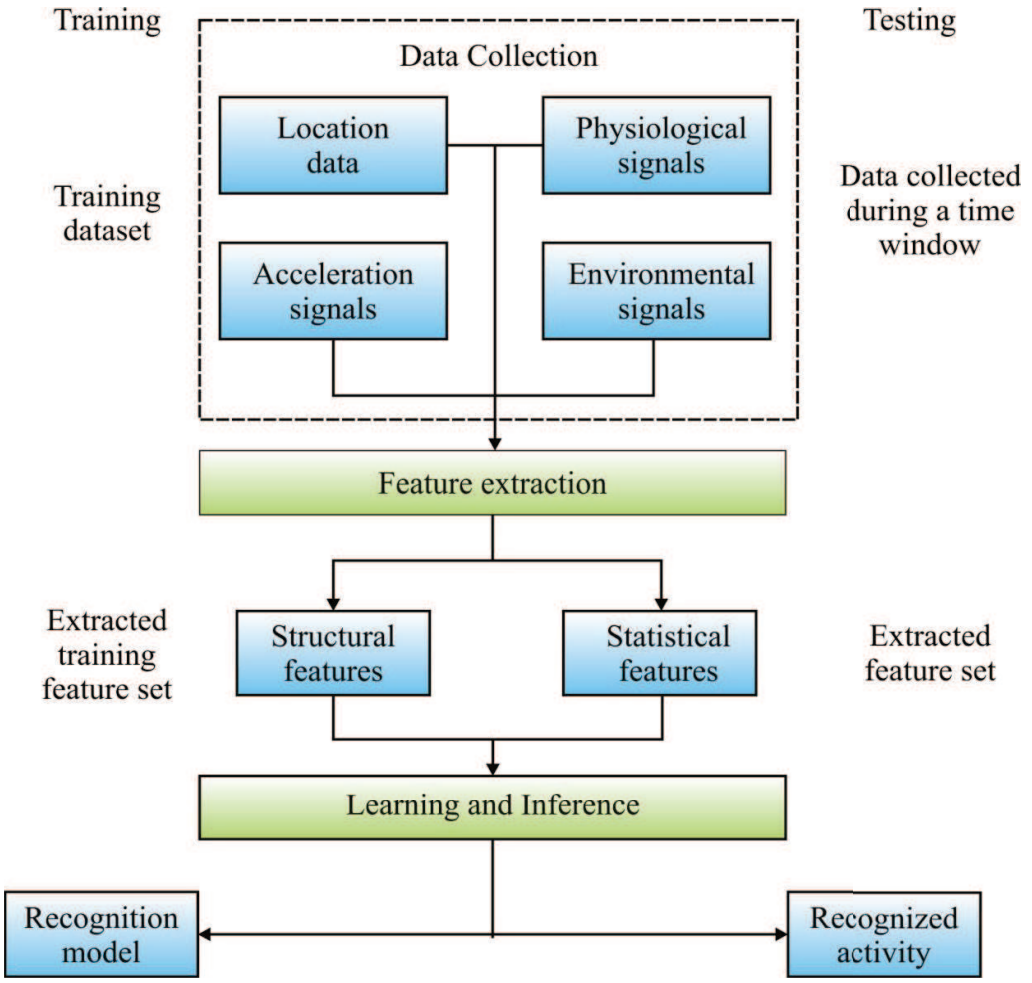

Figure 2.1: General data flow for training and testing HAR systems based on wearable sensors (Miguel A et al., 2014).

Likewise, for testing, data are collected during a time window, which is used to extract features. Such feature set is evaluated in the priory trained learning model, generating a predicted activity label. We have also identified generic data acquisition architecture for HAR systems.

In the first place, wearable sensors are attached to the person's body to measure attributes of interest such as motion, location, and temperature, ECG, among others. These sensors should communicate with an integration device (ID), which can be a cell phone, a PDA, a laptop, or a customized embedded system. The main purpose of the ID is to preprocess the data received from the sensors and, in some cases, send them to an application server for real time monitoring, visualization, and/or analysis.

The communication protocol might be UDP/IP or TCP/IP, according to the desired level of reliability. Notice that all of these components are not necessarily implemented in every HAR system. In, the data are collected offline, so there is neither communication nor server processing. Other systems incorporate sensors within the ID, or carry out the inference process directly on it. The presented architecture is rather general and the systems surveyed in this paper are particular instances of it.

2.2 Design Issues

We have distinguished seven main issues pertaining to human activity recognition, namely, (A) selection of attributes and sensors, (B) obtrusiveness, (C) data collection protocol, (D) recognition performance, (E) energy consumption, (F) processing, and (G) flexibility. The main aspects and solutions related to each one of them are analyzed next.

A. Selection of Attributes and Sensors

Selection of attributes and sensors four groups of attributes are measured using wearable sensors in a HAR context: environmental attributes acceleration, location, and physiological signals.

1) Environmental attributes: These attributes, such as temperature, humidity, audio level, etc., are intended to provide context information describing the individual's surroundings. If the audio level and light intensity are fairly low, for instance, the subject may be sleeping. Various existing systems have utilized microphones, light sensors, humidity sensors, and thermometers, among others.

Those sensors alone, though, might not provide sufficient information as individuals can perform each activity under diverse contextual conditions in terms of weather, audio loudness, or illumination. Therefore, environmental sensors are generally accompanied by accelerometers and other sensors.

2) Acceleration: Triaxial accelerometers are perhaps the most broadly used sensors to recognize ambulation activities (e.g., walking, running, lying, etc.). Accelerometers are inexpensive, require relatively low power, and are embedded in most of today's cellular phones.

However, other daily activities such as eating, working at a computer, or brushing teeth, are confusing from the acceleration point of view. For instance, eating might be confused with brushing teeth due to arm motion. The impacts of the sensor specifications have also been analyzed. In fact, Maurer et al. studied the behavior of the recognition accuracy as a function of the accelerometer sampling rate (which lies between 10 Hz and 100 Hz). Interestingly, they found that no significant gain in accuracy is achived above 20 Hz for ambulation activities. In addition, the amplitude of the accelerometers varies from ±2g, up to ±6g yet ±2g was shown to be sufficient to recognize ambulation activities. The placement of the accelerometer is another important point of discussion: He et al. found that the best place to wear the accelerometer is inside the trousers pocket. Instead, other studies suggest that the accelerometer should be placed in a bag carried by the user, on the belt, or on the dominant wrist.

At the end, the optimal position where to place the accelerometer depends on the application and the type of activities to be recognized.

3) Location: The Global Positioning System (GPS) enables all sorts of location based services. Current cellular phones are equipped with GPS devices, making this sensor very convenient for context-aware applications, including the recognition of the user's transportation mode. The place where the user is can also be helpful to infer their activity using ontological reasoning.

As an example, if a person is at a park, they are probably not brushing their teeth but might be running or walking. And, information about places can be easily obtained by means of the Google Places Web Service, among other tools. However, GPS devices do not work well indoors and they are relatively expensive in terms of energy consumption, especially in real-time tracking applications. For those reasons, this sensor is usually employed along with accelerometers.

Finally, location data has privacy issues because users are not always willing to be tracked. Encryption, obfuscation, and anonymization are some of the techniques available to ensure privacy in location data.

4) Physiological signals: Vital signs data (e.g., heart rate, respiration rate, skin temperature, skin conductivity, ECG, etc.) have also been considered in a few works. Tapia et al. proposed an activity recognition system that combines data from five triaxial accelerometers and a heart rate monitor. However, they concluded that the heart rate is not useful in a HAR context because after performing physically demanding activities (e.g., running) the heart rate remains at a high level for a while, even if the individual is lying or sitting.

In a previous study we showed that, by means of structural feature extraction, vital signs can be exploited to improve recognition accuracy. Now, in order to measure physiological signals, additional sensors would be required, thereby increasing the system cost and introducing obtrusiveness. Also, these sensors generally use wireless communication which entails higher energy expenditures.

B. Obtrusiveness

For Obtrusiveness to be successful in practice, HAR systems should not require the user to wear many sensors nor interact too often with the application. Furthermore, the more sources of data available, the richer the information that can be extracted from the measured attributes. There are systems which require the user to wear four or more accelerometers, or carry a heavy rucksack with recording devices. These configurations may be uncomfortable, invasive, expensive, and hence not suitable for activity recognition. Other systems are able to work with rather unobtrusive hardware. For instance, a sensing platform that can be worn as a sport watch is presented in sentine only requires a strap that is placed on the chest and a cellular phone. Finally, the systems introduced in recognize activities with a cellular phone only.

C. Data Collection Protocol

The procedure followed by the individuals while collecting data is critical in any HAR. In 1999, Foerster et al. demonstrated 95.6% of accuracy for ambulation activities in a controlled data collection experiment, In a natural environment (i.e., outside of the laboratory), the accuracy dropped to 66% (Foerster et al., The number of individuals and their physical characteristics are also crucial factors in any HAR study. A comprehensive study should consider a large number of individuals with diverse characteristics in terms of gender, age, height, weight, and health conditions. This is with the purpose of ensuring flexibility to support new users without the need of collecting additional training data.

D. Recognition Performance

The performance of a HAR system depends on several aspects, such as (1) the activity set (2) the quality of the training data, (3) the feature extraction method, and (4) the learning algorithm. In the first place, each set of activities brings a totally different pattern recognition problem. For example, discrimination among walking, running, and standing still, turns out to be much easier than incorporating more complex activities such as watching TV, eating, ascending, and descending. Secondly, there should be a sufficient amount of training data, which should also be similar to the expected testing data.

Finally, a comparative evaluation of several learning methods is desirable as each dataset exhibits distinct characteristics that can be either beneficial or detrimental for a particular method. Such interrelationship among datasets and learning methods can be very hard to analyze theoretically, which accentuates the need of an experimental study. In order to quantitatively understand the recognition performance, some standard metrics are used, e.g., accuracy, recall, precision, Fmeasure, Kappa statistic, and ROC curves.

E. Energy Consumption

Context-aware applications rely on mobile devices-such as sensors and cellular phones— which are generally energy constrained. In most scenarios, extending the battery life is a desirable feature, especially for medical and military applications that are compelled to deliver critical information. Surprisingly, most HAR schemes do not formally analyze energy expenditures, which are mainly due to processing, communication, and visualization tasks. Communication is often the most expensive operation, so the designer should minimize the amount of transmitted data. In most cases, short range wireless networks (e.g., Bluetooth or Wi-Fi) should be preferred over long range networks (e.g., cellular network or WiMAX) as the former require lower power.

Some typical energy saving mechanisms is data aggregation and compression yet they involve additional computations that may affect the application performance. Another approach is to carry out feature extraction and classification in the integration device, so that raw signals would not have to be continuously sent to the server. Finally, since all sensors may not be necessary simultaneously, turning some of them off or reducing their sampling/transmission rate is very convenient to save energy. For example, if the user's activity is sitting or standing still, the GPS sensor may be turned off.

F. Processing

Another important point of discussion is where the recognition task should be done, whether in the server or in the integration device. On one hand, a server is expected to have huge processing, storage, and energy capabilities, allowing

incorporating more complex methods and models. On the other hand, a HAR system running on a mobile device should substantially reduce energy expenditures, as raw data would not have to be continuously sent to a server for processing.

The system would also become more robust and responsive because it would not depend on unreliable wireless communication links, which may be unavailable or error prone; this is particularly important for medical or military applications that require real-time decision making. Finally, a mobile HAR system would be more scalable since the server load would be alleviated by the locally performed feature extraction and classification computations. However, implementing activity recognition in mobile devices becomes challenging because they are still constrained in terms of processing, storage, and energy. Hence, feature extraction and learning methods should be carefully chosen to guarantee a reasonable response time and battery life. For instance, classification algorithms such as Instance Based Learning and Bagging are very expensive in their evaluation phase, which makes them not convenient for HAR.

G. Flexibility

There is an open debate on the design of any activity recognition model. Some authors claim that, as people perform activities in a different manner (due to age, gender, weight, and so on), a specific recognition model should be built for each individual. This implies that the system should be retrained for each new user. Other studies rather emphasize the need of a monolithic recognition model, flexible enough to work with different users. Consequently, two types of analyses have been proposed to evaluate activity recognition systems: subject-dependent and subject-independent evaluations.

In the first one, a classifier is trained and tested for each individual with his/her own data and the average accuracy for all subjects is computed. In the second one, only one classifier is built for all individuals using cross validation or leave-one individual-out analysis. It is worth to highlight that, in some cases, it would not be convenient to train the system for each new user, especially when (1) there are too many activities; (2) some activities are not desirable for the subject to

carry out (e.g., falling downstairs); or (3) the subject would not cooperate with the data collection process (e.g., patients with dementia and other mental pathologies). On the other hand, an elderly lady would surely walk quite differently than ten years-old boy, thereby challenging a single model to recognize activities regardless of the subject's characteristics. A solution to the dichotomy of the monolithic vs. particular recognition model can be addressed by creating groups of users with similar characteristics.

2.3 Review of Literature

Many efforts have been reported in literature for developing a system for human detection and recognition.

A fair amount of research works have been published in literature for Gabor based image recognition. Lades et al. developed a Gabor wavelet based face recognition system using dynamic link architecture (DLA) framework which recognizes faces by extracting Gabor jets at each node of a rectangular grid over the face image (M. Lades et al., 1993).

Wiskott et al. Subsequently expanded on DLA and developed a Gabor wavelet based elastic bunch graph matching (EBGM) method to label and recognize facial images (L. Wiskott et al., 1997). In the EBGM algorithm, the face is represented as a graph, each node of which contains a group of coefficients, known as jets. However, both DLA and EBGM require extensive amount of computational cost.

Anuj Mohan et al. have developed Example-Based Object Detection in Images by Components (Anuj Mohan et al., 2001). They presented a general example-based framework for detecting objects in static images by components. The technique was demonstrated by developing a system that locates people in cluttered scenes. The system was structured with four distinct example-based detectors which were trained to separately find the four components of the human body namely the head, legs, left arm and right arm. It was ensured that these components were present in the proper geometric configuration. A second example-based classifier then combined the results of the component detectors to classify a pattern as either a

person or a nonperson. This type of hierarchical architecture, in which learning occurs at multiple stages, is called an Adaptive Combination of Classifiers (ACC). They presented that this system performs better than a similar full-body person detector. This suggests that the improvement in Performance is due to the component-based approach and the ACC data classification architecture. The algorithm is also more robust than the full-body person detection method in that it is capable of locating partially occluded views of people and people whose body parts have little contrast with the background.

Kyong I. Chang et al. have developed Multi-biometrics Using Facial Appearance, Shape and Temperature (Kyong I. Chang et al., 2004). They presented results of the first study to examine individual and multi-modal face recognition using 2D, 3D and infrared images of the same set of subjects. Each sensor captures different aspects of human facial features like appearance in intensity representing surface reflectance from a light source, shape data representing depth values from the camera, and the pattern of heat emitted respectively. They employed a database containing a gallery set of 127 images and an accumulated time-lapse probe set of 297 images. Using a PCA-based approach tuned separately for 2D, 3D and IR they found rank-one recognition rates of 90.6% for 2D, 91.9% for 3D and 71.0% for IR. Combining each pair of modalities, they found a multi-modal rank-one recognition rate of about 98.7% for 2D-3D, 96.6% for 2D-IR and 98.0% for 3D-IR. When all three modalities are combined, they obtained 100% recognition. The results shown in their study appear to support the conclusion that the path to higher accuracy and robustness in biometrics involves use of multiple biometrics instead of the best possible sensor and algorithm for a single biometric.

Alrxunder M. et al. have developed fusion of 2d and 3d data in three-dimensional face recognition (Alrxunder M. et al., 2004). They presented the synthesis between the 3D and the 2D data in three-dimensional face recognition. They showed how to compensate for the illumination and facial expressions using the 3D facial geometry and presented the approach of canonical images, which allowed to incorporate geometric information into standard face recognition approaches. The focus of their paper was on the synthesis between the 3D and the

2D data. Besides making the facial surface geometry available, 3D imaging also benefits from the ability to compensate for the illumination in the reflectance images of the face, i.e. estimate the albedo of the face. When the facial geometry together with the albedo is embedded into a plane, a 2D illumination- expression and pose-invariant representation of the face is obtained. Standard techniques can then be employed to carry out the recognition. In their paper, they considered the problem of fusing the 2D and the 3D data in three-dimensional face recognition. As the working framework, they focused on the geometric face recognition approach. The availability of the facial geometry, combined with known illumination direction allowed to extract the albedo of the face, which is invariant to illumination. They highlighted a method to estimate the albedo from photometric stereo and structured light. The albedo image can be then attenuated using the MDS procedure applied to the matrix of geodesic distances on the face measured using the Fast Marching method. The resulting canonical image incorporated the geometric invariants of the face (the geodesic distances), which appeared to be nearly-invariant to facial expression as well as illumination invariant. For recognition purposes, these invariant representations can be compared using the classical techniques like as eigen decomposition.

Dhruv Batra et al. have developed Gabor Filter based Fingerprint Classification Using Support Vector Machines (Dhruv Batra et al.,2004). They have mentioned that fingerprint dassification is important for various practical applications. An accurate and consistent classification can greatly reduce finger print matching time for large databases. Gabor filter based Feature extraction scheme can be used to generate a 384 dimensional feature vector for each finger print image. The classification of these patterns is done through a novel two stage classifier in which K Nearest Neightbour (K“) acts as the first step and finds out the two most frequently represented classes amongst the K nearest pattern, followed by the pertinent SVM classifier choosing the most apt class of the two. 6 SVMs have to be trained for a four class problem, (6Ca), that is, all one against one SVMs. Using this novel scheme and working on the FVC 2000 database (257 final images) they achieved a maxi” accuracy of 98.81% with a rejection percentage of 19596.This was

significantly higher than most reported results in contemporary literature. The SVM training time was 145 seconds, i.e. 24 seconds per SVM on a Pentium III machine.

Malassiotis, et al. have developed Face Localization and Authentication Using Color and Depth Images (Malassiotis, et al., 2005).They proposed a complete face authentication system integrating both two-dimensional (color or intensity) and three-dimensional (3-D) range data, based on a low-cost 3-D sensor, capable of real-time acquisition of 3-D and color images. Novel algorithms were proposed that exploited depth information to achieve robust face detection and localization under conditions of background clutter, occlusion, face pose alteration, and harsh illumination. The well-known embedded hidden Markov model technique for face authentication was applied to depth maps and color images. To cope with pose and illumination variations, the enrichment of face databases with synthetically generated views is proposed.

P. Jonathon et al. have developed Overview of the Face Recognition Grand Challenge (P. Jonathon et al.,2005). Over the last couple of years, face recognition researchers have been developing new techniques. These developments are made possible by advances in computer vision techniques, computer design, sensor design, and interest in fielding face recognition systems. Such advances will help reducing the error rate in face recognition systems by an order of magnitude over Face Recognition Vendor Test (FRVT) 2002 results. The Face Recognition Grand Challenge (FRGC) was designed to achieve this goal by presenting to researchers a six-experiment challenge problem along with data corpus of 50,000 images. The data consists of 3D scans and high resolution still imagery taken under controlled and uncontrolled conditions. Their paper described the challenge problem, data corpus, and presents baseline performance and preliminary results on natural statistics of facial imagery.

Jae-Hak Kim and Richard Hartley have proposed Translation Estimation from Omni directional Images (Jae-Hak Kim and Richard Hartley, 2005). They presented a translation estimation method from omni directional images with missing data. Given omni directional images, it is possible to estimate the rotations

of each camera using a fundamental matrix. They got the translations of each camera using a plane-based projective reconstruction method. However, the plane-based translation estimation method does not give a robust result when there are measurement errors in the data.

J.K. Aggarwal *have* explored Understanding of Human Motion, Actions and Interactions (J. K. Aggarwal, 2005). The efforts to develop computer systems able to detect humans and recognize their activities form an important area of research in computer vision today. The recognition of human activities will lead to a number of applications, like personal assistants, virtual reality, *smart* monitoring and surveillance systems, as well as motion analysis in sports, medicine and choreography. Motion is an important cue for the human visual system and for understanding human actions. It has been the subject of intense study in a number of fields including philosophy, psychology and neurobiology and, of course, computer vision, robotics and computer graphics. In computer vision research, motion has played an important role for the past thirty years. The research included the study of interactions at the gross (blob) level and at the detailed (head, torso, arms and legs) level. The two levels present different problems in terms of observation and analysis. For blob level analysis, they used a modified Hough transform called the Temporal Spatio-Velocity transform to isolate pixels with similar velocity profiles. For the detailed-level analysis, they employed a multi-target, multi-assignment strategy to track blobs in consecutive frames. An event hierarchy consisting of pose, gesture, action and interaction was used to describe human-human interaction. A methodology was developed to describe the interaction at the semantic level. Professor Aggarwal's presentation focused on the contributions from other fields leading to the study of motion in computer vision. Further, it addressed the interaction issue at the blob level and at the detailed level. In addition, addressed the directions of future research in motion and human activity recognition.

Dang-Hui et al. have developed Illumination invariant face recognition (Dang-Hui et al.2005). They presented that the appearance of a face will vary drastically when the illumination changes. Variations in lighting conditions will make face recognition an even more challenging and difficult task. In their paper,

they proposed a novel approach to handle the illumination problem. Their method can restore a face image captured under arbitrary lighting conditions to one with frontal illumination by using a ratio-image between the face image and a reference face image, both of which are blurred by a Gaussian filter. An iterative algorithm was then used to update the reference image, which was reconstructed from the restored image by means of principal component analysis (PCA), in order to obtain a visually better restored image. Image processing techniques are also used to improve the quality of the restored image. To evaluate the performance of their algorithm, restored images with frontal illumination were used for face recognition by means of PCA. Experimental results demonstrate that face recognition using their method can achieve a higher recognition rate based on the Yale B database and the Yale database. Their algorithm has several advantages over other previous algorithms:

(1) it does not need to estimate the face surface normals and the light source directions, (2) it does not need many images captured under different lighting conditions for each person, nor a set of boot strap images that includes many images with different illuminations, and (3) it does not need to detect accurate positions of some facial feature points or to warp the image for alignment, etc.

Shiguang Shan et al. have developed Ensemble of Piecewise FDA Based on Spatial Histograms of Local (Gabor) Binary Patterns for Face Recognition (Shiguang Shan et al.,2006). Spatial histogram* of Local Binary Pattern (LBP) and Local Gabor Binary Pattern (LGBP) has been successfully applied to face recognition and achieved state-of-the-art performance. Both LBP and LGBP Utilize traditional histogram matching method such as histogram intersection for face classification. In their paper, they proposed a statistical extension for L(G)BP similarity computation by introducing Fisher Discriminate Analysis (FDA) of the L(G)BP spatial histogram “features”. More than a simple application of FDA, they have constructed Ensemble of Piecewise FDA (EPFDA) classifiers, each of which is designed using one segment of the entire spatial histogram features. They showed that this extension not only greatly reduces the feature dimension but also brings very impressive performance improvement. Especially, they have made a large step to recognize all the faces in the standard FERET face database.

Bo Wu and Ram Nevatia proposed Tracking of Multiple, Partially Occluded Humans based on Static Body Part Detection **(**Bo Wu and Ram Nevatia, 2006)**.**Tracking of humans in videos is important for many applications. A major source of difficulty in performing this task is due to inter-human or scene occlusion. They presented an approach in which humans are represented as an assembly of four body parts and detection of the body parts in single frame which makes the method insensitive to camera motions. The responses of the body part detectors and a combined human detector provide the "observations" used for tracking. Trajectory initialization and termination are both fully automatic and rely on the confidences computed from the detection responses. An object is tracked by data association if its corresponding detection response can be found; otherwise it is tracked by a mean shift style tracker. Their method can track humans with both inter-object and scene occlusions. The system is evaluated on three sets of Videos and compared with previous methods.

Yu Su,Shiguang et al. have developed Patch-Based Gabor Fisher Classifier for Face Recognition(Yu Su,Shiguang et al.,2006). Face representations based on Gabor features have achieved great success in face recognition, such as Elastic Graph Matching, Gabor Fisher Classifier (GFC), and AdaBoosted Gabor Fisher Classifier (AGFC). In GFC and AGFC, either down-sampled or selected Gabor features are analyzed in holistic mode by a single classifier. In their paper, they proposed a novel patch-based GFC (PGFC) method, in which Gabor features are spatially partitioned into a number of patches, and on each patch one GFC is constructed as component classifier to form the final ensemble classifier using sum rule. The positions and sizes of the patches are learned from a training data using AdaBoost. Experiments on two large-scale face databases (FERET and CAS-PEAL-R1) show that the proposed PGFC with only tens of patches outperforms the GFC and AGFC impressively.

Chengjun Liu has explored "Capitalize Dimensionality Increasing Techniques for Improving Face Recognition Grand Challenge Performance" (Chengjun Liu. 2006).They presented a novel pattern recognition framework by capitalizing on dimensionality increasing techniques. In particular, the framework

integrates Gabor image representation, a novel multiclass Kernel Fisher Analysis (KFA) method, and fractional power polynomial models for improving pattern recognition performance. Gabor image representation, which increases dimensionality by incorporating Gabor filters with different scales and orientations, was characterized by spatial frequency, spatial locality, and orientational selectivity for coping with image variabilities such as illumination variations. The KFA method first performs nonlinear mapping from the input space to a high-dimensional feature space, and then implements the multiclass Fisher discriminate analysis in the feature space. The significance of the nonlinear mapping is that it increases the discriminating power of the KFA method, which is linear in the feature space but nonlinear in the input space. The novelty of the KFA method comes from the fact that 1) it extends the two-class kernel Fisher methods by addressing multiclass pattern classification problems and 2) it improves upon the traditional Generalized Discriminate Analysis (GDA) method by deriving a unique solution (compared to the GDA solution, which is not unique). The fractional power polynomial models further improved performance of the proposed pattern recognition framework. Experiments on face recognition using both the FERET database and the FRGC (Face Recognition Grand Challenge) databases show the feasibility of the proposed framework. In particular, experimental results using the FERET database show that the KFA method performs better than the GDA method and the fractional power polynomial models help both the KFA method and the GDA method to improve their face recognition performance. Experimental results using the FRGC databases show that the proposed pattern recognition framework improves face recognition performance upon the BEE baseline algorithm and the LDA-based baseline algorithm by large margins.

Deva Ramanan et al. have developed Tracking People by Learning their appearance (Deva Ramanan, et al., 2007). An open vision problem is to automatically track the articulations of people from a video sequence. This problem is difficult because one needs to determine both the number of people in each frame and estimate their configurations. But, finding people and localizing their limbs is hard because people can move fast and unpredictably, can appear in a variety of

poses and clothes, and are often surrounded by limb-like clutter. They developed a completely automatic system that worked in two stages; it first build a model of appearance of each person in a video and then it tracked by detecting those models in each frame ("tracking by model-building and detection"). They developed two algorithms that build models; one bottom-up approach groups together candidate body parts found throughout a sequence. They also described a top-down approach that automatically builds people-models by detecting convenient key poses within a sequence. They finally showed that building a discriminative model of appearance is quite helpful since it exploits structure in a background (without background-subtraction). They demonstrated the resulting tracker on hundreds of thousands of frames of unscripted indoor and outdoor activity, a feature-length film ("Run Lola Run"), and legacy sports footage (from the 2002 World Series and 1998 Winter Olympics). Experiments suggest that their system 1) can count distinct individuals, 2) can identify and track them, 3) can recover when it loses track, for example, if individuals are occluded or briefly leave the view, 4) can identify body configuration accurately, and 5) is not dependent on particular models of human motion.

Md. Tajmilur Rahman and Md. Alamin Bhuiyan have developed Face Recognition using Gabor Filters(Md. Tajmilur Rahman and Md. Alamin Bhuiyan, 2008). Gabor based face representation has achieved enormous success in face recognition. This system addressed a novel algorithm for face recognition using neural networks trained by Gabor features. The system commences on convolving some morphed images of particular face with a series of Gabor filter co-efficient at different scales and orientations. Two novel contributions of this paper are: scaling of RMS contrast, and contribution of morphing as an advancement of image recognition perfection. The neural network employed for face recognition is based on the Multi Layer Perception (MLP) architecture with back-propagation algorithm and incorporates the convolution filter response of Gabor jet. The effectiveness of the algorithm has been justified over a morphed facial image database with images captured in different illumination conditions.

Venet Osmani, Sasitbramaniam and Dmitri Botvichharan Balasu. have developed "Human activity recognition in pervasive health-care: Supporting

efficient remote collaboration (Venet Osmani, Sasitbramaniam and Dmitri Botvichharan Balasu. 2008). Technological advancements, including advancements in the medical field have drastically improved our quality of life, thus pushing life expectancy increasingly higher. This has also had the effect of increasing the number of elderly population. More than ever, health-care institutions must now care for a large number of elderly patients, which is one of the contributing factors in the rising health-care costs. Rising costs have prompted hospitals and other health-care institutions to seek various cost-cutting measures in order to remain competitive. One avenue being explored lies in the technological advancements that can make hospital working environments much more efficient. Various communication technologies, mobile computing devices, micro-embedded devices and sensors have the ability to support medical staff efficiency and improve health-care systems. In particular, one promising application of these technologies is towards deducing medical staff activities. Having this continuous knowledge about health-care staff activities can provide medical staff with crucial information of particular patients, interconnect with other supporting applications in a seamless manner (e.g. a doctor diagnosing a patient can automatically be sent the patient's lab report from the pathologist), a clear picture of the time utilization of doctors and nurses and also enable remote virtual collaboration between activities, thus creating a strong base for establishment of an efficient collaborative environment. In their paper, they described activity recognition system that in conjunction with their efficiency mechanism has the potential to cut down health-care costs by making the working environments more efficient. Initially, they outlined the activity recognition process that has the ability to infer user activities based on the self-organization of surrounding objects that user may manipulate. They then used the activity recognition information to enhance virtual collaboration in order to improve overall efficiency of tasks within a hospital environment. They have analysed a number of medical staff activities to guide their simulation setup. Their results showed an accurate activity recognition process for individual users with respect to their behavior. At the same time they supported remote virtual collaboration through tasks allocation process between doctors and nurses with results showing maximum efficiency within the resource constraints.

Linlin Shen1 et al. have developed Data Driven Gabor Wavelet Design for Face Recognition (Linlin Shen1 et al., 2009)**.** They presented a general example-based framework for detecting objects in static images by components. The technique was demonstrated by developing a system that locates people in cluttered scenes. The system was structured with four distinct example-based detectors that were trained to separately find the four components of the human body namely the head, legs, left arm, and right arm. After ensuring that these components are present in the proper geometric configuration, a second example-based classifier combines the results of the component detectors to classify a pattern as either a ªpersonº or a ªnonperson.º They called this type of hierarchical architecture, in which learning occurs at multiple stages, an Adaptive Combination of Classifiers (ACC).

Ahmed Bilal et al. have developed The painful face – Pain expression recognition using active appearance models (Ahmed Bilal et al., 2009). Pain is typically assessed by patient self-report. Self-reported pain, however, is difficult to interpret and may be impaired or in some circumstances (i.e., young children and the severely ill) not even possible. To circumvent these problems behavioral scientists have identified reliable and valid facial indicators of pain. Hitherto, these methods have required manual measurement by highly skilled human observers. In their paper they explored an approach for automatically recognizing acute pain without the need for human observers. Specifically, their study was restricted to automatically detecting pain in adult patients with rotator cuff injuries. The system employed video input of the patients as they moved their affected and unaffected shoulder. Two types of ground truth were considered. Sequence-level ground truth consisted of Likert-type ratings by skilled observers. Frame-level ground truth was calculated from presence/ absence and intensity of facial actions previously associated with pain. Active appearance models (AAM) were used to decouple shape and appearance in the digitized face images. Support vector machines (SVM) were compared for several representations from the AAM and of ground truth of varying granularity. They explored two questions pertinent to the construction, design and development of automatic pain detection systems. First, at what level (i.e., sequence- or frame-level) should datasets be labeled in order to obtain

satisfactory automatic pain detection performance? Second, how important is it, at both levels of labeling, that we non-rigidly register the face.

Thi Duon et al. have developed Efficient duration and hierarchical modeling for human activity recognition (Thi Duon et al., 2009). A challenge in building pervasive and smart spaces is to learn and recognize human activities of daily living (ADLs). In their paper, they addressed this problem and argued that in dealing with ADLs, it is beneficial to exploit both their typical duration patterns and inherent hierarchical structures. They exploited efficient duration modeling using the novel Coxian distribution to form the Coxian hidden semi-Markov model (CxHSMM) and apply it to the problem of learning and recognizing ADLs with complex temporal dependencies. The Coxian duration model has several advantages over existing duration parameterization using multinomial or exponential family distributions, including its denseness in the space of nonnegative distributions, low number of parameters, computational efficiency and the existence of closed-form estimation solutions. Further they combined both hierarchical and duration extensions of the hidden Markov model (HMM) to form the novel switching hidden semi-Markov model (SHSMM), and empirically compare its performance with existing models. The model can learn what an occupant normally does during the day from unsegmented training data and then perform online activity classification, segmentation and abnormality detection. Experimental results showed that Coxian modeling outperforms a range of baseline models for the task of activity segmentation. They also achieved a recognition accuracy competitive to the current state-of-the-art multinomial duration model, while gaining a significant reduction in computation. Furthermore, cross-validation model selection on the number of phases K in the Coxian indicates that only a small K is required to achieve the optimal performance. Finally, their models are further tested in a more challenging setting in which the tracking is often lost and the activities considerably overlap. With a small amount of labels supplied during training in a partially supervised learning mode, their models are again able to deliver reliable performance, again with a small number of phases, making their proposed framework an attractive choice for activity modeling.

Antonios Oikonomopoulos et al. have developed Sparse B-spline polynomial descriptors for human activity recognition (Antonios Oikonomopoulos et al., 2009).*They presented* The extraction and quantization of local image and video descriptors for the subsequent creation of visual codebooks is a technique that has proved very effective for image and video retrieval applications.

Xi Zhao, et al. have developed *AU* Recognition on 3D Faces Based On An Extended Statistical Facial Feature Model (Xi Zhao, et al., 2010). Recognition of facial action units (AU) is one of two main streams in the facial expressions analysis. Action units deform facial appearance simultaneously in landmark locations and local texture as well as geometry on 3D faces. Thus, it is necessary to extract features from multiple facial modalities to characterize these deformations comprehensively. In order to fuse the contribution of the discriminative power from all features efficiently, they proposed to use their extended statistical facial feature models (SFAM) to generate feature instances corresponding to AU class for each feature. Then the similarity between each feature on a face and its instances are evaluated so that a set of similarity scores are obtained. All sets of scores on the face are then weighted for AU recognition. Experiments on the Bosphorus database show its state-of-the-art performance.

Rajeev Shrivastava & Ankita nigam have developed Analysis and performance of face recognition system using Gabor filter bank with HMM model (Rajeev Shrivastava & Ankita nigam,2010). We presented a biometrics system performing identification, of automatic face recognition. This system is based on Gabor features extraction using Gabor filter bank construction. For feature extraction the input image is convolved with Gabor filter bank to select a set of informative and non-redundant Gabor features. The extracted features are again subjected to Discrete Radom Transform (DRT) to extract a sequence of feature vectors. The HMM (Hidden Markov Models) is used for matching the input face image to the stored images.

Jose Gonzalez-Mora et al. have developed Learning a generic 3D face model from 2D image databases using incremental Structure-from-Motion (Jose Gonzalez-

Mora et al., 2010). They presented a Over the last decade 3D face models have been extensively used in many applications such as face recognition, facial animation and facial expression analysis. 3D Morphable Models (MMs) have become a popular tool to build and fit 3D face models to images. Critical to the success of MMs is the ability to build a generic 3D face model. Major limitations in the MMs building process are: (1) collecting 3D data usually involves the use of expensive laser scans and complex capture setups, (2) the number of available3D databases is limited, and typically there is a lack of expression variability and (3) finding correspondences and registering the 3D model is a labor intensive and error prone process.

Xiaoming Liu has proposed Video-based face model fitting using Adaptive Active Appearance Model (Xiaoming Liu, 2010). They have Active Appearance Model (AAM) representing the shape and appearance of an object via two low-dimensional subspaces, one for shape and one for appearance. AAM for facial images is currently receiving considerable attention from the computer vision community. However, most existing work focuses on fitting an AAM to a single image. For many applications, effectively fitting an AAM to video sequences is of critical importance and challenging, especially considering the varying quality of real-world video content. Their paper proposed an Adaptive Active Appearance Model (AAAM) to address this problem, where both a generic AAM component and a subject-specific appearance model component are employed simultaneously in the proposed fitting scheme. While the generic AAM component is held fixed, the subject-specific model component is updated during the fitting process by selecting the frames that can be best explained by the generic model. Experimental results from both indoor and outdoor representative video sequences demonstrate the faster fitting convergence and improved fitting accuracy.

Huimin Qian et al. have developed Recognition of human activities using SVM multi-class classifier (Huimin Qian et al.,2010).Even great efforts have been made for decades, the recognition of human activities is still an unmature technology that attracted plenty of people in computer vision. In their paper, a system framework was presented to recognize multiple kinds of activities from videos by an SVM multi-class classifier with a binary tree architecture. The

framework composed of three functionally cascaded modules: (a) detecting and locating people by non-parameter background subtraction approach, (b) extracting various of features such as local ones from the minimum bounding boxes of human blobs in each frames and a newly defined global one, contour coding of the motion energy image (CCMEI), and (c) recognizing activities of people by SVM multi-class classifier whose structure is determined by a clustering process. The thought of hierarchical classification is introduced and multiple SVMs are aggregated to accomplish the recognition of actions. Each SVM in the multi-class classifier is trained separately to achieve its best classification performance by choosing proper features before they are aggregated. Experimental results both on a homebrewed activity data set and the public Schüldt's data set show the perfect identification performance and high robustness of the system.

Rajeev Shrivastava & Ankita nigam have developed on Analysis and performance of face recognition system using log Gabor filter bank with HMM model (Rajeev Shrivastava & Ankita nigam, 2011). We presented a biometrics system performing identification, of automatic face recognition. This system is based on Gabor features extraction using Log Gabor filter bank construction. For feature extraction the input image is convolved with log Gabor filter bank to select a set of informative and non-redundant Gabor features. The extracted features are again subjected to Discrete Radom Transform (DRT) to extract a sequence of feature vectors. The HMM (Hidden Markov Models) is used for matching the input face image to the stored images. The Log-Gabor filter has a response that is Gaussian when viewed on a logarithmic frequency scale instead of a linear one. This allows more information to be captured in the high frequency areas and also has desirable high pass characteristics.

Jinhui Hu, et al. have developed Fast human activity recognition based on structure and motion (Jinhui Hu, et al.,2011). They presented a method for the recognition of human activities. The proposed approach is based on the construction of a set of templates for each activity as well as on the measurement of the motion in each activity. Templates are designed so that they capture the structural and motion information that is most discriminative among activities. The direct motion

measurements capture the amount of translational motion in each activity. The two features are fused at the recognition stage. Recognition is achieved in two steps by calculating the similarity between the templates and motion features of the test and reference activities. The proposed methodology is experimentally assessed and is shown to yield excellent performance.

Daniel Roggen et al. have developed Activity Recognition in Opportunistic Sensor Environments (Daniel Roggen et al.,2011). They OPPORTUNITY is project under the EU FET-Open funding1 in which we develop mobile systems to recognize human activity in dynamically varying sensor setups. The system autonomously discovers available sensors around the user and self-configures to recognize desired activities. It reconfigures itself as the environment changes, and encompasses principles supporting autonomous operation in open-ended environments. Opportunity mainstreams ambient intelligence and improves user acceptance by relaxing constraints on body-worn sensor characteristics, and eases the deployment in real-world environments.

Óscar D. Lara, et al. have developed Centinela: A human activity recognition system based on acceleration and vital sign data **(**Óscar D. Lara, et al.,2011)**.** This presented Centinela, as a system that combines acceleration data with vital signs to achieve highly accurate activity recognition. Centinela recognizes five activities namely walking, running, sitting, ascending, and descending. The system includes a portable and unobtrusive real-time data collection platform, which only requires a single sensing device and a mobile phone. To extract features, both statistical and structural detectors are applied, and two new features are proposed to discriminate among activities during periods of vital sign stabilization. After evaluating eight different classifiers and three different time window sizes, their results showed that Centinela achieves up to 95.7% overall accuracy, which is higher than current approaches under similar conditions. Their results also indicated that vital signs are useful to discriminate between certain activities. Indeed, Centinela achieves 100% accuracy for activities such as running and sitting, and slightly improves the classification accuracy for ascending compared to the c.

Chun Zhu and Weihua Sheng have developed Motion- and location-based online human daily activity recognition (Chun Zhu and Weihua Sheng, 2011). They proposed an approach to indoor human daily activity recognition which combined motion data and location information. One inertial sensor was worn on the right thigh of a human subject to provide motion data, while an optical motion capture system was used to provide the human location information. Such a combination has the advantage of significantly reducing the obtrusiveness to the human subject at a moderate cost of vision processing, while maintaining a high accuracy of recognition. First, a two-step algorithm was proposed to recognize the activity based on motion data only. In the coarse-grained classification, two neural networks were used to classify the basic activities. In the fine grained classification, the sequence of activities was modeled by an HMM to consider the sequential constraints. The modified short-time Viterbi algorithm was used for real-time daily activity recognition. Second, to fuse the motion data with the location information, Bayes' theorem was used to update the activities recognized from the motion data. They conducted experiments in a mock apartment and the obtained results proved the effectiveness and accuracy of their algorithms.

Jin Wanga, et al. have developed Recognizing Human Daily Activities From Accelerometer Signal (Jin Wanga, et al.,2011).Automated recognition of human daily activities from wearable sensor signals has attracted a great deal of interests in many applications including health care, sports and aged care. They presented a Hidden Markov Model (HMM)-based recognition method to recognize six human daily activities from sensor signals collected from a single waist-worn tri-axial accelerometer. All training signals from the same activity class are modeled as generated by a HMM, while a Gaussian Mixture Model (GMM) was used to model the continuous observation for each hidden state. A new test signal was classified to the activity class corresponding to the HMM that can produce the highest likelihood.

Hirotaka et al. have developed Importance-weighted least-squares probabilistic classifier for covariate shift adaptation with application to human activity recognition (Jin Wanga, et al.,2012).Human activity recognition from accelerometer data (e.g., obtained by smart phones) is gathering a great deal of

attention since it can be used for various purposes such as remote health-care. However, since collecting labeled data is bothersome for new users, it is desirable to utilize data obtained from existing users. In their paper, they formulated this adaptation problem as learning under covariate shift, and propose a computationally efficient probabilistic classification method based on adaptive importance sampling. The usefulness of the proposed method was demonstrated in real-world human activity recognition.

Loren Arthur Schwarz et al. have developed Recognizing multiple human activities and tracking full-body pose in unconstrained environments (Loren Arthur Schwarz et al.,2012). Visual observations, such as camera images, are hard to obtain for long-term human motion analysis in unconstrained environments. In their paper, they present a method for human full-body pose tracking and activity recognition from measurements of few body-worn inertial orientation sensors. The sensors make their approach insensitive to illumination and occlusions and permit a person to move freely. Since the data provided by inertial sensors is sparse, noisy and often ambiguous, they used a generative prior model of feasible human poses and movements to constrain the tracking problem. Their model consists of several low-dimensional, activity-specific manifold embeddings that significantly restrict the search space for pose tracking. Using a particle filter, their method continuously explored multiple pose hypotheses in the embedding space. An efficient activity switching mechanism governs the distribution of particles across the activity-specific manifold embeddings. Selecting a pose hypothesis that best explained incoming sensor observations simultaneously allowed them to classify the activity a person is performing and to estimate the full-body pose. They also derived an effective measure of predictive confidence that enabled detecting anomalous movements. Experiments on a multi-person data set containing several activities showed that their method can seamlessly detect activity switches and accurately reconstruct full-body poses from the data of only six wearable inertial sensors.

Prasad S.Halgaonkar et al. have developed Face Recognition System for Feature Extraction based on Maximum Margin Criteria (Prasad S.Halgaonkar et al.,2012). Face recognition has attracted significant attention because of its wide

range of applications. Recently, more researchers focus on robust face recognition such as face recognition systems invariant to pose, expression and illumination variations. Illumination variation is still a challenging problem in face recognition research area. The same person can appear greatly different under varying lighting conditions. Their paper dealt with developing a Face Recognition System which was invariant to illumination variations. Face recognition system which uses Linear Discriminant Analysis (LDA) as feature extractor have "Small Sample Size (SSS). Their paper consisted of implementation of Feature Extraction Module using Two Dimensional Maximum Margin Criteria which removed "Small Sample Size (SSS)" problem present in existing Face Recognition System. In statistical pattern recognition, high dimensionality is a major cause of the practical limitations of many pattern recognition technologies. Moreover, it has been observed that a large number of features may actually degrade the performance of classifiers if the number of training samples is small relative to the number of features. This fact, which is referred to as the "peaking phenomenon", is caused by the "curse of dimensionality". The dimensionality of images after feature extraction for storing feature database was reduced in their paper. The input for the system is images from standard database. Features are extracted of given images using Two Dimensional Maximum Margin Criteria from row as well as column direction. Finally matching is done using Euclidean Distance for test image and stored image features.

Jose M. Chaquet et al. have developed A survey of video datasets for human action and activity recognition(Jose M. Chaquet et al., 2013). Vision-based human action and activity recognition has an increasing importance among the computer vision community with applications to visual surveillance, video retrieval and human–computer interaction. In recent years, more and more datasets dedicated to human action and activity recognition have been created. The use of these datasets allowed them to compare different recognition systems with the same input data. Their survey introduced in the paper tried to cover the lack of a complete description of the most important public datasets for video-based human activity and action recognition and to guide researchers in the election of the most suitable dataset for benchmarking their algorithms.

F.Chamroukhi et al. have developed Joint segmentation of multivariate time series with hidden process regression for human activity recognition (F.Chamroukhi et al., 2013). The problem of human activity recognition is central for understanding and predicting the human behavior, in particular in a prospective of assistive services to humans, such as health monitoring, well being, security, etc. There is therefore a growing need to build accurate models which can take into account the variability of the human activities over time (dynamic models) rather than static ones which can have some limitations in such a dynamic context. In their paper, the problem of activity recognition was analyzed through the segmentation of the multidimensional time series of the acceleration data measured in the 3-d space using body-worn accelerometers. The proposed model for automatic temporal segmentation was a specific statistical latent process model which assumed that the observed acceleration sequence is governed by sequence of hidden (unobserved) activities. More specifically, the proposed approach was based on a specific multiple regression model incorporating a hidden discrete logistic process which governs the switching from one activity to another over time. The model is learned in an unsupervised context by maximizing the observed-data log-likelihood via a dedicated expectation– maximization (EM) algorithm. They applied it on a real-world automatic human activity recognition problem and its performance was assessed by performing comparisons with alternative approaches including well-known supervised static classifiers and the standard hidden Markov model(HMM).The obtained results were very encouraging and show that the proposed approach is quite competitive even it worked in an entirely unsupervised way and did not require a feature extraction preprocessing step.

Enrique Yeguas et al. have developed Exploring STIP-based models for recognizing human interactions in TV videos (Enrique Yeguas et al. 2013). Human motion recognition – action (HAR) or interaction (HIR) – in real video data is identified as a very challenging task. In the last few years models of increasing complexity have been proposed in order to improve the performance in the task. However, it still remains unclear whether it is the features or the models what deserves the increase in complexity. In their paper an evaluation of such problem

was carried out in the HIR task. For that purpose, they compared the results obtained in their experiments – by using STIP-based features and BOW models as basis and combined with a standard classifier – with some of the more effective and recent approaches that used alternative representation models. They performed a comprehensive experimental study on two state-of-the-art databases in HIR: TV Human interactions and UTinteractions. They compared the results of their experiments with recent results published on these datasets. In addition, they run cross-data experiments on Hollywood-2 dataset in order to study the capability of generalization of the trained models through different datasets. The most relevant result is that the model combining STIP + BOW is competitive in the HIR task in comparison with the most complex ones. It was also shown that the vocabulary learning subtask can be improved by using compression algorithms on large enough initial set of features. In contrast to other categorization tasks the context does not help, the results showed that dense sampling of STIP is the best choice, but only when it is used inside the region of interest of the interaction.

M. Ros et al. have developed Online recognition of human activities and adaptation to habit changes by means of learning automata and fuzzy temporal windows (M. Ros et al..2013).Smart Homes are intelligent spaces that contain resources to collect information about user's activities and ease the assisted living. Abnormal behavior detection has been remarked as one of the most challenging application fields in this research area, due to its possibilities for assisting elders or people with special needs. These systems helped to maintain people's independence, enhancing their personal comfort and safety and delaying the process of moving to a nursing home. In their paper, they described a new approach for the behavior recognition problem based on Learning Automata and fuzzy temporal windows. Their proposal learned the normal behaviors, and used that knowledge to recognise normal and abnormal human activities in real time. In addition, their proposal was able to adapt online to environmental variations, changes in human habits, and temporal information, defined as an interval of time when the behavior should be performed.

Enrique Garcia-Ceja and Ramon Brena *have* explored Long-Term Activity Recognition from Accelerometer Data (Enrique Garcia-Ceja and Ramon Brena,

2013). In the last years, simple activity recognition through wearable sensors has been achieved successfully, however complex activity recognition is still challenging. Simple activities may last just a few seconds, e.g., walking, running, resting, etc. whereas complex activities involve a combination of the former and they may last from a few minutes to several hours. In their work long-term activity recognition was performed and modeled as a distribution of simple activities represented as a histogram. For the experiments, the raw histograms were used for the recognition task and then they added an additional step which consists of extracting features over the histogram and applying a simple threshold to reduce noise. This additional step resulted in an increase on the classification accuracy.

Georgios Goudelis n, et al. have developed "Exploring trace transform for robust human action recognition(Georgios Goudelis n, et al.,2013). Machine based human action recognition has become very popular in the last decade. Automatic unattended surveillance systems, interactive video games, machine learning and robotics are only few of the areas that involve human action recognition. Their paper examined the capability of a known transform, the so-called Trace, for human action recognition and proposed two new feature extraction methods based on the specific transform. The first method extracted Trace transforms from binarized silhouettes, representing different stages of a single action period. A final history template composed from the above transforms, represented the whole sequence containing much of the valuable spatio- temporal information contained in a human action. The second, involves trace for the construction of a set of invariant features that represented the action sequence and can cope with variations usually appeared in video capturing. The specific method takes advantage of the natural specifications of the Trace transform, to produce noise robust features that are invariant to translation, rotation, scaling and are effective, simple and fast to create. Classification experiments performed on two well known and challenging action datasets (KTH and Weizmann) using Radial Basis Function (RBF) Kernel SVM provided very competitive results indicating the potentials of the proposed techniques.

Mehrsan Javan Roshtkhari and Martin D. Levine *have* proposed 'Human activity recognition in videos using a single example (Mehrsan Javan Roshtkhari

and Martin D. Levine, 2013). They presented a novel approach for action recognition, localization and video matching based on a hierarchical code book model of local spatio-temporal video volumes. Given a single example of an activity as a query video, the proposed method finds similar videos to the query in a target video dataset. The method was based on the bag of video words (BOV) representation and does not require prior knowledge about actions, background subtraction, motion estimation or tracking. It was also robust to spatial and temporal scale changes, as well as some deformations. The hierarchical algorithm codes a video as a compact set of spatio-temporal volumes, while considering their spatio-temporal compositions in order to account for spatial and temporal contextual information. This hierarchy was achieved by first constructing a codebook of spatio-temporal video volumes. Then a large contextual volume containing many spatio-temporal volumes (ensemble of volumes) was considered. These ensembles were used to construct a probabilistic model of video volumes and their spatio-temporal compositions. The algorithm was applied to three available video datasets for action recognition with different complexities (KTH, Weizmann, and MSR II) and the results were superior to other approaches, especially in the case of a single training example and cross-dataset action recognition.

Hongqing Fang et al. have developed Influence of time and length size feature selections for human activity sequences recognition (Hongqing Fang et al., 2013). They presented Viterbi algorithm based on a hidden Markov model was applied to recognize activity sequences from observed sensors events. Alternative features selections of time feature values of sensors events and activity length size feature values were tested, respectively, and then the results of activity sequences recognition performances of Viterbi algorithm were evaluated. The results showed that the selection of larger time feature values of sensor events and/or smaller activity length size feature values will generate relatively better results on the activity sequences recognition performances.

Muhammad Shahzad Cheema et al. have developed “Human activity recognition by separating style and content(Muhammad Shahzad Cheema et al., 2013). Studies in psychophysics suggested that people tend to perform different

actions in their own style. The article dealt with the problem of recognizing human actions and the underlying execution styles (actors) in videos. They presented a hierarchical approach that was based on conventional action recognition and asymmetrical bilinear modeling. In particular they employed bilinear factorization on the tonsorial representation of the action videos to characterize styles of performing different actions. Their approach was completely based on the dynamics of the underlying activity. The model was examined on the IXMAS and the Berkeley-MHAD data sets using different modalities based on optical motion capture, Kinect depth videos, and 3D motion history volumes. In each case higher recognition accuracy was achieved in comparison to the symmetric bilinear modeling and the Nearest Neighbor classification.

Jeongmin Yu et al. have developed Weighted feature trajectories and concatenated bag-of-features for action recognition (Jeongmin Yu et al.,2013). Key-point trajectory based approached to recognize human actions in realistic videos have recently shown promising results. However, their coverage of the entire actor was not sufficient to describe human actions, and the trajectories often disappear due to partial occlusions. In their paper, they proposed a novel approach based on weighted feature trajectories and concatenated bag-of-features (BOF) using weighted spatio-temporal descriptors to overcome these limitations. For capturing spatio-temporal regions of interest(ROI), they first build scale-invariant feature transform (SIFT) trajectories based on matching SIFT descriptors. Then, they extracted particle trajectories around the SIFT trajectories using dense optical flow to retain ROI subsidiarily and to give a sound coverage of the actors. To balance the leverages between SIFT and particle trajectories, a weighting scheme was presented according to the accumulated global distribution of feature points. Furthermore, They proposed the concatenated two distinct SIFT and particle BOFs with weighted spatio-temporal descriptors for action recognition. Experimental results demonstrated the effectiveness of this approach in challenging real-world action datasets such as KTH UCF sports, and TV human interaction.

Salah Althloothi, et al. have developed “Human Activity Recognition Using Multi-Features and Multiple Kernel Learning (Salah Althloothi, et al.,2013).They

presented two sets of features, shape representation and kinematic structure, for human activity recognition using a sequence of RGB-D images. The shape features were then extracted using the depth information in the frequency domain via spherical harmonics representation. The other features included the motion of the 3D joint positions (i.e. the end points of the distal limb segments) in the human body. Both sets of features were fused using the Multiple Kernel Learning (MKL) technique at the kernel level for human activity recognition. Their experiments on three publicly available datasets demonstrated that the proposed features are robust for human activity recognition and particularly when there are similarities among the actions.

Ilias Theodorakopoulos et al. have developed Pose-based human action recognition via sparse representation in dissimilarity space (Ilias Theodorakopoulos et al., 2014). Human actions can be considered as a sequence of body poses over time, usually represented by coordinates corresponding to human skeleton models. Recently, a variety of low-cost devices have been released, able to produce marker less real time pose estimation. Nevertheless, limitations of the incorporated RGB-D sensors can produce inaccuracies, necessitating the utilization of alternative representation and classification schemes in order to boost performance. In this context, They proposed a method for action recognition where skeletal data are initially processed in order to obtain robust and invariant pose representations and then vectors of dissimilarities to a set of prototype actions are computed. The task of recognition is performed in the dissimilarity space using sparse representation. A new publicly available dataset was introduced in this system, created for evaluation purposes. The proposed method was also evaluated on other public datasets, and the results are compared to those of similar method.

Ferda Ofli et al. have developed “Sequence of the most informative joints (SMIJ): A new representation for human skeletal action recognition (Ferda Ofli et al.,2014). Much of the existing work on action recognition combines simple features with complex classifiers or models to represent an action. Parameters of such models usually do not have any physical meaning nor do they provide any qualitative insight relating the action to the actual motion of the body or its parts. In their paper, they proposed a new representation of human actions called sequence of

the most informative joints (SMIJ), which was extremely easy to interpret. At each time instant, they automatically select a few skeletal joints that are deemed to be the most informative for performing the current action based on highly interpretable measures such as the mean or variance of joint angle trajectories. They then represented the action as a sequence of these most informative joints. Experiments on multiple databases showed that the SMIJ representation is discriminative for human action recognition and performs better than several state-of-the-art algorithms.

Guoliang Lu and MineichiKudo have explored Learning action patterns in difference images for efficient action recognition (Guoliang Lu and MineichiKudo, 2014). A new framework was presented for single-person oriented action recognition. This framework does not require detection/location of bounding boxes of human body nor motion estimation in each frame. The novel descriptor/pattern for action representation was learned with local temporal self-similarities (LTSSs) derived directly from difference mages. The bag-of-words framework was then employed for action classification taking advantages of these descriptors. They investigated the effectiveness of the framework on two public human action data sets: the Weizmann dataset and the KTH data set. In the Weizmann dataset, the proposed framework achieves a performance of 95.6% in the recognition rate and that of 91.1% in the KTH data set, both of which are competitive with those of state-of-the-art approaches, but it has a high potential to achieve a faster execution performance.

Mathieu Barnachon et al. have developed Ongoing human action recognition with motion capture (Mathieu Barnachon et al., 2014). Ongoing human action recognition is a challenging problem that has many applications, such as video surveillance, patient monitoring, human–computer interaction, etc. They presented a novel frame work for recognizing streamed actions using Motion Capture (MoCap) data. Unlike the after-the-fact classification of completed activities, their work aimed at achieving early recognition of ongoing activities. The proposed method is time efficient as it is based on histograms of action poses, extracted from MoCap data, that are computed according to Hausdorff distance. The histograms are then

compared with the Bhattacharyya distance and warped by a dynamic time warping process to achieve their optimal alignment. This process, implemented by their dynamic programming-based solution, has the advantage of allowing some stretching flexibility to accommodate for possible action length changes. They have shown the success and effectiveness of their solution by testing it on large datasets and comparing it with several state-of-the-art methods. In particular, they were able to achieve excellent recognition rates that have outperformed many well known methods.